Characterization of DNA interference by a minimal Type I-F CRISPR-Cas system

Dissertation

zur

Erlangung des Doktorgrades

der Naturwissenschaften

(Dr. rer. nat.)

dem Fachbereich Biologie

der Philipps-Universität Marburg

vorgelegt von

Hanna Constanza Müller Esparza

aus Antofagasta, Chile

Marburg/Lahn, Juli 2019

Die Untersuchungen der vorliegenden Arbeit wurden von September 2015 bis Juni 2019 unter der Betreuung von Herrn Dr. Lennart Randau am Max-Planck-Institut für terrestrische Mikrobiologie in Marburg durchgeführt.

Vom Fachbereich Biologie

der Philipps-Universität Marburg als Dissertation

angenommen am: 18.09.2019

Erstgutachter: Herr Dr. Lennart Randau

Zweitgutachter: Herr Prof. Dr. Martin Thanbichler

Weitere Mitglieder der Prüfungskommission:

Herr Prof. Dr. Kai Thormann

Herr Prof. Dr. Victor Sourjik

Tag der mündlichen Prüfung am: 20.09.2019

Erklärung

Ich versichere, dass ich meine Dissertation mit dem Titel „Characterization of DNA interference by a minimal Type I-F CRISPR-Cas system“ selbstständig ohne unerlaubte Hilfe angefertigt und mich dabei keiner anderen als der von mir ausdrücklich bezeichneten Quellen und Hilfsmittel bedient habe.

Diese Dissertation wurde in der jetzigen oder einer ähnlichen Form noch bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

Marburg, den 1. August 2019

Hanna Constanza Müller Esparza

„Nothing in life is to be feared, it is only to be understood. “

- Maria Skłodowska-Curie

Teile dieser Arbeit sind in folgenden Artikeln veröffentlicht:

Gleditzsch, D., **Müller-Esparza, H.**, Pausch, P., Sharma, K., Dwarakanath, S., Urlaub, H., et al. (2016). Modulating the Cascade architecture of a minimal Type I-F CRISPR-Cas system. Nucleic Acids Research, 44(12), 5872-5882.

Pausch, P., **Müller-Esparza, H.,** Gleditzsch, D., Altegoer, F., Randau, L., & Bange, G. (2017). Structural Variation of Type I-F CRISPR RNA Guided DNA Surveillance. Molecular Cell, 67(4), 622-632.

Turkowyd B.*, **Müller-Esparza H.***, Climenti V., Steube N., Endesfelder U.*, Randau L.* (2019). Live-cell single-particle tracking photoactivated localization microscopy of Cascade-mediated DNA surveillance. Methods in Enzymology. 616, 133-171.

*equal contribution

Weitere Veröffentlichungen:

Müller-Esparza H., Randau L. (2017). Commentary: Type I CRISPR-Cas targets endogenous genes and regulates virulence to evade mammalian host immunity. Frontiers in Microbiology. 8, 319.

Müller-Esparza, H., Gleditzsch, D., Randau L. (2018). Vielfältige Genscheren: natürliche Aktivitäten von CRISPR-Cas-Systemen. BIOspektrum. 24 (7), 704-706.

Gleditzsch D., Pausch P., **Müller-Esparza H.**, Özcan A., Guo X., Bange G., Randau, L. (2019). PAM identification by CRISPR-Cas effector complexes: diversified mechanisms and structures. RNA Biology. 46 (4), 504-517.

Contents

Summary

Clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) proteins constitute the only known adaptive immune system present in Archaea and Bacteria. This system targets and degrades foreign genetic material through ribonucleoprotein effector complexes carrying CRISPR RNAs (crRNAs), in a process termed interference. CRISPR-Cas systems are classified into 6 different types, with Type I systems being the most widespread in nature. They harbour the Cas3 helicase/nuclease as signature protein, involved in target degradation, and can be divided into 7 subtypes (A-F, U). They elicit interference by ribonucleoprotein complexes (termed CRISPR associated complexes for anti-viral defence, Cascades), formed of Cas proteins and a single crRNA. Cascades are able to discriminate between self and non-self substrates by identifying a short Protospacer Adjacent Motif (PAM) next to the crRNA target (termed protospacer). In Type I-F systems, PAM recognition is carried out by Cas8f (or large subunit).

In the present work, we analysed a smaller variant of a Type I-F CRISPR-Cas system, identified in *Shewanella putrefaciens* CN-32. This system lacks a large subunit and contains two previously uncharacterised Cas proteins, functionally identified as Cas5 and Cas7 homologues. We expressed the complex in *Escherichia coli* BL21-AI and demonstrated its activity against bacteriophages and plasmids in a sequence-, PAM- and Cas3-dependent manner. This heterologous activity indicates that interference can be carried out without the need of a large subunit or additional proteins from *S. putrefaciens*. Furthermore, we were able to identify a unique alpha helical domain in Cas5fv responsible for stringent GG PAM identification from the major groove side, in addition to Cas7fv-mediated non-target strand stabilisation. We also determined the binding affinities of the complex through BioLayer Interferometry (BLI), gaining insights into its target requirements. Moreover, we studied the dynamics of the minimal system through Single-Particle Tracking Photo-activated Localization Microscopy (sptPALM), revealing an unspecific DNA-binding capacity of the alpha helical domain of Cas5fv, as well as RNA interactions of Cas5fv and Cas6f. For Cascades, we determined the binding times to genomic targets, which were directly proportional to the complementarity between crRNA and DNA. Fully complementary targets elicited a binding duration of around 15 seconds.

We propose that this minimised Cascade version is an evolutionary response to the appearance of viral Anti-CRISPR (Acr) proteins, small proteins able to block CRISPR-Cas interference. The effect of a broad-spectrum Acr protein, AcrF9, on the minimal complex was tested. Neither a reduction of interference activity nor physical interactions were detected. This supports the hypothesis of complex reduction as a response to Acr pressure.

Zusammenfassung

Clustered regularly interspaced short palindromic repeats (CRISPR) und CRISPR-assoziierte (Cas) Proteine stellen das einzige in Bakterien und Archaeen bekannte adaptive Immunsystem dar. Hierbei kann ein Organismus mit Hilfe von Ribonukleoproteinkomplexen, die mit kleinen CRISPR RNAs (crRNAs) ausgestattet sind, fremde Nukleinsäuren erkennen und degradieren. Dieser Vorgang wird als Interferenz bezeichnet. CRISPR-Cas Systeme werden in sechs verschiedene Typen klassifiziert. Die Systeme des Typs I sind in der Natur am weitesten verbreitet; sie bestehen aus sieben Untertypen (A-F, U) und ihr gemeinsames Erkennungsmerkmal ist die Ziel-DNA Helikase/Nuklease Cas3. Alle Typ I-Systeme führen Interferenz mit Hilfe eines Ribonukleoproteinkomplexes (genannt CRISPR associated complex for antiviral defence, Cascade) aus, welcher aus mehreren Cas-Proteinen sowie einer einzigen crRNA besteht. Die Cascade-Komplexe können hierbei zwischen eigenen und fremden Nukleinsäuren unterscheiden, indem sie die PAM (Protospacer Adjacent Motif) identifizieren. Dabei handelt es sich um eine kurze, neben dem Protospacer liegende Sequenz, die wiederum neben der zur crRNA komplementären Region liegt. In Systemen des Typs I-F wird die Erkennung der PAM von der großen Untereinheit Cas8f ausgeführt.

Die vorliegende Arbeit analysiert eine minimale Variante des Typ I-F CRISPR-Cas Systems, die in *Shewanella putrefaciens* CN-32 entdeckt wurde. Während dieser minimalen Variante die große Untereinheit fehlt, enthält sie zwei zuvor unbekannte Cas-Proteine, die in einer vorangehenden Arbeit als funktionelle Homologe von Cas5 und Cas7 identifiziert wurden. Der minimale Cascade-Komplex wurde in *Escherichia coli* BL21 AI exprimiert. Seine Interferenz-Aktivität sowohl gegen Bakteriophagen als auch gegen Plasmide konnte als sequenz-, PAM- und Cas3-abhängig nachgewiesen werden. Dies beweist, dass für die Interferenz weder die Cas8-Untereinheit, noch andere Proteine aus *S. putrefaciens* notwendig sind. Durch strukturelle Analysen konnte des Weiteren eine alpha-helikale Domäne des Cascade-Proteins Cas5fv identifiziert werden, welche für die Erkennung der GG PAM aufseiten der großen Furche verantwortlich ist, sowie die Stabilisierung des Nicht-Ziel-Stranges durch Cas7fv-Untereinheiten. Außerdem wurde die Bindungskinetik des Komplexes mit Hilfe von BioLayer Interferometry (BLI) bestimmt. Dies gewährt uns Einsicht in die strukturellen Anforderungen, die der Cascade-Komplex an seine Substrate stellt. Weiterhin wurde Single-Particle Tracking Photo-Activated Localization Microscopy (sptPALM) verwendet, womit eine unspezifische DNA-Bindungskapazität der alpha-helikalen Domäne von Cas5fv aufgedeckt werden konnte, sowie die Fähigkeit von Cas5fv und Cas6f mit RNA zu interagieren. Bei den Cascade-Komplexen wurden die Bindungszeiten an genomische Ziele bestimmt, die sich direkt proportional zur Komplementarität von crRNA und DNA verhielten. Vollständig komplementäre Ziele lösten eine Bindung von 15 Sekunden aus.

Wir schlagen vor, dass dieser minimale Cascade-Komplex eine evolutionäre Antwort auf das Entstehen von viralen Anti-CRISPR-Proteinen (Acr-Proteine) ist. Dies sind kleine Proteine, welche die CRISPR-Cas Interferenz auf verschiedenen Wegen blockieren können. Die Wirkung des vielseitigen Acr-Proteins AcrF9 auf den minimalen Komplex wurde untersucht und es konnte weder eine Abnahme der Interferenz, noch eine anderweitige physikalische Interaktion gezeigt werden. Dies bestätigt die Annahme, dass selektiver Druck durch die Entstehung der Acr-Proteine eine Reduzierung der Cascade-Komplexe zur Folge hat.

Chapter I: Introduction

Introduction

In the environment, Bacteria and Archaea commonly face invasion by foreign DNA, in the form of mobile genetic elements (MGE). Through conjugation, transformation, transduction, transposition and viral infection, pieces of genetic material can be horizontally transferred between microorganisms. The integration or expression of these elements can have deleterious effects in the host, disturbing its genomic arrangement and altering its metabolism. In some cases, this translates into cell death and the collapse of microbial populations (Darmon & Leach, 2014; Maslov & Sneppen, 2017).

In order to avoid undesirable consequences derived from horizontal gene transfer (HGT), organisms have developed several defence mechanisms to either prevent the entry of exogenous DNA, degrade it once inside the cell or induce programmed cell death to protect the population (Makarova, Wolf, & Koonin, 2013). Among these mechanisms, CRISPR (Clustered Regular Interspaced Short Palindromic Repeats)-Cas (CRISPR-associated) systems have been characterised as adaptive immune systems, capable of fighting off invading plasmids and phages through the targeting and posterior degradation of their genetic material by ribonucleoprotein (RNP) complexes (Barrangou et al., 2007; Mojica, Díez-Villaseñor, García-Martínez, & Soria, 2005).

CRISPR-Cas systems are found predominately in Archaea (87% of analyzed genomes) and with less frequency in bacterial genomes (45%) (Grissa, Vergnaud, & Pourcel, 2007). CRISPR-Cas modules consist of arrays of unique spacer sequences, derived from fragments of foreign DNA (protospacers), flanked by short palindromic repeat sequences; plus a cluster of Cas genes, usually found in the vicinity of this array (Makarova et al., 2011). CRISPR arrays are highly dynamic, as new spacers can be added in a process denominated adaptation. During adaptation, Cas1 and Cas2 form a complex either among themselves or with the targeting RNP complexes to bind and cleave fragments between 26 and 72 bp (depending on the type of system) and insert them at the start of the CRISPR array as new spacers, with the subsequent addition of an upstream repeat (Nunez et al., 2014) (Figure I.1).

The CRISPR array gets transcribed into a long precursor RNA that is matured by specific Cas proteins with endoribonuclease activity. This process generates multiple CRISPR RNAs (crRNAs), each carrying a single spacer and part of the flanking repeats. The crRNAs form RNP complexes with the Cas proteins, which are able to screen through both self and non-self DNA until a short sequence, the Protospacer Adjacent Motif

(PAM), followed by the protospacer complementary to the crRNA is found. The PAM is located only in the invader genetic material, preventing the targeting of the CRISPR array. PAM recognition is the first step of the interference process, which ends up with the recruitment of an endonuclease capable of degrading the targeted genetic material (Barrangou et al., 2007; Brouns et al., 2008).

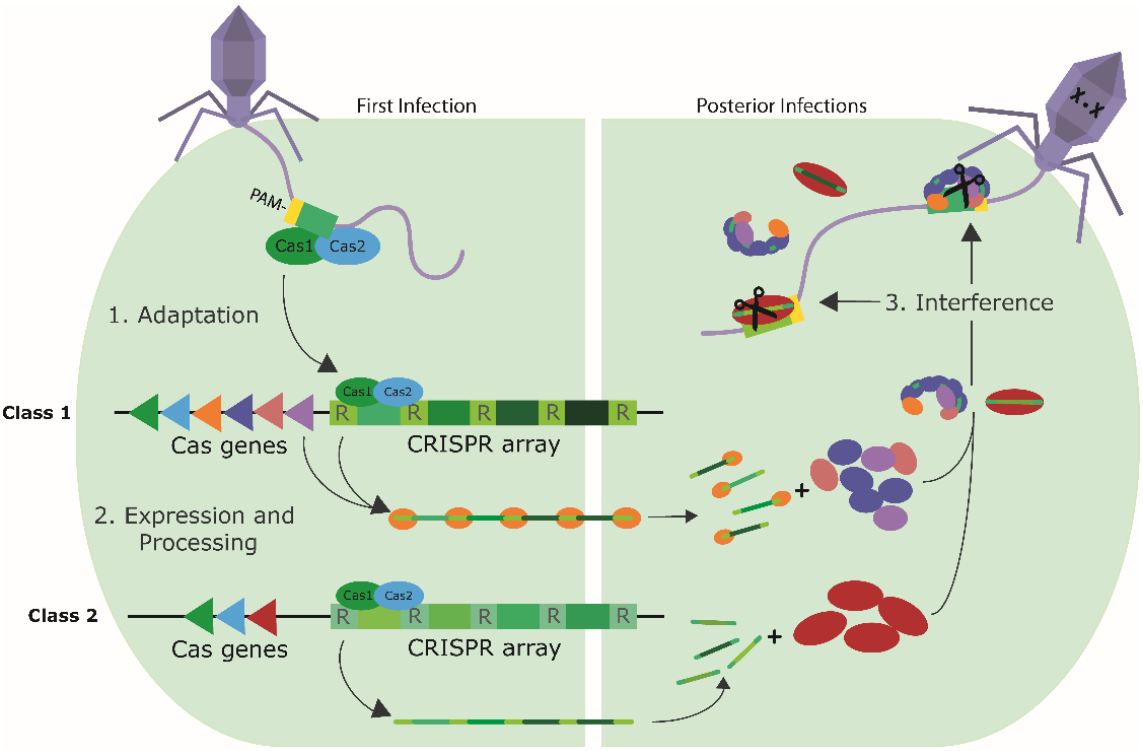


Figure I.1. Establishment of CRISPR-Cas immunity by Class1 and Class 2 systems. Upon viral infection, the systems are able to integrate unique sequences (spacers, in green shades) from the invading DNA into the CRISPR array. This array is then transcribed and processed into individual crRNA, that are then loaded into effector complexes (Class 1) or individual Cas proteins (Class 2). Each crRNA carries a unique sequence (spacer) and fragments of the flanking repeat sequences (R). For interference, the complexes scan the DNA for the corresponding PAM (yellow) and complementarity with the spacer, in order to cleave the target (scissors).

To date, several CRISPR-Cas systems have been described and classified into two classes and six types, according to characteristic proteins and features (Koonin, Makarova, & Zhang, 2017). Class 1 groups Type I, III and IV; while Class 2 groups Type II, V and VI systems, where interference is carried out by a single Cas protein. In the first class, targeting is performed by multi-protein complexes (Plagens, Richter, Charpentier, & Randau, 2015). For Type III systems, the process is coupled to transcription, where the multi-protein complex targets the nascent mRNA in a PAM-independent manner and non-specifically degrades any proximal DNA (Elmore et al., 2016; Estrella, Kuo, & Bailey, 2016; Jia et al., 2019; Kazlauskiene, Tamulaitis, Kostiuk, Venclovas, & Siksnys, 2016;

You et al., 2019). For Type IV systems, complex formation has been described (Ozcan et al., 2019), yet the role of it in interference is uncertain, as the lack of an associated nuclease suggests novel functions for this CRISPR-Cas system (Koonin & Makarova, 2019).

Type I CRISPR-Cas systems are the most abundant in nature. This type is further divided in seven subtypes (I-F, U) that share the conserved helicase/endonuclease Cas3, responsible for target degradation (Huo et al., 2014; Koonin et al., 2017). In this type, the RNP complexes in charge of interference are termed Cascades (CRISPR-associated Complexes for Antiviral Defence) (Brouns et al., 2008).

The Type I-F system found in *Pseudomonas aeruginosa* PA14 is one of the best-characterised systems among Type I subtypes. Here, Cascade is formed by a 60-nt long crRNA and four different Cas proteins with an uneven stoichiometry: $(Cas8f)_1$-$(Cas5f)_1$-$(Cas7f)_6$-$(Cas6f)_1$ (Cady, Bondy-Denomy, Heussler, Davidson, & O'Toole, 2012; Haurwitz, Jinek, Wiedenheft, Zhou, & Doudna, 2010; Wiedenheft et al., 2011). Cas6f is an endoribonuclease involved in pre-crRNA maturation; it generates an 8 nt 5' handle and a 20 nt 3' hairpin to which it stays bound. Cas5f caps the 5' end of the crRNA; the six Cas7fv subunits act as a backbone that binds the 32 nt spacer sequence; Cas8f (also known as large subunit) is liable for PAM recognition and, together with Cas7fv, it also stabilises the opened DNA (Chowdhury et al., 2017; Guo et al., 2017; Rollins et al., 2019). In this system, the Cas3 nuclease is fused to the adaptation protein Cas2 (Makarova et al., 2011).

Interference by Type I-F Cascade starts with the recognition of a GG PAM by the large subunit, as two asparagine residues in the N-terminal domain of Cas5fv interact with both nucleotides from the minor groove of the DNA. Subsequently, the double strand is opened at the first position of the protospacer by a lysine wedge of Cas8f, aiding in crRNA-target strand pairing. The complex then probes for the correct hybridisation of the 8 nts at the PAM-proximal end of the crRNA, termed seed sequence. Mismatches in this region are not tolerated, leading to abortive binding. When complementarity is sufficient, R-loop formation proceeds, with the displaced DNA strand being stabilised by electrostatic interactions with Cas8f, Cas5f and the Cas7f backbone. The fully-matched 3-stranded structure generates a conformational change of Cas8f, which rotates 180º, bringing the complex to a locked state that exposes the recruitment site for the DNA nuclease Cas2-3 (Wiedenheft et al., 2011; Rollins et al., 2019).

When target-crRNA complementarity is only partial or the PAM sequence is mutated, Type I-F Cascade will adopt a looser conformation upon target binding. This open state is unable to recruit Cas2-3, but rather recruits a propeller-shaped complex formed by Cas1 and Cas2-3 (Rollins et al., 2017). Together, the super-complex will generate and add a new spacer to the CRISPR array. The corresponding crRNA will match the target, elicit interference and close the positive feedback loop of the process termed primed adaptation (Redding et al., 2015; Rollins et al., 2017; Xue, Whitis, & Sashital, 2016).

As described, CRISPR-Cas systems are very diverse and spread among a wide number of microbes. By the Red Queen hypothesis, the resistance elicited by these systems would pressure the other interactors in a population, especially bacteriophages, to counter-adapt mechanisms to by-pass this immunity and survive (Paterson et al., 2010; Samson, Magadan, Sabri, & Moineau, 2013; Stenseth & Smith, 1984). Recently, small proteins capable of blocking CRISPR-Cas interference were discovered in prophages of *P. aeruginosa* (Bondy-Denomy, Pawluk, Maxwell, & Davidson, 2013; Pawluk, Bondy-Denomy, Cheung, Maxwell, & Davidson, 2014; Pawluk et al., 2016*).* This bacterium carries active Type I-F and Type I-E systems, with spacers targeting the integrated viral genome. Therefore, the activity of these systems would lead to self-targeting and death, an outcome that does not benefit either of the players.

The survival of these bacteria was first described to depend on the presence of four phage-encoded Anti-CRISPR (Acr) proteins against the Type I-F system (Bondy-Denomy et al., 2013). Although the search for homologues of these proteins, AcrF1-4, did not retrieve more inhibitors, the use of a gene upstream of the *Acrs,* coding for an associated transcriptional regulator called Anti-CRISPR associated gene (*Aca1*), revealed a conserved architecture within counter-defence loci. Using *Aca1* as an anchor to look for more inhibitors, several studies have reported a high variety of Acrs that match the CRISPR-Cas diversity, targeting both Class1 and 2 systems (Borges, Davidson, & Bondy-Denomy, 2017; Marino et al., 2018; Rauch et al., 2017; Watters, Fellman, Bai, Ren, & Doudna, 2018)

To date, 14 Acr proteins that block Type I-F interference have been described, and the mechanism of action has been elucidated for 3 of them. AcrF1 and AcrF2 inhibit DNA recognition by interacting with Cas7f and Cas8f (Chowdhury et al., 2017; Guo et al., 2017), while AcrF3 mimics a domain of Cas8f in order to block Cas2-3 (Rollins et al., 2019). The wide range of Acr discovered so far gives insight into the evolution of CRISPR-Cas systems and the arms-race that guides their diversification.

A recent study from our group has described a Type I-F variant (Type I-Fv) from *Shewanella putrefaciens* CN-32*,* whose Cas module only comprises 5 proteins: Cas1, a fusion of Cas2 and 3 (Cas2-3), Cas6f and two proteins with no sequence similarity to other Cas proteins described to date, but characterised by functional analysis as Cas5fv and Cas7fv (Figure I.2A) (Dwarakanath et al., 2015). From these proteins, only Cas6f, Cas5fv, Cas7fv and a 60 nt crRNA are required to form a stable Cascade (Figure I.2B), arising as one of the simplest Type I effector complexes described so far.

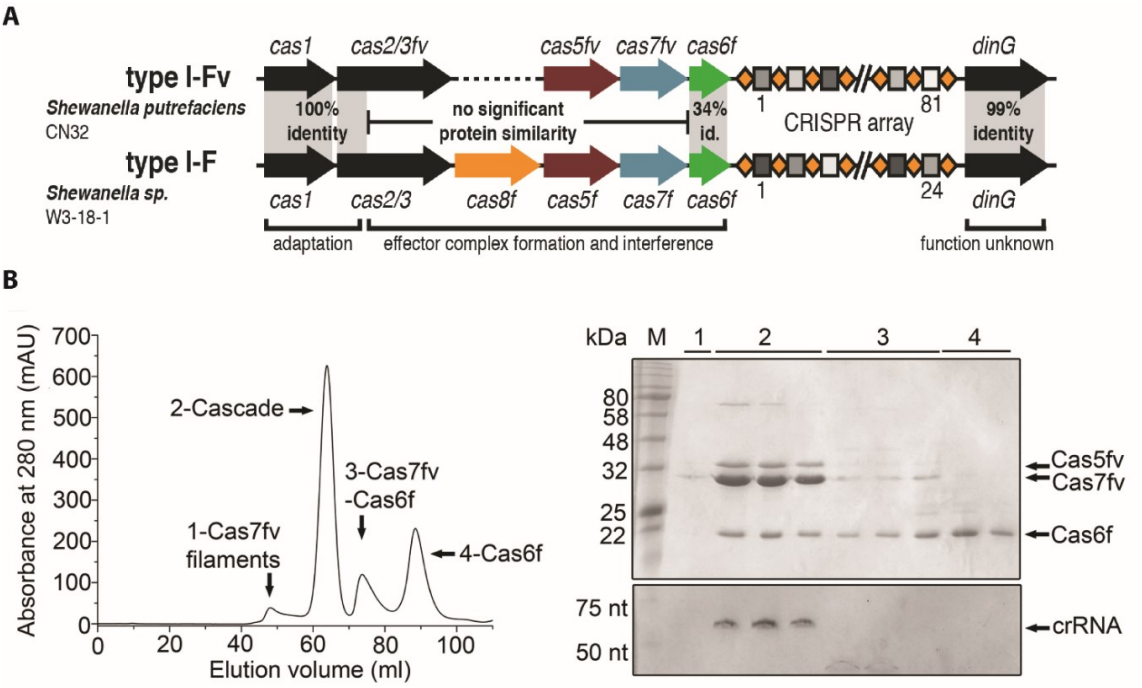


Figure I.2: Representation of the minimal Type I-Fv CRISPR-Cas module and its Cascade components. (A) Comparison between Type I-F and Type I-Fv *S. putrefaciens* CRISPR-Cas modules, evidencing the absence of a large subunit (Cse1) in the Type I-Fv system (modified from (Dwarakanath et al., 2015)). (B) Purification of recombinant Type I-Fv Cascade via size-exclusion chromatography using a His-tagged version of Cas6f (left). Protein SDS-PAGE (top)/UreaPAGE (bottom) of the elution fractions was used to identify Cascade components (right) (Gleditzsch et al., 2016).

This reduced system was initially found in the genome of *Photobacterium profundum* SS9, a bacterium that also carries a plasmid-encoded Type I-E system. Other organisms carrying a Type I-Fv system are found among the gamma- and betaproteobacteria, with several of them clustering in the Alteromonadales and Vibrionales orders (Makarova et al., 2015). *S. putrefaciens* CN-32 carries only the minimal system, with a CRISPR array containing 81 spacers, which suggests that the system is highly active.

Type I-Fv Cascade was shown to interfere with conjugative plasmids *in vivo* in *S. putrefaciens,* when plasmids carried a sequence with complementarity to the crRNA and a flanking GG PAM; a motif identified by inverse search of the spacer sequences on viral databases. This proves that the minimal complex is functional and shares some target requirements with closely related systems (Dwarakanath et al., 2015). As the minimal system lacks a large Cascade subunit, described as essential for interference in other Type I systems, the mechanism behind target recognition could be drastically different.

The aim of the present work was to further characterise this reduced CRISPR-Cas system, with a focus on its ability to target and degrade foreign DNA. In order to fulfil this aim, we used a broad set of techniques ranging from *in vitro* to *in vivo* approaches and described two new methods to study the target interactions of Cascades.

First, we used BioLayer Interferometry (BLI), an *in vitro* technique that allows to measure, in real time and in solution, the binding of CRISPR-Cas complexes to DNA. BLI relies on the immobilization of one of the interactors to a biolayer, through which white light incides and gets reflected. The reflected beam is proportional to the thickness of the biolayer. Therefore, as a second player binds to the immobilized bait, the changes in the density of the biolayer can be detected, resulting in a specific sensogram for the interaction. With this, it is then possible to determine the affinity of complexes for different targets and also the kinetics of the binding (Sultana & Lee, 2015).

Second, we applied Single Particle Tracking Photo-activated Localization Microscopy (sptPALM) to study the target search dynamics of Cascade complexes in live cells. PALM is a super-resolution imaging method that relies on photo-switchable fluorescent proteins to surpass the diffraction limit of light. By stochastically switching the fluorescent particles to a detectable state, it is possible to temporally separate the detection of two closely-located molecules, to later integrate all signals and compose a final image. Furthermore, following the trajectories of the fluorescently-tagged proteins or complexes by SPT allows to determine the speed at which they move, giving insights into the behaviour of the population and helping to predict interactions (Turkowyd, Virant, & Endesfelder, 2016).

These and other methods were used to characterize several aspects of the target interaction by the minimal Cascade. Particularly, we investigated how the minimal Cascade performs PAM recognition in the absence of a large subunit. Furthermore, we analysed how this surveillance complex is able to form and maintain the R-loop structure,

and sought to define its substrate specificity and affinity. In addition, the effect of Acr proteins on the establishment of immunity by the Type I-Fv Cascade was studied. Finally, we worked towards elucidating the different DNA scanning and binding states of the complex, as well as determining the dynamics of these states *in vivo*.

References

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., et al. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science, 315*(5819), 1709-1712.

Bondy-Denomy, J., Pawluk, A., Maxwell, K. L., & Davidson, A. R. (2013). Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature, 493*(7432), 429-432.

Borges, A. L., Davidson, A. R., & Bondy-Denomy, J. (2017). The Discovery, Mechanisms, and Evolutionary Impact of Anti-CRISPRs. *Annu Rev Virol, 4*(1), 37-59.

Brouns, S. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J., Snijders, A. P., et al. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science, 321*(5891), 960-964.

Cady, K. C., Bondy-Denomy, J., Heussler, G. E., Davidson, A. R., & O'Toole, G. A. (2012). The CRISPR/Cas adaptive immune system of Pseudomonas aeruginosa mediates resistance to naturally occurring and engineered phages. *J Bacteriol, 194*(21), 5728-5738.

Chowdhury, S., Carter, J., Rollins, M. F., Golden, S. M., Jackson, R. N., Hoffmann, C., et al. (2017). Structure Reveals Mechanisms of Viral Suppressors that Intercept a CRISPR RNA-Guided Surveillance Complex. *Cell, 169*(1), 47-57 e11.

Darmon, E., & Leach, D. R. (2014). Bacterial genome instability. *Microbiol Mol Biol Rev, 78*(1), 1-39.

Dwarakanath, S., Brenzinger, S., Gleditzsch, D., Plagens, A., Klingl, A., Thormann, K., et al. (2015). Interference activity of a minimal Type I CRISPR-Cas system from Shewanella putrefaciens. *Nucleic Acids Res, 43*(18), 8913-8923.

Elmore, J. R., Sheppard, N. F., Ramia, N., Deighan, T., Li, H., Terns, R. M., et al. (2016). Bipartite recognition of target RNAs activates DNA cleavage by the Type III-B CRISPR-Cas system. *Genes Dev, 30*(4), 447-459.

Estrella, M. A., Kuo, F. T., & Bailey, S. (2016). RNA-activated DNA cleavage by the Type III-B CRISPR-Cas effector complex. *Genes Dev, 30*(4), 460-470.

Gleditzsch, D., Muller-Esparza, H., Pausch, P., Sharma, K., Dwarakanath, S., Urlaub, H., et al. (2016). Modulating the Cascade architecture of a minimal Type I-F CRISPR-Cas system. *Nucleic Acids Res, 44*(12), 5872-5882.

Grissa, I., Vergnaud, G., & Pourcel, C. (2007). CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res, 35*(Web Server issue), W52-57.

Guo, T. W., Bartesaghi, A., Yang, H., Falconieri, V., Rao, P., Merk, A., et al. (2017). Cryo-EM Structures Reveal Mechanism and Inhibition of DNA Targeting by a CRISPR-Cas Surveillance Complex. *Cell, 171*(2), 414-426 e412.

Haurwitz, R. E., Jinek, M., Wiedenheft, B., Zhou, K., & Doudna, J. A. (2010). Sequence- and Structure-Specific RNA Processing by a CRISPR Endonuclease. *Science, 329*(5997), 1355-1358.

Huo, Y., Nam, K. H., Ding, F., Lee, H., Wu, L., Xiao, Y., et al. (2014). Structures of CRISPR Cas3 offer mechanistic insights into Cascade-activated DNA unwinding and degradation. *Nat Struct Mol Biol, 21*(9), 771-777.

Jia, N., Mo, C. Y., Wang, C., Eng, E. T., Marraffini, L. A., & Patel, D. J. (2019). Type III-A CRISPR-Cas Csm Complexes: Assembly, Periodic RNA Cleavage, DNase Activity Regulation, and Autoimmunity. *Mol Cell, 73*(2), 264-277 e265.

Kazlauskiene, M., Tamulaitis, G., Kostiuk, G., Venclovas, C., & Siksnys, V. (2016). Spatiotemporal Control of Type III-A CRISPR-Cas Immunity: Coupling DNA Degradation with the Target RNA Recognition. *Mol Cell, 62*(2), 295-306.

Koonin, E. V., & Makarova, K. S. (2019). Origins and evolution of CRISPR-Cas systems. *Philos Trans R Soc Lond B Biol Sci, 374*(1772), 20180087.

Koonin, E. V., Makarova, K. S., & Zhang, F. (2017). Diversity, classification and evolution of CRISPR-Cas systems. *Curr Opin Microbiol, 37*, 67-78.

Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J., Charpentier, E., Horvath, P., et al. (2011). Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol, 9*(6), 467-477.

Makarova, K. S., Wolf, Y. I., Alkhnbashi, O. S., Costa, F., Shah, S. A., Saunders, S. J., et al. (2015). An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol, 13*(11), 722-736.

Makarova, K. S., Wolf, Y. I., & Koonin, E. V. (2013). Comparative genomics of defense systems in archaea and bacteria. *Nucleic Acids Res, 41*(8), 4360-4377.

Marino, N. D., Zhang, J. Y., Borges, A. L., Sousa, A. A., Leon, L. M., Rauch, B. J., et al. (2018). Discovery of widespread type I and type V CRISPR-Cas inhibitors. *Science, 362*(6411), 240-242.

Maslov, S., & Sneppen, K. (2017). Population cycles and species diversity in dynamic Kill-the-Winner model of microbial ecosystems. *Sci Rep, 7*, 39642.

Mojica, F. J. M., Díez-Villaseñor, C. s., García-Martínez, J., & Soria, E. (2005). Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements. [journal article]. *Journal of Molecular Evolution, 60*(2), 174-182.

Nunez, J. K., Kranzusch, P. J., Noeske, J., Wright, A. V., Davies, C. W., & Doudna, J. A. (2014). Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. *Nat Struct Mol Biol, 21*(6), 528-534.

Ozcan, A., Pausch, P., Linden, A., Wulf, A., Schuhle, K., Heider, J., et al. (2019). Type IV CRISPR RNA processing and effector complex formation in Aromatoleum aromaticum. *Nat Microbiol, 4*(1), 89-96.

Paterson, S., Vogwill, T., Buckling, A., Benmayor, R., Spiers, A. J., Thomson, N. R., et al. (2010). Antagonistic coevolution accelerates molecular evolution. *Nature, 464*(7286), 275-278.

Pawluk, A., Bondy-Denomy, J., Cheung, V. H., Maxwell, K. L., & Davidson, A. R. (2014). A new group of phage anti-CRISPR genes inhibits the type I-E CRISPR-Cas system of Pseudomonas aeruginosa. *MBio, 5*(2), e00896.

Pawluk, A., Staals, R. H., Taylor, C., Watson, B. N., Saha, S., Fineran, P. C., et al. (2016). Inactivation of CRISPR-Cas systems by anti-CRISPR proteins in diverse bacterial species. *Nat Microbiol, 1*(8), 16085.

Plagens, A., Richter, H., Charpentier, E., & Randau, L. (2015). DNA and RNA interference mechanisms by CRISPR-Cas surveillance complexes. *FEMS Microbiol Rev, 39*(3), 442-463.

Rauch, B. J., Silvis, M. R., Hultquist, J. F., Waters, C. S., McGregor, M. J., Krogan, N. J., et al. (2017). Inhibition of CRISPR-Cas9 with Bacteriophage Proteins. *Cell, 168*(1-2), 150-158 e110.

Redding, S., Sternberg, S. H., Marshall, M., Gibb, B., Bhat, P., Guegler, C. K., et al. (2015). Surveillance and Processing of Foreign DNA by the Escherichia coli CRISPR-Cas System. *Cell, 163*(4), 854-865.

Rollins, M. F., Chowdhury, S., Carter, J., Golden, S. M., Miettinen, H. M., Santiago-Frangos, A., et al. (2019). Structure Reveals a Mechanism of CRISPR-RNA-Guided Nuclease Recruitment and Anti-CRISPR Viral Mimicry. *Mol Cell, 74*(1), 132-142 e135.

Rollins, M. F., Chowdhury, S., Carter, J., Golden, S. M., Wilkinson, R. A., Bondy-Denomy, J., et al. (2017). Cas1 and the Csy complex are opposing regulators of Cas2/3 nuclease activity. *Proc Natl Acad Sci U S A, 114*(26), E5113-E5121.

Samson, J. E., Magadan, A. H., Sabri, M., & Moineau, S. (2013). Revenge of the phages: defeating bacterial defences. *Nat Rev Microbiol, 11*(10), 675-687.

Stenseth, N. C., & Smith, J. M. (1984). Coevolution in Ecosystems: Red Queen Evolution or Stasis? *Evolution, 38*(4), 870-880.
Sultana, A., & Lee, J. E. (2015). Measuring protein-protein and protein-nucleic Acid interactions by biolayer interferometry. *Curr Protoc Protein Sci, 79*, 19 25 11-26.
Turkowyd, B., Virant, D., & Endesfelder, U. (2016). From single molecules to life: microscopy at the nanoscale. *Anal Bioanal Chem, 408*(25), 6885-6911.
Watters, K. E., Fellman, C., Bai, H. B., Ren, S. M., & Doudna, J. A. (2018). Systematic discovery of natural CRISPR-Cas12a inhibitors. *Science, 362*(6411), 236-239.
Wiedenheft, B., van Duijn, E., Bultema, J. B., Waghmare, S. P., Zhou, K., Barendregt, A., et al. (2011). RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc Natl Acad Sci U S A, 108*(25), 10092-10097.
Xue, C., Whitis, N. R., & Sashital, D. G. (2016). Conformational Control of Cascade Interference and Priming Activities in CRISPR Immunity. *Mol Cell, 64*(4), 826-834.
You, L., Ma, J., Wang, J., Artamonova, D., Wang, M., Liu, L., et al. (2019). Structure Studies of the CRISPR-Csm Complex Reveal Mechanism of Co-transcriptional Interference. *Cell, 176*(1-2), 239-253 e216.

Chapter II:
Modulating the Cascade architecture of a minimal Type I-F CRISPR-Cas system

Daniel Gleditzsch[1], Hanna Müller-Esparza[1], Patrick Pausch[2,3], Kundan Sharma[4], Srivatsa Dwarakanath[1], Henning Urlaub[4,5], Gert Bange[2,3], and Lennart Randau[1,2,*]

This chapter is written in manuscript style and was published in Nucleic Acids Research in July, 2016. My contribution to this work included designing and performing the experiments *in vivo*, analyzing the obtained data and writing the corresponding parts of the manuscript.

---

[1] Prokaryotic Small RNA Biology Group, Max Planck Institute for Terrestrial Microbiology, D-35043 Marburg, Germany
[2] LOEWE Center for Synthetic Microbiology, Philipps University Marburg, D-35043 Marburg, Germany
[3] Department of Chemistry, Philipps University Marburg, D-35043 Marburg, Germany
[4] Bioanalytics Research Group, Department of Clinical Chemistry, University Medical Centre, D-37075 Göttingen, Germany
[5] Bioanalytical Mass Spectrometry Group, Max Planck Institute for Biophysical Chemistry, D-37077 Göttingen, Germany
* Corresponding Author

Abstract

Shewanella putrefaciens CN-32 contains a single Type I-Fv CRISPR-Cas system which confers adaptive immunity against bacter iophage infection. Three Cas proteins (Cas6f, Cas7fv, Cas5fv) and mature CRISPR RNAs were shown to be required for the assembly of an interference complex termed Cascade. The Cas protein-CRISPR RNA interaction sites within this complex were identified via mass spectrometry. Additional Cas proteins, commonly described as large and small subunits, that are present in all other investigated Cascade structures, were not detected. We introduced this minimal Type I system in *Escherichia coli* and show that it provides heterologous protection against lambda phage. The absence of a large subunit suggests that the length of the crRNA might not be fixed and recombinant Cascade complexes with drastically shortened and elongated crRNAs were engineered. Size-exclusion chromatography and small-angle X-ray scattering analyses revealed that the number of Cas7fv backbone subunits is adjusted in these shortened and extended Cascade variants. Larger Cascade complexes can still confer immunity against lambda phage infection in *E. coli*. Minimized Type I CRISPR-Cas systems expand our understanding of the evolution of Cascade assembly and diversity. Their adjustable crRNA length opens the possibility for customizing target DNA specificity.

Introduction

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)-Cas (CRISPR-associated) modules are adaptive immune systems found in archaea and bacteria. CRISPR arrays consist of short repeat sequences that flank unique spacer sequences, which can be derived from viral genomes and conjugative plasmids (Grissa, Vergnaud, & Pourcel, 2009; Mojica, García-Martínez, & Soria, 2005); Pourcel, Salvignol, and Vergnaud (2005). These arrays are transcribed into long precursor molecules that are further processed into multiple small crRNAs (CRISPR RNAs). Mature crRNAs contain a single spacer sequence and parts of the repeat sequences at their termini (Carte, Pfister, Compton, Terns, & Terns, 2010; Gesner, Schellenberg, Garside, George, & MacMillan, 2011; Rachel E Haurwitz, Jinek, Wiedenheft, Zhou, & Doudna, 2010; H. Richter et al., 2012; Semenova et al., 2011). The acquisition of new genomic spacers requires cleavage of the foreign DNA, which is mediated by the nearly universal Cas proteins Cas1 and Cas2 (Nuñez et al., 2014; Nuñez, Lee, Engelman, & Doudna, 2015; Corinna Richter et al., 2014). Recognition of the source DNA, i.e. the protospacer, depends on the presence of a short conserved 2-5 base pairs (bp) sequence, termed protospacer adjacent motif (PAM). The crRNAs are bound by Cas proteins to form a CRISPR ribonucleoprotein (crRNP) surveillance complex. These complexes utilize crRNA spacers as guides to recognize foreign genetic material, e.g. during a recurring viral infection (Brouns et al., 2008; André Plagens et al., 2014; Rouillon et al., 2013; Sapranauskas et al., 2011; Semenova et al., 2011; van der Oost, Westra, Jackson, & Wiedenheft, 2014). Target DNA recognition and subsequent degradation requires the presence of PAM sequences and Watson-Crick base pairing between the crRNA and the unwound target strand (Jackson et al., 2014; Maier et al., 2013; Sabin Mulepati, Héroux, & Bailey, 2014).

CRISPR-Cas systems are diverse and were divided into three main types, which are defined by their conserved signature proteins Cas3 (Type I), Cas9 (Type II) and Cas10 (Type III) (Kira S Makarova, Aravind, Wolf, & Koonin, 2011; Kira S Makarova, Haft, et al., 2011). DNA interference in Type I CRISPR-Cas systems is mediated by a multi-protein crRNP complex termed Cascade (CRISPR associated complex for antiviral defense) (Brouns et al., 2008; Garneau et al., 2010; Jinek et al., 2012), while Type II systems utilize the single nuclease Cas9 and a trans-activating RNA for target recognition and degradation (Chylinski, Le Rhun, & Charpentier, 2013; Jinek et al., 2012; Jinek et al., 2014). Both Types employ PAM-dependent target recognition mechanisms (Anders, Niewoehner, Duerst, & Jinek, 2014; Maier et al., 2013; Westra et al., 2013).

Type III systems generate crRNP complexes (termed Csm or Cmr) which are able to target ssRNA in a PAM-independent manner (Rouillon et al., 2013; Staples et al., 2014). Recently, two additional CRISPR-Cas Types have been classified (K. S. Makarova et al., 2015). Type IV systems lack Cas1 and Cas2 and the Type V and Type VI signature proteins function as stand-alone DNA nucleases (Shmakov et al., 2015).

Type I CRISPR-Cas systems are most widespread in nature and the Cas protein composition of the employed Cascade interference complexes differs between seven subtypes (A-F,U) (K. S. Makarova et al., 2015). Most insights were obtained for the Type I-E Cascade complex from *Escherichia coli* and several crystal structures are available (Hayes et al., 2016; Jackson et al., 2014; Sabin Mulepati et al., 2014). The Type I-E Cascade crRNP consists of the five Cas proteins Cas5, Cas6, Cas7, Cse1 and Cse2 and a 61 nucleotide (nt) mature crRNA (Jackson et al., 2014; Sabin Mulepati et al., 2014; Wiedenheft et al., 2011). The crescent-shaped structure of this complex has a molecular mass of 405 kDa and displays an uneven stoichiometry: $(Cse1)_1$-$(Cse2)_2$-$(Cas5)_1$-$(Cas7)_6$-$(Cas6e)_1$ (Semenova et al., 2011). The 61 nt mature crRNA contains a 32 nt spacer sequence that is flanked by a 8 nt long handle-region at the 5′-end and a 21 nt long repeat sequence with a terminal hairpin at 3′-end. The crRNA is generated by the RNA endonuclease Cas6. After processing, Cas6 stays tightly associated with the 3′-end, while Cas5 caps the crRNA's 5′-terminal repeat tag (Niewoehner, Jinek, & Doudna, 2014; Semenova et al., 2011). The helical backbone of Type I-E Cascade is formed by six copies of Cas7, which generates a groove to bind and protect the crRNA. The large subunit, Cse1, is responsible for PAM recognition and is involved in recruitment of the target nuclease Cas3 (Hayes et al., 2016; Sashital, Wiedenheft, & Doudna, 2012; Semenova et al., 2011; Wiedenheft et al., 2011). A dimer of two small subunits, Cse2, stabilizes the formation of an R-loop structure and binds the displaced DNA strand (Jackson et al., 2014; Sabin Mulepati et al., 2014; Semenova et al., 2011). The general architecture of Cascade crRNPs appears to be conserved in Type I systems, even though their Cas protein composition can differ (Judith, James, & Malcolm, 2013).

One example of a strikingly altered minimalistic Cascade architecture was recently described as a Type I-F variant (Type I-Fv). This subtype was initially identified in several beta- and gamma-proteobacteria and suggested to rely on a minimal set of five Cas proteins (Kira S Makarova, Aravind, et al., 2011). Type I-Fv systems contain Cas1, the integrase that mediates spacer acquisition (Nuñez et al., 2015; C. Richter, Gristwood, Clulow, & Fineran, 2012), Cas3, the target nuclease and Cas6f, the crRNA endonuclease (Rachel E Haurwitz et al., 2010). The two additional Cas proteins showed no apparent sequence similarity to known Cas protein families and small and large Cascade subunits

were not detected (Kira S Makarova, Aravind, et al., 2011). Our first characterization of this Type I-Fv CRISPR-Cas system revealed that the minimal Cascade can target conjugative plasmids in *Shewanella putrefaciens* CN-32 (Dwarakanath et al., 2015). Recombinant Cascade was found to contain the Cas6f endonuclease, the backbone-forming protein Sputcn32_1821 and the protein of unknown function Sputcn32_1822 (Dwarakanath et al., 2015). With support of our current knowledge on the roles of these two proteins and in analogy to the *E.coli* Cascade nomenclature, Sputcn32_1821 was renamed as Cas7fv (Cas7 backbone protein of the Type I-F variant CRISPR-Cas) and Sputcn32_1822 was renamed as Cas5fv (Cas5 protein of the Type I-F variant CRISPR-Cas). However, it remained unclear whether additional host proteins substitute for the missing small and large subunits. In the present study, we addressed this question and transferred the *S. putrefaciens* Cascade into *E. coli*. Heterologous interference activity against lambda phage infections was observed. The minimal architecture of this active Cascade system allowed for the modulation of the crRNA's spacer length and synthetic Cascade assemblies with elongated and shortened backbones were created. In addition, requirements for proper Cascade assembly were screened. Surprisingly, some of the synthetic Cascade complexes can retain DNA interference activity. These results shed light on the evolution of minimalist Cascade complexes without small and large subunits. The observed possibility of altering the spacer length highlights a potential for the modulation of the DNA target specificity.

Materials and Methods

Bacterial strains and growth conditions

E. coli BL21 (DE3) cells (Novagen) were used for the production of recombinant proteins and cultures were grown in either LB or NZ-amine (1% NZ-amine, 0.5% yeast extract and 1% NaCl) media with respective antibiotics until an OD_{600} of ~0.6 was reached. Gene expression was induced by addition of 1 mM IPTG and cultures were grown overnight at 18 °C.

E. coli BL21-AI strains (F- ompT hsdSB (rB-mB-) gal dcm araB::T7RNAPtetA, Invitrogen) were used for phage assays and grown in 2YTL media (16 g/L tryptone, 10 g/L yeast extract, 5 g/L NaCl, 10 mM $MgSO_4$, 0.2 % maltose) supplemented with appropriate antibiotics at 37°C. A virulent variant of lambda phage (NCCB 3467) was obtained from Centraalbureau voor Schimmelcultures (Utrecht, Netherlands).

Production and purification of recombinant proteins

The gene cassette cas7fv, cas5fv, cas6f was cloned into pRSFDuet-1 which allows for the simultaneous production of all three proteins with Cas7fv having an N-terminal His-tag fusion. The *cas* gene containing plasmids (pCas) were recloned to obtain variants without Cas5fv (pCas3) or with a His-tag fusion at Cas6f (pCas2). The pCas plasmid variants were co-transformed into *E. coli* with a second pUC19 vector containing the repeat-spacer4-repeat sequence of the single *S. putrefaciens* CN-32 CRISPR array downstream of a T7 RNA polymerase promoter. The CRISPR sequences were generated by cloning annealed oligonucleotides into the pUC19 vector using BamHI and HindIII restriction sites. Plasmids harboring extended (+18/+15) or shortened (-18/-15) spacer sequences contained either additional random nucleotides at the 5′-end of the spacer sequence or deleted nucleotides at the 3′-end of the spacer sequence. All sequences are listed in Table S.II.1. Cas protein production and the purification of recombinant Cascade variants via Ni-NTA chromatography and size exclusion chromatography (Superdex 200) in a buffer containing 50 mM HEPES-NaOH pH 7.0, 150 mM NaCl, 1 mM DTT and 1 mM EDTA was performed as described (Dwarakanath et al., 2015).

Identification of protein-RNA interactions via mass-spectrometry analysis

Protein-RNA crosslinking was performed using UV irradiation at 254 nm followed by the enrichment of cross-linked peptide–RNA heteroconjugates as described in (Kramer et al., 2014; Sharma et al., 2015). Around 1 nmol of the Type I-Fv Cascade complex was resuspended in 100 µl of 50 mM HEPES (pH 7.0), 300 mM NaCl, 1 mM EDTA and 1 mM DTT. The complex was incubated at 4°C for 30 min. The samples were then transferred to black polypropylene microplates (Greiner Bio-One) and irradiated at 254 nm for 10 minutes on ice as described in (Sharma et al., 2015). The samples were ethanol precipitated and the pellet was dissolved in 4M Urea and 50 mM Tris-HCl pH 7.9. The final concentration of urea was then adjusted to 1 M with 50 mM Tris-HCl pH 7.9 and the RNA was hydrolysed using 1 µg RNase A and T1 (Ambion, Applied Biosystems) for 2 hours at 52 °C followed by digestion with Benzonase at 37 °C for 1 hr. The proteolysis was carried out using trypsin (Promega) at 37°C, overnight. The sample was desalted to remove non cross-linked RNA fragments using an in-house prepared C18 (Dr. Maisch GmbH) column and the cross-linked peptides were enriched on an in-house prepared TiO_2 (GL Sciences) column using the protocol described in (Sharma et al., 2015). The samples were then dried and resuspended in 5% v/v ACN, 1% v/v FA for mass spectrometry analysis. The sample was injected onto a nano-liquid chromatography system (Dionex, Ultimate 3000, Thermo Fisher Scientific) coupled with Q-Exactive HF

instrument (Thermo Fisher Scientific) as described in (Kirli et al., 2015). Online ESI-MS was performed in data-dependent mode using a TOP20 HCD method. All precursor ions as well as fragment ions were scanned in the Orbitrap, and the resulting spectra were measured with high accuracy (< 5 ppm) both in the MS and MS/MS level. Data analysis was done using a dedicated data base search tool as described in (Kramer et al., 2014).

Small-angle X-ray scattering

Recombinant Cascade complexes for SAXS analysis were produced and purified as described above, snap frozen in liquid nitrogen and stored at -80 °C. SEC buffer aliquots of the individual purifications were kept and treated equally for sample dilution and solvent blanking. SAXS data were collected at the European Synchrotron Radiation Facility (Grenoble, France) on beamline BM29. Detailed SAXS statistics are listed in Table S.II.2. Dilution series of the complexes were measured to assess sample quality in respect to inter-particle interference and aggregation. Measurements exhibiting radiation damage were discarded. Buffer subtracted SAXS curves of 10 mg ml^{-1} for the short variant, 20 mg ml^{-1} for the wild type complex and 25 mg ml^{-1} for the elongated version were subjected to data analysis and *ab initio* shape restoration bead modeling using the Primus implemented program Dammif (Konarev, Volkov, Sokolova, Koch, & Svergun, 2003). 20 independent shape restorations were calculated in slow mode to generate average and filtered models. SAXS data, one individual *ab initio* shape restoration model for each construct and the averaged and filtered models were deposited at the curated repository SASBDB under accession SASDBK4 (short Cascade), SASDBL4 (wild-type Cascade) and SASDBM4 (long Cascade complex). The volume of the Cas7fv protein was estimated using the peptide property calculator tool (Northwestern University) based on (Harpaz, Gerstein, & Chothia, 1994).

Construction of *E. coli* strains with heterologous Type I-Fv CRISPR-Cas systems

A minimal CRISPR was designed with a single spacer targeting the positions 6463-6494 (GA PAM) or 6630-6662 (GG PAM) of the lambda phage gene E, flanked by two repeat sequences and a T7 RNA polymerase promoter (pCRISPRλGA and pCRISPRλ, Table S.II.1). The CRISPR sequence was generated by cloning annealed oligonucleotides into pCDFDuet-1 vector using BamHI and HindIII restriction sites. In addition, a second CRISPR with a spacer corresponding to a random 32 bp sequence of the *S. putrefaciens* genome was generated as a negative control (pCRISPR NT, Table S.II.1). CRISPR variants were generated by adding 1, 6, 9, 12 or 18 nt (complementary to the phage genome) to the targeting spacer; or by removing 6 or 18 nt (pCRISPRλ variants). The spacer sequences have no significant complementarity to the *E. coli* genome. A mutation

in the HD nuclease motif of Cas3 (H156A/D157A) was introduced into pCas6 by QuikChange site-directed mutagenesis according to the manufacturer's protocol (pCas7). *E. coli* BL21-AI cells were transformed with pCDFDuet-1 carrying one variant of the CRISPR array and pCas6 or pCas7. After transformation, single colonies were picked for each strain, grown as a lawn for 12 h in 2YTL agar and colony-forming cells were isolated for phage assays. In parallel, a strain with empty pRSFDuet-1 and pCDFDuet-1 vectors was generated as a control.

Lambda phage-assays for DNA interference analyses

Phage assays were essentially performed as described (Brouns et al., 2008). Briefly, a 1/100 dilution of an *E. coli* BL21-AI overnight culture was grown until OD_{600} nm = 0.3 and induced with 0.2% arabinose (Sigma-Aldrich) and 0.1 mM IPTG for 30 min. Then, cells were pelleted and resuspended in 10 mM $MgSO_4$. A 1:1 mix of cells and serial phage dilutions was incubated for 20 min at 37°C and added to 2YTL top-agar containing the mentioned inducers, plated over 2YTL agar plates with antibiotics and incubated overnight at 37°C. Plates were then stained with 0.2% crystal violet and plaque number was determined. Efficiency of plaquing (EOP) was defined as the ratio between the plaque count of the strain of interest and the plaque count of the strain carrying empty plasmids. Phage assays were performed in triplicate and error bars were calculated as standard error of the mean (SEM).

Electrophoretic mobility shift assays

The recombinant Cascade complex was tested for its ability to bind target DNA in electrophoretic mobility shift assays (EMSAs). Utilized target oligonucleotides are detailed in Table S.II.1. 100 pmol of each substrate was 5′-labeled with [γ-^{32}P]-ATP (5000Ci/mmol, Hartmann Analytic) and T4 PNK (Ambion) for 1 h at 37°C. The reaction was stopped by addition of formamide loading buffer and substrates were separated by denaturing-PAGE (10% polyacrylamide, 8 M Urea). After autoradiographic exposure, RNA bands were cut from the gel, eluted and EtOH precipitated. 20 nM (~20000 cpm) of labeled substrate was incubated with varying concentrations of Cascade (0-60 nM) for 30 min at 30°C in 50 mM HEPES-NaOH pH 7.0, 150 mM NaCl, 1 mM DTT and 1 mM EDTA. Samples were separated via non-denaturing TBE-PAGE (6% polyacrylamide, 1x TBE).

Results

Recombinant production of the minimal *S. putrefaciens* CN-32 Type I-Fv Cascade

In our earlier study, we demonstrated that recombinant Type I-Fv Cascade can be produced from His-tagged Cas7fv, Cas5fv, Cas6f and crRNA in *E.coli* (Dwarakanath et al., 2015). We improved this purification procedure and also shifted the location of the His-tag from the N-terminus of Cas7fv to the C-terminus of Cas6f to be able to evaluate the Cas5fv:Cas7fv protein ratio (Figure II.1A) Affinity chromatography of Cas6f was successfully used to co-purify mature 60 nt long crRNA and the untagged Cas5fv and Cas7fv proteins. The eluted complex was subjected to size-exclusion chromatography, which revealed four distinct peaks (Figure II.1B) that were investigated via SDS-PAGE (Figure II.1B). A small first peak contains Cas7fv filaments that were shown to form due to the unspecific binding of RNA molecules. The largest second peak contains the fully assembled Cascade ribonucleoprotein complex. This verifies that the pre-crRNA was produced, cleaved by Cas6f and incorporated into Cascade. An increased abundance of the backbone protein Cas7fv was evident. Two smaller peaks contained either Cas6f-Cas7fv complexes or Cas6f alone. Mature crRNA was not detected in the corresponding fractions. EMSAs verified that the recombinant Cascade complex specifically binds complementary DNA target-strands (Supplementary Figure S.II.1).

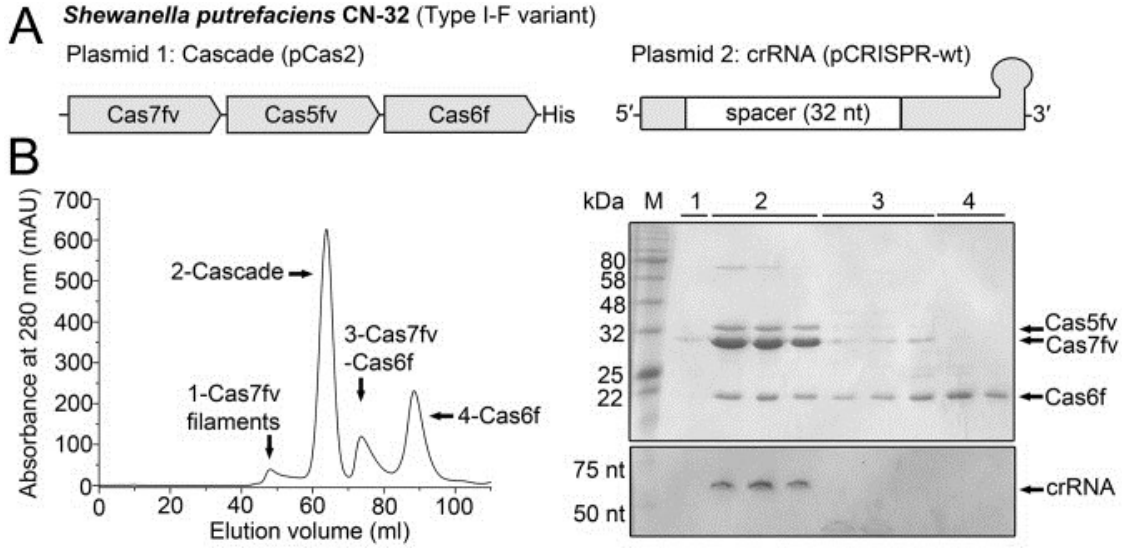


Figure II.1. Purification of recombinant Type I-Fv Cascade interacting with His-Cas6f. (**A**) Schematic overview of the investigated recombinant Cas proteins and crRNA. (**B**) His-tagged Cas6f co-eluted with the Cascade components Cas7fv, Cas5fv and mature crRNA during size-exclusion chromatography (peak 2). Cas7fv filaments were observed in the void volume (peak 1) and Cas7fv-Cas6f complexes (peak 3) and Cas6f (peak 4) were separated. SDS-PAGE (top) and 8 M urea PAGE with toluidine blue staining (bottom) was used to analyse the protein and RNA content of the fractions corresponding to the peaks indicated in the gel-elution chromatogram (left).

The presence of Cas5fv and crRNA is required for proper Cascade assembly

Next, we asked which components are required for proper Type I-Fv Cascade assembly and investigated if a Cas-protein complex forms in the absence of crRNA. The three recombinant proteins Cas6f, Cas5fv and Cas7fv with an N-terminal His-tag were produced in *E. coli* and subjected to affinity- and subsequent size-exclusion chromatography. Two distinct peaks were obtained which were analyzed via SDS-PAGE. The first peak fractions contain the previously observed Cas7fv filaments and the largest second peak corresponds to Cas5fv-Cas7fv dimers (Fig. II.2A). Multimeric Cascade-like structures were not observed, which indicates that the Cas5fv-Cas7fv might be Cascade building blocks that require mature crRNAs as assembly signals. It is evident that Cas6f is necessary to process the pre-crRNA transcripts and Cas7fv is required to form the filamentous Cascade backbone (Dwarakanath et al., 2015). However, the role of Cas5fv during Cascade assembly is not known. Thus, we constructed a plasmid lacking the gene encoding for Cas5fv and attempted to produce recombinant Cascade structures. It was shown that Cascade cannot be reconstituted in the absence of Cas5fv (Fig. S.II.2). Precipitation of Cas7fv was observed which suggests that it requires the interaction partner Cas5fv for stability. The small and indistinct peaks revealed Cas7fv-Cas6f interactions. Finally, we asked if the proteins Cas1 and Cas2-3 are stable subunits of the minimal Cascade. A larger plasmid was constructed that contains genes encoding all five Type I-Fv Cas proteins. The construct contained only one His-tag at the N-terminus of Cas1. Affinity chromatography revealed that Cas1 was co-purified with Cas2-3 and Cas6f (Fig. S.II.3). However, crRNA was not present in the elution fractions and Cas5fv-Cas7fv-Cas6f Cascade-complexes were eluted in the flow-through. Taken together, these results suggest that the minimal DNA surveillance complexes are formed by three Cas proteins interacting with a second Cas1-Cas2-3 complex during target DNA degradation.

Analysis of Cascade protein–RNA interactions via mass-spectrometry

In order to pinpoint specific amino acids and nucleotides that interact within the minimal Type I-Fv Cascade complex, we utilized UV-induced cross-linking and mass spectrometry. Upon UV irradiation at 254 nm a covalent bond forms between the amino acid side chain and nucleic acid bases when both components are in close spatial proximity. The cross-linked peptide-RNA heteroconjugates can then be identified via LC-MS/MS analysis and a dedicated database search (Kramer et al., 2014; Sharma et al., 2015).

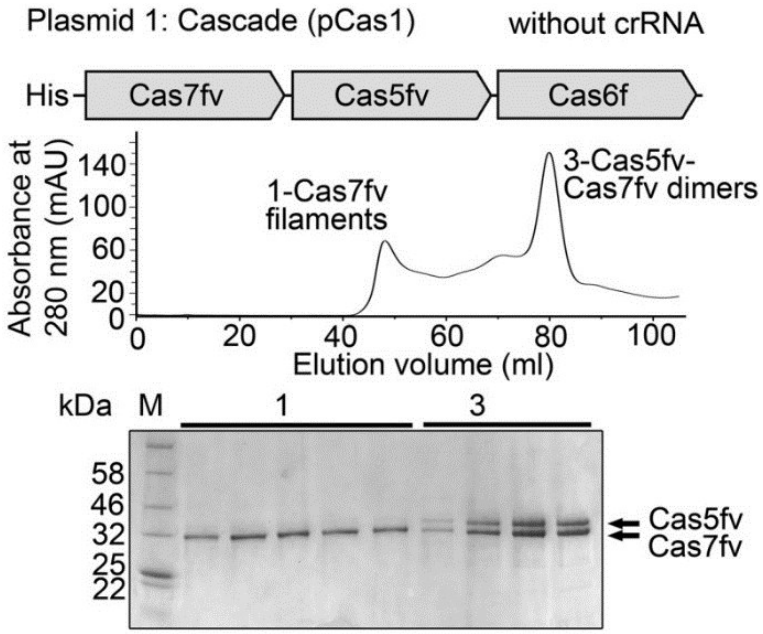


Figure II.2. Size-exclusion chromatography analysis of Cas proteins in the absence of crRNA. The Cas proteins were produced and purified via size-exclusion chromatography (see Figures II.1 and II.4) in the absence of crRNA. Cas7fv filaments (peak 1) and Cas5fv-Cas7fv dimers (peak 3) were observed but the characteristic middle peak (peak 2, elution volume 60–70 ml) was absent indicating the lack of Cascade backbone formation.

This approach allowed for the identification of thirteen peptides with different cross-links between amino acids and RNA fragment, which provides first insights into RNA recognition sites of the Cas proteins (Fig. S.II.4-8). The analyses confirmed extensive interactions of all three Cascade subunits with the crRNA (Fig. II.3). Most interactions were identified for Cas7fv, which is in agreement with the unspecific RNA binding function of this Cascade backbone protein. The crRNA was also found to interact with a distinct C-terminal region of Cas6f and an N-terminal region and the C-terminus of Cas5fv. The shortness of the identified cross-linked RNA moieties does not permit their unambiguous assignment to crRNA nucleotides. However, it is plausible that all identified Cas5fv interactions are with nucleotides (i.e. CU, UU, AAU) that are also found at the 5'-tag of the crRNA. Structural data is not available for Cas5fv and Cas7fv and only the structure of Cas6f can be modeled due its similarity to available Cas6 structures. Here, the observed interaction of the C-terminal domain of Cas6f with the crRNA is in agreement with the Cas6f-crRNA co-crystal structure from *Pseudomonas aeruginosa* (R. E. Haurwitz, Sternberg, & Doudna, 2012). The identified cross-linked nucleotides (i.e. U, GU) are found at the very beginning of the 3'-terminal repeat tag of the crRNA.

Biochemical analysis of synthetic Cascade variants with shortened and elongated crRNAs

Type I-E Cascade complexes form crescent-shaped structures with small subunits located along its belly. These small subunits were shown to interact with Cas7 and Cas6e (Jackson et al., 2014). The absence of the small and large subunits suggests that the minimal Type I-Fv Cascade might facilitate the presence of a more variable backbone. To test if the Cascade backbone can be altered, we created pre-crRNA constructs with shortened and extended spacer sequences. The investigated crRNA contained a 32 nt spacer which was engineered into a drastically shortened crRNA with a 14 nt spacer (-18 nt) and a long 50 nt crRNA with an extended spacer (+18 nt). We verified that all three transcripts were produced in *E. coli* and processed by Cas6f into mature crRNA species (Fig. S.II.9).

Figure II.3. Mass-spectrometry analysis of Cas protein–crRNA interactions in Type I-Fv Cascade. Identified Cas protein peptides (red) cross-linked to the indicated nucleotides were subjected to tandem mass spectrometry fragmentation to identify which amino acid is cross-linked (blue). Potential interaction sites of Cas5fv (5.1–5.4) and Cas6fv (6.1 and 6.2) have been projected onto the crRNA sequence. Repeat tags flanking the spacer are indicated in green.

Recombinant Cascade complexes were produced in the presence of the three crRNA variants. Surprisingly, all three variants resulted in stable ribonucleoprotein complex formation. The complexes were purified and subjected to size-exclusion chromatography. Each complex purification procedure resulted in three distinct peaks that were analyzed via SDS-PAGE (Fig. II.4). The first peak always corresponded to long Cas7fv filaments and the third peak corresponded to the previously observed Cas5fv-Cas7fv dimers. The constant distance between these two peaks demonstrated that the size-exclusion chromatography runs were performed under identical conditions to allow for their comparison. The highest absorbance value was always identified as a distinct middle peak. The corresponding fractions revealed complexes consisting of Cas6f, Cas5fv, multiple subunits of Cas7fv and the mature crRNA. The peak was shifted in accordance to the crRNA size. The gel-filtration column was calibrated to evaluate the approximate sizes of the eluted Cascade complexes (Fig. S.II.9). The wild type complex has a calculated size of 215 kDa, the short crRNA (-18 nt) is incorporated in a synthetic Cascade complex of 125 kDa and the long crRNA (+18 nt) is incorporated in a complex of 280 kDa. SDS-PAGE analysis demonstrated that the intensity of the Cas7fv bands varied. Thus, we conclude that the alteration of the crRNA spacer length can be compensated in synthetic Cascade complexes. Shortened crRNAs can result in the omission of Cas7fv subunits and elongated crRNAs can result in additional Cas7fv subunits. The *S. putrefaciens* Cas7fv has a molecular mass of 36 kDa which suggests that at least two Cas7fv subunits were added in the long Cascade complex (crRNA +18 nt) and three Cas7fv subunits were omitted in the short Cascade complex (crRNA -18 nt). To investigate the importance of the 6 nt periodicity that was observed for the *E. coli* Cascade, we repeated these experiments with crRNAs whose length was altered by 15 nt. Again, stable Cascade complexes were formed with short crRNAs (-15 nt) and long crRNAs (+15 nt) and the size-exclusion peaks correspond to Cascade assemblies with varied numbers of Cas7fv subunits (Fig. S.II.10).

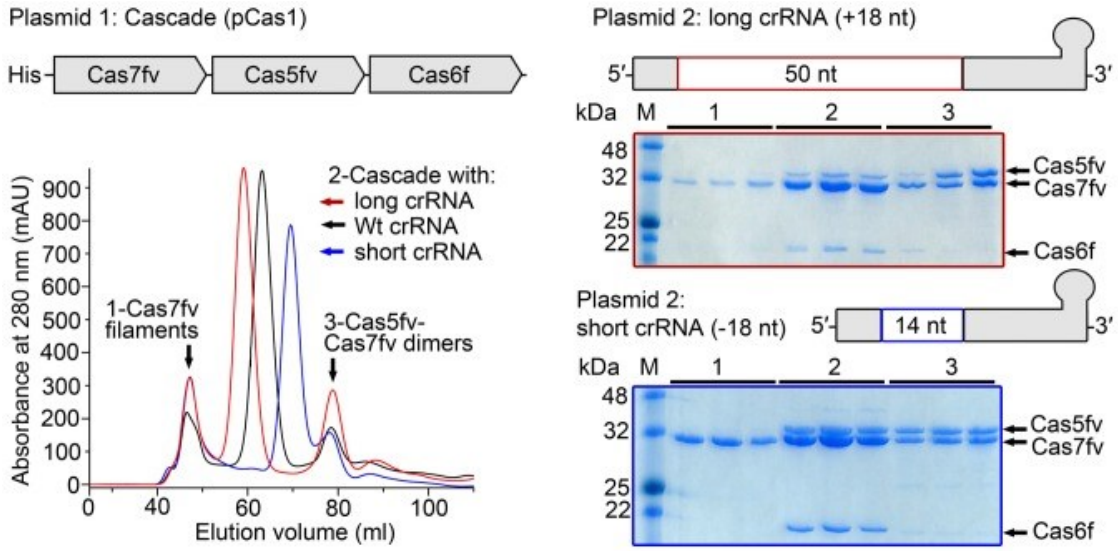


Figure II.4. Purification of recombinant synthetic Cascade variants with shortened and elongated spacers. Variants of crRNAs with wild type (WT) spacer (32 nt), short spacer length (14 nt) and long spacer length (50 nt) were designed. Production and maturation of the crRNA variants in *E. coli* was verified. Recombinant Cascade complexes were produced and purified via size-exclusion chromatography. Cas7fv filaments (peak 1) and Cas5fv-Cas7fv dimers (peak 3) were observed and the middle peak corresponded to fully assembled Cascade ribonucleoproteins (peak 2). The relative shift of this peak during identical size-exclusion chromatography runs and SDS-PAGE revealed that additional spacer nucleotides result in additional Cas7fv subunits.

Structural analyses of synthetic Cascade variants

SAXS analysis was performed to assess the synthetic assemblies on a structural level in comparison to wild type Cascade (Fig. II.5A). *Ab initio* shape restoration bead modeling revealed that distinct Cascade assemblies are formed, depending on the crRNA length (Fig. II.5B). Wild type Cascade forms a crescent assembly (Fig. II.5B, middle), reminiscent in size and shape of the *sea-horse*-shaped *E. coli* Cascade. Truncation of the scaffold crRNA by 18 nt resulted in a contracted structure (Fig. II.5B, left) and elongation of the RNA by 18 nt consequently manifested in an elongated assembly (Fig. II.5B, right). Kratky analysis of the scattering indicates that the long Cascade might be more flexible and structurally less folded, as judged from the increased q^2I in the high q-range, compared to short and wild type Cascade (Fig. II.5A, middle). Porod volume estimation revealed that wild type Cascade features a volume of ~360.28 nm^3, whereas the short version is characterized by a volume of ~207.82 nm^3 and the long version by a volume of ~519.09 nm^3 (Tab. S.II.2). The volume difference of 152.46 nm^3 (wild type-short) and 158.81 nm^3 (long-wild type) suggests that an equal amount of His_6-Cas7fv subunits was removed or added in respect to the crRNA length variation. To address how many subunit less or more were accommodated, we estimated the volume of His_6-Cas7fv to be 44.972 nm^3 ((1.21 x MW)A^3/molecule), arguing for the volume difference to correspond to three subunits.

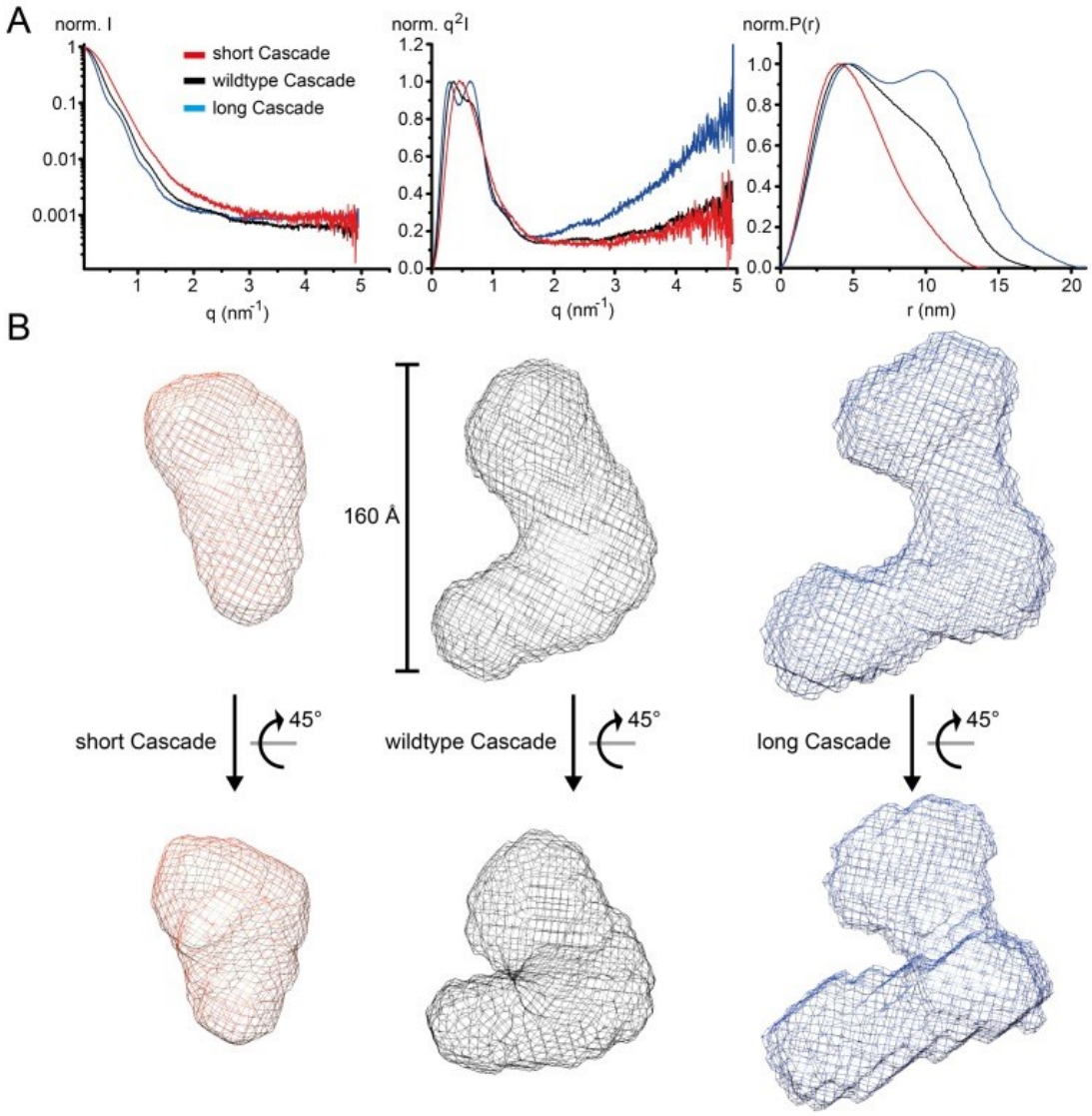


Figure II.5. Small-angle X-ray scattering analyses of synthetic Cascade assemblies. (**A**) SAXS of the short (red), WT (black) and long (blue) Cascade constructs. Left: scattering curve, normalized to max I; middle: Kratky plot, normalized to max q^2I, illustrating the 'random coil likeliness' differences in the high q-range; right: P(r) distance distribution curve, normalized to max P(r), highlighting the different domain organization. (**B**) Surface grid representations of averaged and filtered *ab initio* bead models calculated by Dammif (41).

A minimal Type I-Fv CRISPR-Cas system provides heterologous protection against lambda phage infections

The Type I-Fv CRISPR-Cas system was shown to interfere with plasmid conjugation in *S. putrefaciens* CN-32 (Dwarakanath et al., 2015). However, it remained possible that *S. putrefaciens* contained unidentified Cas proteins that could substitute the roles of the missing small and large subunits. We aimed to transfer the minimal system into *E. coli* to investigate this possibility and to screen if the system is able to target invading viruses. Two plasmids were transformed into *E. coli* BL21-AI, which does not contain endogenous *cas* genes. One plasmid (pCas6) harbored a cassette for the production of Cas1, Cas6f,

Cas5fv, Cas7fv and the DNA nuclease Cas3. The second plasmid (pCRISPRλ) was designed to produce precursor-crRNA with a single spacer targeting gene E lambda phage (Tab. S.II.1, Fig. II.6A). Cell sensitivity to lambda phage was tested using standard plaque assays. The presence of all Type I-Fv Cascade compounds and a crRNA with a target "GG" PAM reduced the efficiency of plaquing (EOP) by 77% in comparison to the strain carrying empty plasmids (Fig. II.6). A crRNA with full spacer complementarity and a target "GA" PAM resulted in an EOP of 0.94, demonstrating PAM recognition. In addition, a crRNA with a random spacer sequence showed an EOP of 0.98, indicating sequence-specific targeting of the lambda phage genome (Fig. II.6B). As a second control, a plasmid (pCas7) was constructed that allows the production of a nuclease-deficient Cas3 mutant (Cas3 H156A/D157A). This inactive Cas3 mutant did not confer immunity against lambda phage (EOP=0.94) (Fig. II.6B). Taken together, these results show that the minimal Type I-Fv CRISPR-Cas subtype can be transferred into *E. coli* where it provides heterologous protection against lambda phage infections. The interference reaction is sequence- and PAM-specific, depends on Cas3 nuclease activity and does not require the presence of small and large subunits.

Cascade assemblies with modulated Cas7 backbones can retain functionality

Finally, we investigated the functionality of the synthetic Cascade variants by lambda phage interference assays. We changed the spacer length of the crRNA provided by the pCRISPRλ plasmid, while retaining all five *cas* genes on the pCas6 plasmid. Added nucleotides were designed to extend base complementarity with the target sequence. The plaque assay results showed that crRNAs with spacers elongated by 1 nt (EOP=0.11), 6 nt (EOP=0.23) or 9 nt (EOP=0.09) retained activity, which was comparable to wild type spacer activity (EOP=0.18). Further elongation of the spacer reduced interference activity (Fig. II.7). Therefore, spacer length can be increased to a certain degree and crRNAs with a spacer length that is not found in native Type I-Fv CRISPRs can be functional. The reduction of the crRNA spacer length by either 6 or 18 nt resulted in Cascade variants that did not confer immunity against lambda phage (Fig. II.7).

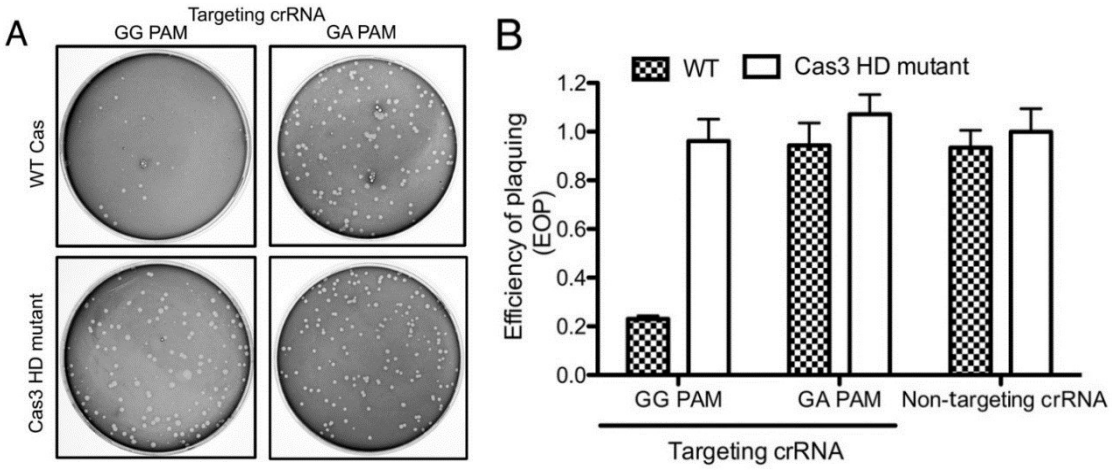


Figure II.6. The *S. putrefaciens* Type I-Fv CRISPR-Cas system provides heterologous protection against lambda phage infection in *E. coli*. (**A**) Plaque formation by lambda phage was observed in *E. coli* BL21-AI strains carrying two plasmids: a first plasmid encoding a crRNA with a spacer of 32 nt complementarity to the lambda phage genome flanked either by a 'GG' or 'GA' PAM (targeting crRNA) and a second plasmid encoding all Cas proteins (pCas6) or pCas7 (Cas3 HD mutant). (**B**) Quantification of plaque formation was performed in triplicate (represented as efficiency of plaquing, EOP) of strains carrying the WT or the Cas3 HD mutant plasmid, in addition to a second plasmid producing either the targeting crRNA (in the presence of a target 'GG' PAM or a target 'GA' PAM) or a crRNA with a 32 nt non-targeting random spacer without complementarity to the phage genome. EOP is defined as the ratio between the plaque count of the strain of interest and the strain carrying empty plasmids. Bars represent mean ± SEM.

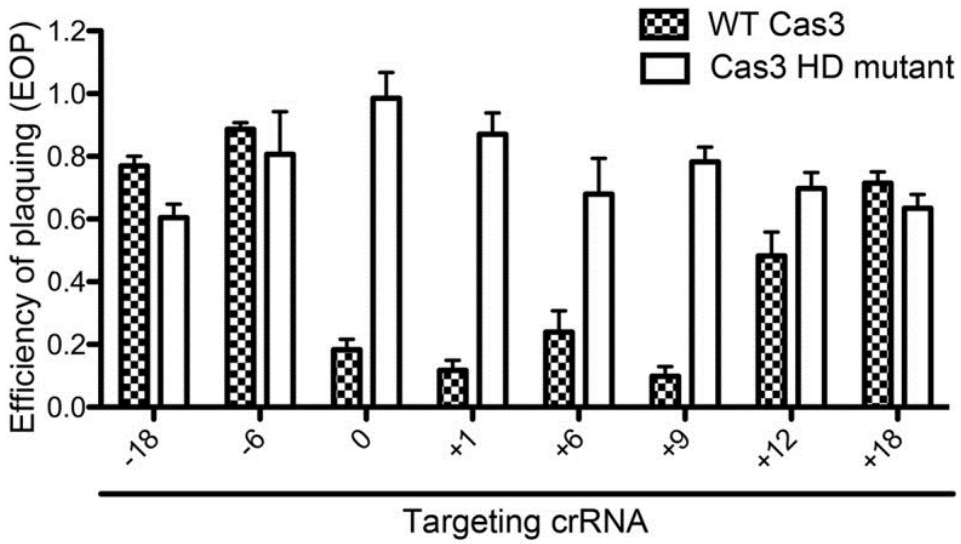


Figure II.7. Cascade-mediated interference activity with extended and shortened crRNAs. Shown is the efficiency of plaquing of Type I-Fv CRISPR-Cas system variants expressed in *E. coli*. The 32 nt spacer sequence (0) was modulated by removal or addition of the indicated number of nucleotides (−18 to +18) while maintaining complementarity to the lambda phage sequence. These arrays were co-expressed in *E. coli* BL21-AI with pCas6 (in checkers) or pCas7 (Cas3 HD mutant, white bars), which was then subjected to plaque assays with lambda phage. Bars represent mean ± SEM of three independent experiments.

Discussion

The three proteins Cas6f, Cas5fv and Cas7fv are sufficient to form a Cascade structure that can interact with Cas3 to target viral DNA in *E. coli*. Heterologous immunity was previously shown to be provided by Type II CRISPR-Cas, e.g. derived from *Streptococcus thermophilus* (Sapranauskas et al., 2011), and CRISPR-Cas systems usually exist as compact modules that have the potential to be transferred between species. However, all investigated Type I systems contain additional subunits, termed large subunits and the best studied system, the Type I-E CRISPR-Cas from *E. coli*, contains both, a large subunit, Cse1, and two small subunits, Cse2. Our results show that a functional Type I Cascade system can be transferred from *S. putrefaciens* into *E. coli* and can provide a DNA targeting mechanism in the absence of these additional subunits. Cse2 is described to be necessary for efficient binding of Cascade to dsDNA and Cse1 was found to be involved in PAM recognition and target DNA duplex destabilization (Hayes et al., 2016; van Erp et al., 2015). A crystal structure of the *E. coli* Cascade bound to a foreign double-stranded DNA target revealed that distinct structural features of the large subunit Cse1 recognize the PAM duplex sequence (Hayes et al., 2016). However, these structures also highlight that Cas5e is located in close vicinity of the PAM. Thus, we hypothesize that the highly divergent Cas5fv protein might play a critical role in DNA target binding and its presence might abrogate the need for a large subunit. Modeling attempts of the Cas5fv structure did not succeed, underlining its dissimilarity to other Cas5 proteins. The minimal Type I-Fv Cascade requires Cas3 HD nuclease activity to function and interactions between Cas1 and Cas3 were observed, in agreement with previous studies of the Type I-F system of *Pectobacterium atrosepticum* (C. Richter et al., 2012) and Type I-A systems (A. Plagens, Tjaden, Hagemann, Randau, & Hensel, 2012). It is suggested that PAM-dependent foreign DNA processing involves the Cascade-mediated recruitment of Cas3, which creates a nick in the R-loop (S. Mulepati & Bailey, 2013; Sinkunas et al., 2013). Our results indicate that the Type I-Fv Cascade is sufficient for this Cas3 recruitment activity and other *S. putrefaciens* host proteins are not required. This minimal Cascade architecture might represent its most basic assembly model as the involved Cas5, Cas6 and Cas7 proteins were all proposed to be derived from a single ancestral RNA-binding protein (Kira S Makarova, Aravind, et al., 2011). It is possible that cellular processes that generate ssDNA are coupled to the Type I-Fv Cascade, which could also explain its moderate activity. These minimal Cascade complexes are only observed in mesophilic bacteria, which might not require the most active defense measures as evidenced by down-regulated or absent CRISPR-

Cas systems. Instead, they might benefit from CRISPR-Cas variants that effectively balance foreign DNA targeting activity and their fitness cost, e.g. by reducing the self-targeting potential.

E. coli Type I-E Cascade was shown to be produced in the absence of crRNAs (Beloglazova et al., 2015). In this system, the Cas proteins Cas5, Cas7 and Cas6e form a stable complex with small and large subunits which could be loaded *in vitro* with synthetic crRNAs to generate functional DNA interference complexes. Similarly, Type I-A Cascade complexes can be purified without bound crRNAs (André Plagens et al., 2014). These observations are in stark contrast to the results obtained for the assembly of the minimal Type I-Fv Cascade, as this complex was shown to require crRNA for proper assembly. Furthermore, the length of the crRNA was identified to define the length of the Cas7fv backbone. It is possible that Type I-E Cascade can assemble without a crRNA scaffold if non-crRNAs provide a substituted scaffold. Thus, the defined length of the Cas7e backbone might be a result of protein-protein interaction between the capping enzymes Cas5e and Cas6e and the large and small subunits along the Cascade belly. These interactions are absent in the minimal Type I-Fv Cascade, which supports the notion that crRNA guidance is necessary for the Cascade assembly process. In addition, these minimal complexes are likely maintained as stable ribonucleoprotein units that do not permit the cycling of crRNAs. Our purification procedures of Cascade complexes without crRNAs yielded Cas5fv-Cas7fv dimers and the isolated Cas proteins are not stable. This suggests that these dimers could function as Cascade building blocks, following the Cas6f-mediated release of the crRNA's 5'-tag. Cas6f remains bound to the 3'-terminal repeat hairpin and would serve as a roadblock to crRNA-guided Cas7fv filament formation. This model entails that the alteration of the crRNA's spacer length results in modulated Cas7fv filaments. In agreement, drastically shortened crRNAs resulted in Cascade complexes with fewer Cas7fv subunits and elongated crRNAs resulted in added Cas7fv subunits. Flexibility of the Cas7 backbone was not yet observed for Type I Cascade systems, but is well documented for Type III systems. In these systems, Csm3 or Cmr4 act as ruler proteins that span the crRNA spacer in 6 nt increments (Hale et al., 2009; Hatoum-Aslan, Samai, Maniv, Jiang, & Marraffini, 2013). Variant crRNAs whose length differs by 6 nt are found in nature, which indicates that both forms can function in active Type III complexes. Thus, crRNA-guided backbone filament formation is a shared feature of multi-Cas protein complexes. In contrast, Type II, V and VI CRISPR-Cas systems each contain a single large Cas protein that spans the crRNA and filaments are absent. Nevertheless, the length of the crRNA in these systems can be altered without adjustment of the Cas protein content. It was shown that the length of the guide RNA-DNA interaction in *Streptococcus pyogenes* Cas9 can be reduced by

3 nt without affecting activity, while a reduction of 5 nt was not tolerated (Fu, Sander, Reyon, Cascio, & Joung, 2014). Interestingly, the truncated guide RNAs improved the specificity of the Cas9 nuclease and the recognition of off-target sites was decreased. Our results demonstrate that the length of a Type I Cascade can be modulated and that only some alterations affect activity. Thus, the degree of modulating the Cascade backbone is likely limited by its requirement for productive Cas3 docking. Nevertheless, Cascade complexes with added spacer nucleotides will influence DNA targeting specificity. It has been shown that Cascade can be used as a gene silencing tool (Rath, Amlinger, Hoekzema, Devulapally, & Lundgren, 2015), but its large size might be a disadvantage for its transfer into heterologous hosts. However, endogenous Type I and Type III CRISPR-Cas systems can be efficiently utilized for genome editing purposes (Li et al., 2015; Luo, Mullis, Leenay, & Beisel, 2015). In these cases, a minimal Cascade system without small and large subunits would provide the most compact CRISPR-Cas system with an adjustable Cas protein filament backbone. The variability of the spacer length of recombinant Type I-Fv Cascade opens the possibility of influencing DNA targeting specificity.

Acknowledgements

The authors thank André Plagens for helpful discussions. The authors also thank the European Synchrotron Radiation Facility (ESRF), Grenoble, France for beamline support.

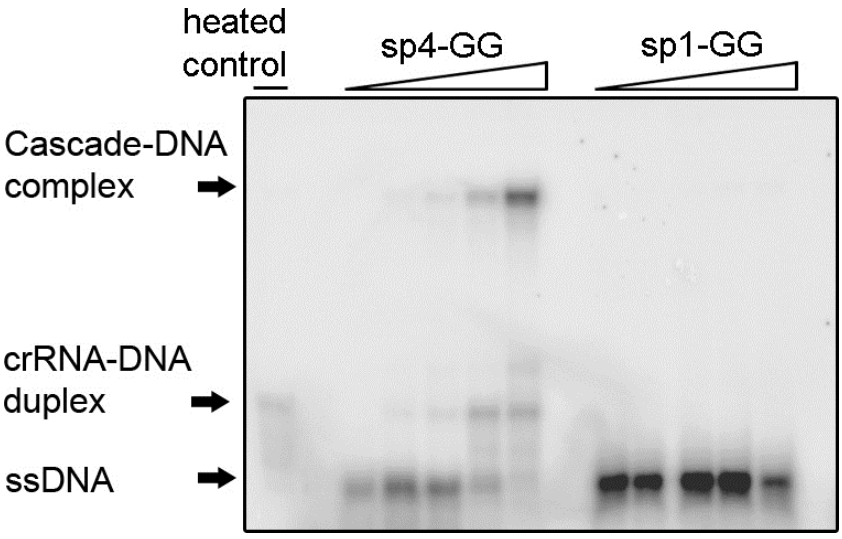
Supplementary Material

Supplementary Table S1. List of constructed plasmids and investigated crRNA sequences

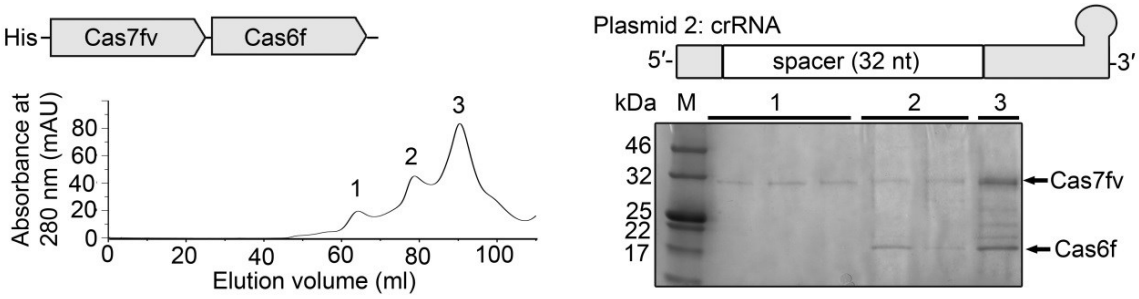
Plasmid	Description
pCDFDu et-1	T7 RNAP polymerase based expression vector, $Spec_R$ (Novagen)
pRSFDu et-1	T7 RNAP polymerase based expression vector, Kan_R (Novagen)
pUC19	Cloning vector, Amp_R (New England Biolabs)
pCas1	pRSFDuet-1 + *cas7fv*, *cas5fv* and *cas6f* with an N-terminal His-tag on *cas7fv*
pCas2	pRSFDuet-1 + *cas7fv*, *cas5fv* and *cas6f* with a C-terminal His-tag on *cas6f*
pCas3	pRSFDuet-1 + *cas7fv* and *cas6f* with an N-terminal His-tag on *cas7fv*
pCas4	pRSFDuet-1 + *cas3*, *cas7fv*, *cas5fv* and *cas6f* with an N-terminal His-tag on *cas3*
pCas5	pRSFDuet-1+ *cas1, cas3*, *cas7fv*, *cas5fv* and *cas6f* with an N-terminal His-tag on *cas1*
pCas6	pRSFDuet-1+ *cas1, cas3*, *cas7fv*, *cas5fv* and *cas6f*
pCas7	pCas3 with a C466G, A467C, A470C mutation in *cas3*
pCRISPR-wt	pUC19 + T7 RNAP promoter and the repeat-spacer4-repeat sequence of *S.putrefaciens* CN-32 Spacer 4: 5′-TATCGCCCAGCAAGACGCGCAAACCTATAACC-3′
pCRISPR+18	pCRISPR-wt with spacer4 extended by 18 random nt: 5′-CACTCAATCCCGCTAATG-spacer4-3′
pCRISPR-18	pCRISPR-wt with spacer4 shortened by 18 nt
pCRISPR+15	pCRISPR-wt with spacer4 extended by 15 random nt: 5′-CACTCAACGCTAATG-spacer4-3′
pCRISPR-15	pCRISPR-wt with spacer4 shortened by 15 nt
pCRISPRλ	pCDFDuet-1+ T7 RNAP promoter and a repeat-spacer-repeat array. The repeat sequence was obtained from *S. putrefaciens* CN-32. The 32 nt spacer is complementary to a phage lambda gene E sequence and flanked by a GG PAM at the 3´end of the target strand: 5′-GGCGGCACGGAGTGGAGCAAGCGTGACAAGTC-3′
pCRISPRλGA	pCDFDuet-1+ T7 RNAP promoter and a repeat-spacer-repeat array. The repeat sequence was obtained from *S. putrefaciens* CN-32.The 32 nt spacer is complementary to phage lambda gene E sequence and flanked by a GA PAM at the 3´end of the target strand: 5′-GCATCATCATGCAGAACATGCGTGACGAAGAG-3′
pCRISPR λ+1	pCRISPR λ with a spacer sequence elongated by 1 nt complementary to phage lambda gene E sequence.
pCRISPR λ+6	pCRISPR λ with a spacer sequence elongated by 6 nt complementary to phage lambda gene E sequence.
pCRISPR λ+9	pCRISPR λ with a spacer sequence elongated by 9 nt complementary to phage lambda gene E sequence.
pCRISPR λ+12	pCRISPR λ with a spacer sequence elongated by 12 nt complementary to phage lambda gene E sequence.
pCRISPR λ+18	pCRISPR λ with a spacer sequence elongated by 18 nt complementary to phage lambda gene E sequence.
pCRISPR λ-6	pCRISPR λ with a spacer sequence shortened by 6 nt
pCRISPR λ-18	pCRISPR λ with a spacer sequence shortened by 18 nt
pCRISPR NT	pCRISPR λ with a spacer sequence without a match in the phage lambda genome: 5′-GCATCATCATGCAGAACATGCGTGACGAAGAG-3′
Sp4-GG target	Oligonucleotide used for EMSAs containing spacer 4 target sequence with a GG-PAM 5′-AAGCTTGAGGGCCCAAGCCGTTATGCTAGGGTTATAGGTTTGCGCGTCTTGCTGGGCGATAGGACTCCCTATAGTGAGTCGTATTAGGATCC-3′
Sp1-GG target	Oligonucleotide used for EMSAs containing spacer 1 target sequence with a GG-PAM 5′-AAGCTTGAGGGCCCAAGCCGTTATGCTAGCAATGTGGTCGCGCAATTTATGATTTGGTTGAGGACTCCCTATAGTGAGTCGTATTAGGATCC-3′

Supplementary Table S2. Small angle X-ray scattering (SAXS) statistics
Primus (44) (ATSAS software package) was used for data analysis. 20 *ab initio* shape restoration models have been calculated with Dammif (implemented in Primus) in slow mode to generate averaged and filtered models.

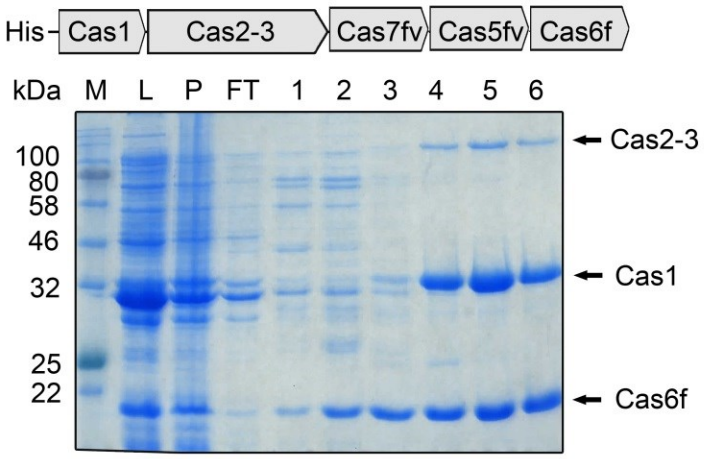
Data collection	**S Cascade**	**WT Cascade**	**L Cascade**
Detector	Pilatus 1M	Pilatus 1M	Pilatus 1M
Beam size at sample (µm µm)	700 x 700	700 x 700	700 x 700
Wavelength (Å)	0.9919	0.9919	0.9919
q range (nm^{-1})	0.025 - 5	0.025 - 5	0.025 - 5
Exposure time (s)	10 (10 x 1)	10 (10 x 1)	10 (10 x 1)
Concentration range (mg ml^{-1})	3 - 10	2.5 - 20	10 - 25
Temperature (°C)	4	4	4
Structural parameters			
Sample concentration (mg ml^{-1})	10	20	25
I(0) (cm^{-1}) from P(r)	95.54 +/- 0.0367	160.4 +/- 0.0622	86.72 +/- 0.04148
R_g (nm) from P(r)	4.173 +/- 0.002564	5.444 +/- 0.002526	6.524 +/- 0.003503
I(0) (cm^{-1}) from Guinier	95.53	160.4	86.72
R_g (nm) from Guinier	4.167	5.435	6.517
D_{max} (nm)	14.19	18.14	21.66
Porod volume estimate (nm^3)	207.82	360.28	519.09
***ab initio* modeling results**			
Normalized spatial distribution (NSD) and NSD variation	0.818 (0.044)	0.822 (0.039)	0.865 (0.063)
Final χ^2 against raw data	1.07	2.180	4.538
Software employed			
Primary data reduction	ESRF BM29 online/ Primus	ESRF BM29 online/ Primus	ESRF BM29 online/ Primus
Data processing	Primus	Primus	Primus
3D graphics representation	Chimera	Chimera	Chimera



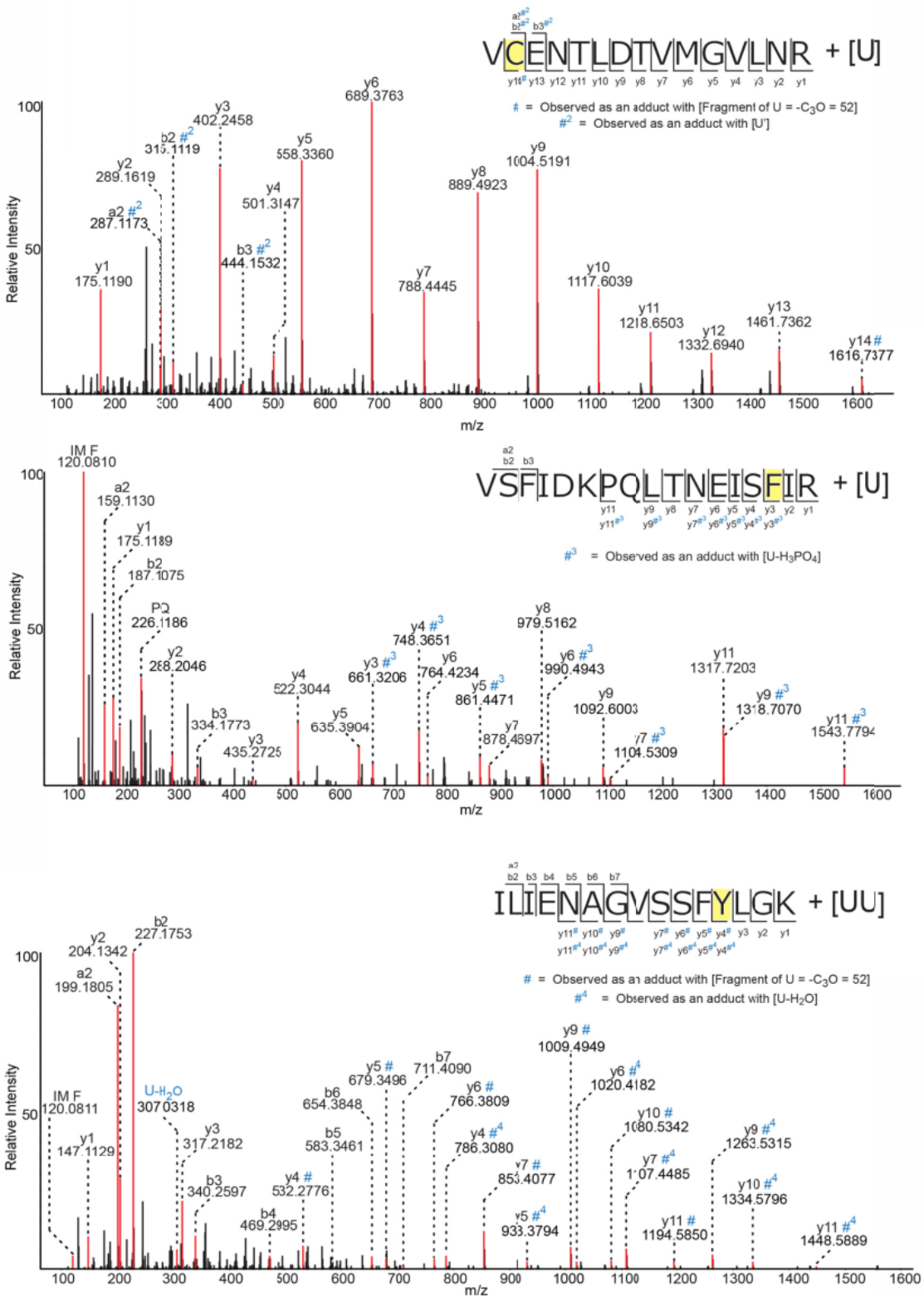
Supplementary Figure II.1. Recombinant Type I-Fv Cascade binds ssDNA. 5'-radioactively labeled target DNA strands (~0.4 pmoles) were incubated with increasing amounts (0, 2, 4, 20, 60 nM) of recombinant S. putrefaciens Type I-Fv Cascade containing a crRNA with spacer 4 of the S. putrefaciens CRISPR array. Binding was observed in electrophoretic mobility shift assays for a target DNA that was complementary to the Cascade-bound crRNA 4 (sp4-GG). A non-complementary sequence containing the target of spacer 1 and a "GG"-PAM (sp1-GG) was not bound. A control was included that contained 60 nM of Cascade after heat-incubation (10 min, 95°C) leading to Cas protein denaturation and crRNA release. This control identified a first smaller shift as a potential crRNA-ssDNA target duplex.



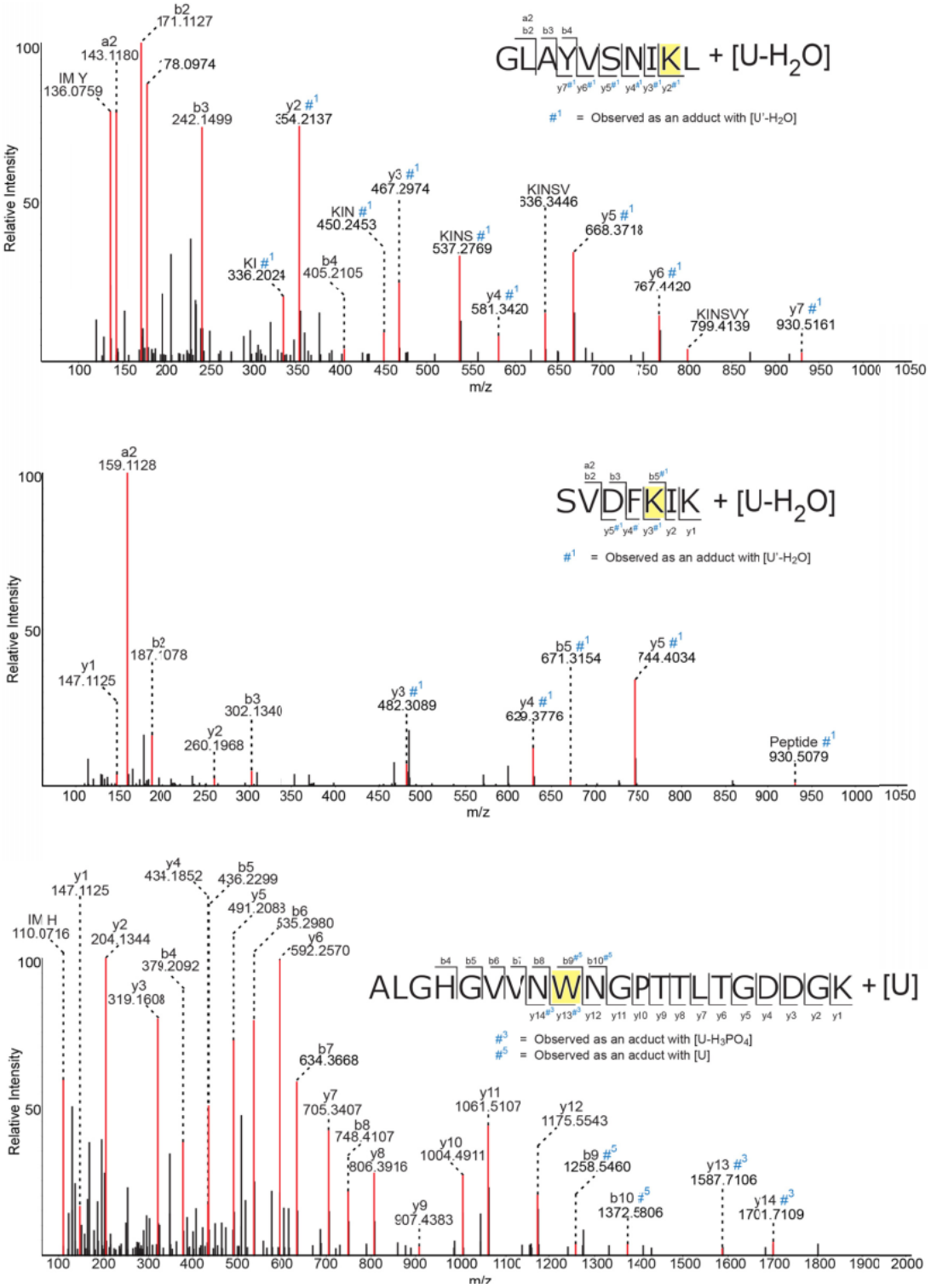
Supplementary Figure II.2. Cascade is not produced in the absence of Cas5fv. (Top) Schematic overview of the investigated recombinant Cas proteins (pCas3) and crRNA. SDS-PAGE (right) was used to analyze the protein and RNA content corresponding to the peaks indicated in the gel-elution chromatogram (left).



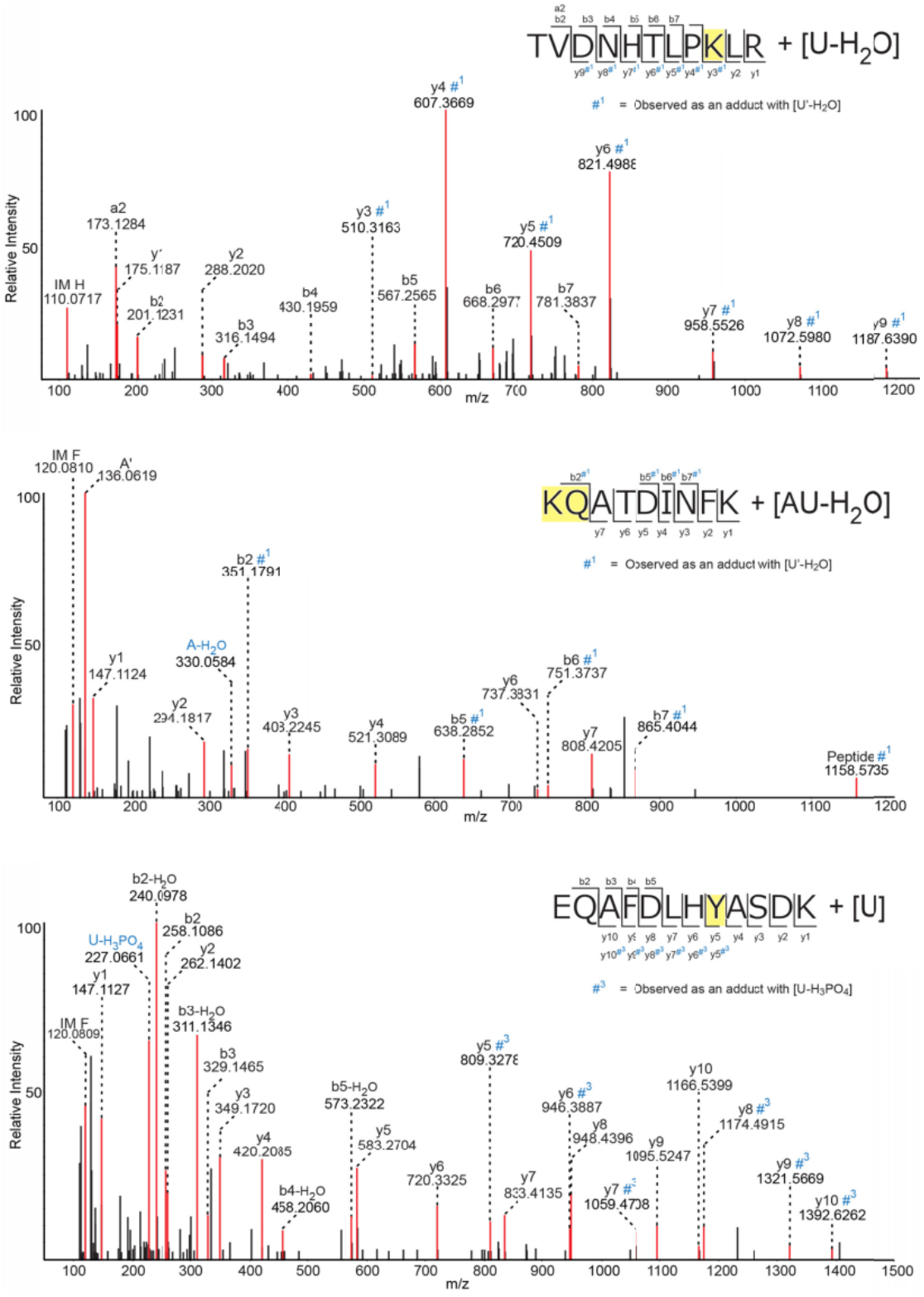
Supplementary Figure II.3. Cas1 interacts with Cas3 and Cas6f but not with the Cascade backbone. (Top) Schematic overview of the investigated recombinant Cas proteins (pCas5) and crRNA. SDS-PAGE (bottom) was used to analyze the protein content of the Ni-NTA purification fractions (Marker, M, Lysate, L, Pellet, P, Flow-through, FT).



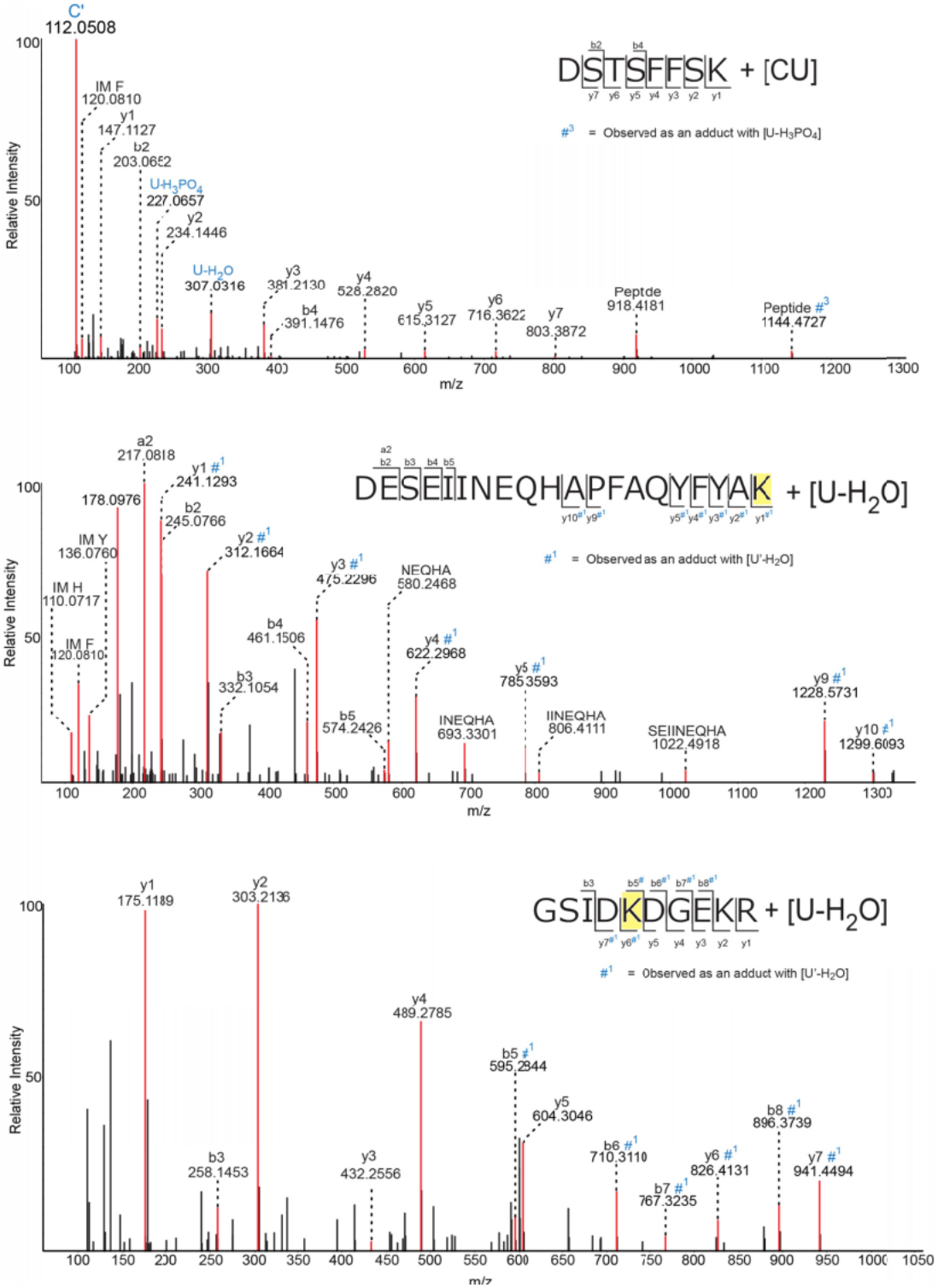
Supplementary Figure II.4.



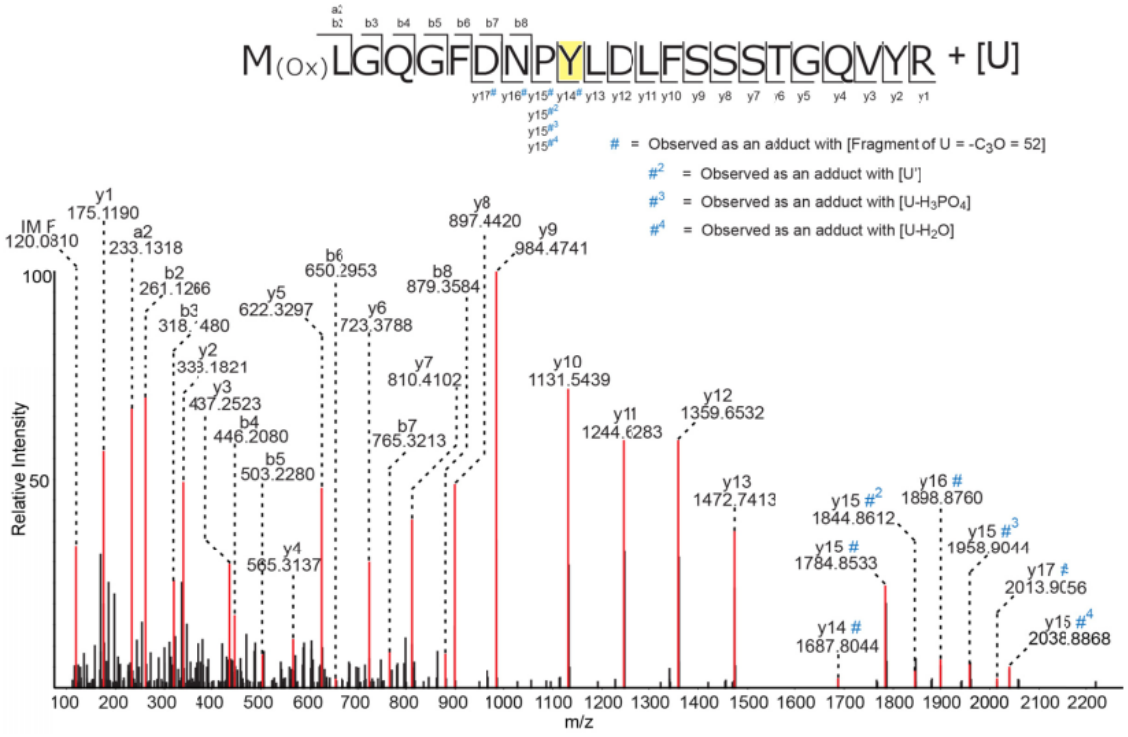
Supplementary Figure II.5.



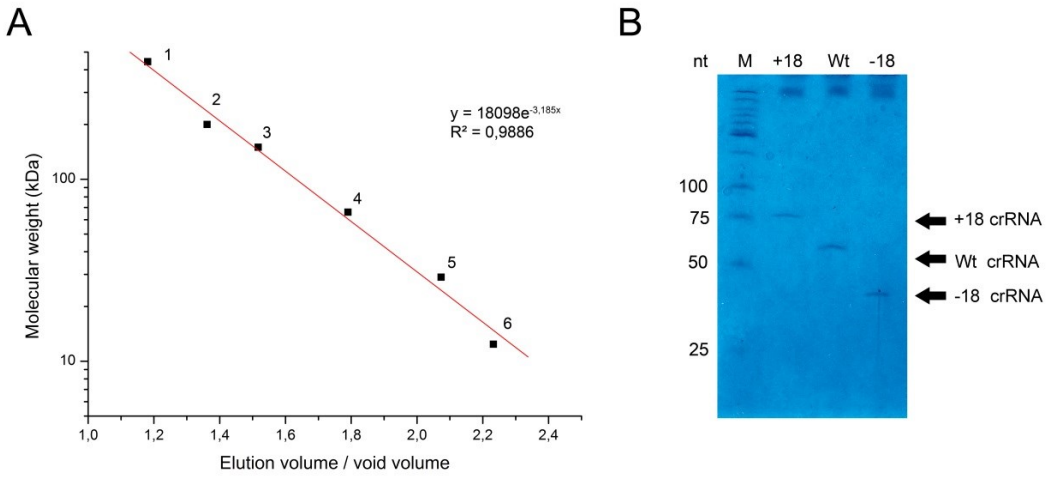
Supplementary Figure II.6.



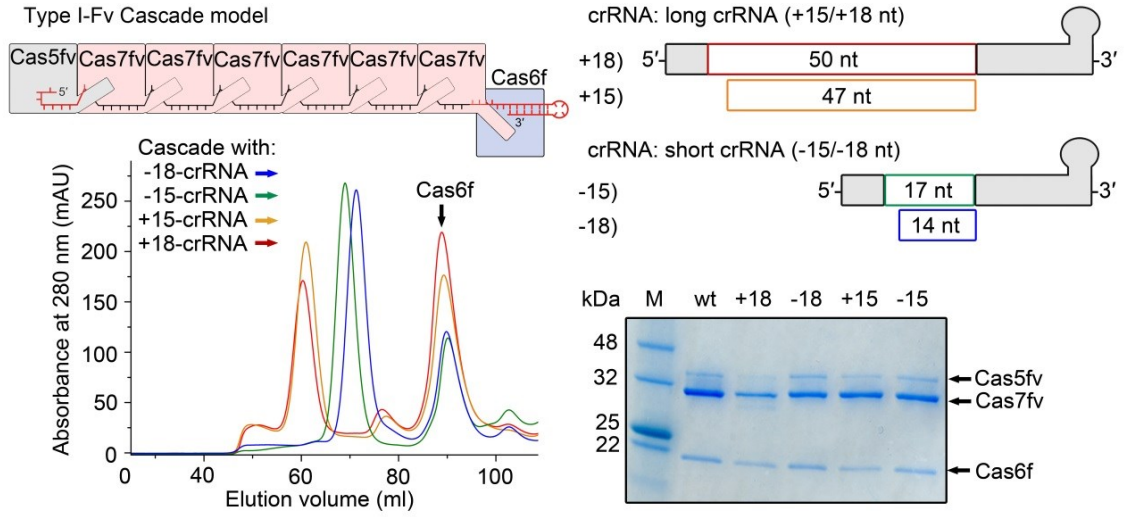
Supplementary Figure II.7.



Supplementary Figures II.4-8. MS/MS fragmentation spectra of the Protein-RNA crosslink identified within the Type I-F Cascade complex from *Shewanella putrefaciens.* Shown are the sample spectra for every peptide observed to be cross-linked with an RNA moiety (see Figure II.3). In each spectrum, the cross-linked peptide sequence and its corresponding y- and b- type fragment ions are indicated at the top, referring to the ions which retain the charge on the N- or C- terminus, respectively. The cross-linked amino-acid residues are highlighted in yellow. Some of the b- and y- ions were observed with a mass shift of #, #1, #2, #3, #4 and #5 corresponding to –C3O (a fragment of Uracil), U´-H2O, U´, U-H3PO4, U-H2O and U, respectively. IM: Immonium ions; C´: Base of C, 112.05 Da; U´: Base of U, 112.0273 Da.



Supplementary Figure II.9. Calibration of the size-exclusion column and analysis of crRNA maturation. (A) Calibration curve of the size-exclusion chromatography column Superdex 200 with the following protein standards: 1) apoferritin (443 kDa), 2) ß-amylase (200 kDa), 3) alcohol dehydrogenase (150 kDa), 4) bovine serum albumin (66 kDa), 5) carbonic anhydrase (29 kDa) and 6) cytochrome C (12.4 kDa). Elution volumes of these proteins were divided by the column's void volume (45 ml). **(B)** 8 M urea PAGE and toluidine blue staining of purified Cascade variants with extended (+18), wildtype (wt) and shortened (-18) crRNA spacers. Processed crRNAs are identified and their length differences correspond to the modulated spacer length.



Supplementary Figure II.10. Purification of recombinant synthetic Cascade variants with shortened and elongated spacers. Top left: Cartoon model of Type I-Fv Cascade. Cas5fv and Cas6f are proposed to cap the crRNA 5′- and 3′-repeat tags. Multiple subunits of Cas7fv are suggested to form a filamentous backbone along the crRNA backbone in analogy to the Type I-E Cas7 backbone of *E. coli*. Additional large and small subunits are missing. Top right: Variants of the crRNAs wildtype spacer (32 nt) with short spacer length (14 & 17 nt) and long spacer length (50 & 47nt) were designed. Bottom left: Recombinant Cascade complexes were produced and purified via size exclusion chromatography. Cas6f monomer peaks were observed. The relative shift of the major second peak during identical size-exclusion chromatography runs revealed efficient Cascade production and Cas7fv subunit number modulation in response to the crRNA's spacer length. Bottom right: SDS-PAGE analysis of purified Cascade variants confirms stable complex formation and reveals a varying Cas7fv band intensity in relation to Cas5fv and Cas6f band intensities.

References

Anders, C., Niewoehner, O., Duerst, A., & Jinek, M. (2014). Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature, 513*(7519), 569-573.

Beloglazova, N., Kuznedelov, K., Flick, R., Datsenko, K. A., Brown, G., Popovic, A., et al. (2015). CRISPR RNA binding and DNA target recognition by purified Cascade complexes from Escherichia coli. *Nucleic Acids Res, 43*(1), 530-543.

Brouns, S. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J., Snijders, A. P., et al. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science, 321*(5891), 960-964.

Carte, J., Pfister, N. T., Compton, M. M., Terns, R. M., & Terns, M. P. (2010). Binding and cleavage of CRISPR RNA by Cas6. *Rna, 16*(11), 2181-2188.

Chylinski, K., Le Rhun, A., & Charpentier, E. (2013). The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems. *RNA biology, 10*(5), 726-737.

Dwarakanath, S., Brenzinger, S., Gleditzsch, D., Plagens, A., Klingl, A., Thormann, K., et al. (2015). Interference activity of a minimal Type I CRISPR-Cas system from Shewanella putrefaciens. *Nucleic Acids Res, 43*(18), 8913-8923.

Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M., & Joung, J. K. (2014). Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol, 32*(3), 279-284.

Garneau, J. E., Dupuis, M.-È., Villion, M., Romero, D. A., Barrangou, R., Boyaval, P., et al. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature, 468*(7320), 67-71.

Gesner, E. M., Schellenberg, M. J., Garside, E. L., George, M. M., & MacMillan, A. M. (2011). Recognition and maturation of effector RNAs in a CRISPR interference pathway. *Nature structural & molecular biology, 18*(6), 688-692.

Grissa, I., Vergnaud, G., & Pourcel, C. (2009). Clustered regularly interspaced short palindromic repeats (CRISPRs) for the genotyping of bacterial pathogens *Molecular Epidemiology of Microorganisms* (pp. 105-116): Springer.

Hale, C. R., Zhao, P., Olson, S., Duff, M. O., Graveley, B. R., Wells, L., et al. (2009). RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell, 139*(5), 945-956.

Harpaz, Y., Gerstein, M., & Chothia, C. (1994). Volume changes on protein folding.
Structure, 2(7), 641-649.

Hatoum-Aslan, A., Samai, P., Maniv, I., Jiang, W., & Marraffini, L. A. (2013). A ruler protein in a complex for antiviral defense determines the length of small interfering CRISPR RNAs. *J Biol Chem, 288*(39), 27888-27897.

Haurwitz, R. E., Jinek, M., Wiedenheft, B., Zhou, K., & Doudna, J. A. (2010). Sequence-and structure-specific RNA processing by a CRISPR endonuclease. *Science, 329*(5997), 1355-1358.

Haurwitz, R. E., Sternberg, S. H., & Doudna, J. A. (2012). Csy4 relies on an unusual catalytic dyad to position and cleave CRISPR RNA. *EMBO J, 31*(12), 2824-2832.

Hayes, R. P., Xiao, Y., Ding, F., van Erp, P. B., Rajashankar, K., Bailey, S., et al. (2016). Structural basis for promiscuous PAM recognition in type I-E Cascade from E. coli. *Nature, 530*(7591), 499-503.

Jackson, R. N., Golden, S. M., van Erp, P. B., Carter, J., Westra, E. R., Brouns, S. J., et al. (2014). Crystal structure of the CRISPR RNA–guided surveillance complex from Escherichia coli. *Science, 345*(6203), 1473-1479.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *Science, 337*(6096), 816-821.

Jinek, M., Jiang, F., Taylor, D. W., Sternberg, S. H., Kaya, E., Ma, E., et al. (2014). Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science, 343*(6176), 1247997.

Judith, R., James, H. N., & Malcolm, F. W. (2013). CRISPR interference: a structural perspective. *Biochemical Journal, 453*(2), 155-166.

Kirli, K., Karaca, S., Dehne, H. J., Samwer, M., Pan, K. T., Lenz, C., et al. (2015). A deep proteomics perspective on CRM1-mediated nuclear export and nucleocytoplasmic partitioning. *Elife, 4*.

Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J., & Svergun, D. I. (2003). PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *Journal of Applied Crystallography, 36*, 1277-1282.

Kramer, K., Sachsenberg, T., Beckmann, B. M., Qamar, S., Boon, K. L., Hentze, M. W., et al. (2014). Photo-cross-linking and high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins. *Nat Methods, 11*(10), 1064-1070.

Li, Y., Pan, S., Zhang, Y., Ren, M., Feng, M., Peng, N., et al. (2015). Harnessing Type I and Type III CRISPR-Cas systems for genome editing. *Nucleic Acids Res*.

Luo, M. L., Mullis, A. S., Leenay, R. T., & Beisel, C. L. (2015). Repurposing endogenous type I CRISPR-Cas systems for programmable gene repression. *Nucleic Acids Res, 43*(1), 674-681.

Maier, L.-K., Lange, S. J., Stoll, B., Haas, K. A., Fischer, S. M., Fischer, E., et al. (2013). Essential requirements for the detection and degradation of invaders by the Haloferax volcanii CRISPR/Cas system IB. *RNA biology, 10*(5), 865-874.

Makarova, K. S., Aravind, L., Wolf, Y. I., & Koonin, E. V. (2011). Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. *Biol Direct, 6*(1), 38.

Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J., Charpentier, E., Horvath, P., et al. (2011). Evolution and classification of the CRISPR–Cas systems. *Nature Reviews Microbiology, 9*(6), 467-477.

Makarova, K. S., Wolf, Y. I., Alkhnbashi, O. S., Costa, F., Shah, S. A., Saunders, S. J., et al. (2015). An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol, 13*(11), 722-736.

Mojica, F. J., García-Martínez, J., & Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of Molecular evolution, 60*(2), 174-182.

Mulepati, S., & Bailey, S. (2013). In vitro reconstitution of an Escherichia coli RNA-guided immune system reveals unidirectional, ATP-dependent degradation of DNA target. *J Biol Chem, 288*(31), 22184-22192.

Mulepati, S., Héroux, A., & Bailey, S. (2014). Crystal structure of a CRISPR RNA–guided surveillance complex bound to a ssDNA target. *Science, 345*(6203), 1479-1484.

Niewoehner, O., Jinek, M., & Doudna, J. A. (2014). Evolution of CRISPR RNA recognition and processing by Cas6 endonucleases. *Nucleic acids research, 42*(2), 1341-1353.
Nuñez, J. K., Kranzusch, P. J., Noeske, J., Wright, A. V., Davies, C. W., & Doudna, J. A. (2014). Cas1–Cas2 complex formation mediates spacer acquisition during CRISPR–Cas adaptive immunity. *Nature structural & molecular biology*. 21(6),528-534.
Nuñez, J. K., Lee, A. S., Engelman, A., & Doudna, J. A. (2015). Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. *Nature*. *519*(7542), 193-198.
Plagens, A., Tjaden, B., Hagemann, A., Randau, L., & Hensel, R. (2012). Characterization of the CRISPR/Cas subtype I-A system of the hyperthermophilic crenarchaeon Thermoproteus tenax. *J. Bacteriol, 194*(10), 2491-2500.
Plagens, A., Tripp, V., Daume, M., Sharma, K., Klingl, A., Hrle, A., et al. (2014). In vitro assembly and activity of an archaeal CRISPR-Cas type IA Cascade interference complex. *Nucleic acids research, 42*(8), 5125-5138.
Pourcel, C., Salvignol, G., & Vergnaud, G. (2005). CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology, 151*(3), 653-663.
Rath, D., Amlinger, L., Hoekzema, M., Devulapally, P. R., & Lundgren, M. (2015). Efficient programmable gene silencing by Cascade. *Nucleic Acids Res, 43*(1), 237-246.
Richter, C., Dy, R. L., McKenzie, R. E., Watson, B. N., Taylor, C., Chang, J. T., et al. (2014). Priming in the Type IF CRISPR-Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer. *Nucleic acids research, 42*(13), 8516-8526.
Richter, C., Gristwood, T., Clulow, J. S., & Fineran, P. C. (2012). In vivo protein interactions and complex formation in the Pectobacterium atrosepticum subtype I-F CRISPR/Cas System. *PLoS One, 7*(12), e49549.
Richter, H., Zoephel, J., Schermuly, J., Maticzka, D., Backofen, R., & Randau, L. (2012). Characterization of CRISPR RNA processing in Clostridium thermocellum and Methanococcus maripaludis. *Nucleic acids research, 40*(19), 9887-9896.
Rouillon, C., Zhou, M., Zhang, J., Politis, A., Beilsten-Edmands, V., Cannone, G., et al. (2013). Structure of the CRISPR interference complex CSM reveals key similarities with cascade. *Molecular cell, 52*(1), 124-134.
Sapranauskas, R., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., & Siksnys, V. (2011). The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. *Nucleic Acids Res, 39*(21), 9275-9282.
Sashital, D. G., Wiedenheft, B., & Doudna, J. A. (2012). Mechanism of foreign DNA selection in a bacterial adaptive immune system. *Molecular cell, 46*(5), 606-615.
Semenova, E., Jore, M. M., Datsenko, K. A., Semenova, A., Westra, E. R., Wanner, B., et al. (2011). Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc Natl Acad Sci U S A, 108*(25), 10098-10103.
Sharma, K., Hrle, A., Kramer, K., Sachsenberg, T., Staals, R. H., Randau, L., et al. (2015). Analysis of protein-RNA interactions in CRISPR proteins and effector complexes by UV-induced cross-linking and mass spectrometry. *Methods, 89*, 138-148.
Shmakov, S., Abudayyeh, O. O., Makarova, K. S., Wolf, Y. I., Gootenberg, J. S., Semenova, E., et al. (2015). Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. *Mol Cell, 60*(3), 385-397.
Sinkunas, T., Gasiunas, G., Waghmare, S. P., Dickman, M. J., Barrangou, R., Horvath, P., et al. (2013). In vitro reconstitution of Cascade-mediated CRISPR immunity in Streptococcus thermophilus. *EMBO J, 32*(3), 385-394.
Staples, J., Qiao, D., Cho, M. H., Silverman, E. K., Nickerson, D. A., & Below, J. E. (2014). PRIMUS: rapid reconstruction of pedigrees from genome-wide estimates of identity by descent. *Am J Hum Genet, 95*(5), 553-564.
van der Oost, J., Westra, E. R., Jackson, R. N., & Wiedenheft, B. (2014). Unravelling the structural and mechanistic basis of CRISPR-Cas systems. *Nature Reviews Microbiology*.
van Erp, P. B., Jackson, R. N., Carter, J., Golden, S. M., Bailey, S., & Wiedenheft, B. (2015). Mechanism of CRISPR-RNA guided recognition of DNA targets in Escherichia coli. *Nucleic Acids Res, 43*(17), 8381-8391.
Westra, E. R., Semenova, E., Datsenko, K. A., Jackson, R. N., Wiedenheft, B., Severinov, K., et al. (2013). Type IE CRISPR-cas systems discriminate target from non-target DNA through base pairing-independent PAM recognition. *PLoS genetics, 9*(9), e1003742.
Wiedenheft, B., van Duijn, E., Bultema, J. B., Waghmare, S. P., Zhou, K., Barendregt, A., et al. (2011). RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proceedings of the National Academy of Sciences, 108*(25), 10092-10097.

Chapter III:
Structural Variation of Type I-F CRISPR RNA Guided DNA Surveillance.

Patrick Pausch[1], Hanna Müller-Esparza[2], Daniel Gleditzsch[2], Florian Altegoer[1], Lennart Randau[2,*], and Gert Bange[1,*]

This chapter is written in manuscript style and was published in Molecular Cell in August, 2017. My contribution to this work included designing and performing the experiments *in vivo*, analyzing the obtained data and writing the corresponding parts of the manuscript.

---

[1] LOEWE Center for Synthetic Microbiology and Faculty of Chemistry, Philipps University Marburg, D-35043 Marburg, Germany
[2] Prokaryotic Small RNA Biology Group, Max Planck Institute for Terrestrial Microbiology, D-35043 Marburg, Germany
* Corresponding Authors

Summary

CRISPR-Cas systems are prokaryotic immune systems against invading nucleic acids. Type I CRISPR-Cas systems employ highly diverse, multi-subunit surveillance Cascades facilitating duplex formation between crRNA and complementary target DNA for R-loop formation, retention and DNA degradation by the subsequently recruited nuclease Cas3. Typically, the large subunit recognizes *bona fide* targets through the protospacer adjacent motif (PAM) and the small subunit guides the non-target DNA strand. Here, we present the Apo- and target DNA-bound structures of the type I-F variant (I-Fv) Cascade lacking the small and large subunits. Large and small subunits are functionally replaced by the 5′-terminal crRNA cap Cas5fv and the backbone protein Cas7fv, respectively. Cas5fv facilitates PAM recognition from the DNA major groove site, in contrast to all other hitherto described type I systems. Comparison of the type I-Fv Cascade with an anti-CRISPR protein bound I-F Cascade reveals that type I-Fv structure differs substantially at known anti-CRISPR protein target sites and might therefore be resistant to viral Cascade interception.

Introduction

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)-Cas (CRISPR-associated) systems are present in archaea and bacteria. Their hallmark elements are small crRNAs (CRISPR RNAs), which contain a spacer sequence guiding Cas surveillance and effector complexes towards complementary nucleic acid molecules, termed protospacers (Grissa et al., 2009; Mojica et al., 2005; Pourcel et al., 2005). Native targets are usually viral genomes and CRISPR-Cas modules act as adaptive immune systems in prokaryotes (Barrangou et al., 2007). The presence of small protospacer adjacent motifs (PAMs) is required for efficient target recognition and discrimination between viral and host genomes (Marraffini and Sontheimer, 2010; Mojica et al., 2009). Designer crRNAs can be utilized for highly specific DNA targeting and the guided Cas protein cleavage activity allows for efficient genome engineering applications (Jinek et al., 2012). CRISPR-Cas systems are divided into two classes and highly variable multisubunit effector complexes characterize the prevalent class I systems, which was further subclassified into three types (Makarova et al., 2015; Mohanraju et al., 2016). Type I CRISPR-Cas systems are most widespread in nature and the Cas protein composition of the employed CRISPR ribonucleoprotein (crRNP) complexes, termed Cascade (CRISPR associated complex for antiviral defense), differs between seven subtypes (A to F,U) (Makarova and Koonin, 2015; Makarova et al., 2015). Structures of type I-E and I-F Cascade complexes are available and reveal crescent-shaped crRNPs consisting of a crRNA spine to which the Cas proteins align. Cascade I-E and I-F compare well in structure and composition of Cas5, Cas6, Cas7 and their large subunit Cse1 (hereafter called: Cas8) (Chowdhury et al., 2017; Hayes et al., 2016; Jackson et al., 2014; Mulepati et al., 2014; Wiedenheft et al., 2011). Both complexes differ within crRNA backbone pitch and a small subunit protein that is absent in type I-F (Chowdhury et al., 2017; Makarova et al., 2015) and required for non-target strand stabilization in type I-E Cascade (Hayes et al., 2016). However, both complexes are comparable in their PAM recognition mechanisms primary proceeding via double stranded (ds) DNA interactions by the Cas8 subunit (Hayes et al., 2016; Rollins et al., 2015). A signature R-loop structure is formed by hybridization of crRNA spacer and target DNA protospacer, while the displaced non-target strand is presented for degradation. Finally, Cascade recruits the DNA helicase/nuclease Cas3, which mediates destruction of the foreign DNA (Garneau et al., 2010; Westra et al., 2012). Type I-F and I-E systems are vulnerable to inhibition by anti-CRISPR proteins as a defense mechanism of phages (Bondy-Denomy et al., 2013; Pawluk et al., 2016). In type I-F, the anti-CRISPR proteins AcrF1, 2 and 3 target the Cas7 backbone, the large subunit/Cas5 and Cas3, respectively, in order to

inhibit effector complex formation via multiple mechanisms (Bondy-Denomy et al., 2015; Bondy-Denomy et al., 2013). AcrF1 targets I-F Cascade and blocks crRNA:DNA hybridization by its interaction with the thumb and web domain of two neighboring Cas7 molecules (Bondy-Denomy et al., 2015; Chowdhury et al., 2017). AcrF2 mimics dsDNA (Chowdhury et al., 2017), and blocks the dsDNA recruitment site formed by Cas8 and Cas7 (Bondy-Denomy et al., 2015). AcrF3 inhibits the action of the DNA helicase/nuclease Cas3 by direct binding to Cas3 (Wang et al., 2016). The ongoing arms race between bacteria and viruses should have contributed to the evolution of mechanisms and/or features against Acr proteins to sustain CRISPR-Cas mediated defense. One hint for the existence of such defense mechanisms is the observation that type I-F systems are highly diverse and contain several derivatives, most prominently the type I-F variant (type I-Fv) (Makarova et al., 2011a; Makarova et al., 2011b). Furthermore, type I-Fv CRISPR-Cas systems have been identified not only in bacterial genomes, but also on mobile auxiliary genetic elements, such as plasmids and genomic islands that are subject to rapid horizontal gene transfer (Makarova et al 2015). Recent studies highlighted the presence of the I-Fv CRISPR-Cas system on novel virulence factor harboring genomic islands in several *Vibrio cholera* strains and suggested an associated competitive advantage through the supplied I-Fv CRISPR-Cas system (Carpenter et al. 2017, Labbate et al. 2016). Cascade I-Fv specifically recognizes a 'GG'-PAM (Dwarakanath et al., 2015) and is able to identify crRNA-complementary protospacer sequences (Fig. S.III.1). However, striking differences to other Cascade assemblies include its extremely low sequence conservation of putative Cas5 and Cas7 homologues and the absence of small and large subunits, which yields a minimal 3 Cas protein crRNP (Gleditzsch et al., 2016). Anti-CRISPR proteins have not been identified for type I-Fv CRISPR-Cas systems. All together, these changes make the I-Fv system an extraordinary candidate to study structures and mechanisms that might have evolved in response to anti-CRISPR proteins during the path of molecular evolution.

Here we describe the structures of the type I-Fv Cascade alone and in complex with target DNA. We discover extensive structural differences between I-Fv and the I-F and I-E Cascades. DsDNA recruitment and PAM recognition in type I-Fv Cascade is facilitated from the major groove side by Cas5fv instead of the large subunit Cas8 and the finger domain of Cas7. A unique feature of the type I-Fv Cas7 variant, which forms a secondary helix parallel to the primary helical backbone, guides the non-target strand. Taken together, type I-F and I-Fv Cascades differ the most at their DNA interaction sites, which are also prime target of anti-CRISPR proteins.

Results

Structure of type I-Fv Cascade reveals a unique architecture

Type I-Fv Cascades differ from other type I Cascades in their lack of the large and/ or small subunits and strong sequence deviations in the Cas5 and Cas7 homologues (Gleditzsch et al., 2016). To gain structural insights into this highly divergent Cascade type, we determined the X-ray structure of I-Fv Cascade from *Shewanella putrefaciens* CN32. Because crystals of the full length I-Fv Cascade showed poor diffraction power, we decided to investigate a contracted variant in which the crRNA spacer length was reduced by 18 nts, resulting in three Cas7fv binding sites (Fig. S.III.2), while retaining the interactions between the different components (Gleditzsch et al., 2016). The X-ray structure of this type I-Fv Cascade was determined by single anomalous dispersion (SAD) and refined to a resolution of 3 Å (Table III.1). The structure of the type I-Fv Cascade revealed an elongated, roughly crescent-shaped crRNP with the overall dimensions of 75, 100 and 130 Å (Fig. III.1), which is in good agreement with a previous, low-resolution small angle X-ray scattering (SAXS) analysis (Gleditzsch et al., 2016).

Table III.1. Crystallographic Table

Data Collection[a]	**SpCascade-I-Fv (SeMet)**	**SpCascade-I-Fv-R-Loop**
Space group	I2	P 3_221
Cell dimensions		
a, *b*, and *c* (Å)	157.002	143.316
-	65.894	143.316
-	160.682	172.698
α, β, and γ (°)	90.00	90.00
-	98.61	90.00
-	90.00	120.00
Energy (Å)	0.979	0.991
Resolution (Å)	47.97 - 3.00	45.27 - 3.25
-	(3.10 - 3.00)	(3.36 - 3.25)
R_{merge}	0.0547 (0.356)	0.168 (1.922)
I / σI	13.18 (2.05)	11.66 (1.80)
Completeness (%)	100.0 (100.0)	100.0 (100.0)
Redundancy	10.0 (9.8)	11.1 (11.4)
CC(1/2)	0.99 (0.72)	0.99 (0.85)
Anomalous completeness (%)	99.5 (99.8)	-
Anomalous redundancy	5.2 (5.3)	-

Refinement		
Resolution (Å)	49.28 - 3.00	46.91 - 3.25
No. reflections	32,875 (3,271)	32,783 (3,192)
R_{work}/ R_{free}	19.4	21.8
-	24.9	27.7

No. atoms	11,803	13,131
Macromolecule	11,803	13,131
Ligand	0	0
Water	0	0
R.m.s deviations	-	-
Bond lengths (Å)	0.015	0.010
Bond angles (°)	1.34	1.29
Ramachandran (%)	-	-
Preferred	95.93	96.76
Allowed	3.92	3.1
Outliers	0.15	0.14

[a] Statistics for the highest-resolution shell are shown in parentheses.

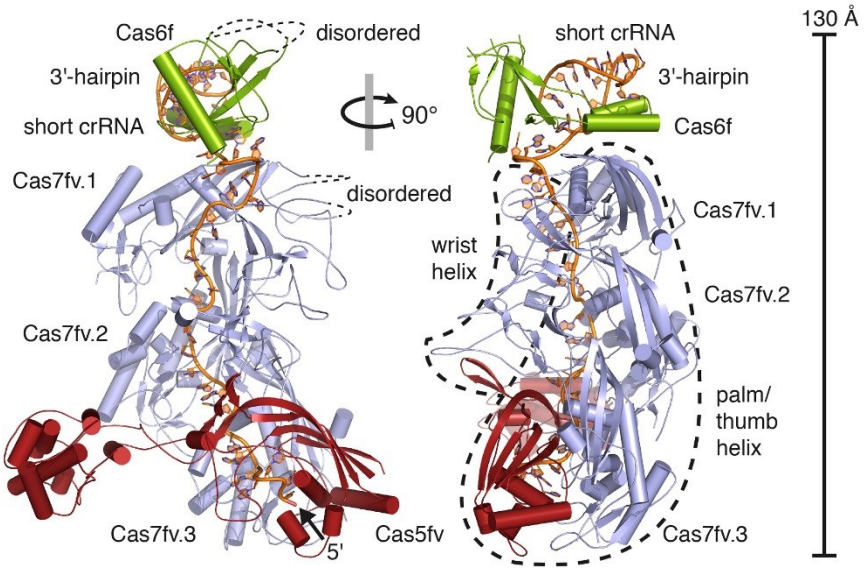


Figure III.1. Crystal structure of the short type I-Fv Cascade. Cartoon representation of the short I-Fv Cascade X-ray crystal structure from *S. putrefaciens* CN32 in two, 90° rotated orientations. Short crRNA, Cas5fv, Cas7fv and Cas6f are coloured in orange, dark red, blue and green, respectively. The crRNA 3′ hairpin and 5′ end are indicated and the Cascade subunits are labelled. Disordered sections are labelled and indicated by thin dotted lines. The two parallel right-handed wrist and palm/thumb helices are labelled accordingly and are indicated by thick dotted lines. Right: scale bar illustrating the total height of 130 Å.

One end of the complex is formed by Cas6f bound to the 20 nt long crRNA 3′-hairpin tag. Three copies of the backbone protein Cas7fv assemble along the shortened crRNA spine to form the body. The Cas6f-opposing end of the I-Fv Cascade is formed by an unusual Cas5f variant (Cas5fv), which interacts with the 8 nt long 5′-handle (Fig. III.1). While the overall shape of type I-Fv Cascade appears roughly similar to other type I Cascades, it clearly differs in several aspects. Type I-Fv contains structurally unique homologues of Cas5 and 7 proteins that shape the complex into a significantly more open configuration than observed for the almost closed type I-F counterpart (Fig. III.2A). Interestingly, the curvature of type I-Fv is more reminiscent to those of the more distantly related type I-E (Fig. S.III.3).

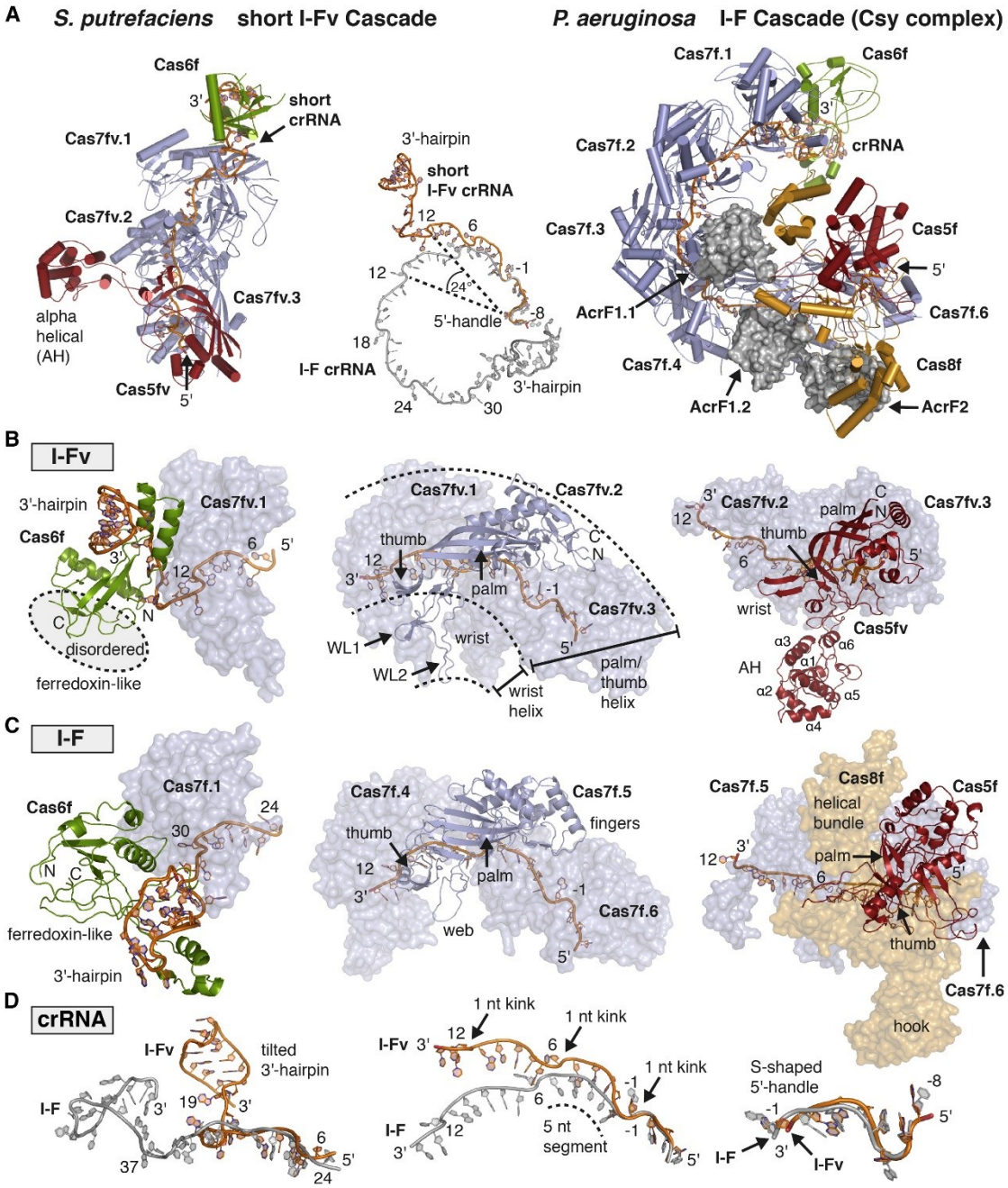


Figure III.2. Structural comparison of type I-F and type I-Fv Cascade. (A) Left: X-ray crystal structure of the short S. putrefaciens I-Fv Cascade shown in a cartoon representation. Color scheme and labeling as in fig. III.1. Middle: crRNA spine superimposition of the short I-Fv (orange) and I-F (grey) crRNA on the basis of a 5′ handle alignment. Nucleotide positions upstream of the first spacer nucleotide are labeled with negative values and the positions of downstream nucleotides are indicated by positive values. The angle of 24° between nucleotide position -6 and 12 illustrates the different crRNA spine pitch. Right: cryo-EM structure of the AcrF1/2 bound P. aeruginosa I-F Cascade (PDB ID: 5ZU9; (Chowdhury et al., 2017)). Components are labeled according to the current nomenclature for type I-F. Color scheme of the I-F Cas homologs is according to I-Fv. The additionally present large subunit protein Cas8f is shown in yellow and the activity inhibiting AcrF1/2 proteins are shown in a grey surface representation. **(B)** and **(C)**: Side-by-side comparison of the I-Fv (B) and I-F (C) Cascade subunits. Color is according to fig. III.2A. The Cas6f proteins are compared in the left panel, Cas7 homologs in the middle panel and Cas5 homologs in the right panel. Adjacent Cascade subunits are shown as transparent surfaces and labeled respectively. Grey circles indicate disordered regions. N and C indicate N- and C-termini, respectively. **(D)** crRNA spine comparison of I-F (orange) and I-Fv (grey). Left: crRNA arrangement at the tilted head structure. Superimposition according to I-Fv nucleotides 6 to 12. Middle: crRNA architecture of the Cas7 backbone bound segment. Superimposition according to I-Fv nucleotides -1 to -4. At every sixth position, the nucleotide is splayed out from the base stacking segments (‘kink’). Right: Superimposition of the similar S-shaped 5′-handles.

Type I-Fv Cascade significantly differs from the closely related type I-F Cascade

To gain a better understanding of the unique architecture of type I-Fv Cascade (this study), we compared its structure to that of the closely related type I-F Cascade recently published (Chowdhury et al., 2017). Both structures are highly similar within the Cas6f subunit bound to its cognate 20 nt long crRNA 3′-hairpin tags (Fig. S.III.4). In both cases, Cas6f interacts with the palm domain of the adjacent Cas7 subunit (Fig. III.2B, C). However, fundamental differences are found in the Cas6f/Cas7 interaction between type I-F and I-Fv. While the ferredoxin-like domain of Cas6f interacts with the palm of Cas7f in type I-F, it is the crRNA-binding α-helical hairpin of Cas6f in the type I-Fv that establishes the interaction (Figs. III.2B, C). This deviating arrangement is accompanied by an approximately 90° tilted reorientation of the 3′-hairpin (Fig. III.2D).

Cas7 proteins, central to all type I Cascades, form a right-handed helical filament along the crRNA and facilitate crRNA spacer presentation for target identification. Cas7 architecture is often described in analogy to a ‘right hand’ with the central RNA recognition motif (RRM), a pronounced loop and a helix-rich domain defined as the palm, thumb and fingers, respectively (Fig. S.III.5). In type I-E and I-F Cascades, all elements of the Cas7 hand are present and fulfill the same functions: The palm provides the binding cavity for 6 nts of a crRNA spacer, while the thumb of an adjacent Cas7 pins the crRNA against the palm. The fingers extrude on the convex side of the helical Cas7 backbone and assist in dsDNA recruitment together with the large subunit at the Cascade base (Hayes et al., 2016). Our structural analysis of type I-Fv shows that while the palm and thumb domains are present in Cas7fv, the fingers are strongly reduced. Instead, two extensive loops (aa 25 to 77) localize next to the thumb at the bottom of the palm (Figs. III.2B, S.III.5). Following the right-hand analogy, we termed these loops: ‘wrist’-loop 1 and 2 (WL1, WL2) (Fig. III.2B). Most surprisingly, the WLs of each of the Cas7fv subunits connect to each other and assemble into a unique helical filament at the concave site of the complex (named: wrist helix). Of note, the wrist helix runs parallel to the helix formed by the palm and thumbs (i.e. palm/thumb helix, (Figs. III.1 and III.2B).

The RRM domain of Cas5fv (palm) encases the crRNA 8 nt long S-shaped 5′-tag together with the adjacent Cas7fv (Fig III.2B, D), similar to the situation in type I-E and I-F (Chowdhury et al., 2017). In contrast to its type I-E and I-F counterparts, Cas5fv contains a wrist that connects to the wrist helix formed by Cas7fv (Figs. III.1 and III.2B). Most importantly, Cas5fv differs in an additional domain consisting of six α-helices termed the α-helical domain (AH, residues: 110 – 266; Figs. III.2A, B). The AH domain extrudes from the convex side of the type I-Fv Cascade at the tip of the thumb that pins the crRNA against the palm of the adjacent Cas7fv (Fig. III.2B).

Taken together, type I-Fv dramatically differs from its counterparts with respect to architecture and arrangement of the Cas7 and Cas5 subunits. Interestingly, major changes in Cas7 and Cas5 localize at positions that are occupied by the small and large subunits in other type I Cascades, and are critical for DNA recruitment and interference.

Type I-Fv and I-F differ at places critical for anti-CRISPR protein binding

The presence of viral Acr proteins provides one plausible cause for the diversification of CRISPR-Cas systems. To understand the obtained differences between I-Fv Cascade and the closely related type I-F complex, we compared both structures with respect to the Acr binding sites at the type I-F Cascade, which were recently determined by cryo-electron microscopy (Chowdhury et al., 2017). In the anti-CRISPR protein bound type I-F Cascade structure, two copies of AcrF1 and one AcrF2 are bound the type I-F Cascade and inhibit association of target DNA. AcrF1 inhibits the function of the type I-F Cascade through interaction with the protospacer-accommodating channel formed by the crRNA spacer, the web and thumb domains of the Cas7f backbone (Chowdhury et al., 2017). In our type I-Fv Cascade, inhibition by AcrF1 would not be possible as the web domain of Cas7f is replaced by the wrist loops of Cas7fv, severely restructuring the AcrF1 binding site (Fig. III.3). The shortened thumb in Cas7fv would further contribute to prevent AcrF1 binding by a reduced contact surface (Fig. III.3). AcrF2 interacts with the large subunit Cas8f of the type I-F Cascade and the thumb of the adjacent Cas7f. In type I-Fv Cascade, both structures are not present, which forbids binding of this anti-CRISPR protein (Fig. III.3). Taken together, our structural comparison shows that the described structural variations should protect type I-Fv Cascade from association with and inhibition by the anti-CRISPR proteins AcrF1 and 2.

X-ray structure of the type I-Fv Cascade in complex with target dsDNA

Next, we aimed to understand how a type I-Fv our Cascade lacking the small/ large subunits and containing highly divergent Cas5/7 proteins achieves PAM-dependent recognition of foreign DNA. Therefore, we reconstituted the type I-Fv Cascade in complex with a DNA target. Specifically, two DNA oligonucleotides (i.e. target and non-target) were designed with the target containing the protospacer sequence in its 5′ region (Fig. III.4B). The remaining 14 nts of the 3′ sequence of the target strand were complementary to the non-target strand to allow for duplex formation.

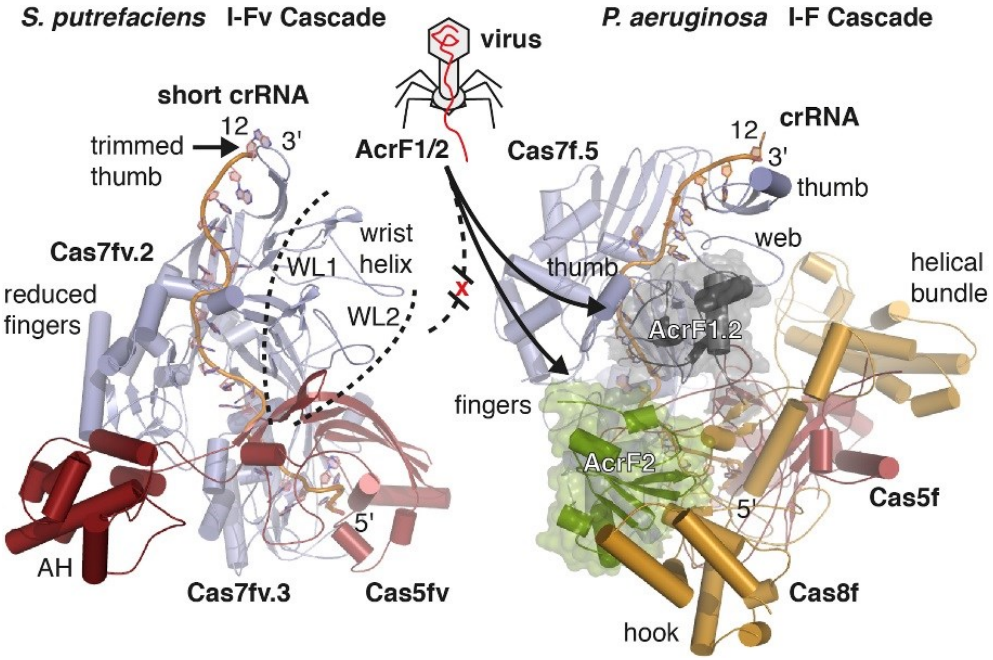


Figure III.3. Structural Cascade variation in response to AcrF1 and AcrF2. Side-by side comparison of the I-Fv Cascade base structure and the homologous and AcrF1/2 vulnerable I-F Cascade base structure (PDB ID: 5ZU9; (Chowdhury et al., 2017). Both structures are shown in cartoon representation and the Cascade body and head structures were removed for clarity. The Cas proteins were colored according to figure III.2. The anti-CRISPR proteins AcrF1.2 (grey) and AcrF2 (green) are highlighted by transparent surfaces. Regions to which AcrF1/2 associate in I-F Cascade (i.e. Cas7f thumb and web; Cas8f) differ drastically in their in their structure to I-Fv, not allowing binding of AcrF1/2 to the I-Fv Cascade.

The duplex contained the 'GG' PAM recognized by type I-Fv Cascade in *S. putrefaciens* CN32 (Gleditzsch et al., 2016) (Fig. S.III.1). After annealing of target and non-target strands and incubation with type I-Fv Cascade, the resulting R-loop/Cascade complex was separated from the individual compounds by size exclusion chromatography. The R-loop/Cascade complex was crystallized and its X-ray structure was determined at 3.25 Å resolution by molecular replacement using Cascade type I-Fv (Table III.1). The phosphate backbone and the nucleobases of the DNA are well defined in the electron density map and both DNA strands could be unambiguously traced (Fig. S.III.6). At the base of the short I-Fv R-loop/Cascade, the dsDNA section downstream of the GG-PAM is pinched in between the RRM fold and the AH domain of Cas5fv (Fig. III.4A). Upstream of the PAM, target and non-target strand bifurcate and are guided along the Cas7fv backbone in two spatially separated routes (Fig. III.4A) with a maximum distance of approximately 25 Å (non-target strand T22/ target strand A7). The target strand protospacer region associates with the short crRNA spacer and the non-target strand is aligned to the wrist helix, leading to structural reorganizations that are detailed below.

At the head of the R-loop/Cascade, the 5′ region of the target DNA protospacer is associated with the respective crRNA spacer nucleotides and pinches the thumb of Cas7fv.2 in between crRNA and target (Fig. III.4A, C). The thumb of Cas7fv.2 is in turn

stabilized and establishes a salt bridge by the interaction of arginine 155 to glutamate 17 of the ferredoxin-like domain of Cas6f. In comparison to the Apo structure, the ferredoxin-like domain is rotated by approximately 7 Å and stabilized (Figs. III.4A, III.5A) as evidenced by significantly reduced B-factors. These conformational rearrangements might be relevant for target DNA association and Cascade stalling via R-loop retention.

Cas7fv accommodates the protospacer and guides the non-target strand in a trench-route along the wrist helix

Along the body of the type I-Fv Cascade, Cas7fv assembles two parallel helices that recruit the separated target and non-target DNA strands (Fig. III.4A). Target association in the positively central channel formed by the palm/thumb helix is similar to the mode observed for the E. coli type I-E Cascade (Hayes et al., 2016; Mulepati et al., 2014) and relies on a set of DNA sequence independent interactions. Following the target DNA path from the PAM, the thumbs of Cas7fv splay out every sixth protospacer nucleotide, analogously to the respective crRNA nucleotides (Fig. III.4A, D). Interaction of target protospacer and crRNA relies on hybridization of the nucleotides in between those kinks and is further stabilized by aromatic residues emanating from the Cas7fv thumb (Y149, F160, F161) that stack the nucleobases in place, which lie in 5′ adjacent to the crRNA and DNA kinks (Fig. III.4D). Similar to the interaction at the Cascade head structure between the ferredoxin-like fold of Cas6f and the thumb of Cas7fv.2, arginine 155 at the thumb tip forms a salt bridge with the adjacent Cas7fv aspartate 192 residue (Fig. III.4D). To our surprise, we found that Cas7fv facilitates not only target DNA interactions, but also parallel guiding of the non-target strand via the novel wrist helix formed by the extensive loop structure at the concave site of the complex (Fig. III.4A, E). The non-target strand is passed along a path formed between the WL1 and WL2, stabilized by sequence unspecific interactions via tyrosine 62 and 64 (Fig. III.4E). Taken together, our R-loop/Cascade structure shows that the wrist helix establishes the trench route for the non-target strand and therefore compensates for the loss of the large and small subunits in type I-Fv.

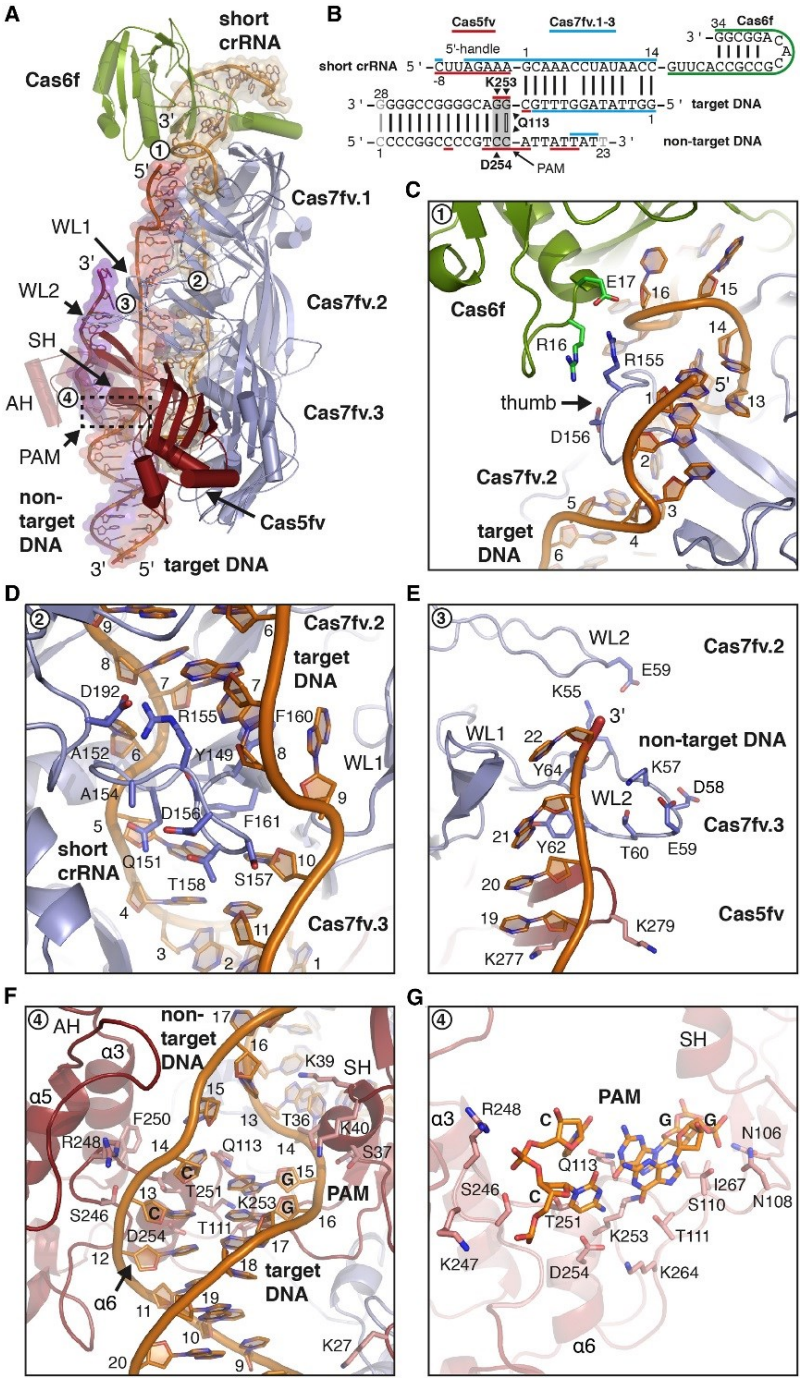


Figure III.4. X-ray crystal structure of the short I-Fv Cascade R-loop complex. (A) Overview of the R-loop Cascade crystal structure. Components are shown in cartoon representation and colored and labeled according to figs. III.1 and III.2. Nucleic acid components are highlighted for clarity by transparent sufaces (orange: short crRNA; red: DNA target strand; violet: non-target strand). Important regions for nucleic acid interaction, detailed in subfigs. III. 3C to G, are tagged with numbers in white circles (1-4) for orientation. **(B)** Design of target and non-target primers for the reconstitution of the R-loop/ I-Fv Cascade complex. The blue, red and green lines indicate the interface between Cas7fv, Cas5fv and Cas6f and the nucleic acids as observed in the R-loop/ I-Fv Cascade structure. Arrows indicate amino acids of Cas5fv interacting with the 'GG' PAM. Grey letters indicate disordered nucleotides. **(C) to (F)** Close up view of the DNA interacting regions close to the Cas6f head structure (C), at the thumb of Cas7fv.3 (D), the base of the wrist helix (E) and the PAM recognition site in between the AH and SH of Cas5fv (F). Amino acid side chains in close proximity to nucleic acids are shown as sticks and are labeled according to their identity and position. Nucleic acids are labeled according to subfig. III.3B. G: Detailed view on the GG-PAM, shown in stick representation. Adjacent nucleotides were removed for clarity.

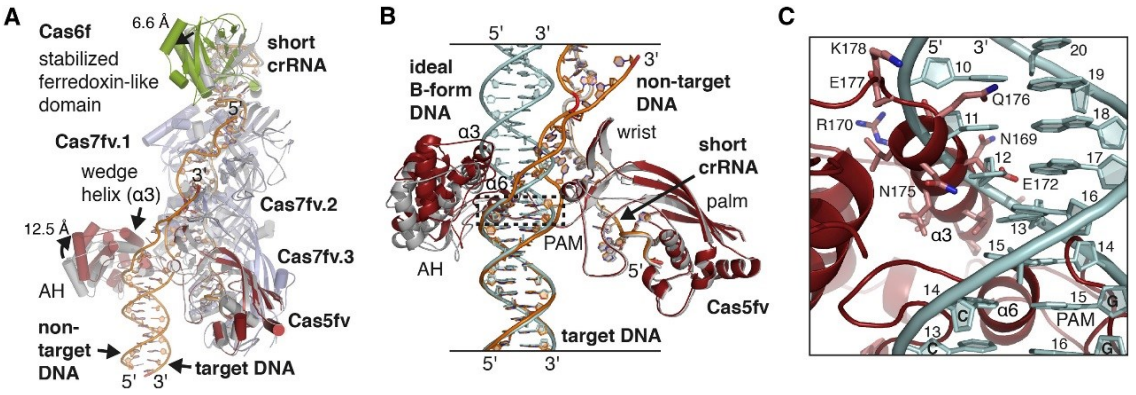


Figure III.5. Structural reorganization of type I-F Cascade upon R-loop formation. (A) Superimposition of the short Apo Cascade (grey cartoon) and the R-loop associated short Cascade (colored cartoon). I-Fv Cascade undergoes structural rearrangements at the Cas6f head and Cas5fv AH domain, indicated by arrows and distances. **(B)** Superimposition along dsDNA segment of the aligned apo (grey) and R-loop bound (colored) short Cascades with an ideal B-form DNA (light blue). The PAM region is highlighted with a dashed line. **(C)** Close up on the superimposition with an ideal B-form DNA (fig. III.4B) upstream of the GG PAM, emphasizing steric clashes that would occur between the target strand DNA and α-helix 3 (wedge helix) of the Cas5fv AH domain upon target and non-target strand association. Residues are labeled according to their identity and position.

Cas5fv recognizes the 'GG' PAM from the major groove site

In our R-loop/Cascade structure, the PAM containing DNA duplex is pinched between the AH domain and a small helix (SH) at the wrist of Cas5fv and recruited by a set of polar interactions (Figs. III.4A, F). In contrast to type I-E, the 'GG' PAM is recognized in the duplex form from the major groove side by the N-terminal linker and α-helix 6 of the AH of Cas5fv (Figs. III.4A, F). In the center, glutamine 113 distorts the first PAM guanosine (G15) of the target and its corresponding cytosine on the non-target (C14) (Figs. III.4F, G). Guanosine G15 is furthermore contacted by lysine 252. The second guanosine/cytosine (G16/C13) pair of the PAM interacts with lysine 252 and aspartate 253 of the C-terminal helix of the AH, respectively (Figs. III.4F, G). Those central GG-PAM recognizing residues are flanked at the AH domain by the helix α6 emanating side chains of threonine 251 and aspartate 254, which might also contribute to PAM recognition (Figs. III.4F,G). Superimposition of the Apo and the R-loop states revealed that the AH domain is reoriented by approximately 12 Å towards the Cascade body (Fig. III.5A). This results in placement of α-helix 3 in a position that would not allow re-association of the complementary non-target and target protospacer regions at the seed region, as illustrated by an ideal B-form dsDNA aligned to the dsDNA PAM downstream region (Fig. III.5B). Polar side-chains (N178-K178) might assist in DNA strand separation (Fig. III.5C) and α-helix 3 was therefore termed 'wedge helix'. Taken together, type I-Fv recognizes target DNA from the major groove site via Cas5fv, which is in clear contrast

to all other Cascades characterized so far. Cas5fv evolved from a 5' crRNA capping protein into a multifunctional subunit compensating for the loss of the large subunit Cas8.

Cas5fv is critical for type I-Fv Cascade-mediated interference in vivo

To evaluate our conclusions drawn from structural analysis, we analyzed the in vivo relevance of Cas5fv domains and –residues by the 'Efficiency of Transformation' (EOT) assay. EOT assay is set up, by providing one plasmid that facilitates Cascade production and a second plasmid containing a mini-CRISPR, consisting of two repeats that flank a spacer complementary to the β-lactamase sequence (Fig. III.6A). This set-up generates active Cascade ribonucleoproteins that target the β-lactamase gene of a pETDuet-1 vector upon transformation. We verified that the EOT of this target vector was drastically reduced in comparison to an inactive Cas2/3 variant with a mutation in its catalytic HD domain (Figs. III.6B, C). EOT assay revealed that mutation of K39/K40, Q113 and K277 to alanine did not result in reduction of transformation efficiency compared to WT Cas5fv (Fig. III.6C). A modest reduction in activity was observed for the mutation of K253, T251 and D254 to alanine, while the triple mutant of the respective residues was not active, comparable to the activity of the Cas2/3 HD mutant control (Fig. III.6C). Complete deletion of the Cas5fv AH domain (ΔAH) via substitution of aminoacids 122-259 by a 'GGSGGS' linker completely abolished Cascade activity, providing further evidence for the essential function of the Cas5fv AH domain in I-Fv Cascade mediated interference (Figs. III.6B, C). To rule out potentially misassembled Cascade complexes, size exclusion chromatography (SEC) of heterologously expressed and affinity-purified I-Fv Cascade ΔAH was performed. SEC clearly shows that deletion of the AH domain of Cas5fv did not affect formation of the Cascade complex (Fig. III.6D).

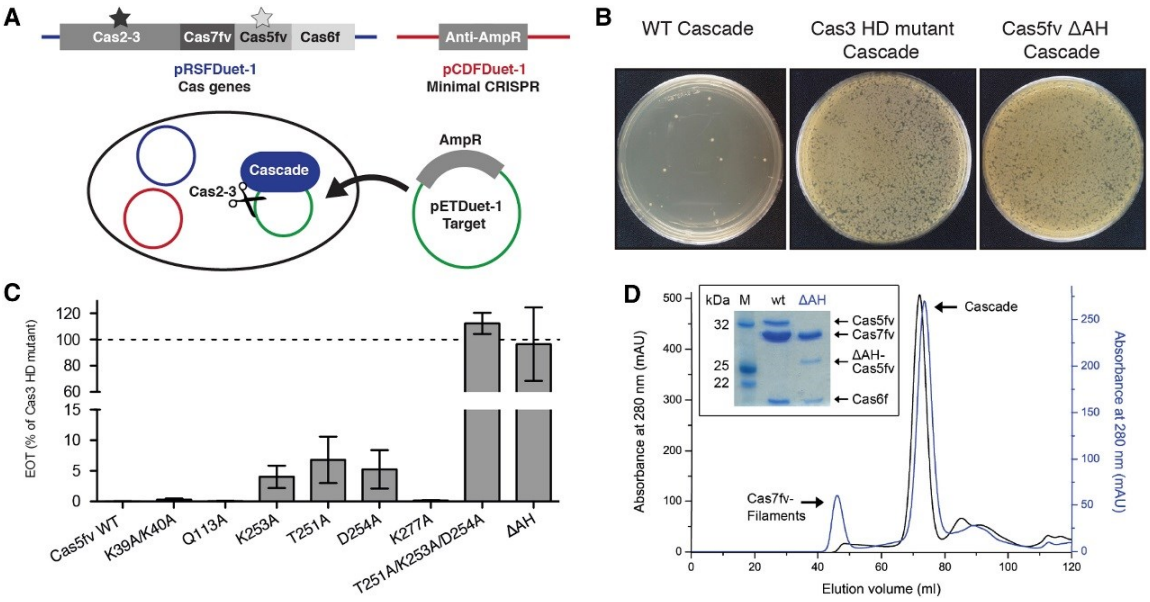


Figure III.6. Efficiency of transformation (EOT) of Cas5fv mutants. (A) Schematic representation of the EOT assay. BL21-AI E. coli cells carrying the wild type or mutant variants (represented with stars) of the cas genes (in pRSFDuet-1) and a minimal CRISPR array (in pCDF-Duet-1) are transformed with a third target plasmid (pETDuet-1) and plated in the presence of the target plasmid resistance antibiotic. After over night incubation, the activity of the wild type and mutant complexes is determined by calculating the EOT as a ratio of the colonies of the strain of interest versus the corresponding Cas HD domain mutant. **(B)** Exemplary EOT plates of the positive control (wild type Cascade), its matching negative control (Cas3 HD mutant Cascade) and the Cas5fv ΔAH plates are shown after over night growth. **(C)** EOT calculation for the investigated mutants. Assays were performed in triplicate and error bars were calculated as standard error of the mean (SEM). **(D)** Stability of the ΔAH domain mutant. Recombinant Cascade variants (containing either His-Cas5fv or His-Cas5fv ΔAH) were purified via Ni-NTA affinity chromatography and gel-filtration. Fractions of the dominant gel-filtration peak were separated via SDS-PAGE and revealed stable Cascade complex formation.

Discussion

Novel mechanism of foreign DNA recognition by type I-Fv Cascade

CRISPR-Cas mediated immunity relies on the discrimination of self from foreign nucleic acids on the basis of PAM recognition. PAM recognition by type I-E Cascade from *E. coli* relies on minor groove DNA interactions of the large subunit protein Cas8e (Hayes et al., 2016). Cryo-EM analysis of the minimal type I-C Cascade from *Desulfovibrio vulgaris* suggested a very similar mechanism of PAM recognition by Cas8c (Hochstrasser et al., 2016). These observations implied that minor groove recognition of a PAM in double-stranded form is likely a common feature of all type I systems. However, our study shows that the type I-Fv Cascade deviates from other characterized type I systems, in that PAM recognition proceeds from the major groove side of the dsDNA (Fig. III.7). In the *E. coli* type I-E Cascade, PAM readout from the minor groove side is required for promiscuous PAM recognition to respond to several distinct PAM sequences. The type I-Fv Cascade,

however, is strictly dependent on the 'GG' PAM for interference activity (Gleditzsch et al., 2016). Reading the PAM from the major groove side of the dsDNA might result in improved accessibility and specificity of dsDNA sequence readout also allowing shorter PAM sequences (e.g. 'GG' PAM).

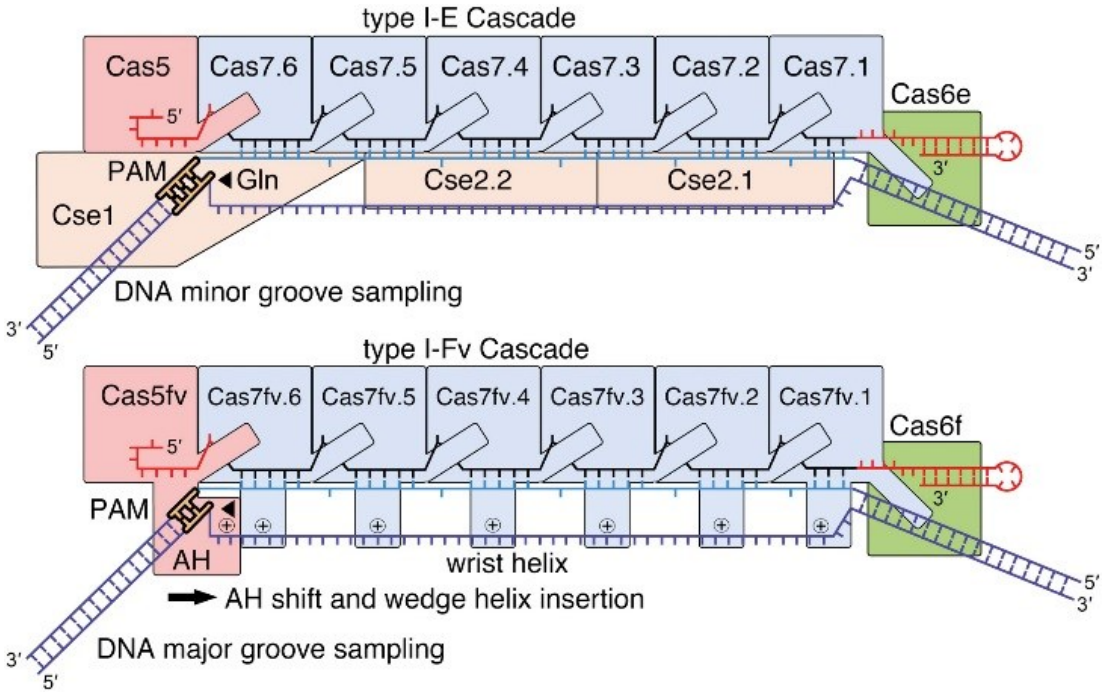


Figure III.7. R-loop formation by type I-E and I-Fv Cascade. Side-by-side comparison of components and mechanism of target DNA recognition by type I-E and I-Fv Cascade as deduced from structural analyses. Upper panel: Mechanism of surveillance and R-loop formation by the *E. coli* type I-E Cascade (compare to Ref: (Hayes et al., 2016)). A glutamine wedge (arrow) of the large subunit Cse1 (Cas8e) opens the R-loop and the non-target strand is guided by small subunits Cse2. Lower panel: The minimal *S. putrefaciens* type I-Fv Cascade structure revealed a shift of the AH domain upon target DNA binding, which results in wedge helix insertion (arrow). The non-target strand is guided along a trench route formed by Cas5fv and Cas7fv wrist loops.

Following PAM recognition, R-loop formation has to be initialized to sample the DNA for spacer complementarity. In E. coli type I-E Cascade, Cas8e-mediated PAM recognition is coupled to the insertion of a glutamine wedge that is present within the N-terminal domain of Cas8e, and disruption of the first two nucleotides of the protospacer (Fig. III.7) (Hayes et al., 2016). This step was proposed to lead to a directional melting of the dsDNA, which in turn allows for target/crRNA heteroduplex formation. This initial duplex and favorable sugar-phosphate contacts of the non-target strand lead to stabilization of the so-called seed bubble. Further DNA unwinding and protospacer association to the crRNA is accompanied by non-target DNA strand sequestration to the trench route formed by the small and large subunits to stabilize the R-loop/Cascade complex. Interestingly, the R-loop conformation in the minimal type I-Fv Cascade is roughly comparable to the one seen in the type I-E Cascade (Fig. S.III.7). Therefore, we

wondered how type I-Fv Cascade might initialize and execute R-loop formation in the absence of the small and large subunits and a completely different trench route formed by the wrist helix of Cas5fv/Cas7fv. We hypothesize that R-loop formation in type I-Fv Cascade proceeds by a very similar mechanism (Fig. III.7): i. Cascade samples dsDNA for a valid PAM 'GG' sequence through the major groove side via the AH and wrist of its Cas5fv, ii. recognition of the 'GG' PAM is coupled to a 12.5 Å-distanced, conformational rearrangement of the AH of Cas5fv resulting in the insertion of its α-helix 3 (wedge helix; Fig. III.7) into the major groove leading to seed bubble formation, iii. protospacer sampling for crRNA spacer complementarity leads to segmented crRNA/target heteroduplex formation, and iv. the simultaneous association of the non-target DNA strand to the trench route formed by the wrist helix of type I-Fv Cascade. Productive R-loop formation and locking is a prerequisite for effector complex formation with Cas3 to degrade the bound DNA.

Structural diversification of type I-F Cascade mediated DNA surveillance

Type I-Fv CRISPR-Cas systems are present in β- and γ- proteobacteria species and mobile genetic elements (Makarova et al., 2015). From an evolutionary point of view, type I-Fv CRISPR-Cas is one variant of the type I-F system. Interestingly, type I-F and I-Fv systems are found in closely related species (Dwarakanath et al., 2015). E.g. in the genus Shewanella, a type I-F system is present in the genome of *Shewanella* W3-18-1, whereas *Shewanella putrefaciens* CN32 contains a type I-Fv system (Fig. S.III.8). Similar scenarios are found in e.g. in Vibrio or Pseudomonas species. While the adaption modules (i.e. Cas1, Cas2) and Cas6f of type I-F and I-Fv share a high degree of sequence similarity, the effector module is strikingly different (Fig. S.III.8). While non-canonical functions of CRISPR-Cas systems, e.g. transcriptional control, stress response and pathogenicity development might contribute to the evolution of CRISPR-Cas systems, the continuous arms race between bacteria and invading nucleic acids is considered to be a key factor (Koonin and Wolf, 2015, Mohanraju et al. 2016, Takeuchi et al. 2012, Westra et al. 2014 , Westra et al. 2016). Hence, it is plausible that the diversification of the Cascade architecture is influenced by the selective evolutionary pressure exerted by the presence of anti-CRISPR proteins that are produced by viruses to paralyze the CRISPR-Cas mediated defense. Our structural comparison of type I-Fv Cascade with a type I-F structure bound to AcrF1/AcrF2 supports this hypothesis (Fig. III.3). The anti-CRISPR proteins AcrF1, 2 and 3 target type I-F effector complex formation via multiple mechanisms (Bondy-Denomy et al., 2015; Bondy-Denomy et al., 2013; Chowdhury et al., 2017; Wang et al., 2016). Our structural and mechanistic analyses show how the type I-Fv system might have evolved in order to counter viral predation.

All Cascade components that are vulnerable to AcrF1 and AcrF2 in type I-F Cascade are in the type I-Fv structure either removed (the large subunit) or drastically modified (Cas7 and Cas5). It is evident that the Cas3 helicase/nuclease must have also changed in response to the varied I-Fv Cascade interface and potentially the presence of Acr3 protein, which is underlined by very limited sequence homology of Cas3 between the type I-F and I-Fv systems (Dwarakanath et al., 2015) (Fig. S.III.8). However, the structural changes that occurred in Cas3 during the molecular arms race underlying the evolution of CRISPR-Cas systems have to be further analyzed. Moreover, it is yet unknown whether anti-CRISPR proteins evolved against the type I-Fv complex. Different AcrF proteins have been identified in Shewanella species, but their modes of action remain to be characterized (Pawluk et al., 2016). Anti-CRISPR proteins against type I-Fv complexes have not been found. Taken together, our study sheds light on the molecular evolution of CRISPR-Cas systems and reveals a surprisingly high level of structural diversity that might have occurred in response to the ongoing arms race between bacteria and viruses.

Materials and Methods

Production and purification of Cascade

Cascade was produced by co-production of N-terminally hexa-his tagged Cas7fv, Cas5fv and Cas6fv from pRSFDuet-1 (pCas) and the simultaneous expression of a pre-crRNA substrate, containing a 18 nucleotide spacer, from pUC19, as described previously (Gleditzsch et al., 2016). Co-transformed *E. coli* BL21 (DE3) cells (Novagen) were inoculated into two liters of LB medium, supplemented with 25 g lactose, ampicillin (100 mg/l) and kanamycin (50 mg/l). Cells were incubated at 30°C overnight under rigorous shaking (150 revolutions per minute (rpm)). Cells were harvested by centrifugation (3,500 x g, 20 min, 4°C) and resuspended in 20 mL buffer A (20 mM HEPES-Na pH 8.0, 250 mM NaCl, 20 mM KCl, 20 mM MgCl2, 40 mM imidazole) before lysis in a M-110L Microfluidizer (Microfluidics). The lysate was cleared at 47,850 x g for 20cmin at 4°C and the supernatant was applied onto two 1 mL HisTrap HP columns (GE Healthcare) for Ni-NTA affinity chromatography. After a wash step with ten column volumes (CV) of buffer A, proteins were eluted with three CV of buffer B (20cmM HEPES-Na pH 8.0, 250cmM NaCl, 20cmM KCl, 20cmM MgCl2, 500cmM imidazole). The Cascade complex was further purified by size exclusion chromatography (SEC) using a HiLoad 26/60 Superdex 200 column (GE Healthcare) equilibrated in buffer C (20cmM HEPES-K pH 7.3, 100cmM

NaCl). The main peak fractions were concentrated to an absorbance at 280 nm of 35 arbitrary units (AU) as determined by a NanoDrop Lite Spectrophotometer.

Seleno-methionine labeled Cascade was produced following a protocol described earlier for other proteins (Bange et al., 2010, Pausch et al., 2015). Shortly, the co-transformed *E. coli* BL21(DE3) cells were grown in M9-Se-Met labeling media (46.7 g Na2HPO4, 16.5 g KH2PO4, 2.75 g NaCl, 5.5 g NH4Cl, 500 mg lysine, 500 mg threonine, 500 mg phenylalanine, 250 mg leucine, 250 mg isoleucine, 250 mg valine, 250 mg L-seleno-methionine, 10 mM MgCl2, 0.5 mM CaCl2 and 20 g glucose in 5 l ddH2O). Cultures were grown shaking at 150 rpm and 37°C to an optical density at 600 nm (OD600) of 0.5 before induction of Cascade production by the addition of 1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) and incubation at 150 rpm and 37°C overnight. Se-Met labeled Cascade was purified as described above for its native counterpart.

Reconstitution of the R-loop/Cascade complex

Cascade was produced and purified as described above, concentrated to an A280 of 70 AU (NanoDrop Lite Spectrophotometer). Target and non-target DNA primers synthesized by Sigma-Aldrich were dissolved in ddH2O to a final concentration of 1 mM. Next, target and non-target were mixed in a 1:2 molar ratio, respectively, and heated at 95°C for ten minutes. Thereafter, the mixture was slowly cooled down to room temperature. The annealed primers were combined with short Cascade in a 1.5 to 1 stoichiometry and incubated at 20°C for 30 min to allow for complex formation. Another step of SEC (HiLoad 26/60 Superdex 200, GE Healthcare) equilibrated in buffer C removed non-bound DNA fragments. R-loop/Cascade complex containing peak fractions were concentrated to an A280 of 40 AU (NanoDrop Lite Spectrophotometer) and directly used for crystallization.

Crystallization

Purified Cascade was concentrated to an absorbance at 280 nm of 35 AU (NanoDrop Lite Spectrophotometer) and subjected to crystallization by hanging drop vapor-diffusion at 20°C. Crystallization of the Se-Met labeled Cascade was achieved in a two-step protocol, based on the initial formation of seed-crystals derived from native Cascade complexes. Seed crystals were generated by mixing equal volumes (1 µl) of Cascade sample and crystallization buffer (16% w/v PEG6000, 0.1 M Tris pH 8.0 and 20 mM 5-amino-2,4,6-triiodoisophthalic acid). Sword shaped seed crystals grew over night and were subsequently used for streak seeding with a cat whisker into crystallization drops containing Se-Met labeled Cascade. Crystals grew within days in drops containing 1 µL of Se-Met labeled Cascade sample and 1 µL crystallization buffer (18% w/v PEG6000,

0.1 M Tris pH 7.6 and 20 mM 5-amino-2,4,6-triiodoisophthalic acid). Crystals were transferred into crystallization buffer containing 20% v/v glycerol as cryo-protectant, subsequently flash frozen and stored in liquid nitrogen. R-loop/Cascade samples were used at an absorbance at 280 nm of 30 AU (NanoDrop Lite Spectrophotometer) and crystallized by hanging drop vapor-diffusion at 20°C. Needle shaped crystals grew within days in drops containing 1 µL of R-loop/Cascade and 1 µL crystallization buffer (22.5% w/v PEG4000, 15% v/v glycerol, 153 mM ammonium acetate and 85 mM sodium citrate pH 5.6). R-loop/Cascade crystals were flash-frozen and stored in liquid nitrogen.

Data collection and structure determination

Diffraction data were collected at beamline ID29 of the European Synchrotron Radiation Facility (ESRF), Grenoble, France. Data were processed with the XDS program package for data reduction (Kabsch, 2010), Crank2 for experimental phasing of Se-Met labeled Cascade (CCP4 package, Winn et al., 2011), coot (Emsley and Cowtan, 2004) in combination with Refmac5 (CCP4 package) and phenix.refine (PHENIX package) for iterative model building and refinement (Adams et al., 2010). The R-loop/Cascade dataset was solved by molecular replacement using the Cascade structure via the CCP4 implemented program Phaser (McCoy et al., 2007). Figures were prepared in Pymol.

Electrophoretic mobility shift assays (EMSA)

Radioactive labeling of DNA substrates and subsequent analysis by electrophoretic mobility shift assays was performed as previously described (1). R-loop constructs were created by hybridizing the target strand sp4-tar:

5′-GGTTATAGGTTTGCGCGTCTTGCTGGGCGATAGGACTCCCTATAGTGAG-3′,

or sp1-tar:

5′-CAATGTGGTCGCGCAATTTATGATTTGGTTGAGGACTCCCTATAGTGAG-3′

and the radioactively labeled non-target strand:

5′-CTCACTATAGGGAGTCCATTATTATTT-3′

utilizing the 17 nt complementary region at the 3′ end of the target strand. A mixture of non-target and target strand (1:1.5) was heated at 95°C for 10 min and then slowly cooled down to room temperature for 2 hr.

Bacterial strains and growth conditions for EOT Assay

E. coli BL21-AI strains (F- ompT hsdSB (rB-mB-) gal dcm araB::T7RNAPtetA, Invitrogen) were used for efficiency of transformation (EOT) assays and grown in 2YTL media (16 g/l tryptone, 10 g/l yeast extract, 5 g/l NaCl, 10 mM MgSO4, 0.2% maltose) supplemented with appropriate antibiotics at 37°C.

Efficiency of Transformation Assays

The pRSFDuet-1 plasmid encoding Cas7fv, Cas5fv and Cas6f and wild-type Cas2-3 (pCas) or an inactive Cas2-3 HD domain mutant (pCasHD, containing C466G, A467C, A470C mutations in cas2-3) were used as templates to introduce point mutations in the sequence of Cas5fv. Alanine replacements of amino acids in Cas5fv were performed by QuikChange site-directed mutagenesis, according to the manufacturer's protocol. In parallel, the deletion of a fragment between amino acids 121 and 259 in Cas5fv (ΔAH mutant) was obtained by inverse PCR on pCas and pCasHD. A short linker (GGSGGS) was inserted at the site of deletion. Additionally, a minimal CRISPR was designed with two repeat sequences flanking a single 32 nt spacer targeting the ampicillin resistance cassette of pETDuet-1 (pCRISPR Anti-AmpR) and cloned into pCDFDuet-1 by annealed oligonucleotides using NcoI and NotI restriction sites. *E. coli* BL21-AI cells were transformed with pCRISPR Anti-Amp plus pCas, pCasHD or the Cas5fv variants. Single colonies of the obtained strains were cultured overnight, diluted 1:500 on a main culture of 2YTL supplemented with 10 mM of MgCl2 and MgSO4 and grown until OD600 = 0.3. Subsequently, 0.1 mM of IPTG and 0.2% of arabinose (Sigma-Aldrich) were added for induction. After OD600 = 0.6 was reached, rubidium chloride competent cells were prepared. 100 µl aliquots of each strain were transformed via heatshock with 50 ng of pETDuet-1 and, after 1 hr of recovery in LB, 100 µl were plated in 2YTL plates with ampicillin (50 µg/ml), kanamycin (25 µg/ml) and spectinomycin (25 µg/ml). After overnight incubation at 37°C, the amount of colonies per plate was counted and the Efficiency of Transformation (EOT) was determined as the ratio between colony count of the strains of interest carrying the wild-type Cas3 and the corresponding ones with the Cas3 HD mutation. Assays were performed in triplicate and error bars were calculated as standard error of the mean (SEM).

Quantification and statistical analysis

Statistics of the X-ray crystallographic data collection and processing, refinement, and structure validation are reported in Table III.1. Errors in the EOT assay were derived from three independent experiments and are reported as standard error of the mean (SEM).

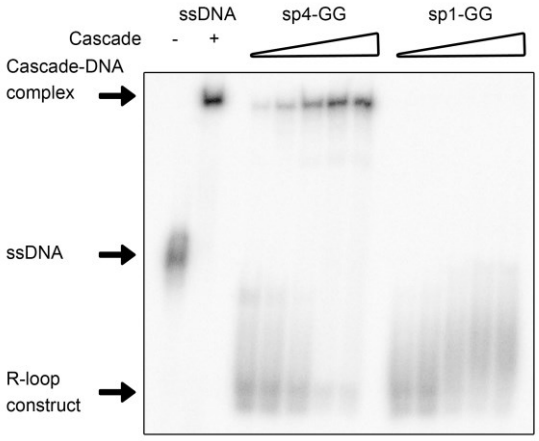
Data and software Availability

The accession numbers for the coordinates and structure factors for the Cascade type I-Fv alone and R-loop bound reported in this paper are PDB: 5O7H and 5O6U, respectively.

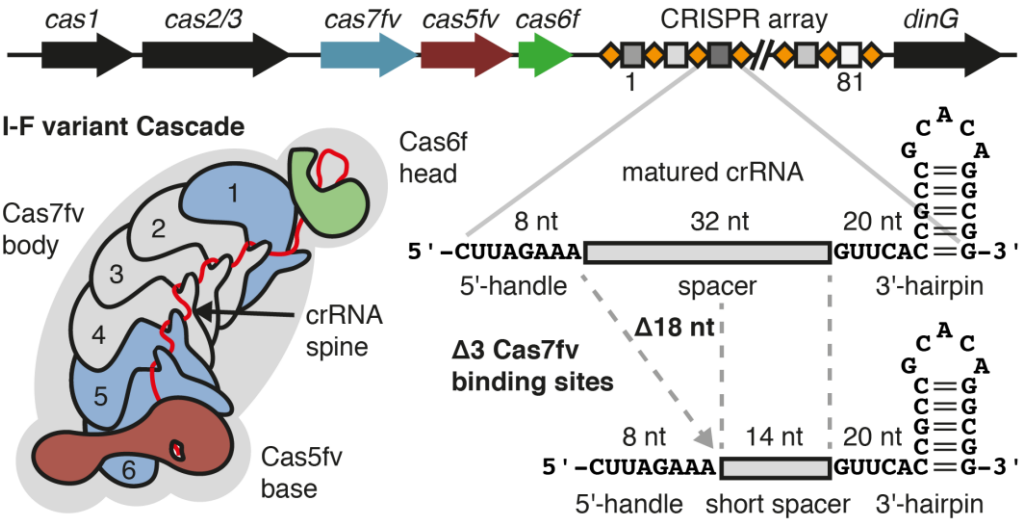
Acknowledgements

P.P. and H.M.-E. are recipients of a fellowship of the International Max-Planck-Research School for Environmental, Cellular, and Molecular Microbiology (IMPRS-Mic). D.G. and L.R. are supported by DFG FOR1680. G.B. thanks the LOEWE excellence initiative for financial support. G.B., P.P., and F.A. acknowledge the always excellent support by the European Synchrotron Radiation Facility (ESRF), Grenoble, France. We thank B. Wiedenheft (Montana State University, USA) for sharing the coordinates of type I-F bound to AcrF1 and AcrF2 prior to publication.

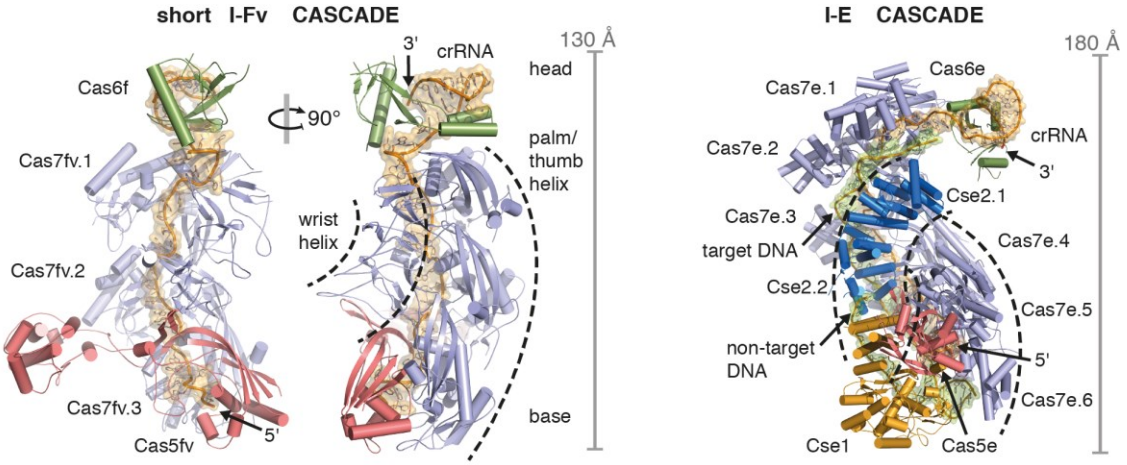
Supplementary Material



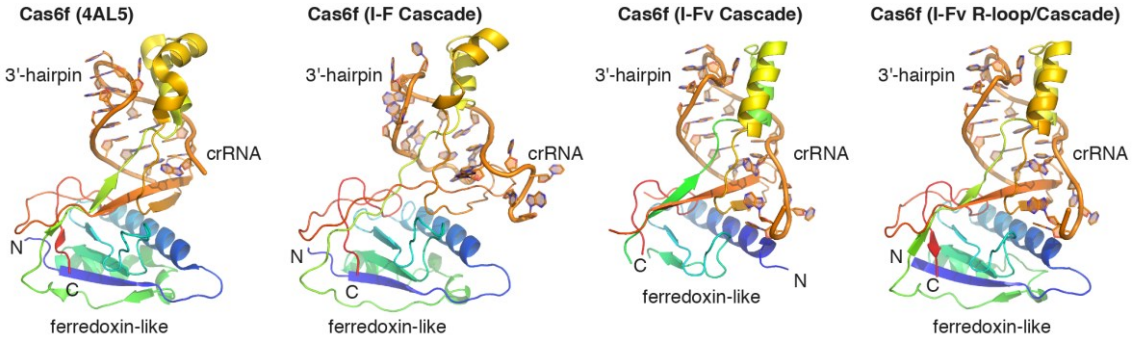
Supplementary Figure III.1. Related to Figure III.1. Recombinant Type I-Fv Cascade binds R-loop DNA duplexes. 5′-radioactively labeled DNA duplexes (~2 pmol) were incubated with increasing amounts (0, 10, 25, 50, 75 mM) of recombinant full-length Type I-Fv Cascade containing a crRNA with spacer 4 of the single S. putrefaciens CRISPRarray. DNA duplexes were constructed by hybridizing the target strand and the radioactively labeled non-target strand (5 min 95 °C, cooling to room temperature over 2h), utilizing a 17 nt complementary region at the 3 -end in the target strand adjacent to the crRNA-matching sequence. Binding of these DNA duplexes was observed by electrophoretic mobility shift assays for the construct featuring a complementary sequence to the Cascade bound crRNA and a 'GG'-PAM in the target strand (sp4-GG). No binding was observed for the construct containing the non-matching target sequence of spacer 1 (sp1-GG). A control (C) was included to confirm Cascade ssDNA substrate (~1 pmol) binding. The ssDNA was incubated without Cascade (-) or with excess Cascade (+).



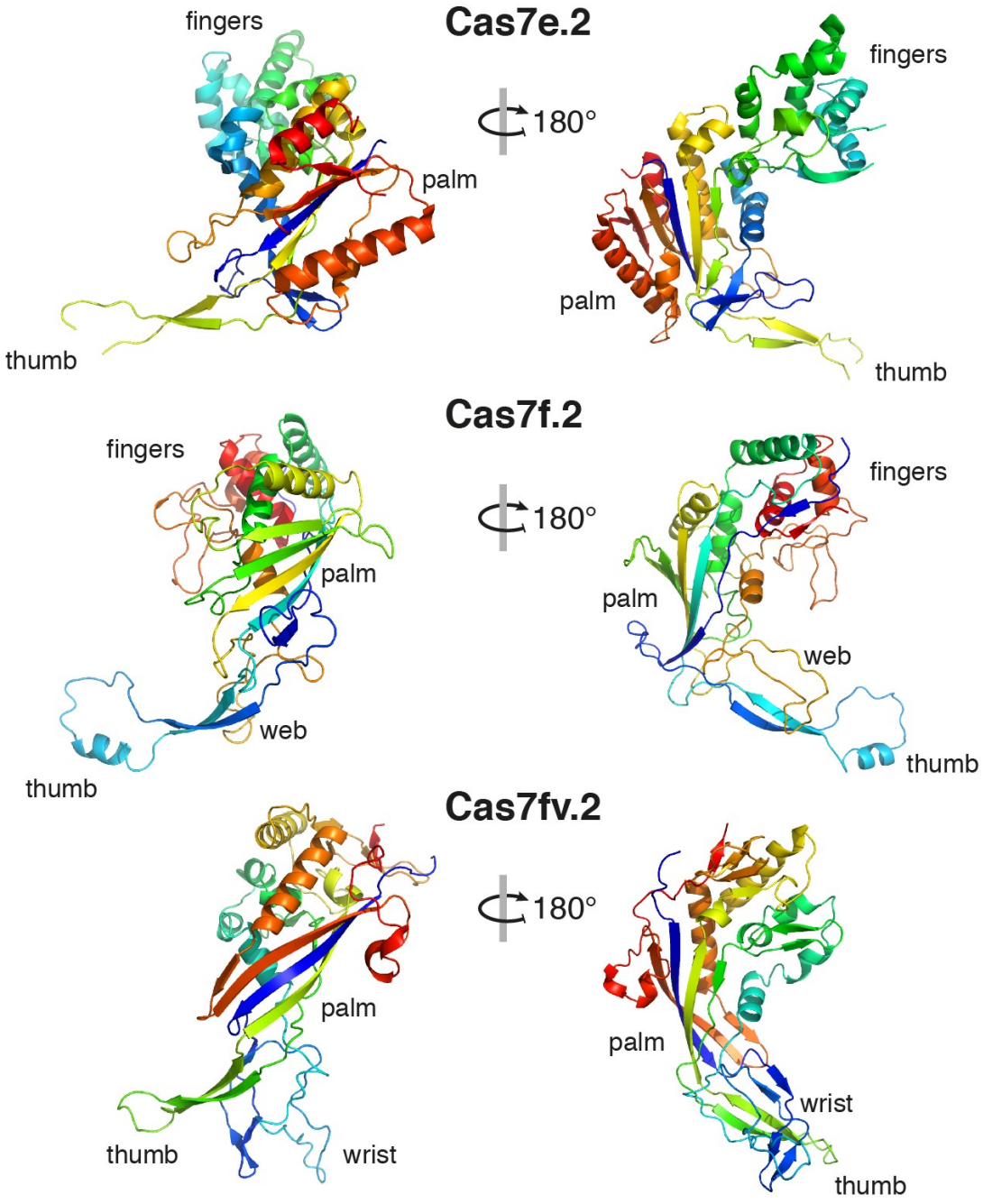
Supplementary Figure III.2. Related to Figure III.1. Design of the short type I-Fv Cascade. Above: S. putrefaciens CN32 CRISPR/Cas locus gene arrangement. Arrows indicate the respective cas genes and the Cascade genes cas5fv, cas7fv and cas6f are colored in dark red, blue and green, respectively. The 81 spacer spanning CRISPR array is indicated as alternating rhombi (orange, crDNA repeat) and squares (grey, crDNA spacer). Left: Schematic representation of the type I-Fv Cascade. The crRNA spine is shown in orange, Cas6f at the head is shown in green, Cas7fv proteins are shown in blue and labeled according to the number of Cas7fv subunits and Cas5fv at the base is shown in dark red. Grey color indicates the number of reduced Cas7fv subunits for the short Cascade construct. Right: crRNA design for the short Cascade that assembles along the 18 nt contracted crRNA.



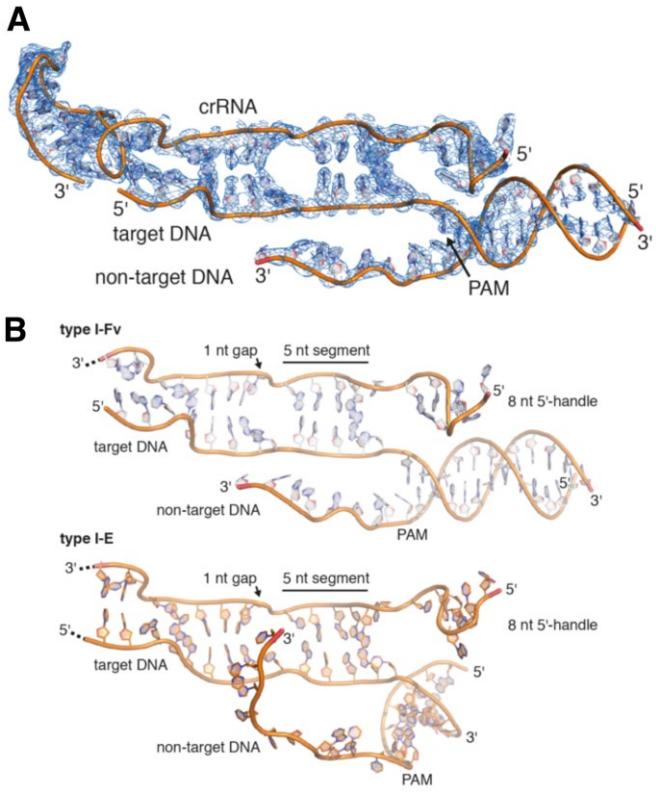
Supplementary Figure III.3. Related to Figures III.1 and III.2. Comparison of the short type I-Fv and type I-E Cascade. Left: The cartoon shows the crystal structure of the type I-Fv Cascade complex from *S. putrefaciens* CN32 in two different orientations. The crRNA, Cas6f, the three Cas7fv proteins and Cas5fv are shown in orange, green, blue and red, respectively. The 3′ and 5′ ends of the crRNA are indicated. Dotted lines indicate the wrist helix and palm/phumb helix formed by Cas7fv.1-3/Cas5fv. Right: Cartoon of the R-loop/type I-E Cascade from *E. coli*. The Cas7e.1-6, Cas6e, Cas5e, Cse1 (Cas8e) and Cse2.1-2 are in light blue, green, red, yellow and darkblue, respectively. The crRNA is represented as orange cartoon/surface. Target and non-target DNA are shown as light-green cartoon/surface. The coordinates are derived from PDB-ID: 5H9A.



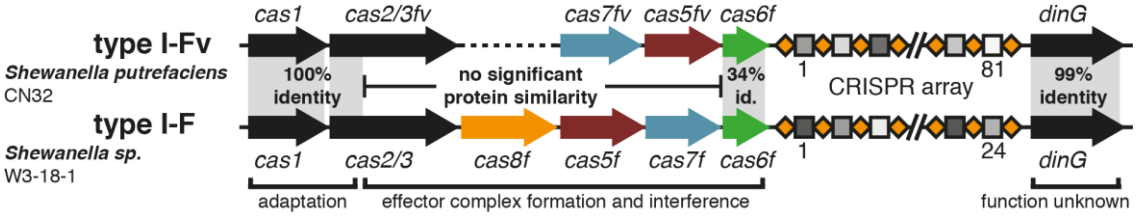
Supplementary Figure III.4. Related to Figure III.2. Comparison of available Cas6f structures. Cas6f and the cognate crRNA hairpin structure are shown as cartoon representations. The crRNA is colored in orange and the Cas6f proteins are rainbow-colored from blue (N-terminus, indicated as N) to red (C-terminus, indicated as C). From left to right: Crystal structure of *P. aeruginosa* type IF Cas6f bound to the crRNA hairpin in absence of the remaining Cascade components (PDB-ID: 4AL5); cryo-EM derived structure of *P. aeruginosa* type IF Cas6f bound to the crRNA hairpin in presence of the remaining Cascade components bound to AcrF (PDB-ID: 5ZU9); Crystal structure of *S. putrefaciens* type IFv Cas6f bound to the crRNA hairpin in presence of the remaining short Cascade components (this study); Crystal structure of *S. putrefaciens* type IFv Cas6f bound to the crRNA hairpin in presence of the remaining short Cascade components and hybridized target DNA (this study).



Supplementary Figure III.5. Related to Figure III.2. Structural comparison of type I Cascade associated Cas7. Cas7 proteins are shown as cartoon representations in two different, 180 ° rotated, orientations and are rainbow colored from blue (N-terminus) to red (C-terminus). Upper panel: Crystal structure of the type I-E Cascade associated Cas7e.2 from *E. coli* (PDB-ID: 4TVX). Middle panel: Cryo-EM derived structure of the type I-F Cascade associated Cas7f.2 from *P. aeruginosa* (PDB-ID: 5ZU9). Crystal structure of the type I-Fv Cascade associated Cas7fv.2 from *S. putrefaciens* (this study). Cas7 proteins are topologically divided according to the 'right-hand' analogy into: fingers, palm, thumb and in case of the novel Cas7fv, wrist. A web structure is furthermore found at the base of the palm in Cas7f.



Supplementary Figure III.6. Related to Figure III.4. A. Observed R-loop electron density. 2Fo-Fc electron density map (blue) of the R-loop contoured at 1.5 σ after refinement. DNA and RNA segments (cartoon) are labeled accordingly and termini are indicated. B. Comparison of R-loops formed by type I Cascades. DNA and RNA segments (cartoon) are labeled accordingly and termini are indicated. Upper panel: R-loop structure formed by the type I-Fv Cascade (grey colored nucleotides). Lower panel: R-loop structure formed by the type I-E Cascade (orange colored nucleotides; PDBID: 5H9E).



Supplementary Figure III.7. Related to Figure III.2. Comparison of the closely related I-Fv and I-F CRISPR/Cas loci from Shewanella. Schematic representation of the CRISPR/Cas loci in *S. putrefaciens* CN32 (type I-Fv) and *Shewanella sp.* W3-18-1 (type I-F). Arrows indicate the respective *cas* genes and the Cascade genes cas5, cas7, cas6f and cas8 are colored in dark red, blue, green and orange, respectively. The CRISPR arrays are indicated as alternating rhombi (orange, crDNA repeat) and squares (grey, crDNA spacer). Grey emphasized areas indicate sections of high protein sequence similarity. No significant protein sequence similarity is found between the components, which are required for effector formation and Cascade mediated interference, except for Cas6f.

References

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010).PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. *66*, 213–221.

Bange, G., Kümmerer, N., Engel, C., Bozkurt, G., Wild, K., and Sinning, I. (2010). FlhA provides the adaptor for coordinated delivery of late flagella building blocks to the type III secretion system. Proc. Natl. Acad. Sci. USA *107*, 11295–11300.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science *315*, 1709–1712.

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

Bondy-Denomy, J., Garcia, B., Strum, S., Du, M., Rollins, M.F., Hidalgo-Reyes, Y., Wiedenheft, B., Maxwell, K.L., and Davidson, A.R. (2015). Multiple mechanisms for CRISPR-Cas inhibition by anti-CRISPR proteins. Nature *526*, 136–139.

Carpenter, M.R., Kalburge, S.S., Borowski, J.D., Peters, M.C., Colwell, R.R., and Boyd, E.F. (2017). CRISPR-Cas and contact-dependent secretion systems present on excisable pathogenicity islands with conserved recombination modules. J. Bacteriol. *199*, 199.

Chowdhury, S., Carter, J., Rollins, M.F., Golden, S.M., Jackson, R.N., Hoffmann, C., Nosaka, L., Bondy-Denomy, J., Maxwell, K.L., Davidson, A.R., et al. (2017). Structure reveals mechanisms of viral suppressors that intercept a CRISPR RNA-guided surveillance complex. Cell *169*, 47–57 e11.

Dwarakanath, S., Brenzinger, S., Gleditzsch, D., Plagens, A., Klingl, A.,Thormann, K., and Randau, L. (2015). Interference activity of a minimal Type I CRISPR-Cas system from Shewanella putrefaciens. Nucleic Acids Res. *43*, 8913–8923.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. *60*, 2126–2132.

Garneau, J.E., Dupuis, M.E., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadan, A.H., and Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature *468*, 67–71.

Gleditzsch, D., Müller-Esparza, H., Pausch, P., Sharma, K., Dwarakanath, S., Urlaub, H., Bange, G., and Randau, L. (2016). Modulating the Cascade architecture of a minimal Type I-F CRISPR-Cas system. Nucleic Acids Res. *44*, 5872–5882.

Grissa, I., Vergnaud, G., and Pourcel, C. (2009). Clustered regularly inter-spaced short palindromic repeats (CRISPRs) for the genotyping of bacterial pathogens. In Molecular Epidemiology of Microorganisms, D.A. Caugant, ed. (Springer), pp. 105–116.

Hayes, R.P., Xiao, Y., Ding, F., van Erp, P.B., Rajashankar, K., Bailey, S., Wiedenheft, B., and Ke, A. (2016). Structural basis for promiscuous PAM recognition in type I-E Cascade from E. coli. Nature *530*, 499–503.

Hochstrasser, M.L., Taylor, D.W., Kornfeld, J.E., Nogales, E., and Doudna, J.A. (2016). DNA targeting by a minimal CRISPR RNA-guided cascade. Mol. Cell *63*, 840–851.

Jackson, R.N., Golden, S.M., van Erp, P.B., Carter, J., Westra, E.R., Brouns, S.J., van der Oost, J., Terwilliger, T.C., Read, R.J., and Wiedenheft, B. (2014). Structural biology. Crystal structure of the CRISPR RNA-guided surveillance complex from Escherichia coli. Science *345*, 1473–1479.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816–821.

Kabsch, W. (2010). Xds. Acta Crystallogr. D Biol. Crystallogr. *66*, 125–132.

Koonin, E.V., and Wolf, Y.I. (2015). Evolution of the CRISPR-Cas adaptive immunity systems in prokaryotes: models and observations on virus-host coevolution. Mol. Biosyst. *11*, 20–27.

Labbate, M., Orata, F.D., Petty, N.K., Jayatilleke, N.D., King, W.L., Kirchberger, P.C., Allen, C., Mann, G., Mutreja, A., Thomson, N.R., et al. (2016). A genomic island in Vibrio cholerae with VPI-1 site-specific recombination characteristics contains CRISPR-Cas and type VI secretion modules. Sci. Rep. *6*, 36891.

Makarova, K.S., Aravind, L., Wolf, Y.I., and Koonin, E.V. (2011a). Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. Biol. Direct *6*, 38.

Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J., Wolf, Y.I., Yakunin, A.F., et al. (2011b). Evolution and classification of the CRISPR-Cas systems. Nat. Rev. Microbiol. *9*, 467–477.

Makarova, K.S., and Koonin, E.V. (2015). Annotation and classification of CRISPR-Cas systems. Methods Mol. Biol. *1311*, 47–75.

Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., Costa, F., Shah, S.A., Saunders, S.J., Barrangou, R., Brouns, S.J., Charpentier, E., Haft, D.H., et al. (2015). An updated evolutionary classification of CRISPR-Cas systems. Nat. Rev.Microbiol.*13*, 722–736.

Marraffini, L.A., and Sontheimer, E.J. (2010). Self versus non-self discrimination during CRISPR RNA-directed immunity. Nature *463*, 568–571.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J. Appl. Cryst. *40*, 658–674.

Mohanraju, P., Makarova, K.S., Zetsche, B., Zhang, F., Koonin, E.V., and van der Oost, J. (2016). Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. Science *353*, aad5147.

Mojica, F.J., Díez-Villaseñor, C., García-Martínez, J., and Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. J. Mol. Evol. *60*, 174–182.

Mojica, F.J., Díez-Villaseñor, C., García-Martínez, J., and Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology *155*, 733–740.

Mulepati, S., Héroux, A., and Bailey, S. (2014). Structural biology. Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target. Science *345*, 1479–1484.

Pausch, P., Singh, U., Ahmed, Y.L., Pillet, B., Murat, G., Altegoer, F., Stier, G., Thoms, M., Hurt, E., Sinning, I., et al. (2015). Co-translational capturing of nascent ribosomal proteins by their dedicated chaperones. Nat. Commun. *6*, 7494.

Pawluk, A., Staals, R.H., Taylor, C., Watson, B.N., Saha, S., Fineran, P.C., Maxwell, K.L., and Davidson, A.R. (2016). Inactivation of CRISPR-Cas systems by anti-CRISPR proteins in diverse bacterial species. Nat. Microbiol.1, 16085.

Pourcel, C., Salvignol, G., and Vergnaud, G. (2005). 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.
Rollins, M.F., Schuman, J.T., Paulus, K., Bukhari, H.S., and Wiedenheft, B.(2015). Mechanism of foreign DNA recognition by a CRISPR RNA-guided surveillance complex from Pseudomonas aeruginosa. Nucleic Acids Res. *43*, 2216–2222.
Takeuchi, N., Wolf, Y.I., Makarova, K.S., and Koonin, E.V. (2012). Nature and intensity of selection pressure on CRISPR-associated genes. J. Bacteriol.*194*, 1216–1225.
Wang, X., Yao, D., Xu, J.G., Li, A.R., Xu, J., Fu, P., Zhou, Y., and Zhu, Y. (2016).Structural basis of Cas3 inhibition by the bacteriophage protein AcrF3. Nat. Struct. Mol. Biol. *23*, 868–870.
Westra, E.R., van Erp, P.B., Künne, T., Wong, S.P., Staals, R.H., Seegers, C.L., Bollen, S., Jore, M.M., Semenova, E., Severinov, K., et al. (2012). CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. Mol. Cell *46*, 595–605.
Westra, E.R., Buckling, A., and Fineran, P.C. (2014). CRISPR-Cas systems: beyond adaptive immunity. Nat. Rev. Microbiol.*12*, 317–326.
Westra, E.R., Downing, A.J., Broniewski, J.M., and von Houte, S. (2016). Evolution and ecology of CRISPR. Annu. Rev. Ecol. Evol. Syst. *47*, 307–331.
Wiedenheft, B., van Duijn, E., Bultema, J.B., Waghmare, S.P., Zhou, K.,Barendregt, A., Westphal, W., Heck, A.J., Boekema, E.J., Dickman, M.J., and Doudna, J.A. (2011). RNA-guided complex from a bacterial immune sys-tem enhances target recognition through seed sequence interactions. Proc.Natl. Acad. Sci. USA *108*, 10092–10097.
Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans,P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G., McCoy, A., et al. (2011). Overview of the CCP4 suite and current developments. Acta Crystallogr. D Biol. Crystallogr. *67*, 235–242

Chapter IV:
BioLayer Interferometry for studying the target binding of CRISPR-Cas effector complexes

Hanna Müller-Esparza[1], Manuel Osorio-Valeriano[3], Niklas Steube[1], Martin Thanbichler[2, 3], Lennart Randau[1, 2, *]

This chapter is written in manuscript style and is currently under preparation for publishing. My contribution to this work included conceiving, designing and performing the experiments, analyzing the obtained data and writing the manuscript.

---

[1] Prokaryotic Small RNA Biology Group, Max Planck Institute for Terrestrial Microbiology, D-35043 Marburg, Germany
[2] LOEWE Center for Synthetic Microbiology, Philipps University Marburg, D-35043 Marburg, Germany
[3] Faculty of Biology, Philipps University Marburg, D-35043 Marburg, Germany
* Corresponding Author

Abstract

CRISPR-Cas systems exhibit a high diversity of effector complexes with different target requirements and binding affinities. Streamlined analysis techniques of the interactions between nucleic acids and protein complexes allow for the characterization and comparison of CRISPR-Cas effectors. BioLayer Interferometry (BLI) is a technique that measures the interference pattern of white light reflected from a biolayer in real time and in solution. As streptavidin-coated biolayers and biotinylated oligonucleotides are commercially available, this method allows for studying the interaction of CRISPR-Cas complexes with different targets both qualitatively and quantitatively in a simple manner. Here, we present a general method to carry out binding assays with the Type I-Fv complex from *Shewanella putrefaciens* and the Type I-F complex from *Shewanella baltica* as model effectors. We report target specificities, dissociation constants and interactions with an Anti-CRISPR (Acr) protein, showcasing some of the possible applications of this technique.

Introduction

CRISPR-Cas systems are adaptive immune systems found in Archaea and Bacteria. They are widespread and diverse, with 6 types (I-VI) and 33 different subtypes described so far (Faure et al., 2019). They are able to establish immunity against invading genetic material through ribonucleoprotein (RNP) complexes, formed by a CRISPR RNA (crRNA) and CRISPR-associated (Cas) proteins, in a process called interference (Barrangou et al., 2007).

In line with the described diversity of effector complexes, different systems have different target requirements. Type I, II and V complexes scan for DNA targets and first recognize a 2-5 bp short motif next to the region with complementarity to the crRNA (protospacer), termed protospacer adjacent motif (PAM) (Anders, Niewoehner, Duerst, & Jinek, 2014; Hayes et al., 2016; Yang, Gao, Rajashankar, & Patel, 2016). This motif is not present in the DNA sequence that codes for the crRNA, the CRISPR array, acting as a safety lock that avoids self-targeting. On the other hand, RNA-interacting Type III and VI complexes employ an exclusion mechanism by which they interfere with complementary sequences flanked by any motif but the one present in the CRISPR array (Abudayyeh et al., 2016; Marraffini & Sontheimer, 2010). A single system might be able to interact with several motifs with varying efficiencies, resulting in interplays that not necessarily lead to interference (Hayes et al., 2016).

The affinity of effector complexes for correct targets is also very diverse. The reported dissociation constant (K_D) for Cas9 (Type II) is 0,5 nM (Sternberg, Redding, Jinek, Greene, & Doudna, 2014), for Type III-A is 0,1 nM (Mogila et al., 2019), for Type I-F is 1 nM (Rollins, Schuman, Paulus, Bukhari, & Wiedenheft, 2015), while for Type I-E is at least ten times higher (13 nM, 20,7 nM) (Beloglazova et al., 2015; Westra et al., 2012). In addition, CRISPR-Cas systems show varying tolerance towards mismatches between the target and the crRNA. Although a general principle, i. e. the further the mismatch is from the PAM the more likely it is to be allowed, has been established, each complex type shows variations of their specific mechanism.

Due to this high variability, target interactions should be studied in a case-by-case basis. Several methods can be used to analyse how the effector complexes interact with their targets. Among the *in vitro* methods, Electrophoretic Mobility Shift Assay (EMSA) is widely used, but new technologies involving label-free samples and real-time measuring of interactions in solution have opened up new possibilities.

One of these methods is BioLayer Interferometry (BLI), which measures the interference pattern of white light reflected from a biolayer in real time. The variety of available biolayers allows studying a wide range of interactions, such as the one between CRISPR-Cas complexes and DNA or RNA (Sultana & Lee, 2015). BLI can provide information about the affinity, stability and speed of the binding, plus the interplay with other proteins, such as anti-CRISPRs (Acrs) or nucleases, as it allows measuring sequential binding events and super-complex formation.

Here, we describe the use of BLI to study the target interactions of a model Type I-Fv effector complex (Cascade) from *Shewanella putrefaciens* CN-32 (Dwarakanath et al., 2015). This complex lacks a large subunit, reported to be responsible for PAM recognition and dsDNA separation in I-E and I-F systems (Chowdhury et al., 2017; Hayes et al., 2016; Xiao et al., 2017). It features two Cas proteins, 5fv and 7fv, with no sequence similarity to other described Cas proteins. Previous work showed that these diversified proteins fulfil the tasks of the missing large subunit (Pausch et al., 2017), but it is unknown whether this affects the way the complex binds to its targets, and the affinity of the interaction.

We also study a Type I-F Cascade from *Shewanella baltica* OS195, to compare how the differences in complex architecture affect the interaction with targets. This system has the same CRISPR array structure as the Type I-Fv, with 32 nucleotide spacers and identical repeat sequences.

In our setup, we use a BLItz system (FortéBio) with single-use High Precision Streptavidin (SAX) Biosensors (Dip and Read™, FortéBio). We measure the interaction as follows: after removing the background signal from the buffer, we immobilize one interactor, a biotinylated oligonucleotide, to the streptavidin-coated biolayer and set a baseline. Then, we measure the shift in wavelength of the reflected light upon binding of the second interactor (a CRISPR-Cas effector) until the equilibrium is reached. Lastly, the biolayer is incubated in buffer to follow the dissociation. The change in reflected light is the readout for the variation in thickness of the biolayer due to the binding or detachment of complexes (Figure IV.1).

This method allowed (i) obtaining dissociation constants (K_D) for the interactions between Cascades and dsDNA, (ii) qualitative analyses of PAM recognition and sequence complementarity and (iii) studying the effect of a predicted Acr inhibitor protein on Cascade binding.

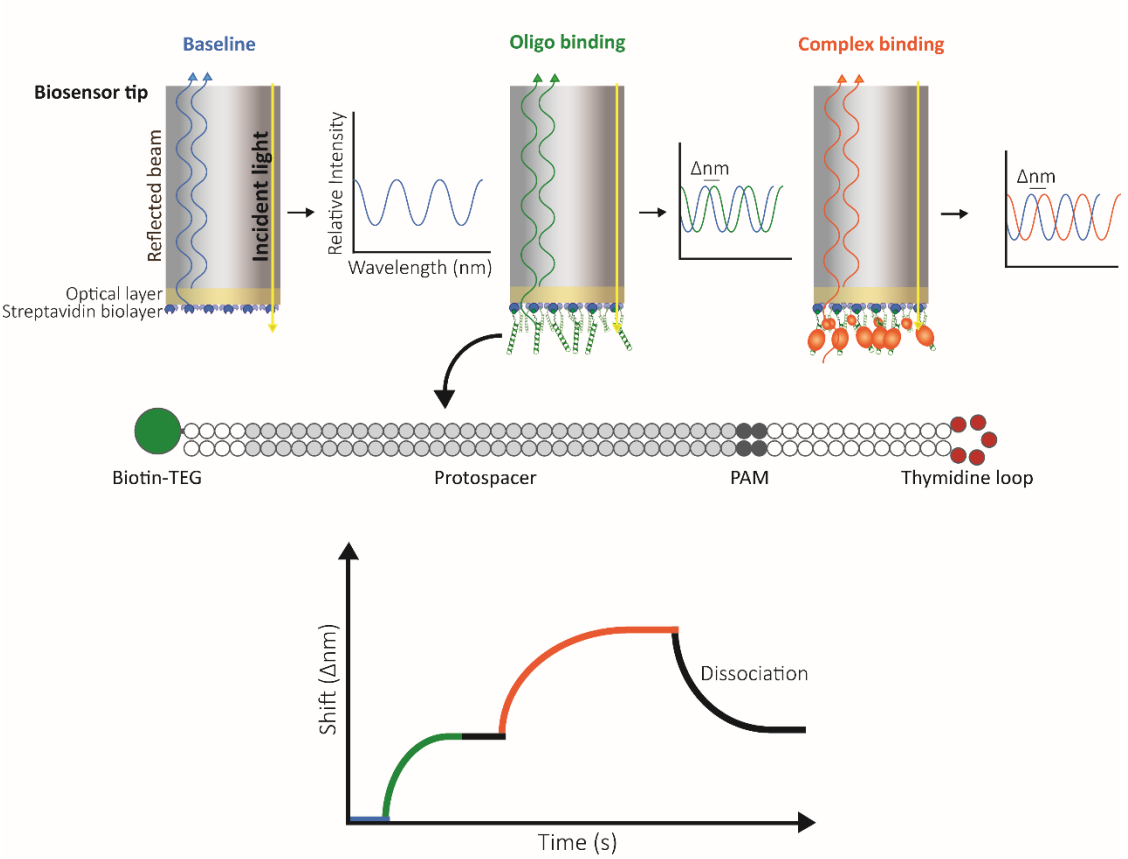


Figure IV.1. Schematic representation of Biolayer Interferometry (BLI). BLI measures the shifts in reflected white light upon changes in the biolayer thickness. The measured shift depends both on the size and the affinity of the interactors. Biolayers are coated with molecules (e.g streptavidin) that allow the immobilization of one of the interactors. For this study, 5´-biotin-tagged oligonucleotides were used. The long ssDNA oligonucleotides have complementary arms separated by a 5 thymidine loop, to allow for dsDNA hybridization. In addition, they contain a complementary sequence to the crRNA carried by the effector complex (protospacer) and a protospacer adjacent motif (PAM). After baselining the light signal on buffer (blue), the oligonucleotide is added, binding to the streptavidin and generating a shift (green) that is also baselined by incubating the biolayer back in buffer (black straight line). After signal stabilization, the label-free ribonucleoprotein complex is added, and the generated shift is recorded until equilibrium is reached (orange). Dissociation of the complex is then measured by incubating the biolayer back in buffer (black curve).

Oligonucleotide design

To follow the binding of the effector complex to the target, one of the players needs to be immobilized to the biolayer as bait. Choosing the smaller molecule is recommended, as immobilizing a big protein complex could result in clashes that impair proper binding. Also, if using the bigger molecule as bait, the shift elicited by the binding of the smaller molecule could be difficult to measure and visualize, since the thickness of the biolayer would not change greatly. Therefore, for the following experiments the target oligonucleotide was used as bait.

There are several options when choosing a biolayer to immobilize the substrate. In this case, a streptavidin-coated biolayer was chosen because biotin-labelled oligonucleotides are commercially available. In order to give the oligonucleotide flexibility, we used a triethylenglycol (TEG) spacer between the biotin and the first nucleotide. The Biotin-TEG tag was placed at the 5'-end of the oligonucleotide.

The target oligonucleotide should be designed taking the requirements of the CRISPR-Cas complex under study into consideration. For Type I-Fv and I-F, the effector complexes bind dsDNA complementary to a 32 nt spacer located on the crRNA, and recognize a GG PAM on the 3'-end of the target strand (Dwarakanath et al., 2015; Wiedenheft et al., 2011). Consequently, a dsDNA target is needed. This could be obtained either by the annealing of two single stranded oligonucleotides, with one being biotinylated, or by the design of a long single stranded oligonucleotide with complementary arms. The first option, although cheaper, has the downside of needing posterior clean up (for example Exonuclease I digestion) to remove non-annealed oligos, while the second option has size limitations, since most biotechnology companies do not offer synthesis of biotinylated oligonucleotides longer than 120 nucleotides.

For initial experiments, we decided on the long oligonucleotide (Figure IV.1), as a length of 120 nt is sufficient to place the 32 bp targeting region and the PAM, plus extra nucleotides at each side. To promote the formation of the intramolecular duplex, a central flexible 5-nt thymidine loop was added. Secondary structure prediction, for example using RNAstructure (Reuter & Mathews, 2010), is recommended to rule out the formation of other unwanted structures. Modifications can be introduced to this dsDNA oligonucleotide, depending on which characteristics of the binding need to be studied (e.g. different PAMs or mismatches to the crRNA).

Oligonucleotide stocks were diluted in water to a concentration of 100 µM and stored at -20° C. Further dilutions should be made before each assay in the buffer that matches the protein sample, as differences in buffer composition could affect protein stability and baseline measurements. Here, oligonucleotides were further diluted to 50 µM with the 2X protein purification buffer that corresponds to each Cascade (Supplementary Table IV.1). Further serial dilutions were made with 1X buffer.

Protein sample preparation

As for the second player, the CRISPR-Cas complex needs to be produced in sufficient amount and purity. The highest concentration used was 5 μM of Cascade to test the limits of saturation. The required concentration will depend on the amount of bait used, as well as the affinity of each complex tested (see section 3). The use of concentrations 10 times higher than the expected K_D is suggested for kinetic measurements (FortéBio, application note: https://www.moleculardevices.com/en/assets/app-note/biologics/biomolecular-binding-kinetics-assays-on-the-octet-platform)

The BLItz system has a 4 μL pocket where samples are placed. When doing qualitative analyses, a single measurement will suffice. When performing kinetic assays, at least 6 measurements at different concentrations are needed. Therefore, the required sample amount depends on the experimental design, taking also in consideration the measurement of technical replicates.

When working with multi-protein complexes, the formation of intermediate products besides the expected effector complex is possible. These sub-complexes can still retain the ability to interact with targets, as most Cas proteins have charged surfaces or pockets (Chowdhury et al., 2017; Hayes et al., 2016; Pausch et al., 2017). To avoid non-specific binding that would alter the kinetic results, the sample needs to be as pure as possible.

First, it is needed that each component of the complex is produced in the cell in sufficient amounts and an excess of crRNA is provided. Adding more spacers to the CRISPR array and using a high copy plasmid to express it can improve the formation of viable complexes. If side products cannot be avoided, after affinity chromatography it is recommended to perform size exclusion chromatography (SEC), with a high resolution column to isolate properly formed complexes.

The Type I-Fv complex from *S. putrefaciens* CN-32 was purified as previously described (Gleditzsch et al., 2016), by heterologous expression in *E. coli* BL21(DE3) of the Cascade components (Cas5fv, N-His-Cas7fv, Cas6f and crRNA) (plasmids in Supplementary Table IV.2). Cascade peak fractions were separated from other sub-complexes, including Cas7fv filaments and Cas7fv-Cas5fv dimers (Supplementary Figure IV.1).

The Type I-F complex from *S. baltica* OS195 was obtained in similar fashion, with a Cas7f-his-tagged complex expressed in *E. coli* BL21(DE3) and purified by immobilized metal affinity chromatography (IMAC) and subsequent SEC. The purification of this

system yields fewer sub-complexes in comparison to Type I-Fv (Supplementary figure IV.1). Buffers used for this purification can be found in the Supplementary Table IV.1.

Effector complex formation was corroborated by SDS-PAGE and mass spectrometry. Samples were concentrated to 10 µM.

Optimization of BLI conditions

Once both interactors are available, experimental conditions need to be optimized. This corresponds to: setting the measuring times, minimizing non-specific interactions and determining the appropriate concentration of bait and complex.

As a starting point, we modified the measuring protocol given by the BLItz Pro software on the Advanced Kinetics experiment as follows:

Step Type	Sample Type	Position	Duration (s)
Initial Baseline	Buffer	Tube	30
Loading	Oligonucleotide	Drop	120
Baseline	Buffer	Tube	30
Association	Protein complex	Drop	300
Dissociation	Buffer	Tube	180

Prior to measuring, the biosensors need to be hydrated for 10 minutes in the same buffer that the interactors are in, as changes in buffer composition might lead to false shifts.

At first, unspecific interactions of the protein complex against the biolayer have to be tested. For this, on the loading step, no oligonucleotide is added. If an increase in signal is observed in the association step, it can be corrected by the addition of blocking agents. Here, we added 0.1 µM of BSA and 0.01% of Triton X-100 to the samples and buffer, which prevented undesired associations.

Secondly, the concentration of oligonucleotide to get a proper readout needs to be determined. For this, we measured the binding of 1 µM of Type I-Fv Cascade to a range of oligonucleotide concentrations. The interaction resulting in the highest binding curve at the lowest concentration, with no deviations from an expected normal curve, should be selected (Sultana & Lee, 2015). In this case, 100 nM of oligonucleotide are sufficient (Figure IV.2A).

Finally, for each experiment, the saturating protein concentration has to be stablished. This is important both for qualitative experiments that need to be performed at non-

saturating concentrations, and quantitative experiments, where saturation needs to be reached in order to determine the kinetics of the binding. Here, to illustrate saturated curves, using 100 nM of oligonucleotide we tested concentrations from 50 μM to 2 μM (Figure IV.2B). The signal obtained for concentrations above 2.5 μM exhibits a deviation from a normal curve, as the signal first increases and subsequently decreases. This abnormal sensogram indicates non-ideal binding due to aggregation of the protein complexes. Other deviations and a troubleshooting guide can be found in (Sultana & Lee, 2015). The fact that the signal does not return to 0 during dissociation indicates that the binding is highly stable.

When performing technical replicates with the same protein sample, it is important to monitor the stability of the complex, as degradation or dissociation leading to sample heterogeneity results in altered measurements. For this, it is suggested to minimize the time between experiments (here, we measure in the two subsequent days after Cascade purification) and compare the obtained shifts. A loss in signal between replicates or the obtainment of curves with non-ideal binding, such as in Figure IV.2B, might indicate complex degradation.

Determination of dsDNA-effector complex binding kinetics

In order to determine the K_D, several protein concentrations need to be measured, ranging from saturation to the detection limit. From this data, the shifts at a time point where the interaction reached the equilibrium are selected and plotted against the corresponding protein concentration. The obtained concentration vs. shift curve can be used to determine the constant by adjusting it to a non-linear regression.

For this, we measured the shift elicited by the binding of Type I-Fv and Type I-F Cascade to 100 nM of complementary dsDNA. In the first column of Figure IV.3 it is possible to see how we measured several concentrations to cover the dynamic range of the interaction. The interaction of the Cascades to the biosensor (Complex Alone) was measured in order to determine the lower detection limit.

We took the shift at 425 seconds as the point where equilibrium was reached, and use the obtained values to plot the graph on the second column of Figure IV.3. As more concentrations are measured, the resolution of the derived curve will increase, giving more precise K_D values.

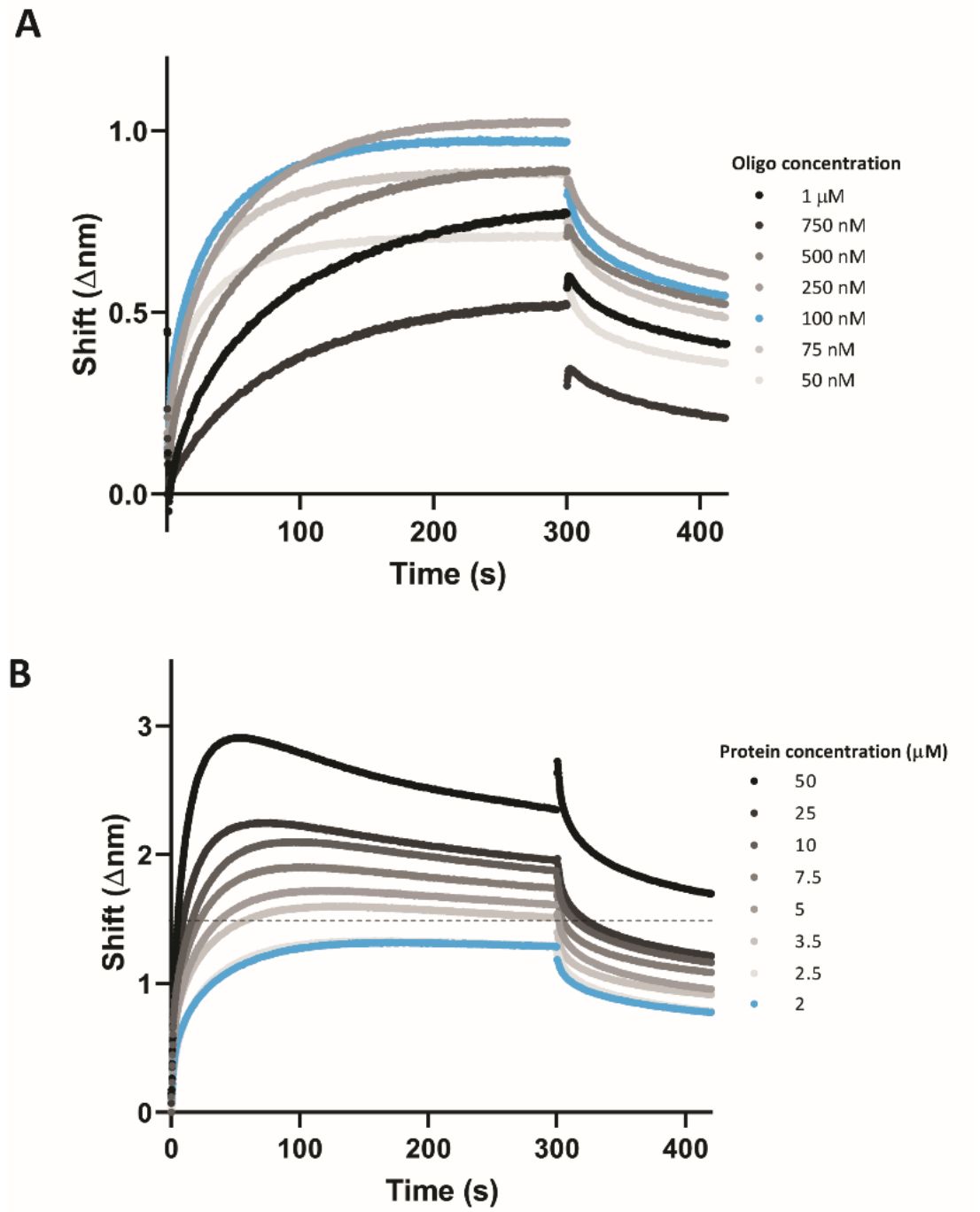


Figure IV.2. Determination of optimal oligonucleotide concentration and protein saturating concentration. (A) The wavelength shift (Δnm) generated by the addition of 1 μM Type I-Fv Cascade to different concentrations of complementary dsDNA oligonucleotide was recorded for 5 minutes. At a concentration of 100 nM of oligonucleotide (blue line) the resulting binding curve gave the highest signal while also reaching equilibrium (plateau). (B) Different concentrations of Type I-Fv Cascade were tested for binding to 100 nM of complementary dsDNA for 5 minutes. Saturating concentration was stablished as the lowest concentration at which a binding curve reaching plateau could be obtained (blue line). Higher concentrations exhibit either no change in the shift, or deviations from an expected binding and dissociation curve such as increases above the final equilibrium shift (depicted for 3.5 μM above the dotted line). Saturation should be determined individually for each experiment.

The obtained shift vs. concentration curves for both systems do not reach a plateau and seem to have two components: an exponential and a linear phase. This suggests the existence of both specific and non-specific interactions with the target. Due to this observation, we adjusted the data to the following non-linear regression using GraphPad Prism 8.0:

$$Y = B_{max} * X / (K_D + X) + NS * X$$

In this formula, Y is the binding (shift), B_{max} is the maximum binding, X is the effector complex concentration, K_D is the equilibrium dissociation constant and NS is the slope of the non-linear regression (accounting for non-specific binding).

The selection of the fitting equation should be made taking in consideration the shape of the obtained curve. It is possible that instead of an exponential binding, the Cascade-dsDNA interaction exhibits cooperativity. In this cases, the adjustment to a sigmoidial model would provide more rigorous results.

Using the exponential regression we obtained a K_D of 46.65 nM for Type I-Fv Cascade (R^2= 0.97) and of 170.8 nM for Type I-F (R^2=0.98). Interestingly, the obtained K_D values are higher than what has been reported for the Type I-F system of *Pseudomonas aeruginosa*, where the dissociation constant is 1 nM (Rollins et al., 2015). This difference in target affinity might be given by diversification of the Cascade components, as *S. putrefaciens* carries disctinct Cas5fv and Cas7fv subunits and *S. baltica* complexes do not have major aminoacid sequence identity when compared to the *P. aeruginosa* system (Supplementary figure IV.2).

Analysis of target requirements for binding

For Type I systems, the effector complex recognizes two main features on the target: first, a correct PAM (Gleditzsch et al., 2019), and second, a complementary sequence to the crRNA carried by the complex (Barrangou et al., 2007). The optimal PAM described for Type I-F Cascades is a GG pair on the 3'-end of the target strand (Chowdhury et al., 2017; Guo et al., 2017). But other motifs are also recognized by Type I complexes, albeit with lower efficiency, triggering not only interference but also the acquisition of new spacers into the CRISPR array (primed acquisition) (Hayes et al., 2016). In the same line, mismatches between the crRNA and the target are allowed (Xue et al., 2015). This provides robustness against escape mutations of the targeted elements.

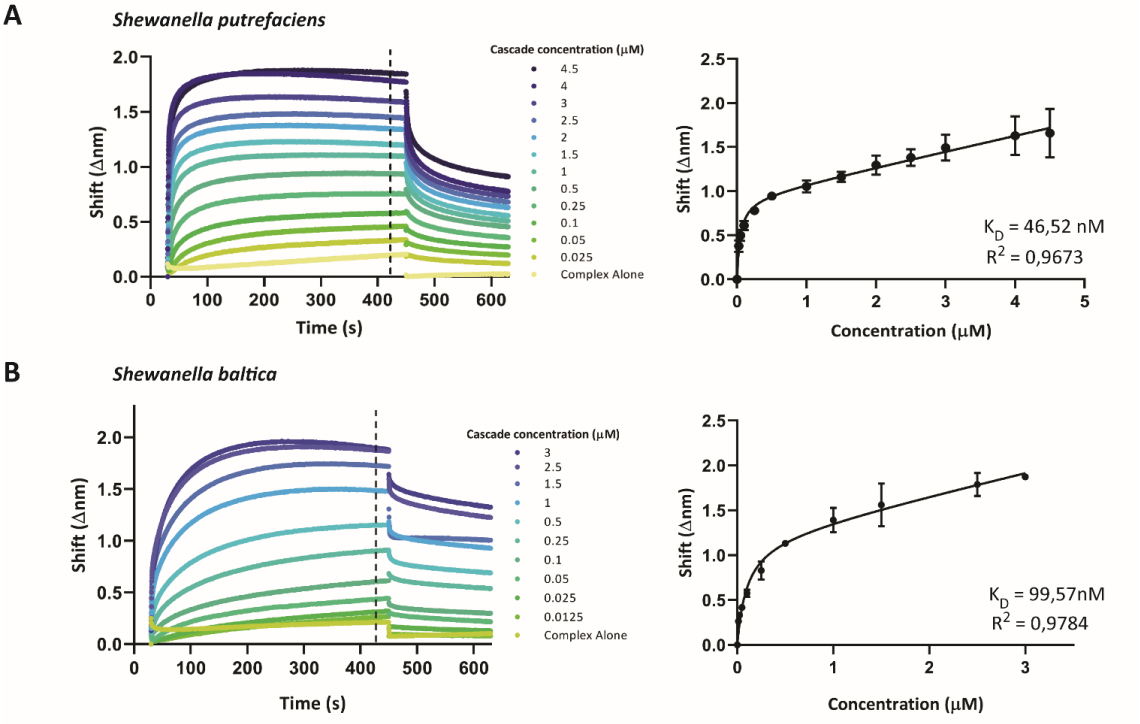


Figure IV.3. Determination of the dissociation constant (K_D) of Type I-Fv and I-F CRISPR-Cas complexes. Wavelength shift (Δnm) generated by the addition of Type I-Fv Cascade from *S. putrefaciens* CN-32 (A) or Type I-F from *S. baltica* OS195 (B) to 100 nM of complementary dsDNA oligonucleotide pre-bound to a streptavidin biolayer. Binding was followed for 7 min in order to reach equilibrium. Afterwards, the protein-bound biolayer was incubated for 3 min on buffer to measure dissociation. The interaction of the complexes to the dsDNA-free biolayer is shown as control (Complex alone). Shifts recorded at 390 sec after start of binding were plotted against the complex concentration in order to calculate the respective K_Ds. Data were fitted to the non-linear regression: $Y=B_{max}*X/(K_D+X) + NS*X$, with B_{max} being the maximum shift and NS the slope of the non-linear regression, representing non-specific binding. Coefficients of determination (R^2) and K_Ds are reported for each complex. Binding assays were performed in duplicates. Error bars correspond to Standard Error of the Mean (SEM).

In order to test the recognition of these two features, we measured the binding of Type I-Fv Cascade to three different dsDNA oligonucleotides: the first carrying the correct PAM and a complementary sequence, the second one a TT PAM and a complementary sequence, and the last no sequence complementary to the crRNA and no GG pairs. A TT PAM was selected, as it is the motif present at the end of the repeat on the CRISPR array. Therefore, TT recognition would lead to self-targeting.

BLI was carried out as described above, following the interaction of 100 nM of the respective oligonucleotide with 1.5 μM of Type I-Fv Cascade. As indicated before, the complex interacts with the complementary target carrying the GG PAM. In contrast, the exchange of the motif to TT leads to a reduced interaction that does not resemble an ideal binding, showing instead non-specific interactions that are washed away on the dissociation step (425 sec). Similarly, the interaction with a non-complementary

oligonucleotide does not appear to have a specific component, and the recorded signal also disappears upon dissociation (Figure IV.4).

As a second qualitative test, we set to determine the effect of dsDNA mismatches (bubbles) on the binding of the Type I-Fv Cascade. After the PAM is recognized during interference, the effector complex proceeds to unwind the dsDNA, allowing for crRNA-DNA pairing, leading to the formation of an R-loop structure. The opening of the target has a directionality, starting from the PAM-proximal site (Huo et al., 2014; Rutkauskas et al., 2015; Xiao et al., 2017). This opening requires the large subunit, but it has been described that Type I-E complexes without the large subunit can still bind dsDNA targets with mismatched sequences (Huo et al., 2014).

Therefore, in order to confirm the directionality of the opening and to test the affinity of the Type I-Fv complex for different bubbled substrates, we designed variants of the complementary dsDNA oligonucleotide. The alternative targets carry 5 nt mismatches either right next to the PAM, in the centre of the protospacer, or at the end of it. In addition, as the dsDNA oligonucleotide has a 5nt thymidine loop, we also inverted the protospacer to test whether the proximity to the ssDNA region would affect the binding (Figure IV.5A).

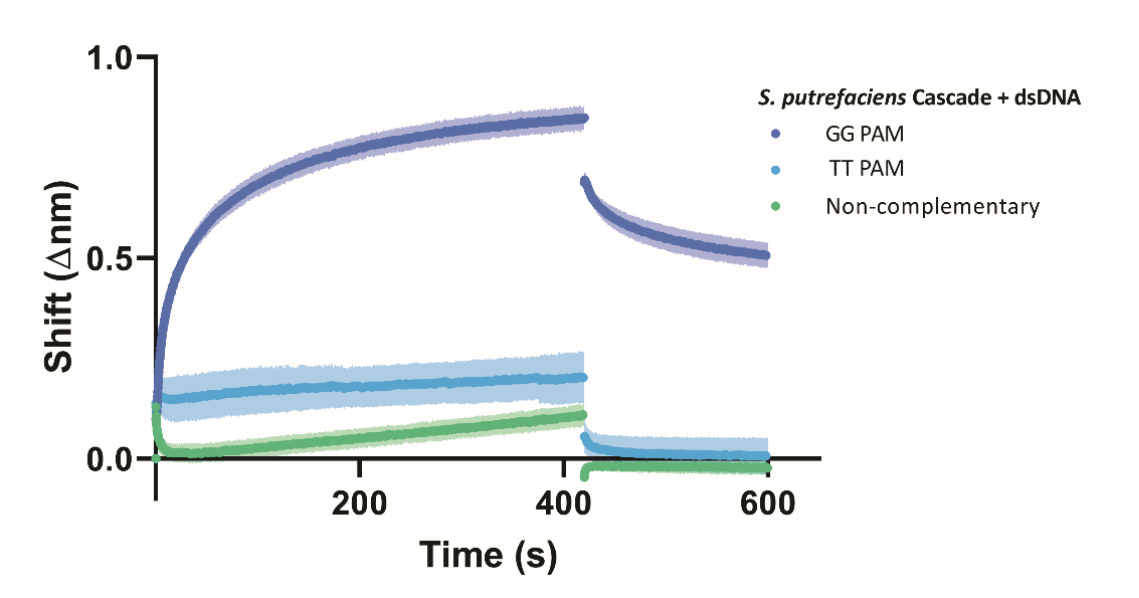


Figure IV.4. Qualitative analysis of PAM preference by Type I-Fv Cascade. Wavelength shift (Δnm) generated by the addition of 1.5 μM of Type I-Fv Cascade from *S. putrefaciens* CN-32 to 100 nM of dsDNA oligonucleotides containing a matching spacer to the crRNA and either a GG or a TT PAM. GG PAM is recognized by the complex, as shown by the shift elicited, while the TT PAM is not recognized by the complex, giving shifts similar to a non-complementary target. Assays were performed in duplicates, the lighter outline represents the SEM.

The Type I-Fv Cascade showed higher affinity for the target with the PAM-proximal bubble, with sequentially lower signals when binding the ones with central and distal bubbles. Nevertheless, all of them produce a higher shift than the fully annealed substrate, indicating that mismatches along the target facilitate binding. Furthermore, it also corroborates the directionality of the binding for this minimal system. The binding to the target with an inverted protospacer elicited a shift similar to the one of the regular protospacer, ruling out an effect of the thymidine loop (Figure IV.5B).

Other variables could also be studied using BLI by simply modifying the oligonucleotides, such as the stringency for mismatches between crRNA and dsDNA, leading to the identification of seed sequences. Furthermore, it would be possible to get further insights into the PAM-recognition mechanism by designing targets with modified nucleosides. For example, the exchange of the guanine pair by inosines or other base-pairing alternatives could help elucidate whether there is a strand bias on PAM recognition, as tested for the Type I-F Cascade (Rollins et al., 2015).

Evaluation of the effect of Anti-CRISPR proteins on the binding of effector complexes

BLI gives a real-time measurement of the interaction with the target. This particularity can be exploited to study how other players influence the binding. Exemplarily, we show the effect of an Anti-CRISPR protein, AcrF9 Vpa (from *Vibrio parahaeomlyticus*) on the binding of both the Type I-Fv and Type I-F Cascade.

AcrF9 Vpa has been classified as a broad range Type I-F Acr, being able to block the I-F system of both *P. aeruginosa* and *Pseudomonas atrosepticum* (Pawluk et al., 2016). So far, the mechanism behind this inhibition is unknown. Here, we show that it is also able to affect the plasmid targeting by the *S. baltica* I-F system, when co-expressed in *E. coli* BL21-AI. On the contrary, this Acr is not able to significantly obstruct the Type I-Fv complex *in vivo* (Figure IV.6A). Efficiency of transformation assays were performed as previously described (Pausch et al., 2017), co-expressing Cascades with AcrF9 VPa (plasmids in Supplementary Table IV.2).

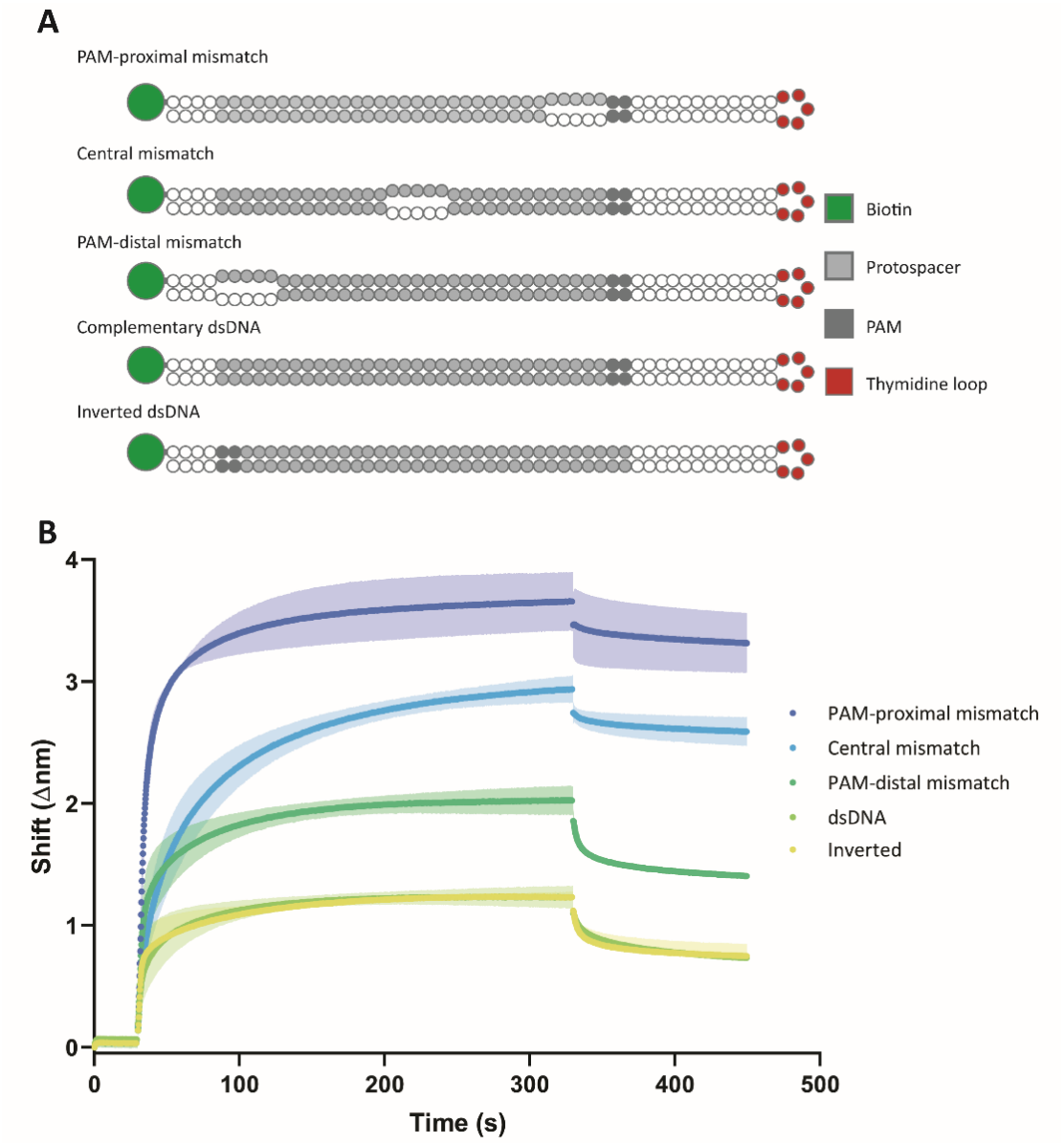


Figure IV.5. Influence of dsDNA mismatches on Type I-Fv Cascade binding. Binding of 1.5 μM of Type I-Fv Cascade to 100 nM of dsDNA oligonucleotides either fully hybridized or with 5 nucleotide mismatches at positions on the protospacer close to the PAM, in the centre or at the end (represented in blue in the oligo scheme). Assays were performed in duplicates, the lighter outline represents the SEM.

In order to get further insights into this interaction, we expressed a His-tagged version of AcrF9 in *E. coli* BL21 (DE3) and purified it by IMAC (buffers in Supplementary table IV.1). The obtained sample was dialysed overnight at 4ºC into the buffer used for Cascade SEC and taken to a final stock concentration of 10 μM. From this sample, working concentrations of 500 nM, 1 μM and 5 μM were prepared using the SEC buffer plus 0.1 μM of BSA and 0.01% of Triton X-100.

First, as a control, the interaction of 5 μM of Acr with the biolayer was tested, showing no unspecific binding. As a second control, we measured the shift elicited by 500 nM of

AcrF9 with 100 nM of the dsDNA complementary oligonucleotide. The obtained shift does not represent a specific interaction and it returns to baseline during the dissociation step (Figure IV.6B).

The BLItz platform has a detection limit of 10 kDa, while the other BLI platform, Octet (FortéBio), can measure shifts generated by molecules as small as 150 Da (Sultana & Lee, 2015). AcrF9 Vpa has a molecular weight of 9 kDa including the His-tag, therefore, when using the BLItz system to study the inhibition, the resulting shifts will depend only on the effector complex, as the binding of the Acr might not elicit a measurable shift. Because of this, we measure the effect of AcrF9 by pre-mixing it with Cascade at different molar ratios and comparing the result with the binding of the complex alone.

In agreement with the *in vivo* assays, the binding of Type I-Fv Cascade is not affected by the presence of AcrF9 Vpa (Figure IV.6B, left column). The measured shift for 500 nM of complex alone with 100 nM of dsDNA oligonucleotide does not change after pre-incubating it with either 500 nM or 2.5 μM of AcrF9 for 10 minutes at room temperature. This is the case when measuring the binding both to complementary and non-complementary oligonucleotides.

On the contrary, the binding of the Type I-F complex is altered in the presence of the inhibitor (Figure IV.6B, right column). Interestingly, the presence of AcrF9 appears not to prevent binding, but rather help the interaction with dsDNA, as seen by the faster binding of the 1:1 complex-Acr mix. For the complementary target, although the equilibrium is reached faster, the final shift is not considerably higher than the one for the complex alone.

Furthermore, when the complex is exposed to an excess of AcrF9 (1:5 ratio), the binding curve exhibits a different shape from the 1:1 interaction. Although the initial binding still happens in a fast manner, the subsequent signal starts decreasing, reaching levels below the equilibrium of the complex alone. The signal continues to decrease during dissociation, not reaching a stable level on neither phase. A loss on signal on BLI indicates the thinning on the biolayer, and could be interpreted as the release of complexes from their targets. As this effect is only seen when AcrF9 is in excess, it is possible that more than one inhibitor binds per complex, and that this is required to actually impair the binding.

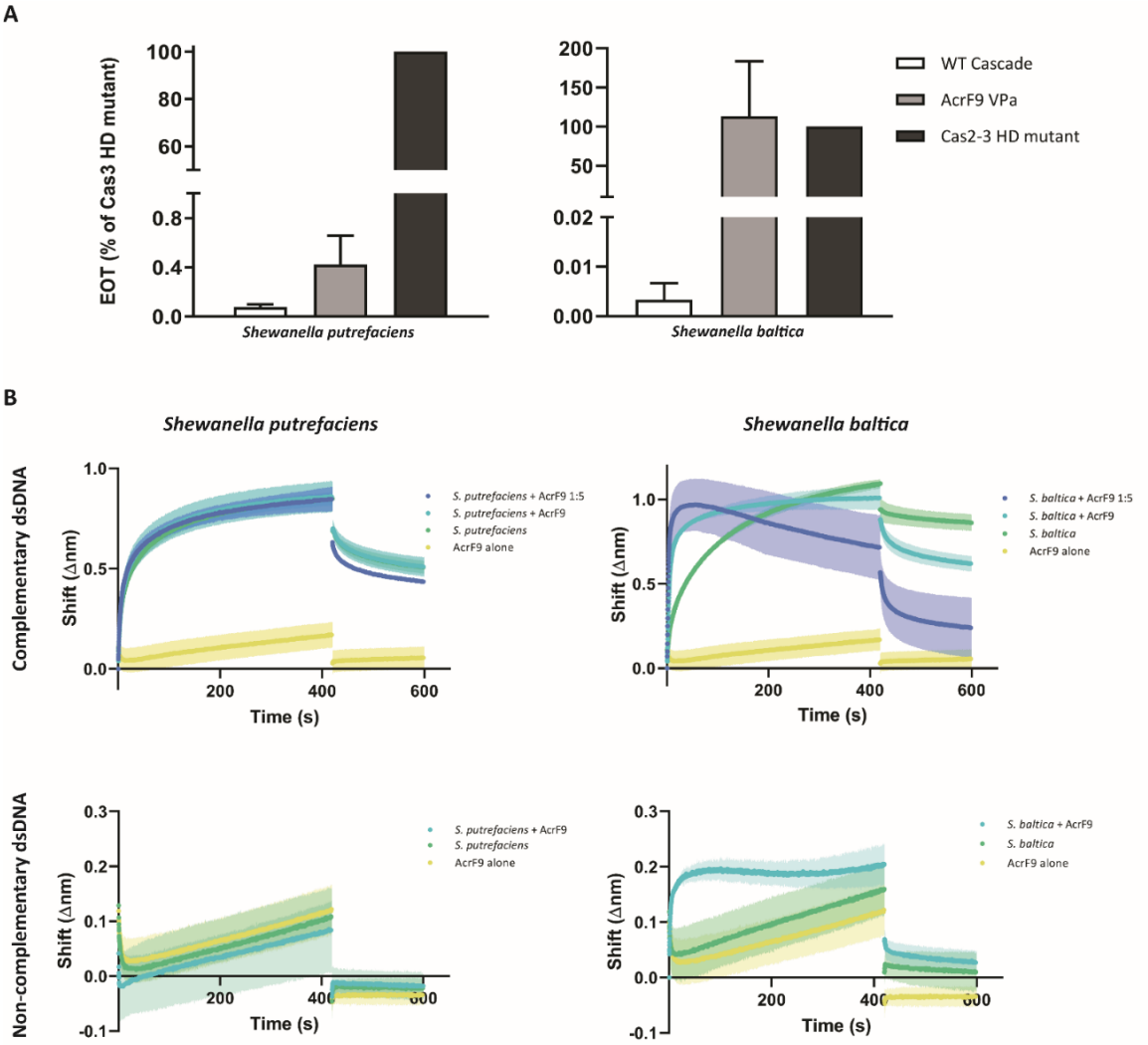


Figure IV.6. Effect of Anti-CRISPR (Acr) proteins on the binding of CRISPR-Cas complexes to dsDNA. (A) Efficiency of Transformation Assays (EOTs) of CRISPR-Cas systems I-Fv from *S. putrefaciens* CN-32 (left) and I-F from *S. baltica* OS195 (right) expressed in *E. coli* BL21-AI. The activity of the effector complexes was tested when expressed alone (WT Cascade) or co-expressed with AcrF9 from *Vibrio parahaemolyticus* (Vpa, WP_031500045.1) (Pawluk et al., 2016). EOT equals to the colony ratio between the strain of interest and its corresponding Cas2/3 HD mutant strain, presented as percentages. Error bars represent the SEM. Three replicates were quantified. (B) Wavelength shift (Δnm) generated by the binding of 500nM of either Type I-Fv (left) or Type I-F (right) Cascades to 100 nM of complementary (top) or non-complementary (bottom) dsDNA oligonucleotides. In order to test the anti-CRISPR effect, AcrF9 was mixed with Cascades at a 1:1 molar ratio or 5 times in excess (1:5). Samples were incubated at room temperature for 10 min before binding analyses. The shift elicited by the Acr protein binding to the dsDNA is shown as a control (AcrF9 alone). Assays were performed in duplicates, the lighter outline represents the SEM.

The binding of the I-F complex to non-complementary targets is also affected when AcrF9 is present. This small protein, similarly as for the complementary target, allows the complex to rapidly bind, even though the final shift at equilibrium is not higher than for the isolated complex. In any case, the final shift is lower than for the complementary target. Moreover, the shape of the Acr-Cascade binding curve does not resemble an

ideal binding, with a decrease in signal at mid-association. This might indicate heterogeneity among the recorded binding events.

Taken together, these results corroborate the evasion of AcrF9 blockage by the Type I-Fv complex and provide insights into the interplay between the protein and the Type I-F complex. The inhibition of binding occurs only when an excess of the Acr protein is present, while with similar amounts the Acr helps the complex bind dsDNA, regardless of the sequence complementarity. Therefore, it might prevent interference by misdirecting complexes to wrong targets, until enough protein is produced to prevent all binding. This would constitute a new mechanism for an Acr protein, highlighting the usefulness of BLI to characterize these inhibitors.

Conclusions

As future studies uncover more effector complexes and their inhibitors, the development of techniques to study these proteins in a fast and reliable manner is needed. BLI has the advantage of measuring interactions with only one labelled player, providing real-time results. Here, we showed how BLI can be used to answer qualitative questions, such as PAM identification, but also give information on the kinetics of the interactions and the mechanisms behind bindings and inhibitions. It is suggested that the obtained kinetic values are corroborated with a second technique, such as MicroScale Thermophoresis (MST), as each platform has different levels of accuracy depending on the experimental design. The use of several techniques is highly recommended when precise values are needed. In conclusion, BLI offers a practical alternative for studying the interaction between CRISPR-Cas complexes and nucleic acids, and was used to describe the target binding of model Type I-Fv and a Type I-F effector complexes.

Acknowledgements

H.M.E. and M.O are recipients of a fellowship from the International Max-Planck-Research School for Environmental, Cellular, and Molecular Microbiology (IMPRS-Mic). L.R. is supported by SPP 2141. We thank Kai Thormann (Justus-Liebig-Universität Gießen, Germany) for kindly facilitating the strain *S.baltica* OS195. We also thank Alan Davidson (University of Toronto, Canada) for the AcrF9 Vpa plasmid and helpful discussions.

Supplementary Material

Supplementary Table IV.1. Buffers for protein purification.

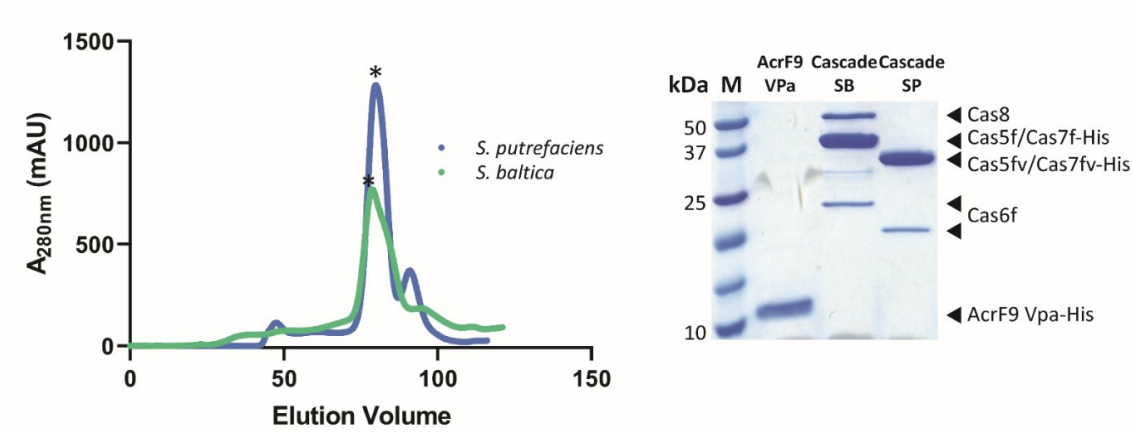
Protein Sample	Buffer Name	Components
S. putrefaciens *S. baltica* Cascades	Lysis Buffer	50 mM Tris-HCl pH=7.0
		300 mM NaCl
		10 mM $MgCl_2$
		1mM Dithiothreitol (DTT)
		10% Glycerol
	Wash Buffer	Lysis Buffer + 20 mM Imidazole
	Elution Buffer	Lysis Buffer + 500 mM Imidazole
AcrF9	Lysis Buffer	20 mM Tris-HCl pH=7.5
		250 mM NaCl
		1 mM DTT
	Wash Buffer	Lysis Buffer + 20 mM Imidazole
	Elution Buffer	Lysis Buffer + 500 mM Imidazole
All samples	SEC Buffer	50 mM HEPES-KOH pH=7.3
		150 mM NaCl
		1mM DTT

Supplementary Table IV.2. Plasmids used in this study. Plasmids in white were used for protein purification, while plasmids in grey were used for EOT assays.

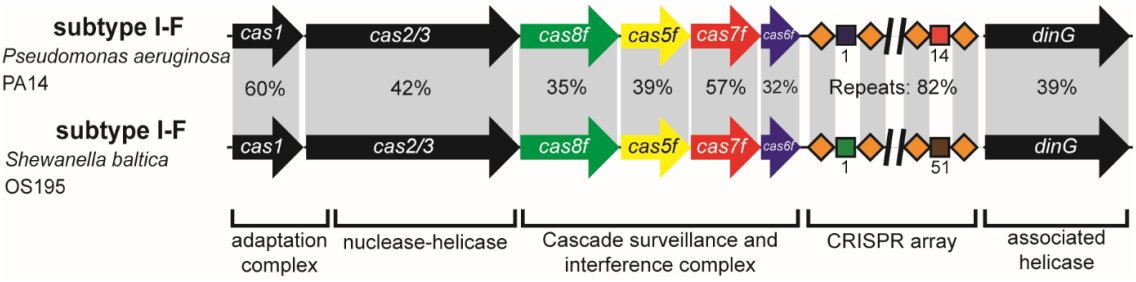
Plasmid Name	Description	Source/ Reference
pCascadeSP	pRSFDuet-1 + N-His-Cas7fv+ Cas5fv + Cas6f from *S. putrefaciens* CN-32	Gleditzsch et al., 2016
pCRISPRsp4	pUC19 + T7 polymerase promoter + repeat-spacer4-repeat sequence from *S. putrefaciens* CN-32. The 8nt handle for Cas5fv was deleted from the second repeat	Gleditzsch et al., 2016
pCascadeSB	pRSFDuet-1 + Cas8f + Cas5f + N-His-Cas7f + Cas6f from *S. baltica* OS195	This work
pAcrF9His	pRSFDuet-1 + N-His- AcrF9 Vpa (WP_031500045.1)	This work
pCasSP	pRha+ Cas2-3 + Cas7fv + Cas5fv + Cas6f from *S. putrefaciens* CN-32	This work
pCasHDSP	pRha+ Cas2-3 HD mutant + Cas7fv + Cas5fv + Cas6f from *S. putrefaciens* CN-32	This work
pCRISPRamp	pCDFDuet-1 + repeat-spacer anti-amp-repeat sequence. Repeats are from the CRISPR array of *S. putrefaciens* CN-32	Pausch et al., 2017
pCascadeSBCam	pACYCDuet-1 + Cas8f + Cas5f + Cas7f + Cas6f from *S. baltica* OS195	This work
pCas3SB	pCRISPRamp + Cas2-3 from *S. baltica* OS195	This work
pCas3HDSB	pCRISPRamp + Cas2-3 HD mutant from *S. baltica* OS195	This work
pAcrF9	pRSFDuet-1 + AcrF9 Vpa (WP_031500045.1)	This work
pETDuet-1	Target Plasmid	Novagen

Supplementary Table IV.3. Biotinylated oligonucleotides used on this study. The thymidine loop is shown in blue.

Oligonucleotide Name	Sequence (with 5' Biotin-TEG tag)
Complementary	CGCCGGTTATAGGTTTGCGCGTCTTGCTGGGCGATAGGACGTGGGATATCTTTTTGATATCC CACGTCCTATCGCCCAGCAAGACGCGCAAACCTATAACCGGCG
Non-complementary	GCTGAACATGATACGCTTCAATAGTTAGATGACGCAGTAGTTATGATATCTTTTTATAACTA CTGCGTCATCTAACTATTGAAGCGTATCATGTTACG
TT PAM	CGCCGGTTATAGGTTTGCGCGTCTTGCTGGGCGATATTACGTGGGATATCTTTTTGATATCC CACGTAATATCGCCCAGCAAGACGCGCAAACCTATAACCGGCG
PAM-proximal mismatch	CGCCGGTTATAGGTTTGCGCGTCTTGCTGGGCGATAGGACGTGGGATATCTTTTTGATATCC CACGTCCGCGTACCCAGCAAGACGCGCAAACCTATAACCGGCG
Central mismatch	CGCCGGTTATAGGTTTGCGCGTCTTGCTGGGCGATAGGACGTGGGATATCTTTTTGATATCC CACGTCCTATCGCCCAGCAAGATAAAAAAACCTATAACCGGCG
PAM-distal mismatch	CGCCGGTTATAGGTTTGCGCGTCTTGCTGGGCGATAGGACGTGGGATATCTTTTTGATATCC CACGTCCTATCGCCCAGCAAGACGCGCAAACCTAGGCTTGGCG
Inverted protospacer	CGCCCCTATCGCCCAGCAAGACGCGCAAACCTATAACCACGTGGGATATCTTTTTGATATCC CACGTGGTTATAGGTTTGCGCGTCTTGCTGGGCGATAGGGGCG



Supplementary Figure IV.1. Purification of Type I-F/I-Fv Cascades and AcrF9 VPa. Chromatogram obtained after IMAC and SEC purification of Type I-F Cascade from *S. baltica* OS195 and Type I-Fv Cascade from *S. putrefaciens* CN-32, heterologously expressed in *E. coli* BL21-AI. For both complexes, the Cas7 protein was N-terminally His-tagged. SEC was performed using a HiLoad Superdex 16/600 200 pg column (GE Healthcare). AcrF9 Vpa was purified by IMAC. The composition of the peak samples (asterisks) was analyzed by SDS-PAGE, where the presence of all Cascade components and AcrF9 Vpa was confirmed. SB = *S. baltica*, SP= *S. putrefaciens*.



Supplementary Figure IV.2. Comparison of the CRISPR-Cas systems from *P. aeruginosa* PA14 and *S. baltica* OS195. Schematic representation of the CRISPR-*cas* loci. The arrows indicate the *cas* and related genes in comparative lengths. Cascade genes are multicolored. The CRISPR arrays are indicated as alternating rhombi (orange, CRISPR repeats) and squares of various colors (spacers). Grey highlights indicate sections of high protein sequence identity and high sequence complementarity (for the repeats).

References

Abudayyeh, O. O., Gootenberg, J. S., Konermann, S., Joung, J., Slaymaker, I. M., Cox, D. B., et al. (2016). C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science, 353*(6299), aaf5573.
Anders, C., Niewoehner, O., Duerst, A., & Jinek, M. (2014). Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature, 513*(7519), 569-573.
Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., et al. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science, 315*(5819), 1709-1712.
Beloglazova, N., Kuznedelov, K., Flick, R., Datsenko, K. A., Brown, G., Popovic, A., et al. (2015). CRISPR RNA binding and DNA target recognition by purified Cascade complexes from Escherichia coli. *Nucleic Acids Res, 43*(1), 530-543.
Chowdhury, S., Carter, J., Rollins, M. F., Golden, S. M., Jackson, R. N., Hoffmann, C., et al. (2017). Structure Reveals Mechanisms of Viral Suppressors that Intercept a CRISPR RNA-Guided Surveillance Complex. *Cell, 169*(1), 47-57 e11.
Dwarakanath, S., Brenzinger, S., Gleditzsch, D., Plagens, A., Klingl, A., Thormann, K., et al. (2015). Interference activity of a minimal Type I CRISPR-Cas system from Shewanella putrefaciens. *Nucleic Acids Res, 43*(18), 8913-8923.
Faure, G., Shmakov, S. A., Yan, W. X., Cheng, D. R., Scott, D. A., Peters, J. E., et al. (2019). CRISPR-Cas in mobile genetic elements: counter-defence and beyond. *Nat Rev Microbiol*.
Gleditzsch, D., Muller-Esparza, H., Pausch, P., Sharma, K., Dwarakanath, S., Urlaub, H., et al. (2016). Modulating the Cascade architecture of a minimal Type I-F CRISPR-Cas system. *Nucleic Acids Res, 44*(12), 5872-5882.
Gleditzsch, D., Pausch, P., Muller-Esparza, H., Ozcan, A., Guo, X., Bange, G., et al. (2019). PAM identification by CRISPR-Cas effector complexes: diversified mechanisms and structures. *RNA Biol, 16*(4), 504-517.
Guo, T. W., Bartesaghi, A., Yang, H., Falconieri, V., Rao, P., Merk, A., et al. (2017). Cryo-EM Structures Reveal Mechanism and Inhibition of DNA Targeting by a CRISPR-Cas Surveillance Complex. *Cell, 171*(2), 414-426 e412.
Hayes, R. P., Xiao, Y., Ding, F., van Erp, P. B., Rajashankar, K., Bailey, S., et al. (2016). Structural basis for promiscuous PAM recognition in type I-E Cascade from E. coli. *Nature, 530*(7591), 499-503.
Huo, Y., Nam, K. H., Ding, F., Lee, H., Wu, L., Xiao, Y., et al. (2014). Structures of CRISPR Cas3 offer mechanistic insights into Cascade-activated DNA unwinding and degradation. *Nat Struct Mol Biol, 21*(9), 771-777.
Marraffini, L. A., & Sontheimer, E. J. (2010). Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature, 463*(7280), 568-571.
Mogila, I., Kazlauskiene, M., Valinskyte, S., Tamulaitiene, G., Tamulaitis, G., & Siksnys, V. (2019). Genetic Dissection of the Type III-A CRISPR-Cas System Csm Complex Reveals Roles of Individual Subunits. *Cell Rep, 26*(10), 2753-2765 e2754.
Pausch, P., Muller-Esparza, H., Gleditzsch, D., Altegoer, F., Randau, L., & Bange, G. (2017). Structural Variation of Type I-F CRISPR RNA Guided DNA Surveillance. *Mol Cell, 67*(4), 622-632 e624.
Pawluk, A., Staals, R. H., Taylor, C., Watson, B. N., Saha, S., Fineran, P. C., et al. (2016). Inactivation of CRISPR-Cas systems by anti-CRISPR proteins in diverse bacterial species. *Nat Microbiol, 1*(8), 16085.
Reuter, J. S., & Mathews, D. H. (2010). RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics, 11*, 129.
Rollins, M. F., Schuman, J. T., Paulus, K., Bukhari, H. S., & Wiedenheft, B. (2015). Mechanism of foreign DNA recognition by a CRISPR RNA-guided surveillance complex from Pseudomonas aeruginosa. *Nucleic Acids Res, 43*(4), 2216-2222.
Rutkauskas, M., Sinkunas, T., Songailiene, I., Tikhomirova, M. S., Siksnys, V., & Seidel, R. (2015). Directional R-Loop Formation by the CRISPR-Cas Surveillance Complex Cascade Provides Efficient Off-Target Site Rejection. *Cell Rep*.
Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C., & Doudna, J. A. (2014). DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature, 507*(7490), 62-67.
Sultana, A., & Lee, J. E. (2015). Measuring protein-protein and protein-nucleic Acid interactions by biolayer interferometry. *Curr Protoc Protein Sci, 79*, 19 25 11-26.
Westra, E. R., van Erp, P. B., Kunne, T., Wong, S. P., Staals, R. H., Seegers, C. L., et al. (2012). CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Mol Cell, 46*(5), 595-605.
Wiedenheft, B., Lander, G. C., Zhou, K., Jore, M. M., Brouns, S. J. J., van der Oost, J., et al. (2011). Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature, 477*(7365), 486-489.
Xiao, Y., Luo, M., Hayes, R. P., Kim, J., Ng, S., Ding, F., et al. (2017). Structure Basis for Directional R-loop Formation and Substrate Handover Mechanisms in Type I CRISPR-Cas System. *Cell, 170*(1), 48-60 e11.
Xue, C., Seetharam, A. S., Musharova, O., Severinov, K., Brouns, S. J., Severin, A. J., et al. (2015). CRISPR interference and priming varies with individual spacer sequences. *Nucleic Acids Res, 43*(22), 10831-10847.
Yang, H., Gao, P., Rajashankar, K. R., & Patel, D. J. (2016). PAM-Dependent Target DNA Recognition and Cleavage by C2c1 CRISPR-Cas Endonuclease. *Cell, 167*(7), 1814-1828 e1812.

Chapter V:
Live-cell single-particle tracking photoactivated localization microscopy of Cascade-mediated DNA surveillance

Bartosz Turkowyd[1,2,*], Hanna Müller-Esparza[3,*], Vanessa Climenti[1,2], Niklas Steube[3], Ulrike Endesfelder[1,2,*], Lennart Randau[2,3,*]

This chapter is written in manuscript style and was published in Methods in Enzymology in January, 2019. My contribution to this work included designing and cloning all the constructs, testing the formation and activity of the fluorescent complex and writing the corresponding parts of the manuscript.

---

[1] Department of Systems and Synthetic Microbiology, Max Planck Institute for Terrestrial Microbiology, D-35043 Marburg, Germany
[2] LOEWE Center for Synthetic Microbiology (SYNMIKRO), D-35043 Marburg, Germany
[3] Prokaryotic Small RNA Biology Group, Max-Planck-Institute for Terrestrial Microbiology, D-35043 Marburg, Germany
* Authors contributed equally

Abstract

Type I CRISPR-Cas systems utilize small CRISPR RNA (crRNA) molecules to scan DNA strands for target regions. Different crRNAs are bound by several CRISPR-associated (Cas) protein subunits that form the stable ribonucleoprotein complex Cascade. The Cascade-mediated DNA surveillance process requires a sufficient degree of base-complementarity between crRNA and target sequences and relies on the recognition of small DNA motifs, termed Protospacer Adjacent Motifs (PAMs). Recently, super-resolution microscopy and single-particle tracking methods have been developed to follow individual protein complexes in live cells. Here, we described how this technology can be adapted to visualize the DNA scanning process of Cascade assemblies in *Escherichia coli* cells. The activity of recombinant Type I-Fv Cascade complexes of *Shewanella putrefaciens* CN-32 serves as a model system that facilitates comparative studies for many of the diverse CRISPR-Cas systems.

Introduction

Prokaryotes have evolved several mechanisms to evade phage infections, including restriction–modification systems (R/M) that are able to inactivate target DNA by cleavage (Pingoud, Fuxreiter, Pingoud, & Wende, 2005), toxin-antitoxin modules that result in abortive phage infection and CRISPR–Cas systems (Unterholzner, Poppenberger, & Rozhon, 2013). The acronym CRISPR-Cas is generated from the abbreviations for Clustered Regularly Interspaced Short Palindromic Repeats and CRISPR-associated proteins (Jansen, van Embden, & Gaastra, 2002). CRISPR arrays were first discovered as unusual structures in the *Escherichia coli* K12 genome (Ishino, Shinagawa, Makino, Amemura, & Nakata, 1987), years before the detection of associated *cas* genes (Jansen et al., 2002). CRISPR-Cas was suggested to function as an adaptive immune system in prokaryotes in 2005 (Mojica, Díez-Villaseñor, García-Martínez, & Soria, 2005; Pourcel, Salvignol, & Vergnaud, 2005), which was experimentally confirmed two years later (Barrangou et al., 2007).

CRISPR-Cas RNA-guided immune systems are widespread in prokaryotes and play a major role in microbial evolution (Jore et al., 2011; Makarova et al., 2011). These systems can be found in about 50 % and 90 % of all sequenced bacteria and archaea, respectively (Grissa, Vergnaud, & Pourcel, 2007b). CRISPR-Cas systems offer prokaryotes a sequence-specific defense mechanism against invasion of foreign nucleic acids from viruses, plasmids or other mobile genetic elements (MGEs) (Barrangou et al., 2007; Edgar & Qimron, 2010; Garneau et al., 2010; Marraffini & Sontheimer, 2008). In this process, these immune systems do not rely on antigen-antibody interactions, but work directly at the nucleic acid level. Small prokaryotic CRISPR-RNA molecules (crRNA) are able to guide a single Cas protein effector or a multi-subunit Cas protein effector complex towards invading DNA or RNA stretches that exhibit base-complementarity to the crRNA. Subsequently, Cas nucleases destroy these foreign nucleic acids, disarming viral threats (Barrangou et al., 2007; Brouns et al., 2008; Mohanraju et al., 2016).

CRISPR-Cas immunity is established in three defined steps; (1) spacer acquisition or adaptation, (2) crRNA expression or biogenesis and (3) interference. The CRISPR-*cas* locus usually describes a cluster of *cas* genes found next to one or several CRISPR arrays. These arrays contain several repetitions of identical sequences (repeats) and variable sequences (spacers). Spacer regions constitute sequences from viruses or MGEs that were integrated via CRISPR-Cas activity. Since integration is directional, the

first spacers in a CRISPR array often correspond to the most recent interactions (Barrangou et al., 2007; Jansen et al., 2002; Pourcel et al., 2005).

The spacer acquisition step (adaptation) is similar for most CRISPR-Cas systems (Koonin, Makarova, & Zhang, 2017). An adaptation module consists of two Cas proteins: the endonuclease Cas1 and the structural subunit Cas2 (Amitai & Sorek, 2016). These two proteins form a stable complex (Nunez et al., 2014; J. Wang et al., 2015) that is responsible for the recognition of foreign DNA or RNA as targets, following integration of a 20 bp to 40 bp sequence into the CRISPR array. In this process, the first repeat sequence is duplicated. To identify foreign nucleic acids as a target, a specific sequence motif is recognized (Sun et al., 2013). This so-called Protospacer Adjacent Motif (PAM) consists of two to five nucleotides and is located directly adjacent to the sequence that gets inserted into the CRISPR array, the protospacer (Brouns et al., 2008).

In the subsequent biogenesis step the CRISPR array is transcribed into a long precursor crRNA (pre-crRNA) that is processed into individual, mature crRNA pieces by type-dependent Cas proteins. Each of these mature crRNAs consists of a single spacer sequence and fragments of the flanking repeat regions. In addition, the *cas* genes are transcribed, translated and assembled into protein complexes with single crRNAs (Carte, Wang, Li, Terns, & Terns, 2008; Haurwitz, Jinek, Wiedenheft, Zhou, & Doudna, 2010).

The final step, interference, exhibits most variation in regard to the involved Cas proteins and mechanisms (Koonin et al., 2017; Makarova et al., 2011; Makarova et al., 2015). Different crRNA-guided effector complexes enable recognition and destruction of foreign genetic material. Hallmark features of this process are (i) crRNAs that utilize base-complementarity between their spacer sequence and nucleic acids to identify targets and (ii) Cas protein complexes that identify PAM sequences near the potential target to prevent harmful self-targeting (Marraffini & Sontheimer, 2010). The PAM sequences differ strongly between subtypes of CRISPR-Cas systems, but are highly conserved within subtypes (Chylinski, Makarova, Charpentier, & Koonin, 2014; Makarova et al., 2011). The diversification of Cas protein interference complexes required an elaborate classification approach for the different CRISPR-Cas systems.

In the current classification, the largest distinction was made in regard to the number of Cas proteins that are present in the effector complex that binds the crRNA. Class II encompasses Type II, V and VI systems that rely on a single large Cas protein for interference. In Type II systems, this effector is called Cas9, a widely used tool for genome editing approaches (Cong et al., 2013; Komor, Badran, & Liu, 2017). Type V

systems are characterized by Cas12 effector proteins that target DNA molecules, while Type VI systems have Cas13 effector proteins that specifically target RNA molecules (Koonin et al., 2017).

Class I CRISPR-Cas systems are differentiated from class II systems by the use of multi-subunit effector complexes. Among them, Type I systems are the most common and diverse in nature. Type III systems are mostly found in archaea and act together with Type I systems. Similarly, Type IV systems appear to rely on the presence of Type I CRISPR-Cas systems as they usually lack adaptation modules. All of these systems form multi-protein complexes with a backbone of several Cas7 subunits proteins that contain a signature RNA recognition motif. Cas7 covers the spacer sequence of the associated crRNA. RNA interactions are also mediated via Cas5 and Cas6 proteins that cap the 5′ and 3′ termini of the crRNA, respectively. Additional proteins, called large and small subunits, are mostly responsible for DNA interactions in the CRISPR-ribonucleoprotein complexes (crRNPs) (Hochstrasser et al., 2014; Jackson et al., 2014; Staals et al., 2014; van der Oost, Westra, Jackson, & Wiedenheft, 2014; Zhao et al., 2014).

In Type I CRISPR-Cas systems, these crRNPs are termed CRISPR-associated complex for antiviral defense (Cascade). Cascade assemblies of different subtypes differ in the number of involved Cas proteins as well as in size and shape of the crRNP (Makarova et al., 2015; Pausch et al., 2017). The seahorse-shaped Cascade architecture of subtype I-E systems is generated by Cas5, Cas6, Cas7, Cas8 and Cas11 in a stoichiometry of 1:1:6:1:2 (Hayes et al., 2016; Jackson et al., 2014; Mulepati, Heroux, & Bailey, 2014; Wiedenheft et al., 2011). The Cascade of subtype I-F systems has a similar architecture, but lacks the small subunit protein Cas11 (Chowdhury et al., 2017; Makarova et al., 2015). Nonetheless, both subtypes exhibit similar structures of shared Cas proteins and comparable PAM recognition mechanisms via dsDNA interactions by the large subunit (Cas8).

A subtype I-Fv Cascade was identified in *Shewanella putrefaciens* CN-32 and shown to lack both Cas11 and Cas8, resulting in a minimal Cascade architecture of three different proteins and one mature crRNA (Dwarakanath et al., 2015; Gleditzsch et al., 2016; Pausch et al., 2017) (**Figure V.1**).

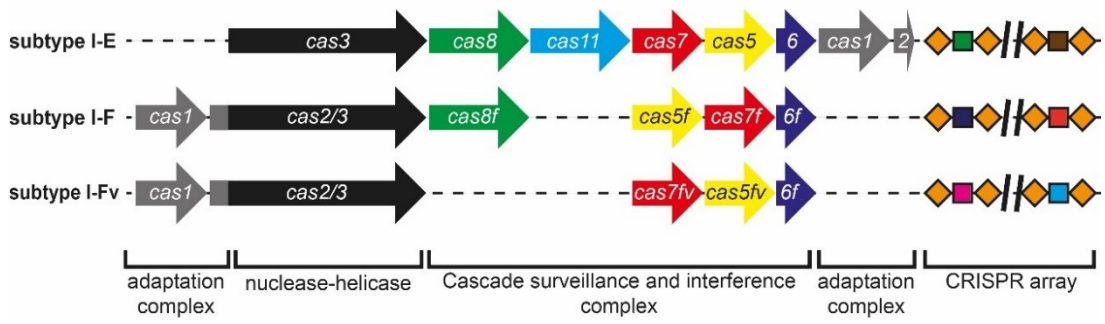


Figure V.1. Gene and CRISPR array arrangement of subtype I-E, I-F and I-Fv CRISPR-Cas systems. A schematic representation of CRISPR-Cas loci is shown. Arrows indicate *cas* genes of comparative length. The Cascade surveillance and interference complex genes are multicolored and consist of five, four and three different genes, respectively. The helicase-nuclease Cas3 is fused to Cas2 in subtypes I-F and I-Fv. The CRISPR arrays are indicated as alternating rhombi (orange, CRISPR-DNA (crDNA) repeats) and squares of various colors (different crDNA spacers).

The variant structures of Cas5fv and Cas7fv proteins of subtype I-Fv systems result in an “open” Cascade formation that resembles subtype I-E Cascade. The small and large subunits are functionally replaced by Cas7fv and Cas5fv, respectively. Cas5fv interacts with the 8 nt long 5′ handle of the crRNA, whereas Cas6f binds to the 20 nt long 3′ terminal hairpin and interacts with the Cas7fv backbone (Pausch et al., 2017). A single Cas7fv backbone subunit covers a 6 nt sequence stretch of the spacer and extending or shortening of the spacer results in altered Cas7 subunit amounts. The *S. putrefaciens* subtype I-Fv system relies on the recognition of a two nucleotide “GG”-PAM to distinguish between target and self DNA (Dwarakanath et al., 2015). In contrast to other Type I systems, PAM recognition is performed from the DNA major groove side by a protruding alpha-helical domain of Cas5fv (Pausch et al., 2017). Cascade effector complexes of Type I CRISPR-Cas systems are able to bind invading nucleic acids by generating R-loop structures as a result of hybridization between crRNA spacers and target DNA protospacers (Garneau et al., 2010; Jore et al., 2011; Westra et al., 2012). Subsequently, the Cas3 helicase-nuclease is recruited to the R-loop region to unwind and finally degrade the foreign nucleic acids (Hochstrasser et al., 2014; van der Oost et al., 2014).

Dynamics of Cas protein-mediated DNA targeting

CRISPR-Cas systems, both Type I and II, were studied on a single-molecule level *in vitro,* e.g. by DNA curtains, single-molecule FRET assays or using magnetic tweezers, providing detailed knowledge about their kinetics and the underlying molecular mechanisms. Here, it could be shown that target binding and R-loop formation proceeds directionally along the target sequence and that full target sequences with incorrect

PAMs are ignored (Sternberg, Redding, Jinek, Greene, & Doudna, 2014) and that PAM proximal sequences affect the R-loop formation rate, whereas PAM distal sequences are involved in R-loop stabilization (Szczelkun et al., 2014). Further studies revealed that Cas9-RNA/DNA association rates are only weakly dependent on the spacer sequence but the dissociation rates increase significantly upon introduction of mismatches in PAM proximity (Singh, Sternberg, Fei, Doudna, & Ha, 2016), R-loop formation was shown to respond to sequence mismatches and that when the R-loop formation reaches the protospacer end, complexes enter a locked state until Cas3 nuclease recruitment (Rutkauskas et al., 2015). Furthermore, quantitative bionumbers could be retrieved, e.g. single-molecule FRET studies revealed the binding times of Type I-E Cascade during interference. Here, the authors found two binding modes, one longer than 30 minutes, which is considered being an interference mode, and the second with ~25 s lifetime being connected to partial R-loop formation (Blosser et al., 2015). DNA curtains experiments showed that Cascade complexes remain in their locked state for more than 57 minutes and for only approximately 25 seconds, when bound to a partially complementary sequence (Redding et al., 2015). Finally, in both studies, the authors evaluated the DNA-probing time to be in the second range: 0.75 to 3 seconds (Redding et al., 2015) and 1.6 ± 0.4 seconds (Blosser et al., 2015). In summary, the *in vitro* measurements of CRISPR-Cas interference show that Cascade-DNA interaction times cover a wide range: from seconds to tens of minutes. However, *in vitro* conditions investigate CRISPR-Cas systems in isolation from their intracellular environment. Thus, it is important to explore CRISPR-Cas interference *in vivo* as well. Until now, only two *in vivo* studies were published, one investigating Cas9 target search in living mammalian cells (Knight et al., 2015) and one in *E. coli* cells (Jones et al., 2017).

In the following sections, we will provide the methodological details for performing live-cell single-particle tracking of the molecular dynamics of Cas protein-mediated DNA surveillance. Single-molecule localization microscopy (SMLM) techniques (Betzig et al., 2006; Folling et al., 2008; Heilemann et al., 2008; Hess, Girirajan, & Mason, 2006; Rust, Bates, & Zhuang, 2006) can bypass the diffraction limit of light by utilizing photochromic fluorophores that can be photoswitched between different molecular states. This allows switching a small, spatially separated subset of fluorophores into their fluorescent on-state, resulting in spatially well-separated single fluorescent spots, characteristic point spread function (PSF) profiles that are resolvable within the diffraction limit. After the full readout of this subset, a new subset is switched on and read out. Repeating this procedure over time leads to a read-out of all fluorophores (**Figure V.2a**). As PSFs are spatiotemporally separated, each centroid can be estimated with nanometer precision

(**Figure V.2b**). This procedure enables high-density mapping of fast single-protein dynamics by single-particle tracking photoactivated localization microscopy (sptPALM) (Manley et al., 2008) as sptPALM combines single-molecule tracking with the spatiotemporal separation of fluorophores of SMLM techniques (**Figure V.2c**).

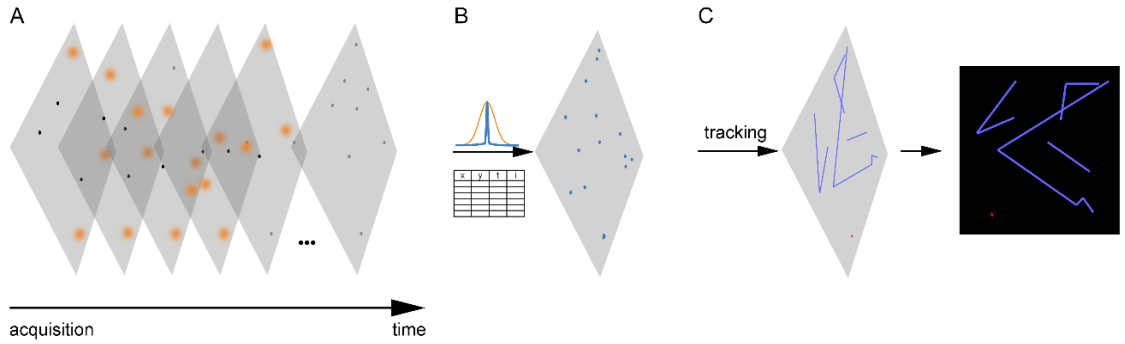


Figure V.2. Workflow of sptPALM. (A) Image acquisition: Fluorescence signals of individual molecules are collected in an acquisition sequence. Fluorophores are photoswitched sequentially to spatially separate the individual signals; (B) localization routine: Each fluorescent spot in form of a PSF is fitted by a localization algorithm to extract its centroid, which determines the fluorophores localization with nanometer precision; (C) tracking routine: localizations are classified into single-molecule trajectories by a tracking algorithm. The dynamics of the single proteins can be inferred from the obtained trajectories.

The described minimal Type I-Fv Cascade of *S. putrefaciens* serves as a model system and is studied in a heterologous *E. coli* host. Particularly, this protocol allows to investigate the different time scales of Cascade interference and to probe their dynamics depending on the complementarity of different crRNAs, as sketched in **Figure V.3**. For example, when imaging the Cascade dynamics using long camera integration times of several seconds, only fluorescence signals from stable interference complexes are detected. The nucleoid, due to its size, is virtually immobile on these time scales, thus complexes bound to genomic targets appear immobile (**Figure V.3**, Cascade (Anti-REP), 16 genomic targets, see section 2.1). Cascade complexes which are only transiently binding or carrying a non-complementary crRNA only show diffusive signals since the transient DNA interactions during target search processes are too short-lived to be visible as clear fluorescent spots at seconds-long exposure times (**Figure V.3**, Cascade (Anti-λ), no genomic targets, see section 2.1). A fast imaging rate, and thus higher temporal resolution, of 30 ms per frame allows to also detect the diffusive molecules. Thereby it is possible to also investigate the transient dynamics of Cascade complexes (**Figure V.3, right)**.

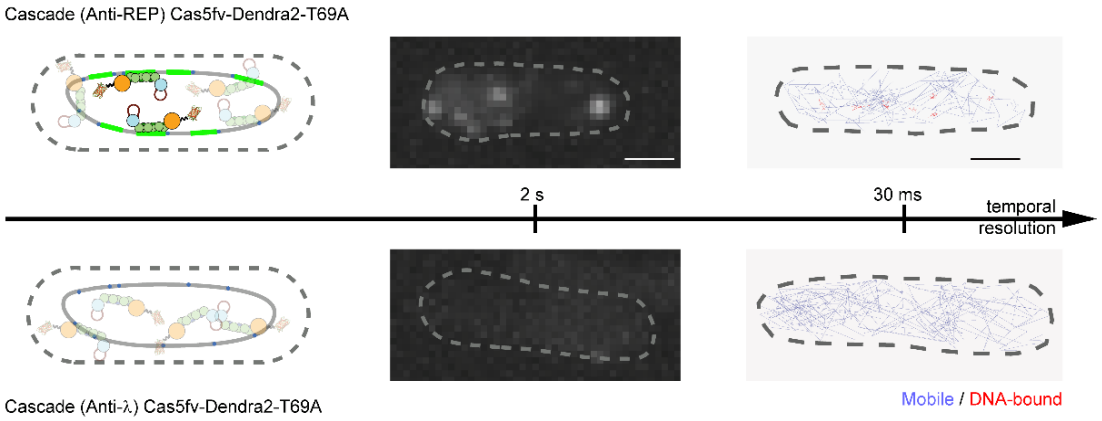


Figure V.3. Two *E. coli* strains expressing Cascade I-Fv complexes carrying different crRNAs. Here the Anti-REP crRNA is targeted against 16 sequences in the genome (top panel), the Anti- λ crRNA has no target sequence present in the host (bottom panel). Imaging with long integration times reveals Cascade complexes bound to their genomic targets while unbound Cascades are not detected. Faster imaging rates at higher temporal resolution visualize the DNA-bound and the mobile fractions of transient Cascade interference dynamics (red (JD ≤ 160 nm) and blue trajectories, respectively). Scale bars: 1 µm.

Generation of Cascade complexes with a fluorescent tag

In order to follow the dynamics of Type I CRISPR-Cas systems *in vivo*, the investigated crRNP complex needs to be tagged with a suitable photoswitchable fluorescent protein. For this purpose, two alternative strategies are possible. First, the CRISPR-Cas system can be studied in the native host organism which has to be suitable for direct imaging. Factors like low autofluorescence and the availability of genetic modification tools need to be considered. As a second option, it can be preferable to transfer the studied CRISPR-Cas system into a suitable host or model system. This protocol covers the imaging of heterologously expressed Type I-Fv CRISPR-Cas complexes from *S. putrefaciens* CN-32 in *E. coli* BL21-AI, which scan be used to test a large variety of CRISPR-Cas systems from mesophilic organisms with similar codon usage. In this case, the transfer of the system was preferred due to the pink coloration of the host organism, a feature that interferes with fluorescent imaging. BL21-AI is a commonly used *E. coli* strain for the heterologous production of proteins and for studies on CRISPR-Cas activity, since it lacks endogenous *cas* genes (Brouns et al., 2008) and prophages. Furthermore, it carries the T7 polymerase gene under the control of the Ara promoter, allowing the use of T7 promoters for the controlled transcription of genes.

The selected fluorescent protein, Dendra2-T69A, has several advantages for sptPALM, such as high photon budget and well-controlled photophysics (Berardozzi, Adam, Martins, & Bourgeois, 2016), which is critical for precise localization and detection sensitivity when it comes to identifying highly mobile proteins. Additionally, it is a monomeric fluorescent protein like its direct ancestor Dendra2 and it is not suspected to

oligomerize or form artificial aggregates (S. Wang, Moffitt, Dempsey, Xie, & Zhuang, 2014).

The Type I-Fv CRISPR-Cas system features a multi-protein effector complex and different Cas proteins are available for fluorescent tagging. In our approach, we determined the single Cas5fv subunit of Cascade to be optimal for tagging with Dendra2-T69A. Proteins with more than one copy per complex should not be selected for fusion, as this setup might lead to errors during data analysis. For example, if Cas7fv subunits were to be tagged, it would not be possible to discern two close-by fluorescent signals as separate Cascades or a single Cascade that was activated more than once.

Due to the large size of the *cas* gene operon of *S. putrefaciens* CN-32, we separated the genes into two different constructs to facilitate cloning into Duet vectors. Furthermore, the parallel generation of both amino- and carboxy-terminal protein fusions is suggested, since both the Cas protein and the fluorophore activity depends on proper folding, a property that is difficult to predict *in silico.* To minimize the risk of misfolding, we introduced a “GGGGS”-linker between the two proteins. Other linker options can be assessed if none of the fusions exhibits appropriate activity (Chen, Zaro, & Shen, 2013).

The use of genomic targets requires the generation of a nuclease-deficient system for *in vivo* imaging. For Type I systems, the HD nuclease domain is highly conserved for Cas3 proteins (Koonin, Makarova, & Zhang, 2017), and the disruption of it is sufficient to abolish target cleavage for the Type I-Fv system (Dwarakanath et al., 2015).

Finally, for the successful imaging of target-bound Cascade, previous knowledge on the crRNA and target requirements of the complex is needed. The length of the spacer, the nature of the targeted molecule (DNA or RNA) and PAM specificity are essential variables to consider during experiment preparation. Further information on seed length and mismatch tolerance is extremely useful, as these parameters are directly related to Cascade retention time on the target and should be taken into account, both for experimental design (e.g. target selection) and data interpretation.

Target selection and oligonucleotide design for the cloning of CRISPR arrays

In order to study the target binding dynamics of Cascade, three types of spacers were designed:

i. **Control spacer with no targets on the genome.** For this purpose, a 32 nt sequence that is complementary to a region of gene E from phage lambda flanked by the I-Fv PAM GG at the 3′ end of the target strand was selected (Gleditzsch et al., 2016). This spacer has no targets in the *E. coli* BL21-AI host genome and the longest observed stretch of complementarity spans only 9 nts.

This crRNA can be used to determine the basal target search dynamics of Cascade and allows for testing the activity of the fluorescently-labeled complex via efficiency of plaquing assays. The following spacer sequence was used: 5′-GGCGGCACGGAGTGGAGCAAGCGTGACAAGTC-3´.

ii. **Targeting spacer with full complementarity to the host genome.** In order to study the DNA-binding dynamics of Cascade complexes, selection of appropriate targets is needed. These sites should be abundant, as it is difficult to study a single binding event in the presence of noise from a larger pool of non-binding complexes looking for targets.

The use of plasmid targets is not recommended, as their polar localization in the cell might overlap with inclusion bodies due to protein over-expression. Furthermore, the reduced cellular size at this position poses a challenge for single-particle tracking, due to higher "bouncing" of the molecules against the cell membrane. An alternative solution is the generation of synthetic target arrays by insertion of large repeat-constructs into the *E. coli* genome, as this allows controlling the number and position of targets. A previous study on Cas9 dynamics utilized an array of LacO sites (Jones et al., 2017), however the short length of these elements (23 bp) makes it inapplicable for Type I CRISPR-Cas systems, where spacers are longer than 30 bp. The insertion of target arrays with longer protospacers would constitute a challenge for genome stability, as repeat sequences over 27 bp are subject to homologous recombination by RecBC (Shen & Huang, 1986). Insertion of individual targets throughout the genome is also a possibility, although very arduous and time-consuming.

A simpler option, suitable for Cascade targeting, is the use of the repetitive elements already present in the *E. coli* genome. Due to the size constrains previously described, we selected the 35 bp long Repetitive Extragenic Palindromic (REP) elements (Tobes & Ramos, 2005). These elements are highly suitable targets, since they are of the appropriate size, contain a GG to act as PAM and are distributed throughout the genome (**Figure V.4**). The natural distribution of targets provides variability, since not all of them might be available for targeting at all times in live cells.

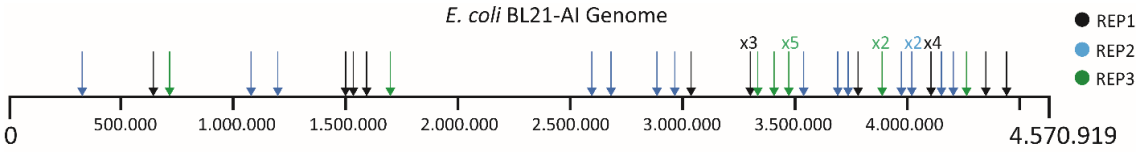


Figure V.4. Distribution of targeted REP elements throughout the *E.coli* BL21-AI genome. Schematic representation of the *E. coli* BL21-AI genome, with arrows indicating the positions of the protospacers targeted by the different spacers used to study the dynamics of Cascade complexes. Black arrows represent the targets for spacer REP1, blue arrows for REP2 and green arrows for REP 3.

The crRNP complex versus target ratio should be taken into consideration when determining the amount of spacers that the CRISPR array should carry. As an optimal ratio of 1:1 is difficult to accomplish, it is suggested to utilize a maximum number of targets to then control the ratio by modulating the induction levels of Cascade. When targeting REP elements with a 32 nt spacer, full targets can be increased from 16, with a single spacer (REP1: 5′-GGATGCGGCGTAAACGCCTTATCCGGCCTACG-3′), to 42 with 3 spacers (REP2: 5′-GGATGCGGCGTAAACGCCTTATCCGGCCTACA-3′, REP3: 5′-GGATGCGGCGTGAACGCCTTATCCGGCCTACG-3′), due to the presence of mismatches at positions 32 and 12 of the protospacers. Imaging of Cascades carrying only 1 spacer variant or all 3 variants gives further insights into the dynamics of the complex, as not only the total number of targets increases, but also the number of incomplete intermediates. Furthermore, as the crRNA is the limiting component for Cascade formation, the increase of spacer number would also promote Cascade formation.

iii. **Targeting spacer with reduced complementarity to regions of the genome.** The design of spacers with partial complementarity allows for studying the stability of the DNA binding events and the retention time of the complex. Since Cas7 subunits cover 6 nt intervals of the crRNA for Cascade and other Class I CRISPR-Cas system, the design of spacers with 6 nt increasing complementarity to the REP elements allows for the study of possible R-loop variants. For this, spacers REP1, REP2 and REP3 were modified to have complementarities of 30 nt, 24 nt, 18 nt and 12 nt to the genomic targets, including randomized sequences with no complementarity to the host DNA to complete the 32 nt spacer stretch. Since REP1 and REP2 sequences are identical apart from the last nucleotide of the spacer (in order to cover targets with a mismatch at position 32), when designing crRNAs with 30 bp of complementarity or less, the CRISPR array can be reduced to only two spacers. **Table 1** shows all the intermediate spacers and

the total amount of incomplete targets in the genome. By shifting the spacer towards an appropriate PAM and modulating the spacer length, the presented sequences can be adjusted to guide other interference complexes, with the REP sequence length of 35 bp as only constraint.

Table 1. Sequence and target number of spacers with partial complementarity to REP elements on the *E. coli* BL21-AI genome.

crRNA-DNA complementarity	Sequence REP1	Sequence REP2	Number of genomic targets
30 bp	GGATGCGGCGTAAACGCCTTATCC GGCCTAAT	GGATGCGGCGTGAACGCCTTATCCG GCCTAGC	75
24 bp	GGATGCGGCGTAAACGCCTTATCC GAAACGTT	GGATGCGGCGTGAACGCCTTATCCC GATAGGT	99
18 bp	GGATGCGGCGTAAACGCCCCTGTT CGTGTCGT	GGATGCGGCGTGAACGCCATCGGTA AATAGCC	107
12 bp	GGATGCGGCGTACCGCGGAGCCA TGTGCCATA	GGATGCGGCGTGCTCGTCTGCGGAG CACTCTG	133

For *S. putrefaciens* CN-32 Type I-Fv systems, the repeat sequence has been defined as 5′-GTTCACCGCCGCACAGGCGGCTTAGAAA-3′ (Dwarakanath et al., 2015). When studying other systems, the repeat sequence can be predicted using the CRISPRCasfinder tool (Couvin et al., 2018; Grissa et al., 2007b), if the organism is not already listed in the CRISPRdb database (Grissa, Vergnaud, & Pourcel, 2007a). Array directionality is relevant for determining the proper repeat sequence and can be predicted using CRISPRDetect (Biswas, Staals, Morales, Fineran, & Brown, 2016). Type I repeats usually carry a short adenine stretch at the 3′ end, providing additional support for the correct directionality prediction of the repeat.

For the cloning of the previously selected spacers as CRISPR arrays, two different strategies were utilized. Single spacer arrays were designed on complementary oligonucleotides pairs carrying corresponding restriction sites at their ends (**Figure V.5**, bottom). For multiplexed sequences, synthesis of the arrays was necessary in order to avoid assembly errors due to the amount of repeats. For these constructs, the same restriction sites as for the minimal CRISPR were added to flank the array.

Oligonucleotide design for the cloning of *cas* genes

Cas genes are encoded in large operons. The *S. putrefaciens* CN-32 *cas* gene operon comprises genes encoding for Cas1, Cas2-3, Cas7fv, Cas5fv and Cas6f, at a total size of 6.457 bp. As Cas1 is needed for adaptation, but its role in interference was deemed

negligible (Brouns et al., 2008; Dwarakanath et al., 2015), this gene was excluded from the cloning strategy. In order to facilitate the assembly of the required plasmids, the operon was split and oligonucleotides were designed to introduce all the components into two multiple cloning sites of Duet vectors. In order to promote complex formation in *E. coli*, the genes *cas7fv* and *cas6f* were introduced into pRSFDuet-1, a high copy plasmid (around 100 copies per cell), while *cas2-3* and the *cas5fv-dendra2-T69A* fusion were designed to be inserted in the second multiple cloning site of pCDFDuet-1, a plasmid of lower copy number (approximately 20 copies per cell). This plasmid was also designed to carry the minimal CRISPR arrays. For the fluorescent fusion, the sequence coding for a "GGGGS" linker was added to the corresponding oligonucleotides for amplifying *cas5fv* and *dendra2-T69A,* either upstream or downstream of the *cas* gene to generate amino- or carboxy-tagged versions, respectively.

An overview of the oligonucleotides designed for cloning these constructs, as well as the resulting plasmids is shown in **Figure V.5**. Gibson assembly (Gibson et al., 2009) requires a cleaved backbone in addition to PCR fragments of the components to be assembled together. The proper arrangement of the fragments is achieved by annealing of the fragments with overlapping regions of at least 30 bp. Incorporation of these sequences occurs during the PCR reaction, since the oligonucleotides used to amplify each unit carry these extra nucleotides. For oligonucleotide design, corresponding 30 nt overlapping sequences (diagonal sections of the black arrows in **Figure V.5**) were added to the 20 nt sequence stretch needed for regular PCR amplification. Randomized nucleotides were added between genes, in addition to ribosome binding sites to ensure translation of all the components in the operon. Restriction sites were chosen to not cleave the previously designed CRISPR arrays, since these would be cloned first into the plasmids, and added to the corresponding oligonucleotides.

As genomic targeting requires the expression of a nuclease-deficient Cas2-3 protein, oligonucleotides for site-directed mutagenesis were designed by using the online tool "Quik Change Primer Design" from Agilent (https://www.agilent.com/store/primerDesignProgram.jsp), in order to mutate the HD domain by alanine replacement (H156A/D157A). These mutations were previously described to be sufficient for abolishing CRISPR-Cas interference activity (Dwarakanath et al., 2015).

Plasmid generation

As a first approach, the Anti- λ CRISPR array was introduced into the BamHI and HindIII sites of pCDFDuet-1 by oligonucleotide annealing **(Figure V.5, bottom)** (Zoephel, Dwarakanath, Richter, Plagens, & Randau, 2012). Synthesized oligonucleotides were phosphorylated, mixed at equal molar ratio, heated up to remove secondary structures, and cooled down slowly to promote annealing. The duplex, which carries sticky ends generated by incomplete annealing at the oligonucleotide termini, was then ligated into a restricted pCDFDuet-1 vector. Array insertion was corroborated by Sanger sequencing, and downstream cloning of *cas2-3* and the *cas5fv* fluorescent fusions was performed by Gibson assembly. In order to determine the activity of amino- and carboxy-fusions of Cas5fv and Dendra2-T69A and to select the most appropriate option, both constructs and *cas2-3* were cloned into the second multiple cloning site of the Anti- λ pCDFDuet-1 using KpnI and XhoI restriction sites. The genes were amplified with the corresponding oligonucleotides carrying overlapping regions, and the resulting PCR fragments were mixed with the cleaved backbone at a ratio of 3:1. After addition of 2X Gibson Assembly Master Mix (NEB), samples were incubated for 1 hr at 50ºC and transformed into chemically competent *E. coli* DH5α cells. Following overnight growth at 37ºC, plasmids were extracted and correct plasmid assembly was confirmed by Sanger sequencing.

Cas2-3 nuclease deficient constructs were generated by site-directed mutagenesis by use of the QuikChange Site Directed Mutagenesis Kit (Agilent) and the previously described oligonucleotides, following the manufacturer's protocol.

Once fluorescence readout and Cascade interference was determined, the best performing fusion plasmid was modified by insertion of the remaining REP-targeting arrays into the plasmid encoding for the Cas2-3 HD mutant.

The CRISPR array carrying a single spacer against the REP elements was cloned by oligonucleotide annealing **(Figure V.5, bottom)**, while multiplexed arrays previously synthesized (Sigma-Aldrich) were introduced by restriction enzyme digestion and ligation. These constructs were cloned into the first multiple cloning site of pCDFDuet-1 using the restriction sites NcoI and NotI, excluding the ribosome binding site and the His-tag present on the empty backbone. Array insertions were corroborated by Sanger sequencing.

In parallel, empty pRSFDuet-1 was digested by BamHI and NotI restriction enzymes, and *cas7fv* plus *cas6f* were cloned by Gibson Assembly into this site, as depicted on **Figure V.5**, right.

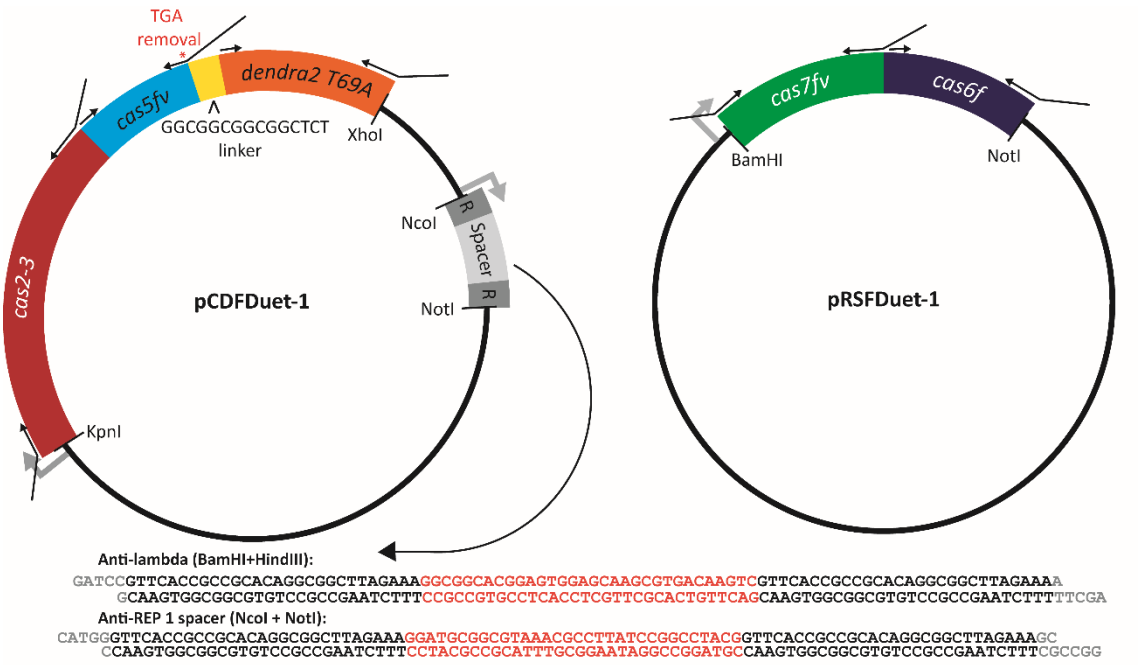


Figure V.5. Cloning strategy for the generation of fluorescently-labelled Cascade. T7 promoters are shown as gray arrows, while Gibson Assembly primers are shown as small black arrows including in diagonal the overlapping regions needed for assembly. R stands for repeat sequence. Due to size constraints, the *cas* operon was divided and re-arranged into two fragments. **Left:** *cas2-3* and the *cas5fv-dendra2-T69A* fusion (carboxy-terminus with the corresponding linker shown as example) were introduced into the second multiple cloning site of pCDFDuet-1. On the first multiple cloning site of the same backbone, variants of a minimal CRISPR array were introduced either by oligonucleotide annealing or restriction enzyme digestion and ligation. Both array versions generated by the first technique are shown, with the generated restriction sites in light gray, the repeat sequence in black and the spacer in red (bottom). **Right**: *cas7fv* and *cas6f* were cloned together as an operon by Gibson Assembly onto the first multiple cloning site of pRSFDuet-1.

Heterologous expression of Cascade

The described plasmids were used to transform the expression strain, in order to test the activity of the fluorescent protein and Cascade. In addition, optimal expression conditions were screened.

Transformation into BL21-AI

I. Generation of strains for the evaluation of Cascade activity. In order to determine the activity of the tagged complex, 100 ng of plasmid combinations were introduced by heat-shock into 100 µl aliquots of chemically-competent *E. coli* BL21-AI (Invitrogen). Strains were designed to carry one version of pRSFDuet-1 and one of pCDFDuet-1, which exhibit compatible antibiotic resistance against kanamycin and spectinomycin.

i. Positive lambda control: pCDFDuet-1 and pRSFDuet-1 empty backbones. This strain, when subjected to infection, displays the full plaquing activity of the bacteriophage and was utilized as a normalization strain for the experiment and imaging.

ii. Positive WT Cascade control: pCDFDuet-1 carrying the Anti- λ CRISPR array and pRSFDuet-1 encoding for the non-tagged Cas operon (excluding Cas1). This strain was utilized as a comparison point for the activity of the Cascades carrying a fluorescent fusion of Cas5fv.

iii. Negative WT Cascade control: pCDFDuet-1 carrying the Anti- λ CRISPR array and pRSFDuet-1 encoding for the non-tagged Cas operon with the Cas2-3 HD mutant (excluding Cas1). This strain allows linking the reduction of plaquing efficiency (if any) to Cascade activity.

iv. Positive Cascade-Dendra2-T69A: pCDFDuet-1 carrying the Anti- λ CRISPR array plus one of the Cas2-3/Cas5fv-Dendra2-T69A fusion version (Cas5fv amino- or carboxy-tagged), in addition to pRSFDuet-1 encoding for Cas7fv and Cas6f.

v. Negative Cascade-Dendra2-T69A: pCDFDuet-1 carrying the anti-lambda CRISPR array plus one of the Cas2-3 HD mutant/Cas5fv-Dendra2-T69A fusion version (Cas5fv amino- or carboxy-tagged), in addition to pRSFDuet-1 encoding for Cas7fv and Cas6f.

II. Generation of strains for sptPALM imaging. 100 µl aliquots of chemically-competent *E. coli* BL21-AI cells were transformed by heat shock with 100 ng of a pCDFDuet-1 plasmid variant carrying the different CRISPR arrays and the *cas2-3* HD mutant/*cas5fv-dendra2-T69A* cassette; plus 100 ng of pRSFDuet-1 plasmid encoding for Cas7fv and Cas6f. A total of six strains were generated, one per each pCDFDuet-1 plasmid with a CRISPR array described in section 2.1.1: four with partial complementarity to the REP elements, one with full complementarity, and one complementary to gene E of the bacteriophage lambda.

Evaluation of the fluorescent signal of Cascade-Dendra2-T69A

As a first approach to corroborate the activity of the tagged complex, the fluorescent signal of the fusions was tested by fluorescent microscopy, taking advantage of the pre-converted state of Dendra2-T69A. In this conformation, the fluorescent protein exhibits a similar excitation and emission spectrum as GFP (excitation: 502 nm, emission: 518 nm) (Berardozzi et al., 2016), allowing for the imaging with microscope setups programmed to visualize this commonly used reporter protein.

To guarantee that the obtained signal is Dendra2-T69A-dependent, strains carrying empty plasmids and expressing the fluorescent protein alone were generated as negative and positive controls, respectively. Single colonies of the strains expressing both plasmids necessary for fluorescent Cascade formation plus the controls were grown

over night at 37ºC, with shaking at 200 rpm in 5 ml of 2YTL media (16 g/L tryptone, 10 g/L yeast extract, 5 g/L NaCl, 10 mM $MgSO_4$, 0.2 % maltose) plus 25 µg/ml of spectinomycin and kanamycin.

Cells were reinoculated into 20 ml of fresh 2YTL media with antibiotics at a ratio of 1:100 and grown at 37ºC, 200 rpm, until OD_{600nm} = 0.3. All cultures were induced by addition of 0.2% of arabinose and 0.1 mM of IPTG (Sigma-aldrich) and grown until OD_{600nm} = 0.6. For all strains, 1 ml of culture was pelleted at 4000 g for 2 min and resuspended in 50 µl of 1X phosphate-buffered saline (PBS, NaCl 137 mM, KCl 2.7 mM, Na_2HPO_4 10 mM, KH_2PO_4 1.8 mM).

Agarose pads were prepared by pipetting 300 µl of 2% agarose in 1X PBS into a glass slide, avoiding bubbles, and then covering it with an extra glass slide. When dried, the top slide was removed and 10 µl of the cell suspension was placed on the middle of the pad and covered with a glass slip, avoiding bubble formation.

Cells were imaged at 100X magnification with the Zeiss Axioplan II microscope with filter sets for GFP detection (band-pass 470/20, beam splitter 493, band-pass 505–530 nm; Zeiss), using the negative control to set the background noise and the positive control to test the proper selection of excitation and emission wavelengths. Photographs were taken with the charged-coupled camera CoolSNAPHQ (Photometrics). Image processing was done with MetaMorph software, version 6.2r6 (Universal Imaging). For Type I-Fv Cascade-Dendra2-T69A, both amino- and carboxy-fusions displayed the expected fluorescent emission. Imaged cells presented fluorescent cell poles, indicating the formation of inclusion bodies, likely caused by protein over-expression. To minimize this issue, decreased induction conditions were tested, with expression induced by the addition of only 0.1% arabinose (inducing T7 polymerase production, but not lifting LacI repression) was deemed to be optimal for further sptPALM experiments, with clear fluorescent signals and good cell morphology.

Evaluation of Cascade activity by Efficiency of Plaquing assays

To confirm the activity of fluorescent Cascade variants, strains described in section 3.1 I were challenged in Efficiency of Plaquing (EOP) assays with a lytic version of phage lambda (NCCB 3467) carrying an unaltered protospacer targeted by the CRISPR arrays present on the strains.

Single colonies from the required strains were grown over night at 37ºC, 200 rpm shaking, in 5 ml of 2YTL media supplemented with 25 µg/ml of spectinomycin and kanamycin.

Cells were reinoculated into 20 ml of fresh 2YTL media (with antibiotics) at a ratio of 1:100 and grown at 37ºC, 200 rpm, until OD_{600nm} = 0.3. All cultures were induced by addition of 0.2% of arabinose and 0.1 mM of IPTG and grown for 30 min under consistent conditions.

During incubation, serial dilutions of the lambda phage were prepared by dilution of the initial stock on SM buffer (NaCl 100 mM, $MgSO_4$ x 7 H_2O 8 mM, Tris-HCl pH=7.5 50 mM, Gelatin 0.01% w/v) and stored at 4ºC. Dilutions were determined according to the phage titer of the stock in order to obtain a quantifiable amount of plaques on the final plates.

Cell cultures were pelleted by centrifugation at 4.000 g for 10 min at 4ºC, the supernatant was discarded and pellets were resuspended in 8 ml of 10 mM $MgSO_4$. 100 µl of phage dilutions were mixed with 100 µl of cell suspension in 1.5 ml reaction tubes for all of the tested strains/phage dilutions. As controls, 100 µl of 10^{-1} phage dilution was mixed with 100 µl of cell-free 10 mM $MgSO_4$, while 100 µl of the empty plasmid control strain was mixed with 100 µl of phage-free SM buffer. In order to promote phage absorption, samples were incubated for 20 min at 37ºC without shaking.

During this interval, 2YTL soft agar (2YTL media with 7.5 gr/l of agar), supplemented with inducers (0.2% arabinose and 0.1 mM IPTG), was aliquoted into 15 ml glass culture tubes that were pre-warmed at 49ºC. Each tube was filled with 3 ml of soft agar and stored at 49ºC in a water bath until samples were ready. After incubation, 200 µl of phage-cell mixes were added to 3 ml of soft agar, lightly vortexed and poured on top of 2YTL agar plates supplemented with antibiotics. After drying of the soft agar, plates were incubated upside down at 37ºC overnight.

The following day, growth was assessed on control plates and, when no cross-contamination was detected, plaque numbers were determined for plates with appropriate phage dilution. Efficiency of plaquing was determined as the ratio of plaques formed on the lawn of the strain of interest versus the number of plaques for the strain carrying empty plasmids.

In order to confirm the activity of the fluorescent Cascade, this assay was performed in triplicates starting from three different colonies of the initial strains.

For Type I-Fv Cascade, the carboxy-terminal Cas5fv fusion revealed activity that resembled wild type Cascade activity and this construct was chosen for subsequent single-particle tracking experiments.

sptPALM imaging of Cascade interference

Sample preparation for sptPALM imaging

Imaging of single fluorophores in sptPALM experiments demands for highly optimized sample preparations as the sensitivity of the method makes it prone to detect background signals. Background signals stemming from the imaged organism itself are common, e.g. due to colorful pigments (e.g. cyanobacteria possessing chlorophyll or bacteria producing red pigments like *Shewanella putrefaciens* or *Serratia marcescens*). Other background signals can be commonly traced back to particular growth conditions. For sptPALM, only high purity grade chemicals and defined colorless growth media such as EZ Rich Defined Medium (EZRDM, Teknova- M2105) should be used (**Figure V.6a**). Furthermore, changing the growth temperature might reduce the background as dying cells become highly autofluorescent (**Figure V.6b**).

Impurities on coverslips and microscope slides are removed before using them. Commonly used cleaning protocols include cleaning with KOH (optionally with sonication) (Stehbens, Pemble, Murrow, & Wittmann, 2012; Turkowyd et al., 2017; Virant, Turkowyd, Balinovic, & Endesfelder, 2017b), sonication in different organic solvents or detergents (Preciado Lopez et al., 2014; Stehbens et al., 2012), plasma cleaning (Lelek, Di Nunzio, & Zimmer, 2014) and cleaning with piranha solution (Simms, Bowman, & Anseth, 2008) or base pstraciranha solution (Preciado Lopez et al., 2014). For this study, overnight cleaning in 1 M KOH (Sigma- 221473) was selected. It is strongly recommended to use high precision coverslips of well-defined thickness of 170 µm (matching the objective working distance), critical for undistorted PSF detection. Cleaned microscope slides and coverslips (Carl Roth H884.1 and LH25.1, respectively) were prepared in a ratio of 2:1 and were stored in ddH_2O (**Figure V.7**).

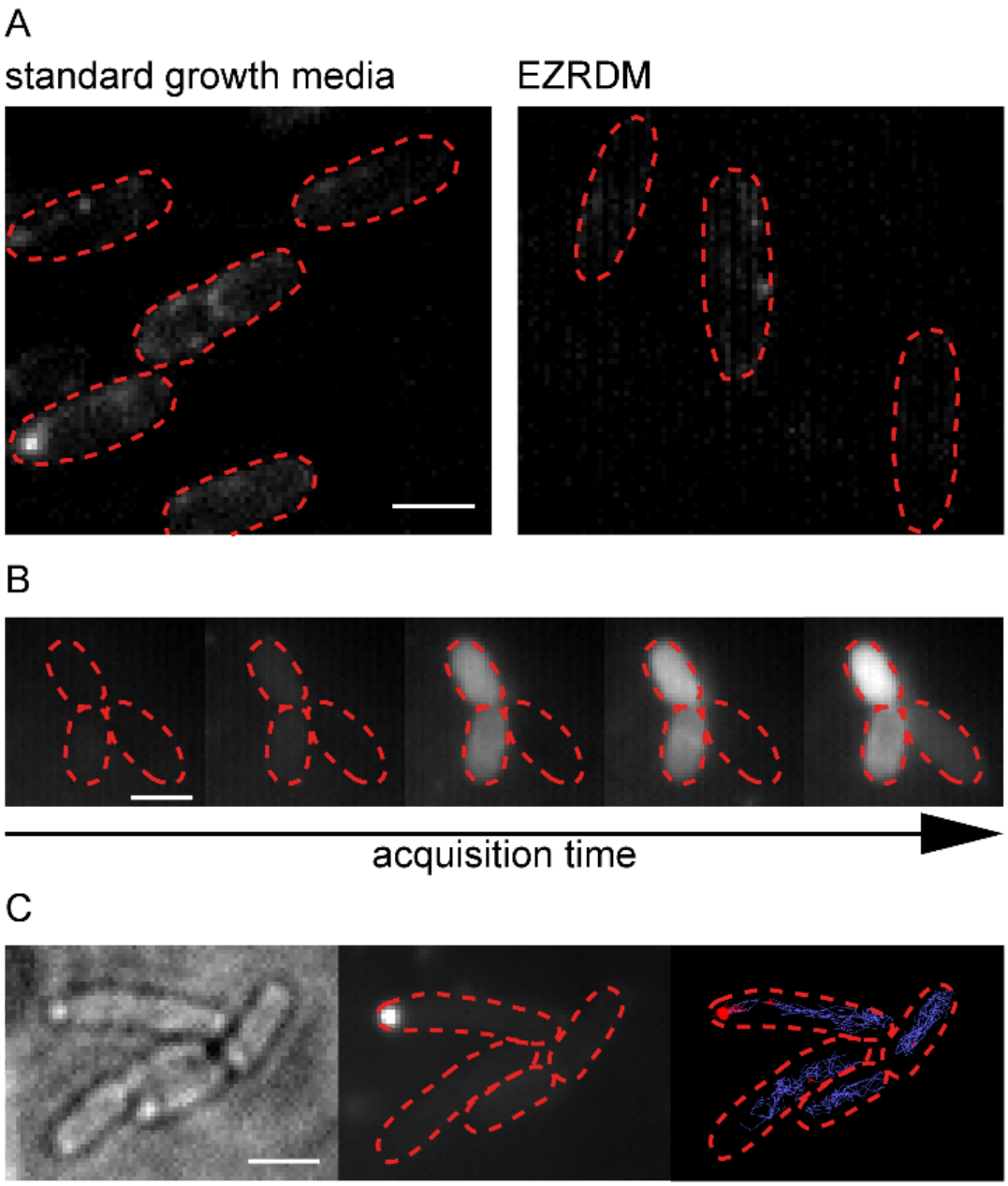


Figure V.6. Background noise and artifacts of different causes. (A) Using colorful growth media might cause cellular background noise (left), defined colorless media reduces unwanted signal (right). (B) High laser intensity might lead to cell death during imaging which is manifested by increasing background intensity over time (left to right, each image is a fluorescence average of each 1000 imaging frames); the cell in the lower right corner reacts later than the other two cells. (C) Overexpression of Cas proteins can cause aggregation of these proteins in inclusion bodies at cell pole, which is clearly visible in bright light (left) as well as in the fluorescence readout channel (middle) and finally in an unfiltered trajectory map (right). Scale bars: 2µm.

Cellular stress leads to the production of metabolites that are often autofluorescent. In order to avoid stress while imaging, the agarose solution for cell mounting was prepared by dilution on growth medium. Here, 1% agarose pads were prepared with EZRDM. Pads can be prepared in two ways: i) 2% agarose solution in ultrapure water is heated up to melt the agarose and afterwards mixed with a room temperature 2X media stock (Uphoff, Reyes-Lamothe, Garza de Leon, Sherratt, & Kapanidis, 2013; Zawadzki et al., 2015). This procedure avoids heating the media, avoiding thermal degradation; ii) 1% of low gelling and fluorescence-free agarose is suspended in media and heated up to 70°C until agarose melts (Turkowyd et al., 2017; Virant, Turkowyd, Balinovic, & Endesfelder, 2017a). The second approach has the advantage of a homogenous consistence (mixing hot agarose with room temperature media can cause agarose clumps) and the solution can be stored at 37°C after melting without the risk of gelling, which is convenient when preparing more agarose pads. In this study, the second approach with low-gelling agarose (Sigma- A9414) was used. Cooled down agarose was loaded onto a cleaned microscope slide with mold and covered with a clean coverslip. Agarose pads should be incubated for 90 minutes before use, in order to allow the agarose to settle. They can be used up to 5 hours after preparation.

Controlled culture cultivation and induction conditions are essential for reproducible sptPALM experiments (**Figure V.7**). Here, initial cultures were prepared in liquid 2YTL media with spectinomycin and kanamycin (both 25 µg/mL) and grown overnight at 37°C under shaking at 210 rpm. On the next day, cells were reinoculated in 1:100 dilutions into fresh 2YTL without drugs and incubated at 37°C. At OD_{600nm} = 0.05, the expression of Cascade proteins and crRNA was induced by adding arabinose to a final concentration 0.1% (w/v) to prevent Cascade proteins accumulation in inclusion bodies at the cell poles (**Figure V.6c**). After incubation for one hour, cultures were pelleted, 2YTL supernatant discarded and replaced by EZRDM with 2% glucose. This sugar inhibits the Ara promoter, blocking induction and hindering protein overexpression. Subsequently, cells were grown for another hour to allow the fluorescent proteins to mature as well as the Cascade complexes to assemble. This is the last step where it is possible to transfer the cells to a defined media suitable for sptPALM imaging (see **Figure V.6a**). 500 µL of culture was centrifuged (3000 x g, 2 min), pellet was washed twice with fresh EZRDM and suspended in 50 µL of EZRDM. After removing the coverslip from an earlier prepared agarose pad, 2 µL of cell suspension was loaded onto the pad, sealed with a new coverslip and gently pressed to remove any excess of liquid. Slides were incubated for 15 minutes on the bench in darkness to allow cells to settle down on the pad.

Afterwards, the sample was gently pressed with a wipe to fully dry the coverslip. Avoid applying too much pressure on the slide, as the coverslip could be invaginated towards the pad, which will cause tension in the agarose and subsequently lead to sample drift while imaging.

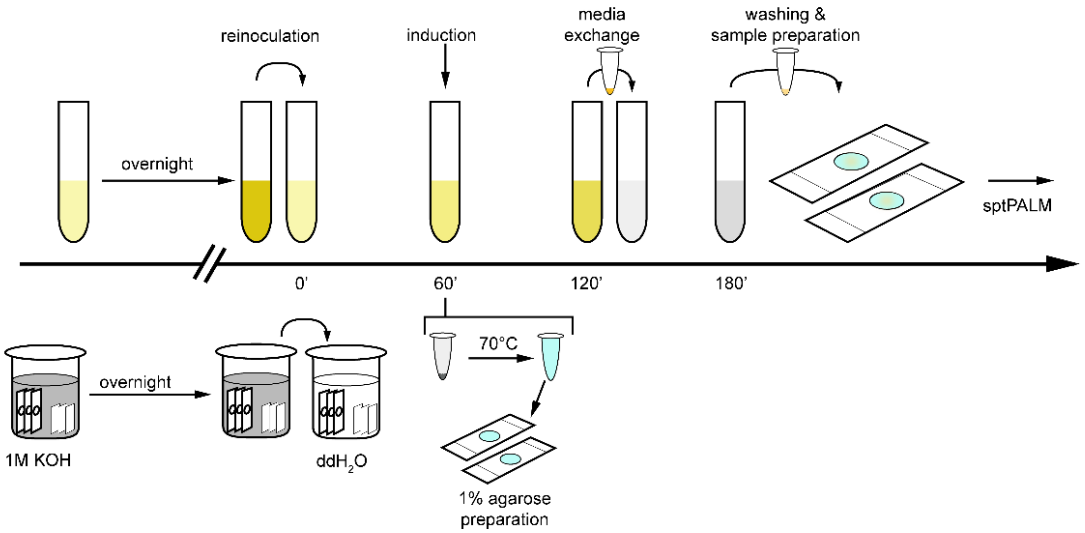


Figure V.7. Timeline of sptPALM sample preparation. Top: culture incubation, induction and cell washing steps. **Bottom**: cleaning microscope slides and coverslips, and agarose pads preparation.

Microscope settings and sptPALM imaging routines

For this study, a SMLM setup consisting of a Nikon Ti Eclipse inverted microscope equipped with a custom-build excitation path and a custom-build heating chamber and heating objective ring was used. 50 mW 405 nm and 150 mW 561 nm lasers (both OBIS Free-space Beam, Coherent Inc.) were controlled by OBIS LX/LS Scientific Remote and were coupled into the microscope optical excitation pathway by using Thorlabs- BB1-E02 set of mirrors and AHF Analysentechnik- F38-M01 and F38-M05 dichroic mirrors. Illumination was controlled (wavelength and intensities) via an acousto-optical tunable filter (AOTF TF525-250-6-3-GH18A, Gooch & Housego) in combination with the ESio AOTF controller (ESio). Beams passing the AOTF were adjusted in spot size using a 10x telescope of two lenses (25 mm focal length and 250 mm focal length). After passing a piezo-sensor controlled TIRF mirror (piezo KC1-PZ/M with MDT693B controller, Thorlabs), the laser beam was focused on the backfocal plane of the objective by a focusing lens (50.8 mm diameter, 300 mm focal length) being reflected by dichroic mirror (AHF Analysentechnik- F68-010) into the immersion objective (CFI Apo TIRF 100x oil objective, NA 1.49, Nikon, immersion oil Type F, Nikon). For detection, fluorescence signals passed through the above mentioned dichroic mirror, a rejection filter (AHF

Analysentechnik- F67-408) and bandpass filter (AHF Analysentechnik- F49-617), and were directed to the EM-CCD camera (iXon Ultra 888, Andor). Most of the setup, except heating chamber and ring and laser remote controller, was controlled via a customized version of Micro-Manager (Edelstein et al., 2014), an opensource plugin for Fiji (ImageJ) (Schindelin et al., 2012).

The fluorescence signal readout quality can be improved by enhancing the signal-to-noise ratio using different illumination modes. For normal epifluorescence illuminations, the laser light passes throughout the entire sample and could induce background signal. To avoid this, one option is the optical sectioning by total internal reflection fluorescence microscopy (TIRF), where laser beam is totally internally reflected at the coverslip-sample boundary (Axelrod, 2003). Only fluorophores in a thin layer above the coverslip (approx. 200 nm) are illuminated and thus fluoresce. Another approach is to use a highly illuminated and laminated optical sheet (HiLo), where laser light passes through the coverslip at a narrow angle (Tokunaga, Imamoto, & Sakata-Sogawa, 2008). For our studies HiLo mode was selected, as Cascade complexes can diffuse throughout the whole cellular thickness of about 1 µm.

As already indicated by *in vitro* single-molecule studies on CRISPR-Cas systems, Cascade-DNA interactions can span from seconds (DNA probing for PAM recognition, R-loop formation) to minutes (locked Cascade state recruiting Cas3) (Blosser et al., 2015). Thereby, to investigate all these interactions, multiple sptPALM imaging experiments spanning different temporal ranges were designed. Fast sptPALM imaging with frame rates of 30 – 100 Hz enables to detect Cascade complexes in different diffusive and binding states. From all trajectories, a spatial diffusion map of each individual cell can be reconstructed, which visualizes the different dynamics within the cellular volume (e.g. shown in **Figure V.3**). sptPALM imaging at slow speed and long camera integration time enables to visualize only immobile and slowly diffusing Cascades as fast diffusive signals blur out. The total fluorescence budget of the fluorophore can be flexibly adjusted to resolve the protein dynamics at different temporal scales: By increasing the imaging interval times introducing a variable gap time without prolonging the camera integration and laser excitation times, it is possible to track individual molecules for a longer time range. Thus, second-long bound times of Cascade complexes on full DNA-targets can also be determined. With this imaging mode, Cascades bound to genomic targets are easily traced until unbinding due to the nearly immobile nature of the nucleoid.

A. Measuring transient Cascade-target interactions at a high spatiotemporal resolution

1. The sample was placed on the heating stage and incubated for 15 minutes to equilibrate the temperature, since diffusion is dependent on this parameter. Before imaging, the stage was switched on for two hours to equilibrate at 25°C.

2. Camera integration times were set to 13 and 30 ms on independent imaging experiments. Recording two temporal regimes is recommended, as this can reveal different diffusive states and different natures of several interactions by changes in the observed dynamics (e.g. the displacement between frames) of stable and transient states with different kinetics.

3. Intensities for 405 nm and 561 nm lasers were set to 5 W/cm² and 800 W/cm², respectively for 13 ms integration time and to 5 W/cm² and 450 W/cm², respectively for 30 ms integration time. For fluorescence readout (561 nm laser), continuous illumination was used. Faster acquisitions require higher readout laser intensities to obtain sufficient signal-to-noise ratios. Excessive intensity, however, should be avoided, as it causes faster bleaching and pronounced blinking of fluorescent proteins, which leads to shorter trajectories (typically, the fluorescence budget of a fluorescent protein can be spread onto on average 10-20 imaging frames). Additionally, this setup might be phototoxic to cells. Photoconversion laser (405 nm) should be used at the lowest suitable intensity and in a pulsed mode (e.g. every 10^{th} frame) to keep the localizations per cell sparse. Importantly, different species in different environmental conditions have different sensitivity to intensive light. Thus, maximal imaging time ranges should be tested beforehand on wild type strains (Turkowyd et al., 2017).

4. A series of controls were prepared and imaged under sptPALM conditions for 10.000 – 20.000 frames:

    i. Negative control without a fluorescent marker. It is mandatory to evaluate if any fluorescent signal from the cells themselves or within the agarose pad might mimic the fluorescence signal of the marker.

    ii. Cytosolic, free diffusing fluorescent marker. This is recommended for two reasons: (i) to check if the chosen fluorescent marker performs well in the cellular environment of the specific organism, as fluorescence signal depends on pH, redox potentials and oxygen and (ii) to evaluate how fast

the fluorescent protein, when not interacting with any cellular compound, diffuses freely. The latter is especially important, as any diffusion distribution of the tagged proteins of interest can be compared to the free diffusing fluorophore.

iii. Immobile fluorescent marker (e.g. fixed cells). Despite the fact that the fluorophore is immobile under this condition, the obtained single-molecule trajectories will show a small residual oscillation due to the finite localization precision of the technique. Like the freely diffusive control, this serves as a suitable reference for nucleoid- or membrane-bound states, and thus immobile distributions of proteins of interest (lower limit of diffusion).

5. Regions of interest (ROIs) were imaged first for 20 frames in bright light mode, to obtain images of the cell outlines needed for cell segmentation (see **Section 4.3 1.**) and afterwards for 10.000-20.000 frames under sptPALM conditions. Multiple ROIs were obtained per sample to maximize the number of imaged cells. Frame number was selected in order to reduce phototoxicity effects, as it is better to collect shorter movies from multiple regions rather than long movies from fewer fields of view.

B. Measuring bound times of DNA-Cascade interference

1. The sample was placed on the heating stage and incubated for 15 minutes to equilibrate the temperature.
2. Interval time was adjusted in reference to the order of magnitude (milliseconds, seconds etc.) expected for the investigated bound state. Camera integration time was set long enough to blur any mobile molecule, but also short enough to keep the readout time as minimal as possible to prevent early fluorescent protein bleaching. For measuring Cascade interference on full targets, the integration time was set to 200 ms with a 500 ms interval.
3. Laser readout intensity was adjusted to the lowest possible intensity which still provided a clear single-molecule fluorescence signal. As before, this prevents fluorophore bleaching and reduces blinking events. 30 W/cm² for the 561 nm laser was used. Intensity of the UV-photoconversion laser was set to a minimum (1 W/cm²) and in pulsed mode (every 50 frames).
4. The following controls were prepared and imaged:

i. Negative control without fluorescent marker to ensure that cells themselves are not generating any background signal.
ii. Cytosolic, free diffusing fluorescent marker. Useful to adjust imaging parameters, as camera integration times should be set at a slow speed which only shows homogenously fluorescent signal building up from freely diffusive molecules within the entire cell but should not reveal any defined fluorescent spots.
iii. Immobile fluorescent marker. Since it is not possible to distinguish whether the disappearance of the fluorescence signal is due to fluorophore bleaching or to an unbinding event, the bleaching probability of the fluorophore under the read-out conditions used has to be measured and used for bleaching correction. Furthermore, as fluorophores still might experience blinking events, these can be characterized on immobilized fluorescent proteins and be used as a parameter for tracking analysis.

5. Cells of interest were imaged for 4.000 frames and several ROIs were taken per sample.

Localization and tracking routines

Data analysis to obtain single-molecule trajectories consists of the following steps (**Figure V.8**):

1. **Cell segmentation**. This step is recommended for two reasons: i) segmented cells provide some useful information about the sample (number of imaged cells, cell dimensions) and allow for sorting cells by their size/cell-cycle stage, number of localizations/trajectories per cell, cell heatmaps, etc.; ii) cell segmentation prevents tracking artifacts by wrongly joining localizations from different, but spatially close cells into a single trajectory. Further, tracking is a computationally demanding process, and segmentation can reduce calculation times. Cell segmentation can be performed with multiple tools. The simplest, most accurate, but also time consuming option is a manual segmentation: drawing regions of interest from brightlight images by hand in ImageJ (Fiji) (Schindelin et al., 2012). Segmentation can also be automatized by using different software like 3D Object Counter (ImageJ) (Bolte & Cordelieres, 2006), MorphoLibJ (Legland, Arganda-Carreras, & Andrey, 2016) or Oufti (Paintdakhi et al., 2016). Recently, machine learning and neural networks have been applied for image segmentation, with the use of softwares like DeepCell (Van Valen et al., 2016) or CDeep3M (Haberl

et al., 2018) However, it should be pointed out that automatized segmentation might need manual correction. For this protocol, manual selection in Fiji was chosen as the most appropriate option.

2. **Localization**. SMLM is a localization-based method, which means that all data analyses are performed on localization coordinates having nanometer precision instead of being image-based. This enables a wide range of analysis approaches like super-resolved image reconstruction, molecule-counting, clustering analysis, co-localization analysis, particle averaging and particle tracking (Endesfelder et al., 2013; Lando et al., 2012; Loschberger, Franke, Krohne, van de Linde, & Sauer, 2014; Malkusch et al., 2012; Manley et al., 2008). Localization coordinates can be extracted from the imaging sequences by many different localization algorithms, which are discussed and compared in the following publications: (Sage et al., 2015; Small & Stahlheber, 2014). In this work we use two localization software packages: rapidStorm (Wolter et al., 2012) and ThunderStorm (Ovesny, Krizek, Borkovec, Svindrych, & Hagen, 2014). In this step, it is important to select appropriate fitting and filtering parameters and thresholds, to sieve most of the noise signal, e.g. by an intensity threshold, as noise usually possesses lower intensities than fluorescence signals from fluorophores. On the other hand, too strong filtering might cause a huge loss of fluorophore detections, which might severely impact obtaining proper single-molecule trajectories. Furthermore, especially for slow frame rate data, where only immobile complexes are investigated, PSFs shape and dimensions are helpful filters. Signals from immobile particles have round PSFs of a typical diffracted spot size. Any asymmetric, elliptic or differently sized PSF can be discarded as it can be attributed either to noise, multiple proximal fluorophores activated at the same time, or moving Cascade complexes.

3. **Tracking**. From the list of localizations, single-molecule trajectories of individual Cas proteins or Cascade complexes can be retrieved. The simplest and fastest method for single-particle tracking is the nearest neighbor (NN) algorithm, in which connections are made between localizations in adjacent frames and the distance between them is the lowest from all possible connections. This solution is sufficient when the density of localizations is low (around 1 event per frame per cell). For higher densities NN might wrongly connect different particles, thus bias the diffusion statistics when mixing different molecular states. This obstacle might be omitted by a simple filtering step, in which only localizations that have a

sufficiently low density and thus only one possible neighbor in reach, are connected. However, in this way some readout is lost, in effect decreasing the statistics. More sophisticated approaches, e.g. based on *a priori* knowledge about fluorophore photophysics and protein behavior (transition probability from immobile to mobile, etc.) can be a solution for dense data and is for this reason implemented into our routine. There are also other tracking algorithms which are used in sptPALM (Chenouard et al., 2014), however, a majority of these algorithms were developed for classical single-particle tracking of bright quantum dots for long observation times or for cell lineage tracking, which differs from sptPALM with fluorescent proteins as markers.

4. **Data quality assessment**. Despite optimal conditions, initial data can still contain some false trajectories. Thus, it is recommended to visually inspect the retrieved data. For obtaining diffusion statistics of high quality, a selection of trajectories of sufficient length can be advisable. Short trajectories might have multiple sources: detection loss due to molecules leaving the focal plane, fluorescent protein blinking or residual noise. Filtering against very long trajectories can also improve data quality, as unusually long trajectories are often caused by impurities in the sample or by aggregates in which multiple fluorescent proteins are brought to their fluorescent state one after another. Finally, trajectories which appear as obvious artifacts, like inclusion bodies caused by protein accumulation (see **Figure V.6c**) and trajectories “sliding” on cell membrane (impurities from the buffer transiently binding to the cell) or protruding the cellular area must be excluded. When inspecting trajectories from slow imaging schemes exploring the complex bound times, most noise can be easily identified by residual, mobile trajectories.

Data analysis and visualization

When recording sptPALM movies at fast read-out rates, the high spatiotemporal data reveals different and highly transient sub-populations of Cascade interactions based on their apparent mobility. In contrast, for slow read-out rates, the fluorescence lifetime of the fluorescent marker is spread over longer imaging times by introducing gap times and thus lowering the spatiotemporal resolution. By doing so, bound times of several seconds for the investigated interactions can be observed but can only be registered for immobile complexes, e.g. bound to the nucleoid.

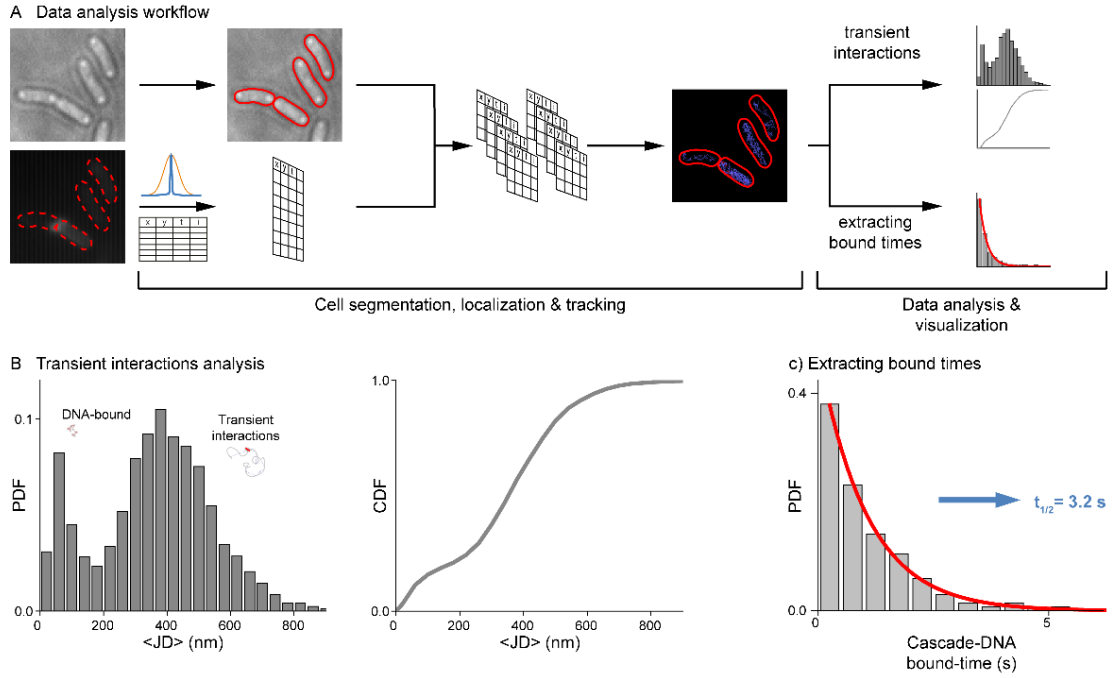


Figure V.8. Data processing workflow. (A) Cells were segmented based on brightlight snapshots taken before sptPALM imaging (top left). In parallel, localizations were extracted based on PSF fitting (bottom left). Next, localizations were split to subarrays representing single cells and tracking was performed (middle). In case of fast imaging data were represented as histograms or cumulative distributions of averaged jump distances (top right). For slow imaging accessing the bound-time of complexes, data histograms representing fluorescence spot bound times were fitted to retrieve the average event lifetime (bottom right). (B) Data from transient interaction analysis represented in form of probability density function (PDF) histogram (bin size = 40 nm) and cumulative distribution function (CDF). Exemplary trajectories for DNA-bound Cascade (red trajectory) and transiently interacting Cascade (blue-red trajectory). (C) PDF representing Cascade bound-time (bin size = 0.5 s) and single exponential model fitted to PDF.

A. Obtaining molecule mobilities

There are different metrics to represent molecule motilities. One method is extracting the diffusion coefficient (*D*) by plotting mean-square displacements (*MSD*) for given time lags *Δt* (Oswald, E, Bollen, & Peterman, 2014; Weimann et al., 2013):

$$MSD(n\Delta t) = 4D * n\Delta t + 4\sigma^2$$

with σ being the localization precision. The diffusion coefficient is calculated individually for each trajectory (usually by using the first four MSD time steps, as otherwise confinement effects become too large) and a distribution of diffusions can be later visualized in histograms as a probability density function (PDF) or as cumulative distribution functions (CDF). The MSD method is a common approach to analyze sptPALM data (Sanamrad et al., 2014; Stracy et al., 2015; Uphoff et al., 2013).

Another way to represent the dynamics of the protein is by the average displacement between adjacent time points within each trajectory (also called jump distances (JD)) (Grunwald et al., 2008; Oswald et al., 2014; Schutz, Schindler, & Schmidt, 1997; Weimann et al., 2013):

$$JD = \sqrt{(r(t + \Delta t) - r(t)}$$

with *r* being the position of the particle at times *t* and *t*+*Δt*. Once the JDs or an average JD for every trajectory (or segment of a trajectory if a change of diffusive state divides the trajectory in several segments) are calculated, they can be as well represented as a PDF or CDF in which different mobile fractions can be distinguished (**Figure V.8b**).

In this study, the second approach was selected. As sptPALM trajectories are generally rather short, MSD values for larger time lag values have poor statistics (Flier et al., 2011; Saxton, 1997). Furthermore, due to the small volume of the bacterial cell, the strong confinement of diffusion affects the diffusion values obtained by MSD analyses (Oswald et al., 2014; Qian, Sheetz, & Elson, 1991). JD values behave more robust for these issues (Chen et al., 2015; Mazza, Abernathy, Golob, Morisaki, & McNally, 2012). Nevertheless, due to confinement and two-dimensional projections of the diffusion (if not imaged in 3D using a microscope with isotropic resolution in all three dimensions), only apparent diffusion coefficients can be retrieved.

Finally, trajectories can be visualized in form of a tracking map per cell, where trajectories are color coded for their dynamics. In this study, fast dynamical sptPALM data was visualized as a function of JD where different fractions of Cascades can be identified (**Figure V.8b**). Peaks at lower JD values resemble complexes which were stably bound to the DNA target, while the broad fast diffusing fraction is attributed to transiently interacting and free diffusing Cascades.

B. Extracting molecular bound times

The dynamics of the target bound Cascade can be extracted by the length of each immobile trajectory, and data can be plotted as a histogram. Correctly filtered data should show a histogram of single exponential decay (assuming only one possible interaction).

This histogram can be described with the following equation:

$$k_{lifetime}e^{-k_{lifetime}t} = (k_{unbinding} + k_{bleaching})e^{-(k_{unbinding}+k_{bleaching})t}$$

with $k_{lifetime}$ being the rate of the average lifetime of the fluorescence signals, $k_{unbinding}$ is the rate of Cascade unbinding events and $k_{bleaching}$ the rate of fluorophore bleaching (Beattie et al., 2017). To extract the unbinding rate, the bleaching rate has to be evaluated. Here, an immobile fluorescent marker control has to be used, e.g. fusing the marker to a protein forming highly stable complexes (significantly longer than imaging times) or by chemically fixing the sample. For fixed samples, as no unbinding events occur, the lifetime rate is equivalent to the bleaching rate. Thus, $k_{unbinding}$ can be retrieved (**Figure V.8c**). Experiments like this are time consuming and thus may suffer from low statistics. Measurement errors can be determined by subsampling (bootstrapping) and performing multiple fits on these subsets (Beattie et al., 2017).

Conclusions

The ability to study Cascade dynamics *in vivo* at high spatio-temporal resolution allows for analyzing target requirements for Cascade binding in parallel with other DNA-interacting proteins, e.g. the replisome, in their natural environment. The observed parameters can be compared with *in vitro* studies to obtain a more complete picture of Cascade-mediated target search processes. Quantitative sptPALM analyses of CRISPR-Cas/DNA interactions could also help to unveil levels of regulation imposed on interference, but also adaptation or crRNP complex assembly stages.

Acknowledgements

We would like to thank Marco Hui for benchmarking some sptPALM parameters and for establishing parts of the data analysis workflow. This work was supported by the Max Planck Society (LR, UE), by SYNMIKRO (UE), SPP 2441 from the Deutsche Forschungsgemeinschaft (LR, UE), by the Fonds der Chemischen Industrie (UE) and by IMPRS-Mic (HME).

References

Amitai, G., & Sorek, R. (2016). CRISPR-Cas adaptation: insights into the mechanism of action. *Nat Rev Microbiol, 14*(2), 67-76.

Axelrod, D. (2003). Total internal reflection fluorescence microscopy in cell biology. *Methods Enzymol, 361*, 1-33.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., et al. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science, 315*(5819), 1709-1712.

Beattie, T. R., Kapadia, N., Nicolas, E., Uphoff, S., Wollman, A. J., Leake, M. C., et al. (2017). Frequent exchange of the DNA polymerase during bacterial chromosome replication. *Elife, 6*.

Berardozzi, R., Adam, V., Martins, A., & Bourgeois, D. (2016). Arginine 66 Controls Dark-State Formation in Green-to-Red Photoconvertible Fluorescent Proteins. *J Am Chem Soc, 138*(2), 558-565.

Betzig, E., Patterson, G. H., Sougrat, R., Lindwasser, O. W., Olenych, S., Bonifacino, J. S., et al. (2006). Imaging intracellular fluorescent proteins at nanometer resolution. *Science, 313*(5793), 1642-1645.

Biswas, A., Staals, R. H., Morales, S. E., Fineran, P. C., & Brown, C. M. (2016). CRISPRDetect: A flexible algorithm to define CRISPR arrays. *BMC Genomics, 17*, 356.

Blosser, T. R., Loeff, L., Westra, E. R., Vlot, M., Kunne, T., Sobota, M., et al. (2015). Two distinct DNA binding modes guide dual roles of a CRISPR-Cas protein complex. *Mol Cell, 58*(1), 60-70.

Bolte, S., & Cordelieres, F. P. (2006). A guided tour into subcellular colocalization analysis in light microscopy. *J Microsc, 224*(Pt 3), 213-232.

Brouns, S. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J., Snijders, A. P., et al. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science, 321*(5891), 960-964.

Carte, J., Wang, R., Li, H., Terns, R. M., & Terns, M. P. (2008). Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev, 22*(24), 3489-3496.

Chen, T. Y., Santiago, A. G., Jung, W., Krzeminski, L., Yang, F., Martell, D. J., et al. (2015). Concentration- and chromosome-organization-dependent regulator unbinding from DNA for transcription regulation in living cells. *Nat Commun, 6*, 7445.

Chenouard, N., Smal, I., de Chaumont, F., Maska, M., Sbalzarini, I. F., Gong, Y., et al. (2014). Objective comparison of particle tracking methods. *Nat Methods, 11*(3), 281-289.

Chowdhury, S., Carter, J., Rollins, M. F., Golden, S. M., Jackson, R. N., Hoffmann, C., et al. (2017). Structure Reveals Mechanisms of Viral Suppressors that Intercept a CRISPR RNA-Guided Surveillance Complex. *Cell, 169*(1), 47-57 e11.

Chylinski, K., Makarova, K. S., Charpentier, E., & Koonin, E. V. (2014). Classification and evolution of type II CRISPR-Cas systems. *Nucleic Acids Res, 42*(10), 6091-6105.

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., et al. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science, 339*(6121), 819-823.

Couvin, D., Bernheim, A., Toffano-Nioche, C., Touchon, M., Michalik, J., Neron, B., et al. (2018). CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res, 46*(1), 246-251.

Dwarakanath, S., Brenzinger, S., Gleditzsch, D., Plagens, A., Klingl, A., Thormann, K., et al. (2015). Interference activity of a minimal Type I CRISPR-Cas system from Shewanella putrefaciens. *Nucleic Acids Res, 43*(18), 8913-8923.

Edelstein, A. D., Tsuchida, M. A., Amodaj, N., Pinkard, H., Vale, R. D., & Stuurman, N. (2014). Advanced methods of microscope control using muManager software. *J Biol Methods, 1*(2).

Edgar, R., & Qimron, U. (2010). The Escherichia coli CRISPR system protects from lambda lysogenization, lysogens, and prophage induction. *J Bacteriol, 192*(23), 6291-6294.

Endesfelder, U., Finan, K., Holden, S. J., Cook, P. R., Kapanidis, A. N., & Heilemann, M. (2013). Multiscale spatial organization of RNA polymerase in Escherichia coli. *Biophys J, 105*(1), 172-181.

Flier, B. M., Baier, M., Huber, J., Mullen, K., Mecking, S., Zumbusch, A., et al. (2011). Single molecule fluorescence microscopy investigations on heterogeneity of translational diffusion in thin polymer films. *Phys Chem Chem Phys, 13*(5), 1770-1775.

Folling, J., Bossi, M., Bock, H., Medda, R., Wurm, C. A., Hein, B., et al. (2008). Fluorescence nanoscopy by ground-state depletion and single-molecule return. *Nat Methods, 5*(11), 943-945.

Garneau, J. E., Dupuis, M. E., Villion, M., Romero, D. A., Barrangou, R., Boyaval, P., et al. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. [Article]. *Nature, 468*, 67.

Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., 3rd, & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods, 6*(5), 343-345.

Gleditzsch, D., Muller-Esparza, H., Pausch, P., Sharma, K., Dwarakanath, S., Urlaub, H., et al. (2016). Modulating the Cascade architecture of a minimal Type I-F CRISPR-Cas system. *Nucleic Acids Res, 44*(12), 5872-5882.

Grissa, I., Vergnaud, G., & Pourcel, C. (2007a). The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics, 8*, 172.

Grissa, I., Vergnaud, G., & Pourcel, C. (2007b). CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res, 35*(Web Server issue), W52-57.

Grunwald, D., Martin, R. M., Buschmann, V., Bazett-Jones, D. P., Leonhardt, H., Kubitscheck, U., et al. (2008). Probing intranuclear environments at the single-molecule level. *Biophys J, 94*(7), 2847-2858.

Haberl, M. G., Churas, C., Tindall, L., Boassa, D., Phan, S., Bushong, E. A., et al. (2018). CDeep3M-Plug-and-Play cloud-based deep learning for image segmentation. *Nat Methods, 15*(9), 677-680.

Haurwitz, R. E., Jinek, M., Wiedenheft, B., Zhou, K., & Doudna, J. A. (2010). Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science, 329*(5997), 1355-1358.
Hayes, R. P., Xiao, Y., Ding, F., van Erp, P. B., Rajashankar, K., Bailey, S., et al. (2016). Structural basis for promiscuous PAM recognition in type I-E Cascade from E. coli. *Nature, 530*(7591), 499-503.
Heilemann, M., van de Linde, S., Schuttpelz, M., Kasper, R., Seefeldt, B., Mukherjee, A., et al. (2008). Subdiffraction-resolution fluorescence imaging with conventional fluorescent probes. *Angew Chem Int Ed Engl, 47*(33), 6172-6176.
Hess, S. T., Girirajan, T. P., & Mason, M. D. (2006). Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophys J, 91*(11), 4258-4272.
Hochstrasser, M. L., Taylor, D. W., Bhat, P., Guegler, C. K., Sternberg, S. H., Nogales, E., et al. (2014). CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. *Proc Natl Acad Sci U S A, 111*(18), 6618-6623.
Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., & Nakata, A. (1987). Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *J Bacteriol, 169*(12), 5429-5433.
Jackson, R. N., Golden, S. M., van Erp, P. B. G., Carter, J., Westra, E. R., Brouns, S. J. J., et al. (2014). Crystal structure of the CRISPR RNA-guided surveillance complex from Escherichia coli. *Science, 345*(6203), 1473-1479.
Jansen, R., van Embden, J. D. A., & Gaastra, W. S., L.M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular Microbiology, 43*(6), 1565-1575.
Jones, D. L., Leroy, P., Unoson, C., Fange, D., Curic, V., Lawson, M. J., et al. (2017). Kinetics of dCas9 target search in Escherichia coli. *Science, 357*(6358), 1420-1424.
Jore, M. M., Lundgren, M., van Duijn, E., Bultema, J. B., Westra, E. R., Waghmare, S. P., et al. (2011). Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat Struct Mol Biol, 18*(5), 529-536.
Knight, S. C., Xie, L., Deng, W., Guglielmi, B., Witkowsky, L. B., Bosanac, L., et al. (2015). Dynamics of CRISPR-Cas9 genome interrogation in living cells. *Science, 350*(6262), 823-826.
Komor, A. C., Badran, A. H., & Liu, D. R. (2017). CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. *Cell, 168*(1-2), 20-36.
Koonin, E. V., Makarova, K. S., & Zhang, F. (2017). Diversity, classification and evolution of CRISPR-Cas systems. *Curr Opin Microbiol, 37*, 67-78.
Lando, D., Endesfelder, U., Berger, H., Subramanian, L., Dunne, P. D., McColl, J., et al. (2012). Quantitative single-molecule microscopy reveals that CENP-A(Cnp1) deposition occurs during G2 in fission yeast. *Open Biol, 2*(7), 120078.
Legland, D., Arganda-Carreras, I., & Andrey, P. (2016). MorphoLibJ: integrated library and plugins for mathematical morphology with ImageJ. *Bioinformatics, 32*(22), 3532-3534.
Lelek, M., Di Nunzio, F., & Zimmer, C. (2014). FlAsH-PALM: super-resolution pointillist imaging with FlAsH-tetracysteine labeling. *Methods Mol Biol, 1174*, 183-193.
Loschberger, A., Franke, C., Krohne, G., van de Linde, S., & Sauer, M. (2014). Correlative super-resolution fluorescence and electron microscopy of the nuclear pore complex with molecular resolution. *J Cell Sci, 127*(Pt 20), 4351-4355.
Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J., Charpentier, E., Horvath, P., et al. (2011). Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol, 9*(6), 467-477.
Makarova, K. S., Wolf, Y. I., Alkhnbashi, O. S., Costa, F., Shah, S. A., Saunders, S. J., et al. (2015). An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol, 13*(11), 722-736.
Malkusch, S., Endesfelder, U., Mondry, J., Gelleri, M., Verveer, P. J., & Heilemann, M. (2012). Coordinate-based colocalization analysis of single-molecule localization microscopy data. *Histochem Cell Biol, 137*(1), 1-10.
Manley, S., Gillette, J. M., Patterson, G. H., Shroff, H., Hess, H. F., Betzig, E., et al. (2008). High-density mapping of single-molecule trajectories with photoactivated localization microscopy. *Nat Methods, 5*(2), 155-157.
Marraffini, L. A., & Sontheimer, E. J. (2008). CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science, 322*(5909), 1843-1845.
Marraffini, L. A., & Sontheimer, E. J. (2010). Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature, 463*(7280), 568-571.
Mazza, D., Abernathy, A., Golob, N., Morisaki, T., & McNally, J. G. (2012). A benchmark for chromatin binding measurements in live cells. *Nucleic Acids Res, 40*(15), e119.
Mohanraju, P., Makarova, K. S., Zetsche, B., Zhang, F., Koonin, E. V., & van der Oost, J. (2016). Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science, 353*(6299), aad5147.
Mojica, F. J. M., Díez-Villaseñor, C. s., García-Martínez, J., & Soria, E. (2005). Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements. *Journal of Molecular Evolution, 60*(2), 174-182.
Mulepati, S., Heroux, A., & Bailey, S. (2014). Structural biology. Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target. *Science, 345*(6203), 1479-1484.
Nunez, J. K., Kranzusch, P. J., Noeske, J., Wright, A. V., Davies, C. W., & Doudna, J. A. (2014). Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. *Nat Struct Mol Biol, 21*(6), 528-534.
Oswald, F., E, L. M. B., Bollen, Y. J., & Peterman, E. J. (2014). Imaging and quantification of trans-membrane protein diffusion in living bacteria. *Phys Chem Chem Phys, 16*(25), 12625-12634.
Ovesny, M., Krizek, P., Borkovec, J., Svindrych, Z., & Hagen, G. M. (2014). ThunderSTORM: a comprehensive ImageJ plug-in for PALM and STORM data analysis and super-resolution imaging. *Bioinformatics, 30*(16), 2389-2390.
Paintdakhi, A., Parry, B., Campos, M., Irnov, I., Elf, J., Surovtsev, I., et al. (2016). Oufti: an integrated software package for high-accuracy, high-throughput quantitative microscopy analysis. *Mol Microbiol, 99*(4), 767-777.
Pausch, P., Muller-Esparza, H., Gleditzsch, D., Altegoer, F., Randau, L., & Bange, G. (2017). Structural Variation of Type I-F CRISPR RNA Guided DNA Surveillance. *Mol Cell, 67*(4), 622-632 e624.

Pingoud, A., Fuxreiter, M., Pingoud, V., & Wende, W. (2005). Type II restriction endonucleases: structure and mechanism. *Cell Mol Life Sci, 62*(6), 685-707.
Pourcel, C., Salvignol, G., & Vergnaud, G. (2005). CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology, 151*(Pt 3), 653-663.
Preciado Lopez, M., Huber, F., Grigoriev, I., Steinmetz, M. O., Akhmanova, A., Dogterom, M., et al. (2014). In vitro reconstitution of dynamic microtubules interacting with actin filament networks. *Methods Enzymol, 540*, 301-320.
Qian, H., Sheetz, M. P., & Elson, E. L. (1991). Single particle tracking. Analysis of diffusion and flow in two-dimensional systems. *Biophys J, 60*(4), 910-921.
Redding, S., Sternberg, S. H., Marshall, M., Gibb, B., Bhat, P., Guegler, C. K., et al. (2015). Surveillance and Processing of Foreign DNA by the Escherichia coli CRISPR-Cas System. *Cell, 163*(4), 854-865.
Rust, M. J., Bates, M., & Zhuang, X. (2006). Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat Methods, 3*(10), 793-795.
Rutkauskas, M., Sinkunas, T., Songailiene, I., Tikhomirova, M. S., Siksnys, V., & Seidel, R. (2015). Directional R-Loop Formation by the CRISPR-Cas Surveillance Complex Cascade Provides Efficient Off-Target Site Rejection. *Cell Rep, 10*(9), 1534-1543.
Sage, D., Kirshner, H., Pengo, T., Stuurman, N., Min, J., Manley, S., et al. (2015). Quantitative evaluation of software packages for single-molecule localization microscopy. *Nat Methods, 12*(8), 717-724.
Sanamrad, A., Persson, F., Lundius, E. G., Fange, D., Gynna, A. H., & Elf, J. (2014). Single-particle tracking reveals that free ribosomal subunits are not excluded from the Escherichia coli nucleoid. *Proc Natl Acad Sci U S A, 111*(31), 11413-11418.
Saxton, M. J. (1997). Single-particle tracking: the distribution of diffusion coefficients. *Biophys J, 72*(4), 1744-1753.
Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat Methods, 9*(7), 676-682.
Schutz, G. J., Schindler, H., & Schmidt, T. (1997). Single-molecule microscopy on model membranes reveals anomalous diffusion. *Biophys J, 73*(2), 1073-1080.
Simms, H. M., Bowman, C. M., & Anseth, K. S. (2008). Using living radical polymerization to enable facile incorporation of materials in microfluidic cell culture devices. *Biomaterials, 29*(14), 2228-2236.
Singh, D., Sternberg, S. H., Fei, J., Doudna, J. A., & Ha, T. (2016). Real-time observation of DNA recognition and rejection by the RNA-guided endonuclease Cas9. *Nat Commun, 7*, 12778.
Small, A., & Stahlheber, S. (2014). Fluorophore localization algorithms for super-resolution microscopy. *Nat Methods, 11*(3), 267-279.
Staals, R. H., Zhu, Y., Taylor, D. W., Kornfeld, J. E., Sharma, K., Barendregt, A., et al. (2014). RNA targeting by the type III-A CRISPR-Cas Csm complex of Thermus thermophilus. *Mol Cell, 56*(4), 518-530.
Stehbens, S., Pemble, H., Murrow, L., & Wittmann, T. (2012). Imaging intracellular protein dynamics by spinning disk confocal microscopy. *Methods Enzymol, 504*, 293-313.
Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C., & Doudna, J. A. (2014). DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature, 507*(7490), 62-67.
Stracy, M., Lesterlin, C., Garza de Leon, F., Uphoff, S., Zawadzki, P., & Kapanidis, A. N. (2015). Live-cell superresolution microscopy reveals the organization of RNA polymerase in the bacterial nucleoid. *Proc Natl Acad Sci U S A, 112*(32), E4390-4399.
Sun, C. L., Barrangou, R., Thomas, B. C., Horvath, P., Fremaux, C., & Banfield, J. F. (2013). Phage mutations in response to CRISPR diversification in a bacterial population. *Environ Microbiol, 15*(2), 463-470.
Szczelkun, M. D., Tikhomirova, M. S., Sinkunas, T., Gasiunas, G., Karvelis, T., Pschera, P., et al. (2014). Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proc Natl Acad Sci U S A, 111*(27), 9798-9803.
Tokunaga, M., Imamoto, N., & Sakata-Sogawa, K. (2008). Highly inclined thin illumination enables clear single-molecule imaging in cells. *Nat Methods, 5*(2), 159-161.
Turkowyd, B., Balinovic, A., Virant, D., Carnero, H. G. G., Caldana, F., Endesfelder, M., et al. (2017). A General Mechanism of Photoconversion of Green-to-Red Fluorescent Proteins Based on Blue and Infrared Light Reduces Phototoxicity in Live-Cell Single-Molecule Imaging. *Angew Chem Int Ed Engl, 56*(38), 11634-11639.
Unterholzner, S. J., Poppenberger, B., & Rozhon, W. (2013). Toxin-antitoxin systems: Biology, identification, and application. *Mob Genet Elements, 3*(5), e26219.
Uphoff, S., Reyes-Lamothe, R., Garza de Leon, F., Sherratt, D. J., & Kapanidis, A. N. (2013). Single-molecule DNA repair in live bacteria. *Proc Natl Acad Sci U S A, 110*(20), 8063-8068.
van der Oost, J., Westra, E. R., Jackson, R. N., & Wiedenheft, B. (2014). Unravelling the structural and mechanistic basis of CRISPR-Cas systems. *Nat Rev Microbiol, 12*(7), 479-492.
Van Valen, D. A., Kudo, T., Lane, K. M., Macklin, D. N., Quach, N. T., DeFelice, M. M., et al. (2016). Deep Learning Automates the Quantitative Analysis of Individual Cells in Live-Cell Imaging Experiments. *PLoS Comput Biol, 12*(11), e1005177.
Virant, D., Turkowyd, B., Balinovic, A., & Endesfelder, U. (2017b). Combining Primed Photoconversion and UV-Photoactivation for Aberration-Free, Live-Cell Compliant Multi-Color Single-Molecule Localization Microscopy Imaging. *Int J Mol Sci, 18*(7), E1524.
Wang, J., Li, J., Zhao, H., Sheng, G., Wang, M., Yin, M., et al. (2015). Structural and Mechanistic Basis of PAM-Dependent Spacer Acquisition in CRISPR-Cas Systems. *Cell, 163*(4), 840-853.

Wang, S., Moffitt, J. R., Dempsey, G. T., Xie, X. S., & Zhuang, X. (2014). Characterization and development of photoactivatable fluorescent proteins for single-molecule-based superresolution imaging. *Proc Natl Acad Sci U S A, 111*(23), 8452-8457.

Weimann, L., Ganzinger, K. A., McColl, J., Irvine, K. L., Davis, S. J., Gay, N. J., et al. (2013). A quantitative comparison of single-dye tracking analysis tools using Monte Carlo simulations. *PLoS One, 8*(5), e64287.

Westra, E. R., van Erp, P. B., Kunne, T., Wong, S. P., Staals, R. H., Seegers, C. L., et al. (2012). CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Mol Cell, 46*(5), 595-605.

Wiedenheft, B., Lander, G. C., Zhou, K., Jore, M. M., Brouns, S. J. J., van der Oost, J., et al. (2011). Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature, 477*(7365), 486-489.

Wolter, S., Loschberger, A., Holm, T., Aufmkolk, S., Dabauvalle, M. C., van de Linde, S., et al. (2012). rapidSTORM: accurate, fast open-source software for localization microscopy. *Nat Methods, 9*(11), 1040-1041.

Zawadzki, P., Stracy, M., Ginda, K., Zawadzka, K., Lesterlin, C., Kapanidis, A. N., et al. (2015). The Localization and Action of Topoisomerase IV in Escherichia coli Chromosome Segregation Is Coordinated by the SMC Complex, MukBEF. *Cell Rep, 13*(11), 2587-2596.

Zhao, H., Sheng, G., Wang, J., Wang, M., Bunkoczi, G., Gong, W., et al. (2014). Crystal structure of the RNA-guided immune surveillance Cascade complex in Escherichia coli. *Nature, 515*(7525), 147-150.

Zoephel, J., Dwarakanath, S., Richter, H., Plagens, A., & Randau, L. (2012). Substrate generation for endonucleases of CRISPR/cas systems. *J Vis Exp* (67).

Chapter VI:
Studying the *in vivo* protein organization and dynamics of a minimal CRISPR-Cas effector complex at high spatiotemporal resolution

Hanna Müller-Esparza[1,*], Bartosz Turkowyd[2,3,*], Lennart Randau[1,3,*], Ulrike Endesfelder[2,3,*]

This chapter is written in manuscript style and is currently under preparation for publishing. My contribution to this work included designing and cloning all the constructs, testing the formation and activity of the fluorescent complex, helping with microscopy sample preparation, writing the sections on the corresponding results and determination of binding times, and editing the manuscript.

---

[3] Prokaryotic Small RNA Biology Group, Max-Planck-Institute for Terrestrial Microbiology, D-35043 Marburg, Germany
[2] Department of Systems and Synthetic Microbiology, Max Planck Institute for Terrestrial Microbiology, D-35043 Marburg, Germany
[3] LOEWE Center for Synthetic Microbiology (SYNMIKRO), D-35043 Marburg, Germany
* Authors contributed equally

Abstract

CRISPR-Cas systems provide adaptive immunity to Archaea and Bacteria, protecting against viruses and mobile genetic elements through a process termed interference. In Type I CRISPR-Cas systems, interference is achieved through multi-ribonucleoprotein complexes (Cascades), which carry crRNAs with sequence complementarity to the invading DNA. In order to successfully overcome an infection, Cascades must scan through the DNA and find the correct target in a timely manner, without neglecting specificity. Mainly, *in vitro* techniques have been used to study the target search dynamics of Cascade complexes, while recent *in vivo* approaches revealed new levels of interaction with other players in the cell. This highlights the relevance of studying these systems in their native context, where crowding and DNA structures might influence the outcome. In this work, we explore the dynamics of both individual Cas proteins and the Cascade complex of a minimal Type I-Fv system from *S. putrefaciens* CN-32. Following a fluorescent version of the complex in *E. coli* by single molecule localization microscopy (SMLM) and single particle tracking (SPT) we report the ability of Cas5fv to bind to RNA and DNA individually, as well as the interaction of Cas6f with RNA. Furthermore, we describe the existence of two main populations for Cascade: target bound, with slow diffusion, and scanning, with fast diffusion. The frequency of these populations varies depending on the degree of crRNA:target complementarity. By adjusting the imaging intervals for Cascade, it was possible to extract the binding times to DNA, which extend up to 15 seconds for fully complementary targets. Sequential reduction of the complementary region led to proportionally shorter binding times. This first account of Cascade residence times *in vivo* shows that the interactions are more transient than the ones described *in vitro*, emphasizing the effect of other players on the stability of target-bound Cascades.

Introduction

The minimal CRISPR-Cas Type I-Fv system of *S. putrefaciens* CN-32

Type I CRISPR-Cas systems recognize and degrade foreign genetic material through multi-protein target recognition complexes, which are formed by Cas proteins and a small CRISPR RNA (crRNA). The crRNA sequence derives from small fragments of foreign DNA that were previously stored in the CRISPR array, in the form of repeat-flanked spacers, by an adaptation machinery. The ribonucleoprotein effector complex, or Cascade (CRISPR associated complex for anti-viral defense), utilizes crRNA to search for complementary sequences in a process termed interference. In a first step of interference, Cascade distinguishes between self and non-self DNA by identifying a short Protospacer Adjacent Motif (PAM) next to the region with complementarity to the crRNA (termed protospacer) (S. J. Brouns et al., 2008; Koonin et al., 2017; Xiao et al., 2017).

One of the best-characterized Type I systems is the *Escherichia coli* Type I-E system. Type I-E Cascade is formed by Cas5, Cas6, Cas7, Cas8 (also termed large subunit) and two small subunits (Cas11) with a stoichiometry of $(Cas8)_1$-$(Cas11)_2$-$(Cas5)_1$-$(Cas7)_6$-$(Cas6)_1$-$(crRNA)_1$ and Cas8 being responsible for PAM recognition. During target verification, the directional formation of an R-loop between the target DNA and the crRNA finally results in a conformational change of the complex, stabilizing the target binding. This locked state then recruits Cas3, a nuclease, for target degradation (Hayes et al., 2016; Redding et al., 2015; Xiao, Luo, Dolan, Liao, & Ke, 2018; Xiao et al., 2017). If the target-crRNA complementarity is partial, the target recognition process might be aborted or lead to primed adaptation. The latter process takes place especially when the PAM or first 8-10 nucleotides of the target (seed sequence) are mutated. In these cases, Cascade is unable to lock on the target and, therefore, cannot recruit Cas3 directly. Instead, a complex is formed by the adaptation machinery, Cas1 and Cas2, which binds Cascade and then recruits Cas3. This process can generate and add further spacers to the CRISPR array, bypassing escape mutations (Blosser et al., 2015; Dillard et al., 2018; Krivoy et al., 2018; Redding et al., 2015; Xue et al., 2016).

Closely related to the Type I-E system is Type I-F, which lacks the small subunits and comprises only 6 Cas genes, yet performs interference in a very similar manner (Chowdhury et al., 2017; Rollins et al., 2019; Rollins et al., 2017; Rollins, Schuman, Paulus, Bukhari, & Wiedenheft, 2015; Xue et al., 2016).

In this work, we focus on a recently identified, even smaller Type I-F variant (I-Fv) from *Shewanella putrefaciens* CN-32, which requires only a minimal set of five Cas proteins (Dwarakanath et al., 2015; Gleditzsch et al., 2016; Pausch et al., 2017). Because of its simplicity, it is well suited as a starting point for the study of Type I systems.

The Type I-Fv system consists of the integrase Cas1, a fusion of Cas2 and 3 (Cas2-3), Cas5fv, responsible for 5´-end capping of the crRNA, Cas6f responsible for maturing crRNAs and 3´-capping and Cas7fv, the Cascade backbone protein. The latter three proteins form in conjunction with a crRNA a minimal Cascade complex in a stoichiometry of $(Cas5fv)_1$-$(Cas7fv)_6$-$(Cas6f)_1$-$(crRNA)_1$. In this minimal complex, the large and small subunits were functionally replaced by domains in Cas5fv and Cas7fv that are not conserved among other Cas proteins (Pausch et al., 2017).

In particular, Cas5fv has a unique domain formed by six α-helixes (termed *Alpha Helical domain*, AH) that extrudes from the complex, occupying the approximate location of the large subunit in Type I-F Cascade. This domain was identified to be responsible for PAM recognition (Pausch et al., 2017). Furthermore, all Cas7fv subunits form a trench route able to stabilize the non-target strand, taking over the role of the missing small subunits. Therefore, despite these structural variations, *S. putrefaciens* Cascade is able to perform PAM-dependent interference, yet it is unknown whether the minimization of this complex also results in altered target search and R-loop formation mechanisms.

Single-molecule experiments exploring CRISPR-Cas systems

Single-Particle-Tracking (SPT) Photoactivated Localization Microscopy (PALM) enables high-density mapping of single-particle dynamics in living cells by combining Single-Molecule Localization Microscopy (SMLM) with SPT (Manley et al., 2008; Turkowyd, Virant, & Endesfelder, 2016). From this spatiotemporal map of individual proteins, information on diffusion behavior, as well as association and dissociation kinetics can be inferred, e.g. of DNA-associated proteins like DNA polymerases (Beattie et al., 2017; Hammar et al., 2014; Moolman et al., 2014).

Single-molecule studies have been integral in elucidating the molecular mechanisms that underlie the establishment of immunity by various CRISPR-Cas systems. They have provided important insights into the binding behavior of Cas9 (Dagdas, Chen, Sternberg, Doudna, & Yildiz, 2017; Globyte, Lee, Bae, Kim, & Joo, 2019; Jones et al., 2017; Josephs et al., 2015; Singh, Sternberg, Fei, Doudna, & Ha, 2016; Sternberg, Redding, Jinek, Greene, & Doudna, 2014; Szczelkun et al., 2014; van Beljouw et al., 2019; M. Yang et al., 2018; Zeng et al., 2018), as well as of Cas12 (Singh et al., 2018; Stella et al., 2018) and Cascade complexes (Blosser et al., 2015; Krivoy et al., 2018; Redding et al., 2015;

Rutkauskas et al., 2015b; Szczelkun et al., 2014; Xue, Zhu, Zhang, Shin, & Sashital, 2017).

In particular, experiments conducted on DNA curtains reported the bound-times of Cascade in a lock state to be longer than 57 minutes. When the complex is probing the DNA for targets, on the other hand, interactions last between 0.75 to 3 seconds. Regions with partial complementarity showed an abortive engagement of the complex of up to 25 seconds (Redding et al., 2015). Single-molecule fluorescence resonance energy transfer (smFRET) data revealed long-lived (> 1000 s) states interpreted as an interference binding mode, a short-lived state of 24.8 ± 8.9 s attributed to only partial R-loop formation and a rather transient interaction of 1.6 ± 0.4 s assigned to transiently bent DNA conformation (Blosser et al., 2015).

Up to now, single-molecule sensitive research has almost exclusively been conducted *in vitro*. The Cas9 target search process has also been studied in mammalian (Knight et al., 2015) and, very recently, in *E.coli* cells (Jones et al., 2017; van Beljouw et al., 2019). For Cascade complexes, a single study is available on the dynamics of a Type I-E complex (Vink et al., 2019). These studies show that single-molecule kinetics of effector complexes studied *in vitro*, where Cascade complex binding is not challenged by DNA structures and other cellular components, differ from Cascade dynamics observed in living cells. Thus, the study of CRISPR-Cas systems and their interactions in the cell should be taken in consideration in order to deepen the knowledge on their functionality.

In this work, we therefore explore the *in vivo* dynamics of the CRISPR-Cas Type I-Fv system with and without provided targets. Here, we explicitly consider and investigate a wide range of time scales, as we expect a large variety of different Cascade-DNA interference interactions, with highly diverse dynamics. These interactions could range from transient DNA-probing to stable binding of Cascade complexes, depending on the provided targets and the cellular environment.

Results and discussion

Studying the *in vivo* kinetics of the CRISPR-Cas Type I-Fv system

A fluorescent version of the Type I-Fv interference complex is needed in order to study Cascade dynamics at high spatio-temporal resolution. For this, *Cas5fv* was C-terminally fused to the green/red photoconvertible fluorescent protein Dendra2, carrying the mutation T69A that increases its brightness (Berardozzi et al., 2016). We heterologously produced this construct, both alone and together with *Cas7fv, Cas6f*, a nuclease-

deficient *Cas2-3* variant (Dwarakanath et al., 2015) and a minimal CRISPR array in *E. coli* BL21-AI. *E. coli* is a well-established prokaryotic model organism for SMLM (Gahlmann & Moerner, 2014) and the laboratory strain BL21-AI has no CRISPR Cas system of its own.

We verified the correct formation of Cascades through affinity purification (Ni-NTA) of the His-tagged Cas5fv-Dendra2-T69A co-expressed with the other complex components, followed by Size-exclusion Chromatography (SEC) (**Suppl. Figure VI.1**). Furthermore, the activity of the fluorescent complexes was assessed *in vivo* using Efficiency of Plaquing (EOP) assays, where interference against phage lambda was shown not to differ from wild type Cascades (0.30±0.19 vs. 0.29 ±0.20, n=3) (**Fig VI.1a**).

Cascade dynamics depend on the number of available targets. In order to study the different behaviors of the complex, we designed two spacers (i) targeting either the repetitive extragenic palindromic (REP) sequence of *E. coli* (Tobes & Ramos, 2005) to obtain Cascade complexes with targets in the *E.coli* genome (termed “Anti-REP”) or (ii) targeting the gene E from lambda phage, yielding Cascades without significant complementarity to the *E. coli* genome (termed “Anti- λ”) (Gleditzsch et al., 2016). We also generated other anti-REP constructs with multiplexed CRISPR arrays of 2 or 3 spacers to increase the number of targets in the genome from 16 to 42. In addition, to further study targeting intermediates, we designed different versions of the anti-REP spacer with varying complementarity to the targets. Since each Cas7fv in a complex binds 6 nt of crRNA, we randomized the 32 bp spacer sequence by multiples of 6 in order to obtain a range of Cascades with different affinities (**Suppl. Table VI.1)**.

Additionally, as the interactions of individual Cas proteins can also be studied through this method, we constructed strains expressing only Cas5fv-Dendra2-T69A, Cas6f-Dendra2-T69A or Cas5fv-ΔAH-Dendra2-T69A (deletion of AH domain, Gleditzsch et al., 2016), as well as a strain expressing an intermediate in the formation of Cascades (crRNA+Cas6f+Cas5fv) (**Suppl. Table VI.2**).

By fine-tuning imaging frequency, frame duration and laser intensity of our pulsed fluorescence read-out to budget the overall fluorescence lifetime, it is possible to observe DNA-Cascade interactions for individual protein complexes at different timescales from the millisecond range to several seconds (**Materials and Methods**) and thus can probe the system for the different interactions. Furthermore, it is also possible to probe their dynamics depending on the complementarity of the different crRNAs. For example, when imaging Cascade complexes with and without targets in living cells at a slow imaging speed (2 seconds long exposure), a significant fraction of Cascade Anti-REP complexes

are visible as immobile spots, being tightly bound to DNA targets. On the other hand, Cascade Anti-λ only show diffusive fluorescent signals (**Figure VI.1b**).

This observation was expected as the Anti-REP (single spacer) crRNA has 16 potential targets of 100% complementarity per *E.coli* genome, whereas Anti-λ has none. Because of this, the transient DNA interactions during target search are expected to be too short-lived to be visible as clear fluorescent spots at seconds-long exposure times. Using a higher temporal resolution, consisting of faster imaging rates with exposure times under 30 ms, the fast diffusing Cascades (for both crRNAs) as well as the DNA-bound, immobile fraction of Anti-REP complexes are revealed (**Figure VI.1b,** blue color-coded single-molecule trajectories have a diffusion of D > 0.03 μm²/s, red color-coded trajectories D < 0.03 μm²/s).

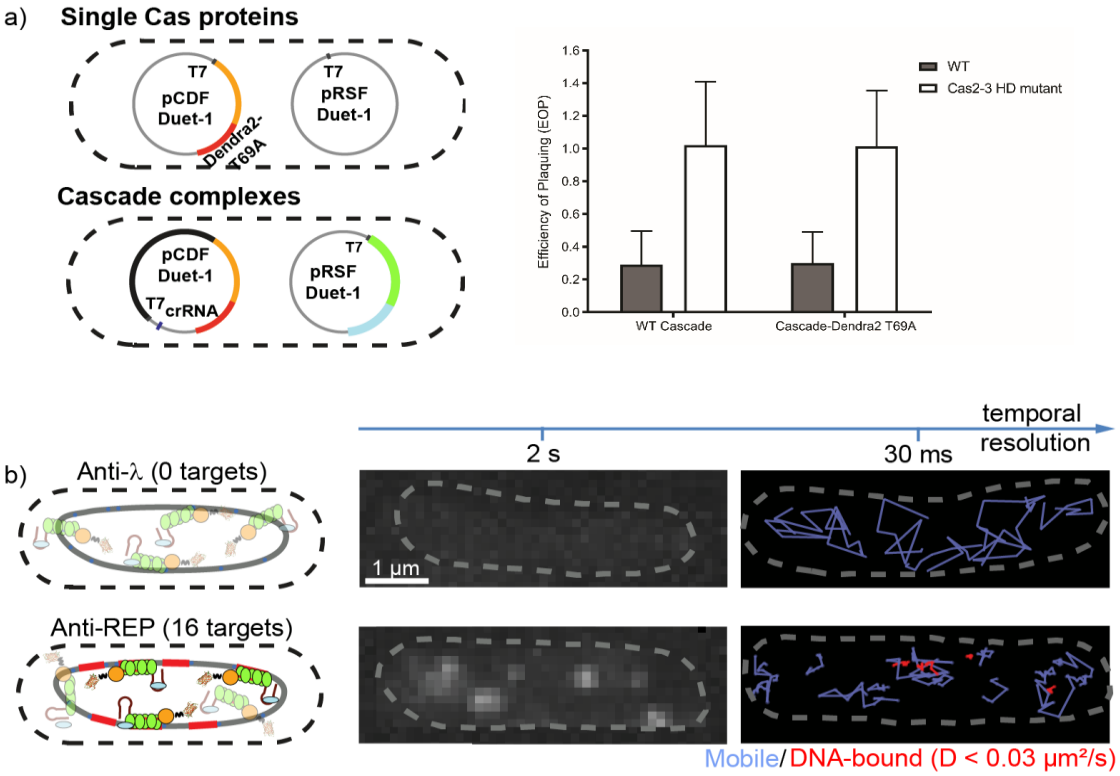


Figure VI.1. Evaluation of the activity of Dendra2-T69A-tagged Cascade complexes. (A) Left: representation of the *E. coli* BL21-AI strains used in this study. Cas5fv (orange) was C-terminally fused to Dendra2-T69A (red) and cloned individually or in an operon with a nuclease deficient mutant of Cas2-3 (black). For Cascade formation, a minimal CRISPR array with a single spacer was also included in this plasmid (blue). Cas7fv (green) and Cas6fv (cyan) were cloned on a second plasmid. Right: the activity of the fluorescent Cascade was assessed *in vivo* by Efficiency of Plaquing (EOP) assays with an active (grey) or a nuclease deficient Cas2-3 (white). EOP is the ratio of plaques formed by virulent lambda phages in the strain of interest versus a strain carrying empty plasmids. Three replicates were quantified. **(B)** Representation of the outcomes of SPT-PALM for two strains with different genomic targets at slow (2 s) and high (30 ms) temporal resolution.

Isolated Cas5fv proteins interact with DNA and RNA *in vivo*

To investigate the dynamics of the Type I-Fv system in detail, we first focused on individual Cas5fv proteins and performed sptPALM imaging of Cas5fv alone at a fast imaging rate of 77 Hz (**Figure VI.2a**). For all conditions measured, we summarized our single-molecule trajectory statistics by calculating the average jump distance (JD, average distance between two consecutive localizations in a trajectory) for each trajectory of statistically significant length (> 6 imaging steps) and plotted JD-probability density histograms weighting the <JD> values by trajectory length. As a SPT calibration standard we expressed cytosolic Dendra2-T69A and imaged it in living cells to obtain a distribution of purely freely diffusing molecules (**Suppl. Figure VI.2a,** median <JD> of diffusing molecules depends on the temporal resolution; 370 nm (13 ms exposure) and 400 nm (30 ms exposure)). As a second control, we imaged Dendra2-T69A in chemically fixed cells to obtain a <JD> distribution of immobile molecules which only show a residual median displacement of about 60 nm. Although particles are static, this result is larger than zero due to the limitations of the technique on localization precision (**Suppl. Figure VI.2b**).

For Cas5fv proteins, the <JD> distribution exhibits two fractions, one minor slow one at about 80 – 120 nm <JD> displacement, and a major faster one at 240 - 360 nm (**Figure VI.2a**). Interestingly, when compared to the Dendra2-T69A controls (grey outlines, **Figure VI.2a** and **Suppl. Figure VI.2**), the slow fraction is not fully immobile and the faster fraction is significantly slower than the freely diffusing fluorophore. As the two measured fractions on a first glance cannot be directly assigned to either of both, we set out to further explore the nature of these two observed fractions.

From the crystal structure of Cas5fv, as well as interference assays, it is known that the alpha-helical domain (AH) of Cas5fv is essential for PAM recognition (Pausch et al., 2017). Thus, we constructed a deletion mutant, Cas5fv-ΔAH, which is expected to show no major interaction with DNA. Interestingly, in our sptPALM data, we observed a shift of the faster fraction from 240 - 360 nm for Cas5fv to 280 – 440 nm for Cas5fv-ΔAH (**Figure VI.2b**). By this shift, this fraction of Cas5fv-ΔAH molecules shows a similar diffusion as cytosolic Dendra2-T69A (**Figure VI.2b**, grey outline) and strongly suggests that for Cas5fv-ΔAH, transient PAM recognitions are strongly reduced. This also lets us suppose that the fast fraction observed for Cas5fv could be explained by very short-lived DNA-protein interactions.

As the nucleoid is a large macromolecule, which at our imaging speed (13 ms per imaging frame) does not exhibit any significant diffusion (Alvarado, Turkowyd,

Endesfelder, & Ringgaard, 2018; Stracy, Uphoff, Garza de Leon, & Kapanidis, 2014; Turkowyd et al., 2017), a stable Cas5fv interaction with a bound time in the range of our imaging resolution would result in an immobile particle. Thus, the timescale of transient Cas5fv-DNA interactions is significantly below our temporal resolution. In consequence, Cas5fv molecules transiently interacting with the nucleoid show a slowed down, but still rather fast apparent diffusion.

For both proteins, Cas5fv and Cas5fv-ΔAH, a minor fraction with a <JD> of about 80 - 120 nm is present. As the crRNA-interaction pocket of Cas5fv is empty, we speculate that this fraction could be explained by a possible interaction of Cas5fv with RNA. This interaction could then lead to a co-diffusion of Cas5fv with RNA, which would account for the slow diffusive behavior for both Cas5fv and Cas5fv-ΔAH. To test this hypothesis, we performed experiments on cells with specific drug treatments.

First, we treated the cells with 10 µg/ml chloramphenicol for 60 minutes, which blocks translation by tethering translating ribosomes to the mRNA (Lopez, Marchand, Yarchuk, & Dreyfus, 1998). As the cell tries to counteract the increasing fraction of non-functional ribosomes, this leads to strongly increased transcription and consequently increases the RNA/protein ratio, in particular in particular resulting in an unusually high amount of ribosomal RNA (Lopez et al., 1998; Scott, Gunderson, Mateescu, Zhang, & Hwa, 2010). Furthermore, the nucleoid is strongly compacted (**Figure VI.2c**, inset, and (Zusman, Carbonell, & Haga, 1973)). In consequence, cells contain large DNA-free regions enriched with RNAs. By selectively analyzing only trajectories from these areas (as indicated by the yellow box in the inset of **Figure VI.2c**), we exclusively observed slow diffusive fraction of Cas5fv (**Figure VI.2c**) which in its nature overlaps with the slow diffusive fraction of Cas5fv in untreated cells (orange outline in **Figure VI.2c**). Importantly, under chloramphenicol treatment, a large fraction of this RNA is rRNA, which is not translated and thus free from ribosomes and accessible for Cas5fv.

We then co-expressed Cas5fv together with Cas6f and the anti-λ crRNA, which does not match any cellular target, in order to fill the crRNA binding site of Casf5v with a Cas6f-matured crRNA. Under chloramphenicol treatment, this should lead to a large fraction of freely diffusing Cas5fv/Cas6f/crRNA complexes and possibly Cas5fv/crRNA in the DNA-free areas of the cells, which thus cannot interact anymore with the RNAs present. Indeed, next to a similar slow fraction as observed for Cas5fv without crRNA (**Figure VI.2d**, orange outline Cas5fv under chloramphenicol), we measured a second, significantly faster fraction of mobile complexes (**Figure VI.2d**), albeit at a bit slower apparent JD peak of 200-320 nm compared to freely diffusive Cas5fv-ΔAH and Dendra2-

T69A. This suggests that this second fast peak corresponds to successfully formed Cas5fv/Cas6f/crRNA complexes, which indeed cannot interact with random RNAs.

We then evaluated the Cas5fv dynamics after the simultaneous treatment with 10 µg/ml chloramphenicol and 5 µg/ml rifampicin for one hour (**Figure VI.2e**). Rifampicin efficiently blocks the transcription of the bacterial RNA polymerase by inhibiting its translocation along the DNA (Campbell et al., 2001; Sippel & Hartmann, 1968). This causes a rapid mRNA depletion in the cells, as the average mRNA lifetime is about 2.5 minutes (Chen, Shiroguchi, Ge, & Xie, 2015; Liang, Ehrenberg, Dennis, & Bremer, 1999; So et al., 2011). Importantly, our expression system used to express Cas5fv from our plasmid system is under the control of a T7 promoter. The viral T7 RNA polymerase is not inhibited by rifampicin (Hesselbach & Nakada, 1977), thus mRNA encoding for Cas5fv can still be synthesized. Moreover, the T7 RNA polymerase yield has been shown to be positively affected under the influence of rifampicin (Kuderova, Nanak, Truksa, & Brzobohaty, 1999). Nevertheless, this mRNA pool should be largely inaccessible due to stalling ribosomes as a result of the chloramphenicol treatment (Lopez et al., 1998). The action of chloramphenicol is rapid, as *de novo* synthesized mRNA reaches a ribosome-rich area in less than one second (Bakshi, Siryaporn, Goulian, & Weisshaar, 2012). Thus, we expected a lowered slow fraction, with the faster DNA-interacting fraction remaining unchanged. Our data is in agreement with these assumptions as the minor peak vanished (**Figure VI.2e**). Nevertheless, as this minor peak is not very significant for Cas5fv, we strived to show this effect in a more pronounced manner and probed Cas6f dynamics.

As Cas6f is involved in pre-crRNA maturation to obtain final crRNA protospacers, we hypothesized it to strongly interact with cellular RNA when trying to find its substrate. Indeed, Cas6f has a rather slowly diffusing JD distribution, which seems to be explained by at least two diffusive fractions (**Figure VI.2f**). Both could be attributed to RNA interactions, yet we did not further investigate them due to the broad range of possibilities. More importantly, the simultaneous drug treatment led to a significant and clearly visible reduction of slow diffusing Cas6f and the overall diffusion dynamics became similar to freely diffusing Dendra2-T69A (**Figure VI.2g**). As a control, rifampicin and chloramphenicol treatment on cytosolic Dendra2-T69A showed no effect (**Suppl. Figure VI.2**). This hints at a strong shortage of RNA interactions. Thus, this clear phenotype of Cas6f dynamics substantiates our hypothesis of almost complete RNA depletion or blockage under double drug treatment and confirms our observations of the diffusive Cas5fv distributions.

We additionally tried to measure the dynamics of Cas5fv-ΔAH/Cas6f/crRNA complexes, as these should interact neither with DNA nor with RNA and remain freely diffusive. Remarkably, we were not able to measure possible Cas5fv-ΔAH/Cas6f/crRNA dynamics, as these complexes were highly excluded from the nucleoid area as indicated by DAPI staining (**Figure VI.2h**, blue signal) and formed concentrated clusters at cell poles or in between nucleoids. This effect cannot be explained by a lack of DNA-interaction only, as we then should observe a similar effect for Cas5fv-ΔAH alone which we found normally distributed and thus present in the nucleoid-rich areas interacting with *de novo* synthesized RNA (**Figure VI.2b**). Together, this strongly indicates that Cas5fv-ΔAH/Cas6f/crRNA also lacks RNA-interaction. In consequence, the incomplete complexes are excluded from the denser cellular mid-cell area by entropic forces and then, due to the high crowding in those excluded areas, are proposed to aggregate.

In summary, our observations strongly indicate that the slowly diffusing native Cas5fv dynamics as measured in **Figure VI.2a** can be attributed to RNA-interacting Cas5fv molecules, whereas the majority of Cas5fv molecules in the main, fast fraction transiently interacts with DNA.

Fast *in vivo* kinetics of Type I-Fv Cascade

We then set out to elucidate the details of the different kinetic sub-populations of Cascade complexes, in addition to assess their interaction dynamics and how they vary depending on their degree of complementarity to the crRNA. First, we determined the amount of Cascades per cell by counting the total of fluorescent molecules until no more photoconversion was detected. In our experimental conditions, an average of 24 Cascades are present per cell (**Suppl. Figure VI.3**), indicating that there are more available targets than Cascade complexes. We performed detailed sptPALM experiments of Cascades carrying different crRNAs (**Figure VI.3a**, details **Suppl. Table VI.1**) and measured 2-4 biological repeats from different cultures and imaging days at large single-molecule statistics for 13 and 30 ms camera exposure times (**Figure VI.3b**). For Cascades, the obtained JD distributions are more complex than for single Cas proteins and vary according to the complementarity of the provided crRNA. Each crRNA revealed an individual and reproducible fingerprint of interaction dynamics (**Figure VI.3b**).

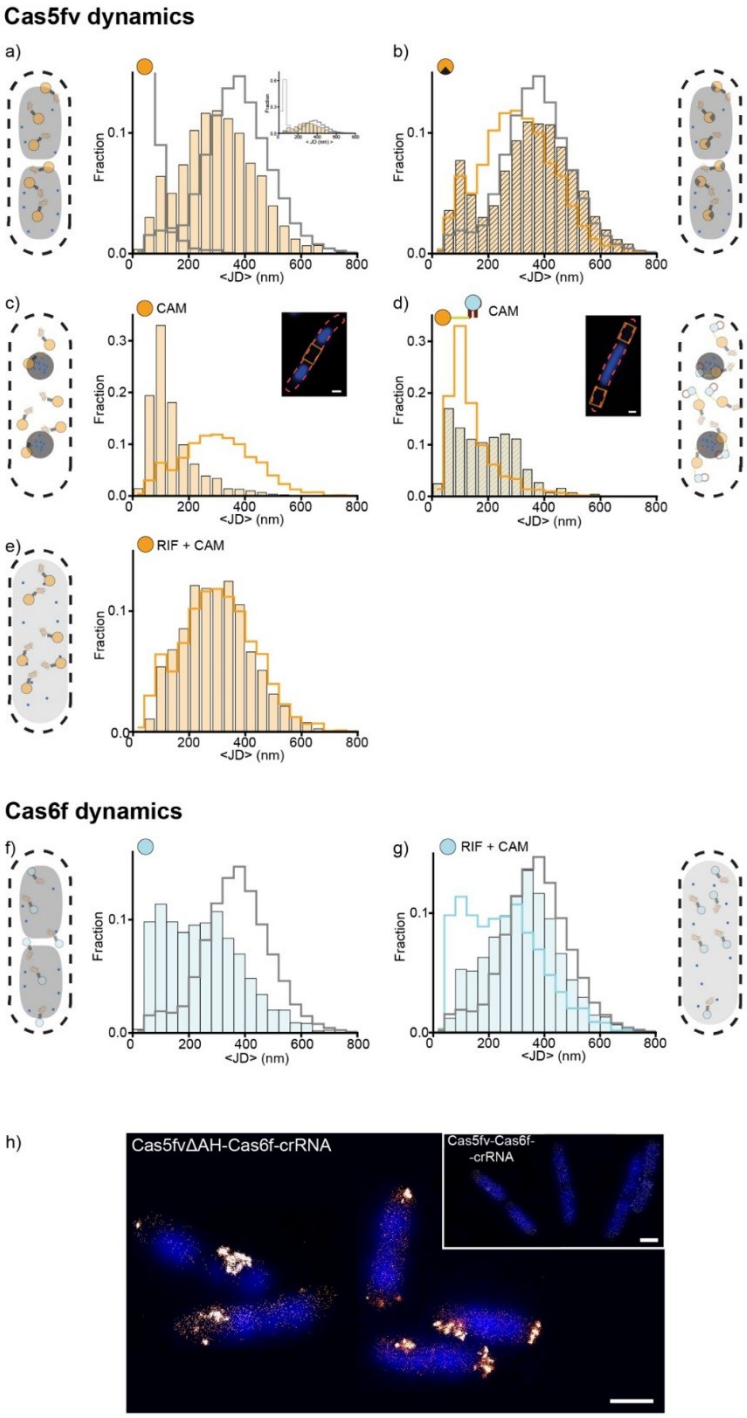


Figure VI.2. Dynamics of individual Cas proteins. Jump Distance (JD) histograms of Cas5fv-Dendra2-T69A and Cas6f-Dendra2-T69A. Grey outlines show the distribution of the immobile (low JD) and freely diffusing (high JD) controls. **(A)** Cas5fv has two fractions showing transient interactions. **(B)** Cas5-ΔAH is mainly freely diffusing but one fraction remains unchanged. The orange outline is the distribution of the wild type Cas5fv **(C)** In DNA-free, RNA-enriched areas (orange box in the inset) Cas5fv exists mostly as a slowly diffusing protein. The orange outline is the distribution of Cas5fv without antibiotics **(D)** When Cas5fv is loaded with the crRNA, the slowly diffusing fraction is decreased when compared to (C) (shown in orange outline). **(E)** In RNA-free cells, Cas5fv exists in a fast, only transiently interacting fraction, similarly to Cas5fv in untreated cells (orange outline). **(F)** Cas6f exists in two slowly diffusing fractions. **(G)** These fractions are not present in RNA-free cells (light blue outline is the distribution in (F)). **(H)** Cas5-ΔAH carrying crRNA is perfectly excluded from the DNA area (blue signal, DAPI) and forms clusters at cell poles and in the middle of the cell (yellow signal). Scale bars: 1 μm.

For 13 ms exposures, Cascade Anti-λ, which has no DNA-targets, is missing a stationary fraction (compared to the grey overlaid outline of immobile Dendra2-T69A) but shows a large peak of high mobility with a maximum around 240 - 320 nm and a smaller peak at slower diffusion with displacements around 80 - 160 nm (**Figure VI.3b, first column**). Importantly, both fractions resemble neither the immobile, fixed Dendra2-T69A nor the freely diffusing Dendra2-T69A distribution. This suggests that we observe some transient interactions that cannot directly be resolved at our temporal resolution (as fully resolved interactions would appear as stably bound fractions similar to chemically fixed Dendra2-T69A).

When imaging at slower, 30 ms exposures, the faster-moving fraction shifts to larger <JD>s with a maximum peak at about 320 - 360 nm, reinforcing the idea that this fraction represents a highly transient state, frequently cycling between on- and off-binding at high rates. Contrarily, the slower transient fraction is not resolvable anymore. This fraction could represent two kinds of interactions: first, a rather slowly frequented state with bound times within the lower millisecond range. Currently, it is not possible to image at faster exposure times to confidently resolve short-lived pauses and prove this hypothesis, as faster imaging results in lower localization precision. Second, this fraction could correspond to a slow free-diffusive state (e.g. large hydrodynamic radius or big mass), which for 30 ms exposures shifts more towards higher <JD>s than the transiently interacting fraction, overlapping with it.

We therefore controlled if this fraction contains complexes interacting with DNA or RNA by depleting the free mRNA pool by a simultaneous treatment with chloramphenicol and rifampicin, as shown for single Cas proteins before. Indeed, for Cascade Anti-λ, the transient peak diminishes like for Casf5v molecules before, indicating a slow free-diffusive state given by RNA interactions (**Figure VI.3c, first column**).

We then allowed for partial 12 bp complementarity with 133 targets on the *E.coli* genome, particularly on the REP sequences (**Figure VI.3b, second column**). These partial targets should elicit the binding of the first two Cas7 subunits of Cascade, leading the complex to slow down upon target recognition. We indeed see a stabilizing effect, as an immobile fraction appears. Furthermore, this fraction does not vary to a large extent at slower imaging speeds of 30 ms exposures. Nevertheless, this sample also exhibits a plateau of transient interactions (around 200 nm displacement), which we do not fully resolve anymore at 30 ms exposures. This indicates that the interaction time of Cascade Anti-REP (12bp) with the DNA is in the tens of millisecond range.

We then explored the dynamics of Cascades carrying a crRNA with full targets (32 bp) in the genome (**Figure VI.3b, third column**). For this, we multiplexed the CRISPR array to carry three spacers targeting the REP sequences in the *E. coli genome*, shifting the previously studied 133 targets of only 12 base pairs complementarity to 42 targets with 100% complementarity to the crRNAs and 91 targets with different degrees of lower complementarity (**Table VI.1, Figure VI.3a**). In consequence, as the total number of targets is still 133 but the complementarity is highly shifted towards higher base pair matching, the ratio of the different fractions shifts in favor of a significantly higher peak of stable DNA interactions. In addition, although the plateau of transient interactions remains similar, the fast fraction at highest mobility drastically decreases.

This distribution can be further shifted to a bimodal fashion, reducing the plateau fraction, by changing the targeting range of the crRNAs. When imaging Cascades carrying a crRNA with less fully complementary targets in the genome (16 instead of 42) and overall less intermediates of lower complementarity (**Figure VI.3b, last column**), we can resolve two clear peaks resembling Cascade complexes. They correspond to complexes either stably bound to highly complementary targets, or interacting only transiently with the DNA, similarly to Cascades with no significant target in the genome (Anti-λ). This suggests that the middle fraction depends on targeting affinities over 12 bp, with lower affinities piling up to the faster moving fractions.

Finally, we cannot detect any significant changes in the distribution (except for a small decrease in the middle, transiently interacting, fraction) when imaging this strain under RNA-depleting conditions, which substantiates the strong DNA-binding nature of Cascades carrying targeting crRNAs (**Figure VI.3c, last column**).

In summary, the interactions of Cascade complexes with the variety of possible genomic targets is very diverse and consequently leads to multi-layered and highly overlapping diffusion statistics. Nevertheless, the distribution patterns are highly specific for each crRNA. We expect the fast moving fraction to group all the interaction intermediates that are too short-lived to be resolved, including PAM scanning. The only component that we can unmistakably assign is the slow fraction (under the respective temporal resolution), that corresponds to target-bound Cascades. Consequently, as the presence of this fraction is constant, we expect to be able to measure the bound time of Cascades to their targets in relation to their crRNA complementarity.

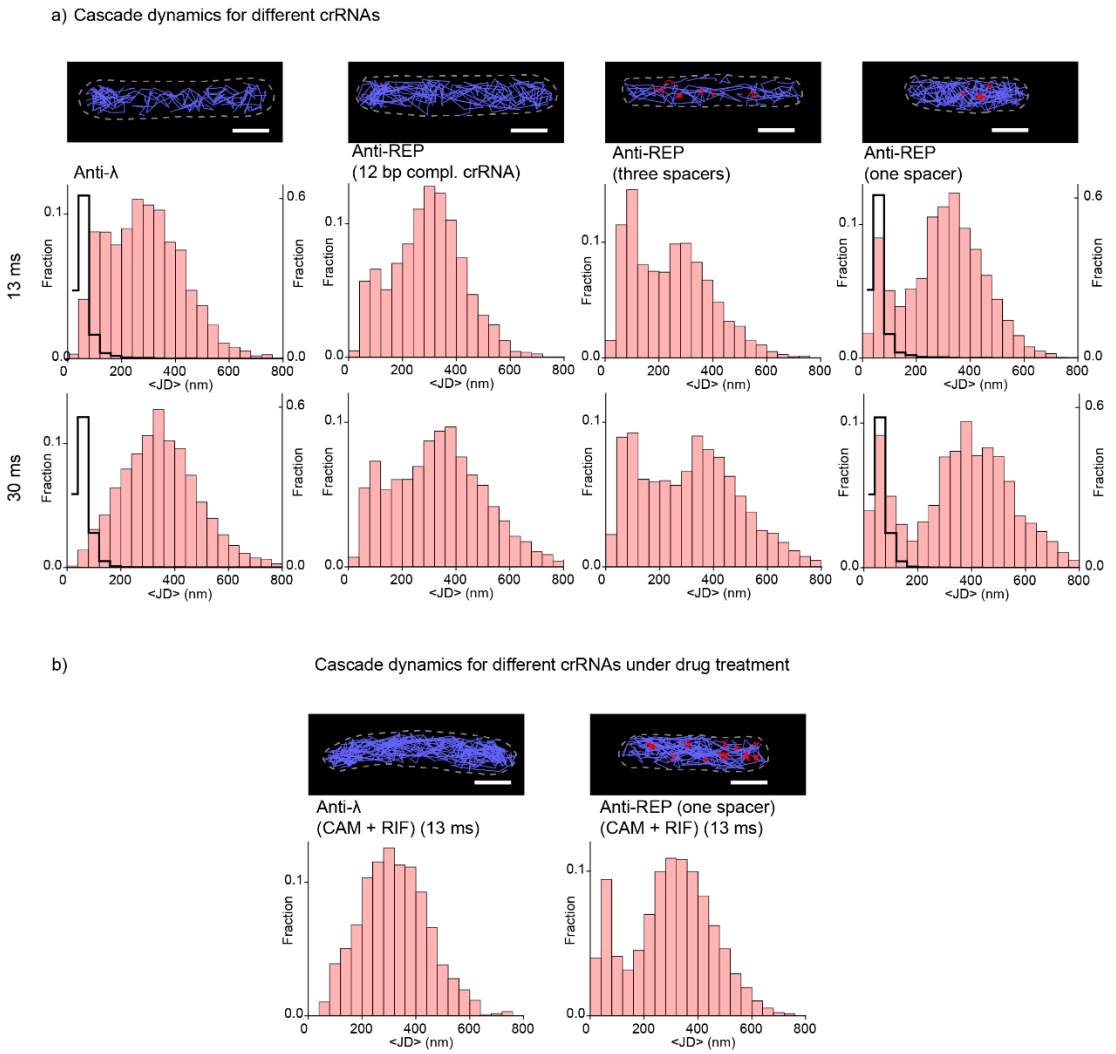


Figure VI.3. Dynamics of Cascades with different target specificities. (**A**) Cascades with different crRNA variants have a unique <JD> distribution. Cascade (Anti-*λ*) (first column) lacks the immobile fraction (compared to grey overlaid outline of immobile Dendra2-T69A) in both, 13 and 30 ms, imaging intervals. Two fractions showing transient interactions are visible with the 13 ms imaging interval and merge into the one fraction under 30 ms exposure. When Cascades can match partial targets (Cascade (Anti-REP 12bp), second column) the immobile fraction appears in the 13 ms imaging interval and remains unchanged when the imaging interval is increased to 30 ms. By providing crRNAs which in most cases can match full (>31 bp complementarity) or almost full (24-30 bp complementarity) targets (Cascade (Anti-REP three spacers), third column), immobile fraction becomes more prominent as well as fraction of the transienly interacting Cascades is significant. When Cascade is supplied with the crRNA which can not match partial (12-24 bp complementarity) targets, <JD> distribution becomes almost bimodal (Cascade (Anti-REP one spacer), last column). (**B**) <JD> distributions for Cascade (Anti-*λ*) and Cascade (Anti-REP (one spacer)) imaged with the 13 ms imaging interval in cells lacking mRNA. In Cascade (Anti-*λ*) slower fraction is not visible anymore, while in Cascade Anti-REP, the <JD> distribution remains unchanged. Trajectory maps for exemplary cells are shown (color coding as in Figure VI.1). Scale bars: 1 *μ*m.

crRNA complementarity dictates Cascade binding times

Determining the binding times by sptPALM is restricted by the photophysics of the selected fluorescent reporter. When tracking immobile foci (e.g DNA-bound Cascades) in time, their disappearance could be explained either by the release of the complex from the DNA or by photobleaching of the fluorescent tag. It is possible to bypass this limitation

by performing non-continuous imaging and introducing a varying dark time in between the read-out frames, which results in time-lapse times (τ_{tl}) of different lengths (Gebhardt et al., 2013). With this approach, the effective off-rate constant (K_{eff}) and, in consequence, the unbinding constant (K_{off}) can be determined. Here, the K_{eff} constant describes the loss of immobile foci and corresponds to the sum of the unbinding constant (K_{off}) and the bleaching constant (K_b). As the bleaching constant does not change due to the consistent read-out illumination in between the varying dark times, a higher K_{eff} would indicate a higher rate of target release. Therefore, K_{eff} is inversely proportional to protein binding time. By measuring a bleaching control, K_b can be determined. Thus, when measuring $K_{eff,}$ K_{off} can be determined for each sample.

In order to evaluate the interaction rates of the complex in context, we first need to stablish the highest and lowest ranges possible. To determine K_b, we imaged a fixed sample of a LacY-Dendra2-T69A fusion at different intervals (T_{tl}). LacY was chosen for this control since it is a membrane bound monomer (Kumar, Mommer, & Sourjik, 2010). As unbinding is not possible in a fixed sample, K_{off} is zero and K_{eff} is the direct equivalent of the bleaching rate (K_b).

On the other hand, imaging of Cascades with no targets in the genome, e.g the Anti- λ complex, would retrieve the highest possible K_{eff}, as no long-lasting bindings occur. The imaging of this sample at long frame intervals results in a small amount of trackable particles, as most of them do not considerably slow down for long periods. This leads to high background noise and few retrievable data points, with consequently higher errors. Nevertheless, we would expect that all the possible complex-DNA interactions fall in between this and the fixed sample values **(Figure VI.4a).**

With upper and lower limits established, we then determined the K_{eff}s of Cascade complexes carrying crRNAS with different complementarities to the REP targets. Imaging the complexes with 12, 18, 25, 30, 32 nt (1 spacer) and 32 nt (3 spacers) of complementarity at intervals of 0.2, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5 and 5 seconds allowed us to plot the relation between K_{eff} and T_{tl}. For all samples, including the fixed and Anti- λ, this relation was adjusted to a linear regression in order to determine the dissociation rate constant (K_{off}), which corresponds to the slope of the function **(Figure VI.4a and Material and Methods).** In the future we expect to refine this determination by imaging at intermediate intervals of the ones here presented.

All Cascade samples with targets in the genome exhibit constants that fall between the previously determined upper and lower limits. Furthermore, an inverse relation exists between the target complementarity and the determined K_{eff}s, with increased sequence complementarity leading to more stable interactions of Cascade. Interestingly, the expected sequential addition of Cas7fv subunits to the R-loop structure generates a continuous increase in $K_{eff.}$ It is possible that an abrupt change in K_{eff} could be observed when studying complexes with complementarities before and after the seed sequence limit, as Cascades have been reported to have a higher stringency for this portion (Semenova et al., 2011; Wiedenheft et al., 2011). As it is still unclear what guides this stringency, it would be interesting to study complexes where a single Cas7fv subunit takes part in the R-loop. On the other hand, we cannot assure that these interactions could be imaged with the current resolution, since they are expected to happen very fast, similar to PAM recognition pauses.

In addition to the complexes with different complementarities, we also imaged and determined the K_{eff}s of Cascades formed with a ΔAH Cas5fv subunit and a fully complementary (32 nt) crRNA. This sample exhibited off-rate constants similar to the ones obtained for the Anti- λ complex, re-inforcing the importance of this domain for target binding.

Using the previous results it was possible to calculate the binding times of the complex to the possible targets in the cell (as k_{off}^{-1}). As pointed out before, the dynamics of the complex depend directly on the target complementarity. Therefore, we need to put the binding times of Cascades in the context of their available targets. Under these experimental conditions there are less Cascades than targets, with an average of 24 complexes per cell (Suppl. Figure VI.3 and Suppl. Table VI.1). For steady state conditions, it is expected to find complexes bound to the best target available. Therefore, most targets with the highest complementarity should be occupied, with leftover Cascades sequentially binding the less-optimal options. Because of this, we calculated the average target length per construct by taking in consideration only the best 24 targets available in the cell (Suppl. Table VI.1), and plotted them against their corresponding binding times.

Minimal complexes carrying crRNAs fully complementary to targets stay bound for an average of 14.6 seconds. In the cells where more perfect targets than complexes are available (anti-REP 3 spacers), the residence time increases (14.0 s for 1 spacer vs 15.2 s for 3 spacers), supporting the hypothesis that complexes bind preferentially to targets with higher matches. When the AH domain of Cas5fv is not present, Cascade interaction

time is highly reduced, with bindings that fall below our detection limit, similar to the Anti-λ sample. These interactions are faster than for complexes with 12 bp of complementarity to targets (9.7 s) **(Figure VI.4b).** We expect to have a higher resolution of these fast-binding states when imaging at shorter intervals ($\mathbf{T}_{tl}$).

The binding times of the complexes with 30 bp of complementarity show the relevance of the complete match between crRNA and DNA, as the increase in binding time given by the pairing of the last two nucleotides (12.2 s vs 15.3 s) supports the existence of a lock state of the complex dependent on the involvement of the sixth Cas7fv subunit, despite the lack of a large subunit.

Furthermore, it was possible to study the effect of intermediate complementarities in the stability of the binding. Here, we used a complex with a target average length of 25 bp. This sample exhibits binding times similar to complexes where 5 Cas7fv subunits are expected to form part of the R-loop (12.1 s vs 12.2 s), even though the complementarity on the portion corresponding to the fifth Cas7fv is of 1 bp. This suggests that the sequential participation of backbone subunits on the dsDNA opening requires a single nucleotide match on the 6 nt stretches that they cover, stabilizing the R-loop and facilitating the pairing of the following nucleotides.

Overall, these results suggest that the further the mismatches are along the crRNA:DNA duplex, the longer it takes for Cascade to terminate the binding, with each additional Cas7fv subunit insertion delaying the release by ~1 second. Therefore, abortive engagement might not be an active process, but rather depend mostly on complexes dislodging due to poor stability.

The interaction times here reported are strikingly shorter than the ones described *in vitro* (Redding et al., 2015). This could be explained by several reasons, with some being intrinsic to the minimal Cascade, and others depending on further cellular components.

As the studies performed *in vitro* have used Type I-E complexes, a first possibility is that the binding of the Type I-Fv Cascade is less stable due to the diversification and reduction of its components. In the future, we expect to explore this possibility by determining the binding times of other complexes, to study the relevance of the small and large subunit in the integrity of the interaction. Furthermore, genomic targets have more dynamic physical features in comparison *to in vitro* targets. The length and topology of the DNA, as well as protein-mediated nucleoid structures, could affect the stability of the R-loops.

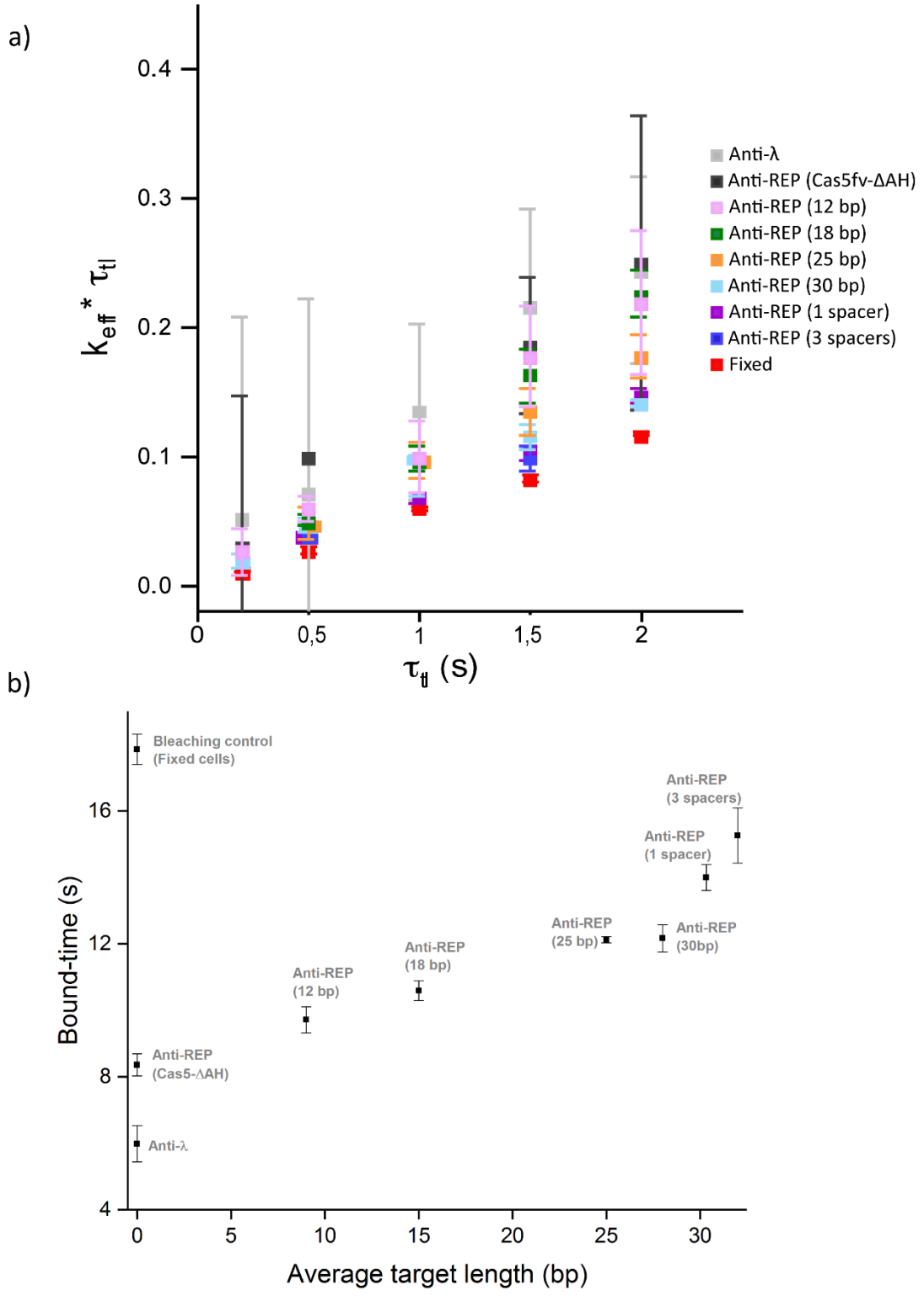


Figure VI.4. Determination of the binding stability of Type I-Fv Cascade complexes. (A) Effective unbinding rates (K_{eff}) multiplied by the imaging interval (τ_{tl}) in function of τ_{tl}. Measurements correspond to the residence time of minimal Cascade complexes to targets with different complementarity. Error bars represent standard error. **(B)** Cascade-DNA binding times represented as a function of the average target length. As the number of Cascades (24 copies) is significantly smaller than the number of potential targets in the host genome, the average target length was calculated from 24 most complementary targets. Error bars represent the standard error.

At the protein level, it has been shown that Type I-E and Type I-F Cascades act as roadblocks for the replication machinery, which are then relieved by the helicase RecG (Heussler, Miller, Price, Collins, & O´Toole, 2016; Ivancic-Bace, Cass, Wearne, & Bolt, 2015; Killelea, Hawkins, Howard, McGlynn, &Bolt, 2019). This is an active mechanism by which cells remove R-loop structures, as they affect genome stability. This variable was not included in the previous *in vitro* determinations, possibly accounting for higher binding times. Finally, other DNA-binding proteins and complexes compete with Cascades for targets under this context. REP sequences, used here as targets, act as binding sites for DNA gyrase, DNA polymerase, Integration Host Factor (IHF) and transposases (Tobes & Pareja, 2006). This competition might lead to lower binding times, depending on the affinity of the complex for its target. As phages exhibit different DNA topology (Ding et al., 2014), as well as a different set of DNA-interacting proteins, the study of a natural target, although technically challenging, could give further insights into how all these players affect Cascade binding.

Materials and Methods

Bacterial strains and growth conditions

E. coli BL21-AI strains (F- ompT hsdSB (rB-mB-) gal dcm araB::T7RNAPtetA, Invitrogen) were used for bio-layer interferometry (BLI) assays and SMLM. All strains were grown in 2YTL media (16 g*/*l tryptone, 10 g*/*l yeast extract, 5 g*/*l NaCl, 10 mM MgSO4, 0.2% maltose) supplemented with appropriate antibiotics at 37ºC.
A virulent variant of lambda phage (NCCB 3467) was obtained from Centraalbureau voor Schimmelcultures (Utrecht, Netherlands).

Construction the Cas5fv-Dendra2-T69A fusion

Gibson assembly was performed using the Gibson Assembly Master Mix (New England BioLabs) following the manufacturer's protocol. Genes coding for Cas2-3 (Sputcn32_1820, wild type and HD nuclease deficient) and a Cas5fv (Sputcn32_1822)-Dendra2-T69A fusion were cloned as an operon on the second multiple cloning site of pCDFDuet-1. Dendra2 T69A amplification oligonucleotides carried overlapping regions to the C-terminal sequence of Cas5fv plus a sequence coding for a GGGGS linker. Following the same procedure, the Cas5fv-Dendra2-T69A fusion was cloned individually for Cascade-independent experiments. On the first multiple cloning site a minimal CRISPR-array was cloned by oligo annealing, as previously described (Gleditzsch et al.,

2016). Arrays carrying multiple spacers targeting the REP sequences of *E.* coli were synthesized by GeneArt and subcloned into pCDFDuet-1.
Production of the full-size Cas5fv-Dendra2-T69A was confirmed by Western Blot of a His-tagged version of the fluorescent fusion (data not shown).

Bio-layer Interferometry (BLI)

Overnight cultures of BL21-AI cells expressing either his-tagged wild type or fluorescent Cascade carrying a crRNA targeting gene E from phage lambda were diluted 1:100 in 1 lt of 2YTL media and grown until OD_{600nm}: 0.6 at 37°C. After induction with 0.2% Arabinose and 0.1 mM IPTG, cells were grown overnight at 18°C. Cascade complexes were purified as previously described (Gleditzsch et al., 2016). The correct assembly of the complex was corroborated by SDS-Page and corresponding fractions were concentrated to 250 nM. Unspecific interactions between the complex and the high precision streptavidin (SAX) bio-layer were avoided by addition of Triton X-100 and bovine serum albumin (BSA) to final concentrations of 0.01% and 10µM, respectively, to the protein purification buffer (15 mM HEPES, 150 mM NaCl, 1 mM DTT, pH=7.0). Two 5'-Biotin-triethylene glycol (TEG)-modified oligos were designed for BLI assays, the first one carrying the GG-PAM and complementary sequence to the Anti- λ crRNA, and the second one with a randomized DNA sequence, without GG pairs. Probes were designed as single stranded 95 nt oligos, with a 5-thymidine loop flanked by an EcoRV restriction site and a randomised sequence flanking the protospacer. They are predicted to adopt the expected dsDNA structure at room temperature (Mathews et al., 2016). Oligos were synthesized by Sigma-Aldrich and diluted to a final concentration of 100 nM in the protein purification buffer plus BSA and Triton X-100. Interactions where measured with the BLItz system from ForteBio. After setting a buffer baseline, the high precision streptavidin bio-sensors were saturated with the respective oligo for 30 s, and a new baseline was set. Cascade binding to dsDNA was measured for 420 s, and dissociation was then measured for another 180 s by incubating the biosensor back in the initial buffer. Data was normalised by establishing the oligo-binding step as the baseline. Experiments were performed in duplicates.

Efficiency of plaquing (EOP) assays

Phage assays were performed as previously described (Gleditzsch et al., 2016). Briefly, a 1/100 dilution of an *E. coli* BL21-AI overnight culture was grown until OD600 nm = 0.3 and induced with 0.2% arabinose (Sigma-Aldrich) and 0.1 mM IPTG for 30 min. Then, cells were pelleted (4000 g, 10 min, 4°C) and resuspended in 10 mM MgSO4. A 1:1 mix

of cells and serial phage dilutions was incubated for 20 min at 37°C and added to 2YTL top-agar containing IPTG and arabinose, plated over 2YTL agar plates with antibiotics and incubated overnight at 37°C, after which the plaque number was determined. Efficiency of plaquing (EOP) was defined as the ratio between the plaque count of the strain of interest and the plaque count of the strain carrying empty plasmids. Phage assays were performed in triplicate and error bars were calculated as standard deviation (SD).

Cell Culture Preparation

E. coli BL21-AI strains carrying the required plasmids (**Suppl. Table VI.1**) were inoculated from cryo stocks into fresh LB medium supplied with kanamycin and spectinomycin (both 25 *µ*g/mL) and grown overnight at 37 °C. Cultures were then reinoculated into fresh 2YTL (1:100 dilution, 3 mL final culture volume), grown for one hour at 37 °C and induced with 0.1 % arabinose for another hour afterwards. The induction of Cas proteins expression was stopped by exchanging medium into 3 mL EZRDM (Teknova, USA).

Cells grew in this medium for one hour to enable protein maturation and Cascade assembly. 500 *µ*L of each culture was centrifuged (2 min, 3000 × *g*), washed twice with fresh EZRDM and suspended in residual 50 *µ*L medium. 2 *µ*L of suspended cells were placed on agarose pads covered with the cleaned coverslip and incubated for 15 minutes on a bench to settle them down on the pad.

BL21-AI carrying the pRSETb-Dendra2-T69A plasmid were inoculated from the cryo stock into fresh LB medium (100 *µ*g/mL ampicillin) and grown overnight at 37 °C. On the next day, the culture was reinoculated into fresh 2YTL medium (1:100 dilution, 3 mL) and after one hour growth at 37 °C induced with 0.05 % arabinose for another hour. Expression was stopped by the replacing medium by 3 mL EZRDM and cells grew for one hour. 500 *µ*L of culture were washed twice with fresh EZRDM and re-suspended in residual 50 *µ*L EZRDM. For fixed samples, 500 *µ*L of the culture was chemically fixed with 3 % formaldehyde for 15 minutes. After fixation, the culture was washed twice and suspended in 50 *µ*L medium. 2 *µ*L of suspended cells were placed on agarose pads covered with a cleaned coverslip and incubated for 15 minutes on a bench to settle them down on the pad.

Agarose Pads Preparation

Low-temperature gelling agarose (Sigma-Aldrich, Germany) was suspended in freshly prepared EZRDM to a final concentration 1 % (w/v) and heated up to 70 °C until agarose

melted completely, and stored later at 37 °C. Cooled to 37 °C agarose was placed on indented microscope slide (Thermo Fisher, Germany) and sealed with a coverslip cleaned overnight in 1 M KOH (Sigma-Aldrich, Germany). After two hours EZRDM-agarose pads were ready to use.

SMLM Setup

Imaging was performed on a custom built setup based on an automated Nikon Ti Eclipse microscope, equipped with appropriate dichroic and filters (ET dapi/Fitc/cy3 dichroic, ZT405/488/561rpc rejection filter, ET610/75 bandpass, all AHF Analysentechnik, Germany) and a CFI Apo TIRF 100x oil objective (NA 1.49, Nikon). 405 and 561 nm laser devices (OBIS, Coherent Inc., Santa Clara, California USA) were modulated via an acousto-optical tunable filter (AOTF, Gooch & Housego, UK). Fluorescence was detected by an emCCD (iXON Ultra 888; Andor, UK) adjusted to 129 nm pixel size. The z-focus was controlled by a commercial perfect focus system (Nikon, Düsseldorf, Germany). Acquisitions were controlled by Micro-Manager (Edelstein et al., 2014). Live cell experiments were performed on a customized heating stage at 25 °C.

sptPALM Data Acquisition

Measuring fast in vivo kinetics.

Living *E. coli* cells growing on EZRDM-agarose pads were placed into the heating chamber of the microscope and imaged with 77 and 33 Hz frame rate for 20.000 frames in HILO illumination mode (Tokunaga, Imamoto, & Sakata-Sogawa, 2008). During the image acquisition, samples were exposed to 561 nm illumination (800 W/cm2) continuously, while a 405 nm laser was pulsed every 20*th* frame with the intensity of 1-3 W/cm2. Before recording the sptPALM movie bright light snapshots of cells were taken.

Measuring in vivo binding kinetics of Cascades.

Cells on agarose pads were imaged with different time intervals (τ tl) (0.2 to 10 s) due the rather low photostability of fluorescent proteins (see Section 1.2). The integration time (τ int) was extended to (200 ms) to collect fluorescent signals only from stably DNA-bound Cascades (Figure VI.1 and Figure VI.3). To prevent Dendra2-T69A from rapid photobleaching, 561 nm laser intensity was set to 30 W/cm2. 405 nm laser was pulsed every 25*th* frame with the intensity of 1-3 W/cm2. Before sptPALM imaging, bright light snapshots were taken. To mimic the fluorescent signal coming from stably bound Cascade-DNA complexes, a chemically fixed *E. coli* BL21-AI strain, expressing LacY-Dendra2-T69A was imaged under the same conditions. With this control we can correct

for photobleaching effects of the measurements. of the microscope and imaged with 77 and 33 Hz frame rate for 20.000 frames in HILO illumination mode [97]. During the image acquisition, samples were exposed to 561 nm illumination (800 W/cm2) continuously, while a 405 nm laser was pulsed every 20*th* frame with the intensity of 1-3 W/cm2. Before recording the sptPALM movie bright light snapshots of cells were taken.
To mimic the fluorescent signal coming from stably bound Cascade-DNA complexes, a chemically fixed *E. coli* BL21-AI strain, expressing LacY-Dendra2 T69A was imaged under the same conditions. With this control we can correct for photobleaching effects of the measurements.

sptPALM Data Analysis

Measuring fast in vivo kinetics.
Cells were segmented manually from bright light images using the open source software Fiji (Schindelin et al., 2012) and stored as ROI areas in zip files. Nanometer precise localizations of diffusing fluorophores were obtained with the rapidSTORM 3.3 software (Wolter et al., 2012). Tracking was performed using custom software, developed in our lab (unpublished), written in C++. For further analysis, only trajectories longer than six steps and shorter than 20 steps were taken. Furthermore, highly autofluorescent cells and obvious noise (e.g. from inclusion bodies at cell poles) were manually discarded from the dataset. For all trajectories which passed the filtering, the average displacement between
adjacent frames (<JD>) was extracted and weighted by the number of steps. Weighted <JD> values were plotted in histogram plots using the software OriginPro 2017 (OriginLab Corporation).
Measuring in vivo binding kinetics of Cascades
In a first step, nanometer precise localizations of chemically fixed LacY-Dendra2-T69A proteins were obtained with the ThunderSTORM software (Ovesny, Krizek, Borkovec, Svindrych, & Hagen, 2014) . As chemically fixed LacY-Dendra2-T69A are immobile, and thus resemble stable Cascade-DNA interactions, parameters characterizing their PSF architecture, like the intensity, xy-dimensions, and uncertainty were used later in the filtering step to exclude mobile Cascade fractions.
Filtered localizations were then connected into trajectories using customized software written in C++. Here, the tracking software makes use of *a priori* knowledge about the fluorophore blinking and bleaching rates, which were obtained from a manual first analysis of the control strain measurements. Additionally, we provided an expected average displacement value for the immobile molecule. Given these *a priori* parameters,

tracking software can evaluate whether a tracked molecule is either static (immobile) or freely diffusing. For further analysis, only the immobile trajectories were taken into consideration. Trajectories coming from inclusion bodies and other artifacts were manually excluded from the dataset. For all trajectories which pass the filtering, trajectory lifetimes (number of frames between first and last localization in a trajectory), representing fluorescence lifetimes, were plotted in histograms, to which single exponential decay model was fitted: $PDF(t) = Ae\text{-}k\text{eff}(\tau\text{tl})t$, where *PDF(t)* is a probability density function, *A* is the amplitude at $t = 0$, *t* is the fluorescence lifetime and *keff(τtl)* is the fluorescence lifetime decay effective rate for a given τtl. *keff(τtl)* values for all tested crRNA variants were plotted in *keffτ*tl(τtl) graphs. Linear fits on plotted data were performed. For fitting, each data point was weighted by its Standard Error (SE) to put more significance on data with better statistics. Slope values of fits are equal to an effective rate (*keff*) for given crRNA variants. Binding times were extracted from K_{off} values (as K_{off}^{-1}) and plotted against the average target length for each sample. Considering that Cascades would bind to the best target available, only the length of the 24 targets with higher complementarity was taken into account.

Acknowledgements

This work was supported by the Max Planck Society (LR, UE), by SYNMIKRO (UE), SPP 2441 from the Deutsche Forschungsgemeinschaft (LR, UE), by the Fonds der Chemischen Industrie (UE) and by IMPRS-Mic (HME).

Supplementary Material

Supplementary Table VI.1. Amount of genomic targets for spacers with different complementarities.

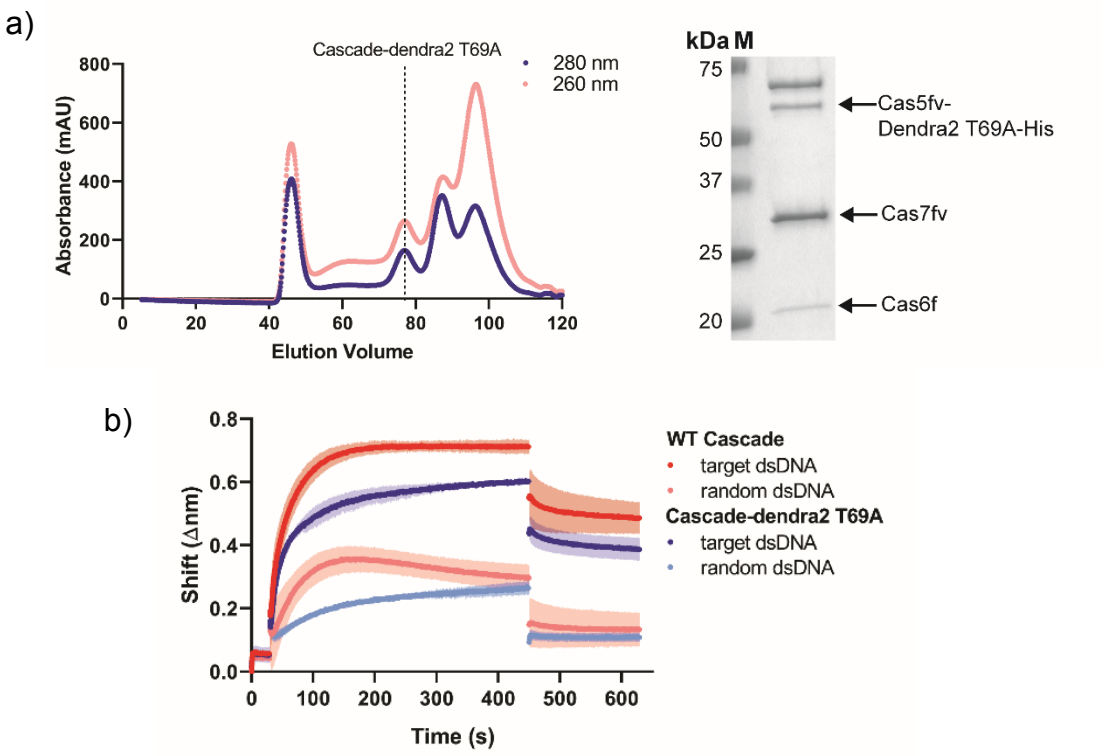
	crRNA	**Anti-REP**					
Target Complementarity (bp)	Anti- λ	12 bp	18 bp	25 bp	30 bp	32 bp 1 spacer	32 bp 3 spacers
>31	0	0	0	0	0	16	42
25-30	0	0	0	39	94	14	33
19-24	0	0	0	66	11	12	24
13-18	0	0	128	13	22	0	8
12	0	133	5	4	5	9	26
11	0	3	3	14	4	85	3
Average target length (bp)	**0**	**9**	**15**	**25**	**28**	**30.33**	**32**

Supplementary Table VI.2. Plasmids used in this study

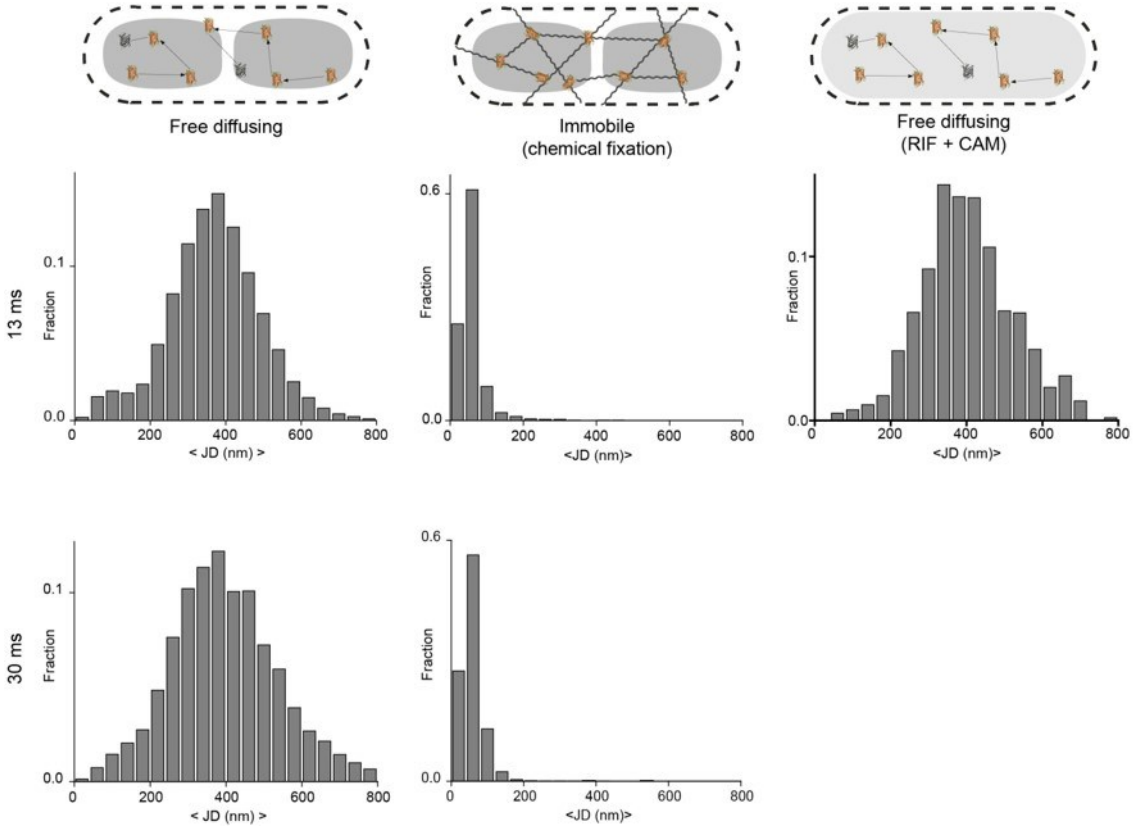
p(n)	Plasmid Name	Description	Source/ Reference
1	pCas5fv	pCDF-Duet1 + Cas5fv-Dendra T69A	This work
2	pCRISPR λ-Cas2-3-Cas5fv	pCDF-Duet1 + repeat-spacer-repeat array + Cas2-3 + Cas5fv-Dendra2-T69A. Spacer is complementary to a 32 bp region of the gene E from phage lambda. Repeat sequences from *S. putrefaciens* CN-32 CRISPR array. No genomic targets	This work
3	pCRISPR λ-Cas2-3 HD-Cas5fv	p(2) carrying point mutations for alanine replacement of the HD nuclease domain of Cas2-3	This work
4	pCRISPR REP	p(3) with the spacer sequence complementary to 16 repetitive extragenic palyndromic (REP) sequences on the *E. coli* genome	This work
5	pCRISPR REP 32	p(3) with a multiplexed CRISPR array carrying 3 spacers against 43 complete targets on the REP sequences on the *E. coli* genome	This work
6	pCRISPR REP 30	p(3) with a multiplexed CRISPR array carrying 2 spacers against 76 30nt-complementary targets in the REP sequences of *E. coli*	This work
7	pCRISPR REP 24	p(3) with a multiplexed CRISPR array carrying 2 spacers against 100 24nt-complementary targets in the REP sequences of *E. coli*	This work
8	pCRISPR REP 18	p(3) with a multiplexed CRISPR array carrying 2 spacers against 108 18nt-complementary targets in the REP sequences of *E. coli*	This work
9	pCRISPR REP 12	p(3) with a multiplexed CRISPR array carrying 2 spacers against 133 12nt-complementary targets in the REP sequences of *E. coli*	This work
10	pCas7-6	pRSFDuet-1 + Cas6f + Cas7fv	This work
11	pCRISPR	pCDFDuet-1 + repeat-spacer-repeat array. Spacer against gene E from phage lambda. Repeat sequences from *S. putrefaciens* CN-32 CRISPR array	Gleditzsch et al., 2016
12	pCas	pRSFDuet-1 + Cas7fv-his + Cas5fv + Cas6f	Gleditzsch et al., 2016
13	pCas5fvΔAH	p(1) expressing a truncated version of Cas5fv-Dendra2-T69A. Amino acids 121-259 replaced by a GGSGGS linker.	This work
14	pCas6f-Dendra2-T69A	pCDFDuet-1 + Cas6f-Dendra2-T69A	This work
15	pCas6f	pRSFDuet-1 + Cas6f	This work
16	pCRISPR λ-Cas5fv	p(2) lacking Cas2-3	This work
17	pRSETb-Dendra2-T69A	Dendra2-T69A	Berardozzi et al., 2016
18	pRha-LacY-Dendra2-T69A	LacY-Dendra2-T69A carboxy-terminal fusion under a rhamnose inducible promoter	This work
19	pCRISPR λ-Cas5fv-his	P(2) with a C-terminal his-tag on Dendra2-T69A	This work
20	pCascade	pRSFDuet-1 + Cas2-3+ Cas7fv+ Cas5fv + Cas6f	Gleditzsch et al., 2016
21	pCascade HD	pRSFDuet-1 + Cas2-3 HD mutant + Cas7fv+ Cas5fv + Cas6f	Gleditzsch et al., 2016

Supplementary Table VI.3. Protospacer sequences on CRISPR arrays of plasmids in Suppl. Table VI.1. Repeat sequence: GTTCACCGCCGCACAGGCGGCTTAGAAA.

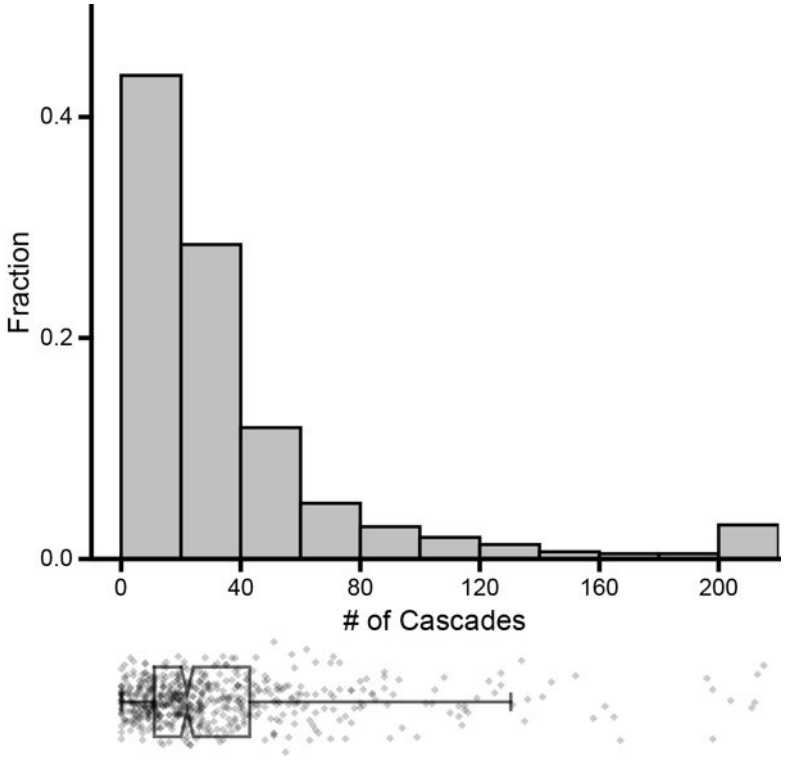
p(n)	Protospacer 1	Protospacer 2	Protospacer 3
2-3	GGCGGCACGGAGTGGAG CAAGCGTGACAAGTC	-	-
4	GGATGCGGCGTAAACGCCT TATCCGGCCTACG	-	-
5	GGATGCGGCGTAAACGCCT TATCCGGCCTACG	GGATGCGGCGTAAACGCC TTATCCGGCCTACA	GGATGCGGCGTGAACG CCTTATCCGGCCTACG
6	GGATGCGGCGTAAACGCCT TATCCGGCCTAAT	GGATGCGGCGTGAACGC CTTATCCGGCCTAGC	
7	GGATGCGGCGTAAACGCCT TATCCGAAACGTT	GGATGCGGCGTGAACGC CTTATCCCGATAGGT	
8	GGATGCGGCGTAAACGCCC CTGTTCGTGTCGT	GGATGCGGCGTGAACGC CATCGGTAAATAGCC	
9	GGATGCGGCGTACCGCGG AGCCATGTGCCATA	GGATGCGGCGTGCTCGTC TGCGGAGCACTCTG	



Supplementary Figure VI.1. Formation and activity of Cascade-dendra 2 T69A fusion. (A) (Left) Chromatogram of the size exclusion purification of Cascade-Dendra2-T69A. The composition of the peak fractions was analysed by SDS-PAGE, identifying the Cascade-dendra2-T69A containing fraction (right). **(B)** BioLayer Interferometry (BLI) of the interaction of wild type and fluorescent I-Fv Cascades with complementary and non-complementary targets.



Supplementary Figure VI.2. Dynamics of free diffusing and immobile Dendra2-T69A. Jump distance (JD) histograms of the tracking of Dendra2-T69A expressed alone in *E. coli* BL21-AI with no treatment (left), chemical fixation (middle) and rifampicin plus chloramphenicol treatment (right). Top row was obtained from imaging at an integration time of 13 ms, while bottom row was imaged at 30 ms. Schemes represent the expected behaviour of the proteins in the cell for each treatment. First two distributions are used as references of free diffusing and immobile particles, respectively, in figures 1 and 2. Third histogram is shown as a control for the effect of antibiotics on the dynamics of Dendra2-T69A.



Supplementary Figure VI.3. Quantification of total fluorescent particles in *E. coli* BL21-AI cells expressing Cascade-Dendra2-T69A fusions. (Above) Cascade amount per cell represented as a histogram. (Below) Box and scatter plots represent the individual data points, with the same x-axis. The box plot shows the range of the second and third quartile, while whiskers mark 5-95% of all the data. The expected number of Cascades is 24 complexes per cell (615 cells analyzed).

References

Alvarado, A., Turkowyd, B., Endesfelder, U., & Ringgaard, S. (2018). Differential diffusion rates of distinct protein states in a diffusion-and-capture mechanism drive protein gradient formation *submitted*.

Bakshi, S., Siryaporn, A., Goulian, M., & Weisshaar, J. C. (2012). Superresolution imaging of ribosomes and RNA polymerase in live Escherichia coli cells. *Mol Microbiol, 85*(1), 21-38.

Beattie, T. R., Kapadia, N., Nicolas, E., Uphoff, S., Wollman, A. J., Leake, M. C., et al. (2017). Frequent exchange of the DNA polymerase during bacterial chromosome replication. *Elife, 6*.

Berardozzi, R., Adam, V., Martins, A., & Bourgeois, D. (2016). Arginine 66 Controls Dark-State Formation in Green-to-Red Photoconvertible Fluorescent Proteins. *J Am Chem Soc, 138*(2), 558-565.

Blosser, T. R., Loeff, L., Westra, E. R., Vlot, M., Kunne, T., Sobota, M., et al. (2015). Two distinct DNA binding modes guide dual roles of a CRISPR-Cas protein complex. *Mol Cell, 58*(1), 60-70.

Brouns, S. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J., Snijders, A. P., et al. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science, 321*(5891), 960-964.

Campbell, E. A., Korzheva, N., Mustaev, A., Murakami, K., Nair, S., Goldfarb, A., et al. (2001). Structural mechanism for rifampicin inhibition of bacterial rna polymerase. *Cell, 104*(6), 901-912.

Chen, H., Shiroguchi, K., Ge, H., & Xie, X. S. (2015). Genome-wide study of mRNA degradation and transcript elongation in Escherichia coli. *Mol Syst Biol, 11*(1), 781.

Chowdhury, S., Carter, J., Rollins, M. F., Golden, S. M., Jackson, R. N., Hoffmann, C., et al. (2017). Structure Reveals Mechanisms of Viral Suppressors that Intercept a CRISPR RNA-Guided Surveillance Complex. *Cell, 169*(1), 47-57 e11.

Dagdas, Y. S., Chen, J. S., Sternberg, S. H., Doudna, J. A., & Yildiz, A. (2017). A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9. *Sci. Adv., 3*(8), eaao0027.

Ding, Y., Manzo, C., Fulcrand, G., Leng, F., Dunlap, D., & Finzi, L. (2014). DNA supercoiling: a regulatory signal for the lambda repressor. *Proc Natl Acad Sci U S A, 111*(43), 15402-15407.

Dillard, K. E., Brown, M. W., Johnson, N. V., Xiao, Y., Dolan, A., Hernandez, E., et al. (2018). Assembly and Translocation of a CRISPR-Cas Primed Acquisition Complex. *Cell*.

Dwarakanath, S., Brenzinger, S., Gleditzsch, D., Plagens, A., Klingl, A., Thormann, K., et al. (2015). Interference activity of a minimal Type I CRISPR-Cas system from Shewanella putrefaciens. *Nucleic Acids Res, 43*(18), 8913-8923.

Edelstein, A. D., Tsuchida, M. A., Amodaj, N., Pinkard, H., Vale, R. D., & Stuurman, N. (2014). Advanced methods of microscope control using muManager software. *J Biol Methods, 1*(2).

Gahlmann, A., & Moerner, W. E. (2014). Exploring bacterial cell biology with single-molecule tracking and super-resolution imaging. *Nat Rev Microbiol, 12*(1), 9-22.

Gebhardt, J. C., Suter, D. M., Roy, R., Zhao, Z. W., Chapman, A. R., Basu, S., et al. (2013). Single-molecule imaging of transcription factor binding to DNA in live mammalian cells. *Nat Methods, 10*(5), 421-426.
Gleditzsch, D., Muller-Esparza, H., Pausch, P., Sharma, K., Dwarakanath, S., Urlaub, H., et al. (2016). Modulating the Cascade architecture of a minimal Type I-F CRISPR-Cas system. *Nucleic Acids Res, 44*(12), 5872-5882.
Globyte, V., Lee, S. H., Bae, T., Kim, J. S., & Joo, C. (2019). CRISPR/Cas9 searches for a protospacer adjacent motif by lateral diffusion. *EMBO J, 38*(4).
Hammar, P., Wallden, M., Fange, D., Persson, F., Baltekin, O., Ullman, G., et al. (2014). Direct measurement of transcription factor dissociation excludes a simple operator occupancy model for gene regulation. *Nat Genet, 46*(4), 405-408.
Hayes, R. P., Xiao, Y., Ding, F., van Erp, P. B., Rajashankar, K., Bailey, S., et al. (2016). Structural basis for promiscuous PAM recognition in type I-E Cascade from E. coli. *Nature, 530*(7591), 499-503.
Hesselbach, B. A., & Nakada, D. (1977). "Host shutoff" function of bacteriophage T7: involvement of T7 gene 2 and gene 0.7 in the inactivation of Escherichia coli RNA polymerase. *J Virol, 24*(3), 736-745.
Heussler, G. E., Miller, J. L., Price, C. E., Collins, A. J., & O'Toole, G. A. (2016). Requirements for Pseudomonas aeruginosa Type I-F CRISPR-Cas Adaptation Determined Using a Biofilm Enrichment Assay. *J Bacteriol, 198*(22), 3080-3090.
Ivancic-Bace, I., Cass, S. D., Wearne, S. J., & Bolt, E. L. (2015). Different genome stability proteins underpin primed and naive adaptation in E. coli CRISPR-Cas immunity. *Nucleic Acids Res, 43*(22), 10821-10830.
Jones, D. L., Leroy, P., Unoson, C., Fange, D., Curic, V., Lawson, M. J., et al. (2017). Kinetics of dCas9 target search in Escherichia coli. *Science, 357*(6358), 1420-1424.
Josephs, E. A., Kocak, D. D., Fitzgibbon, C. J., McMenemy, J., Gersbach, C. A., & Marszalek, P. E. (2015). Structure and specificity of the RNA-guided endonuclease Cas9 during DNA interrogation, target binding and cleavage. *Nucleic Acids Res, 43*(18), 8924-8941.
Killelea, T., Hawkins, M., Howard, J. L., McGlynn, P., & Bolt, E. L. (2019). DNA replication roadblocks caused by Cascade interference complexes are alleviated by RecG DNA repair helicase. *RNA Biol, 16*(4), 543-548.
Knight, S. C., Xie, L., Deng, W., Guglielmi, B., Witkowsky, L. B., Bosanac, L., et al. (2015). Dynamics of CRISPR-Cas9 genome interrogation in living cells. *Science, 350*(6262), 823-826.
Koonin, E. V., Makarova, K. S., & Zhang, F. (2017). Diversity, classification and evolution of CRISPR-Cas systems. *Curr Opin Microbiol, 37*, 67-78.
Krivoy, A., Rutkauskas, M., Kuznedelov, K., Musharova, O., Rouillon, C., Severinov, K., et al. (2018). 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*(8), 4087-4098.
Kuderova, A., Nanak, E., Truksa, M., & Brzobohaty, B. (1999). Use of rifampicin in T7 RNA polymerase-driven expression of a plant enzyme: rifampicin improves yield and assembly. *Protein Expr Purif, 16*(3), 405-409.
Kumar, M., Mommer, M. S., & Sourjik, V. (2010). Mobility of cytoplasmic, membrane, and DNA-binding proteins in Escherichia coli. *Biophys J, 98*(4), 552-559.
Liang, S. T., Ehrenberg, M., Dennis, P., & Bremer, H. (1999). Decay of rplN and lacZ mRNA in Escherichia coli. *J Mol Biol, 288*(4), 521-538.
Lopez, P. J., Marchand, I., Yarchuk, O., & Dreyfus, M. (1998). Translation inhibitors stabilize Escherichia coli mRNAs independently of ribosome protection. *Proc Natl Acad Sci U S A, 95*(11), 6067-6072.
Manley, S., Gillette, J. M., Patterson, G. H., Shroff, H., Hess, H. F., Betzig, E., et al. (2008). High-density mapping of single-molecule trajectories with photoactivated localization microscopy. *Nat Methods, 5*(2), 155-157.
Moolman, M. C., Krishnan, S. T., Kerssemakers, J. W., van den Berg, A., Tulinski, P., Depken, M., et al. (2014). Slow unloading leads to DNA-bound beta2-sliding clamp accumulation in live Escherichia coli cells. *Nat Commun, 5*, 5820.
Ovesny, M., Krizek, P., Borkovec, J., Svindrych, Z., & Hagen, G. M. (2014). ThunderSTORM: a comprehensive ImageJ plug-in for PALM and STORM data analysis and super-resolution imaging. *Bioinformatics, 30*(16), 2389-2390.
Pausch, P., Muller-Esparza, H., Gleditzsch, D., Altegoer, F., Randau, L., & Bange, G. (2017). Structural Variation of Type I-F CRISPR RNA Guided DNA Surveillance. *Mol Cell, 67*(4), 622-632 e624.
Redding, S., Sternberg, S. H., Marshall, M., Gibb, B., Bhat, P., Guegler, C. K., et al. (2015). Surveillance and Processing of Foreign DNA by the Escherichia coli CRISPR-Cas System. *Cell, 163*(4), 854-865.
Rollins, M. F., Chowdhury, S., Carter, J., Golden, S. M., Miettinen, H. M., Santiago-Frangos, A., et al. (2019). Structure Reveals a Mechanism of CRISPR-RNA-Guided Nuclease Recruitment and Anti-CRISPR Viral Mimicry. *Mol Cell, 74*(1), 132-142 e135.
Rollins, M. F., Chowdhury, S., Carter, J., Golden, S. M., Wilkinson, R. A., Bondy-Denomy, J., et al. (2017). Cas1 and the Csy complex are opposing regulators of Cas2/3 nuclease activity. *Proc Natl Acad Sci U S A, 114*(26), E5113-E5121.
Rollins, M. F., Schuman, J. T., Paulus, K., Bukhari, H. S., & Wiedenheft, B. (2015). Mechanism of foreign DNA recognition by a CRISPR RNA-guided surveillance complex from Pseudomonas aeruginosa. *Nucleic Acids Res, 43*(4), 2216-2222.
Rutkauskas, M., Sinkunas, T., Songailiene, I., Tikhomirova, M. S., Siksnys, V., & Seidel, R. (2015). Directional R-Loop Formation by the CRISPR-Cas Surveillance Complex Cascade Provides Efficient Off-Target Site Rejection. *Cell Rep*.
Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat Methods, 9*(7), 676-682.
Scott, M., Gunderson, C. W., Mateescu, E. M., Zhang, Z., & Hwa, T. (2010). Interdependence of cell growth and gene expression: origins and consequences. *Science, 330*(6007), 1099-1102.
Semenova, E., Jore, M. M., Datsenko, K. A., Semenova, A., Westra, E. R., Wanner, B., et al. (2011). Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc Natl Acad Sci U S A, 108*(25), 10098-10103.
Singh, D., Mallon, J., Poddar, A., Wang, Y., Tippana, R., Yang, O., et al. (2018). Real-time observation of DNA target interrogation and product release by the RNA-guided endonuclease CRISPR Cpf1 (Cas12a). *Proc Natl Acad Sci U S A, 115*(21), 5444-5449.
Singh, D., Sternberg, S. H., Fei, J., Doudna, J. A., & Ha, T. (2016). Real-time observation of DNA recognition and rejection by the RNA-guided endonuclease Cas9. *Nat Commun, 7*, 12778.
Sippel, A., & Hartmann, G. (1968). Mode of action of rafamycin on the RNA polymerase reaction. *Biochim Biophys Acta, 157*(1), 218-219.
So, L. H., Ghosh, A., Zong, C., Sepulveda, L. A., Segev, R., & Golding, I. (2011). General properties of transcriptional time series in Escherichia coli. *Nat Genet, 43*(6), 554-560.

Stella, S., Mesa, P., Thomsen, J., Paul, B., Alcon, P., Jensen, S. B., et al. (2018). Conformational Activation Promotes CRISPR-Cas12a Catalysis and Resetting of the Endonuclease Activity. *Cell, 175*(7), 1856-1871 e1821.
Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C., & Doudna, J. A. (2014). DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature, 507*(7490), 62-67.
Stracy, M., Uphoff, S., Garza de Leon, F., & Kapanidis, A. N. (2014). In vivo single-molecule imaging of bacterial DNA replication, transcription, and repair. *FEBS Lett, 588*(19), 3585-3594.
Szczelkun, M. D., Tikhomirova, M. S., Sinkunas, T., Gasiunas, G., Karvelis, T., Pschera, P., et al. (2014). Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proc Natl Acad Sci U S A, 111*(27), 9798-9803.
Tobes, R., & Pareja, E. (2006). Bacterial repetitive extragenic palindromic sequences are DNA targets for Insertion Sequence elements. *BMC Genomics, 7*, 62.
Tobes, R., & Ramos, J. L. (2005). REP code: defining bacterial identity in extragenic space. *Environ Microbiol, 7*(2), 225-228.
Tokunaga, M., Imamoto, N., & Sakata-Sogawa, K. (2008). Highly inclined thin illumination enables clear single-molecule imaging in cells. *Nat Methods, 5*(2), 159-161.
Turkowyd, B., Balinovic, A., Virant, D., Carnero, H. G. G., Caldana, F., Endesfelder, M., et al. (2017). A General Mechanism of Photoconversion of Green-to-Red Fluorescent Proteins Based on Blue and Infrared Light Reduces Phototoxicity in Live-Cell Single-Molecule Imaging. *Angew Chem Int Ed Engl, 56*(38), 11634-11639.
Turkowyd, B., Virant, D., & Endesfelder, U. (2016). From single molecules to life: microscopy at the nanoscale. *Anal Bioanal Chem, 408*(25), 6885-6911.
van Beljouw, S. P. B., van der Els, S., Martens, K. J. A., Kleerebezem, M., Bron, P. A., & Hohlbein, J. (2019). Evaluating single-particle tracking by photo-activation localization microscopy (sptPALM) in Lactococcus lactis. *Phys Biol, 16*(3), 035001.
Vink, J. N. A., Martens K.J.A, Vlot, M., McKenzie, R. E., Almendros, C., Estrada Bonilla, B., et al. (2019). Direct visualization of native CRISPR target search in live bacteria reveals Cascade DNA surveillance mechanism. *bioRxiv*. doi:10.1101/589119.
Wiedenheft, B., Lander, G. C., Zhou, K., Jore, M. M., Brouns, S. J. J., van der Oost, J., et al. (2011). Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature, 477*(7365), 486-489.
Wolter, S., Loschberger, A., Holm, T., Aufmkolk, S., Dabauvalle, M. C., van de Linde, S., et al. (2012). rapidSTORM: accurate, fast open-source software for localization microscopy. *Nat Methods, 9*(11), 1040-1041.
Xiao, Y., Luo, M., Dolan, A. E., Liao, M., & Ke, A. (2018). Structure basis for RNA-guided DNA degradation by Cascade and Cas3. *Science, 361*(6397).
Xiao, Y., Luo, M., Hayes, R. P., Kim, J., Ng, S., Ding, F., et al. (2017). Structure Basis for Directional R-loop Formation and Substrate Handover Mechanisms in Type I CRISPR-Cas System. *Cell, 170*(1), 48-60 e11.
Xue, C., Whitis, N. R., & Sashital, D. G. (2016). Conformational Control of Cascade Interference and Priming Activities in CRISPR Immunity. *Mol Cell, 64*(4), 826-834.
Xue, C., Zhu, Y., Zhang, X., Shin, Y. K., & Sashital, D. G. (2017). Real-Time Observation of Target Search by the CRISPR Surveillance Complex Cascade. *Cell Rep, 21*(13), 3717-3727.
Yang, M., Peng, S., Sun, R., Lin, J., Wang, N., & Chen, C. (2018). The Conformational Dynamics of Cas9 Governing DNA Cleavage Are Revealed by Single-Molecule FRET. *Cell Rep, 22*(2), 372-382.
Zeng, Y., Cui, Y., Zhang, Y., Zhang, Y., Liang, M., Chen, H., et al. (2018). The initiation, propagation and dynamics of CRISPR-SpyCas9 R-loop complex. *Nucleic Acids Res, 46*(1), 350-361.
Zusman, D. R., Carbonell, A., & Haga, J. Y. (1973). Nucleoid condensation and cell division in Escherichia coli MX74T2 ts52 after inhibition of protein synthesis. *J Bacteriol, 115*(3), 1167-1178.

Chapter VII: Discussion

Discussion

How does a minimal Cascade achieve interference?

Type I CRISPR-Cas systems achieve interference through Cascades (Makarova et al., 2011). Although the individual components (Cas proteins and crRNA) might differ in sequence, structure and recognition motifs, the functionality of complete systems remains intact, fending off the invasion of foreign genetic material (Mohanraju et al., 2016).

A first step in interference is PAM recognition, as this process allows Cascades to differentiate between self and non-self DNA (Brouns et al., 2008; Gleditzsch et al., 2019; Hayes et al., 2016). For Type I-A, B, C, E and F, this task is performed by the large subunit Cas8. For type I-D, this function resides on Cas10, a protein also found in Type III systems (Makarova et al., 2015; Plagens, Richter, Charpentier, & Randau, 2015). Strikingly, the Type I-F variant system lacks a homologue to the large subunit. Despite the absence of this key protein, the system found in *S. putrefaciens* CN-32 was shown to interfere with conjugative plasmids in a PAM-dependent manner (Dwarakanath et al., 2015).

This study provides further proof on the interference ability of this minimal complex, in a PAM-, crRNA- and Cas2-3 dependent manner, without the need of other players from the host. This system is able to affect both plasmid and viral propagation through targeting by a minimal Cascade formed only by Cas6f, Cas5fv and six copies of Cas7fv that protect a crRNA.

Interestingly, the number of Cas7fv subunits present in a complex can be modified by altering the spacer length in the crRNA. As Type I-Fv Cas7fv covers 6 nt intervals of the crRNA, increasing or decreasing the length of the protospacer on multiples of six allows to generate smaller and larger complexes. Although stable variants can be formed, activity is maintained only for complexes with 6 to 7 Cas7fv subunits. Therefore, there is a limitation in targeting, explained either by the incapacity of the complex to bind the target stably, or to recruit the Cas2-3 nuclease. Despite this, the synthetic variants hold potential for future applications where protection of small RNAs is needed. This modulation has also been shown for Type I-E complexes, where spacers increased by 6 or 12 nt, as well as decreased by 3 nt could still perform interference (Kuznedelov et al., 2016; Luo et al., 2016).

The formation of a variant complex that is still functional is explained by the diversification of two out of the three main components: the backbone protein Cas7fv, and the 5' crRNA-capping protein, Cas5fv. Each shows no sequence similarity to any Cas protein reported so far, but biochemical and structural results allowed to confirm their classification (Dwarakanath et al., 2015).

As elucidated by its crystal structure, Cas5fv acts as the Swiss army knife of this minimal complex. Similar to other Cas5 proteins, it is able to recognize and bind the 8 nt 5'-handle of the crRNA (Dwarakanath et al., 2015). Furthermore, it has an alpha helical domain that protrudes from the body of the complex, occupying an equivalent position to the large subunits in Type I-F and I-E complexes. This additional domain consists of six alpha helixes that clamp the dsDNA against the body of the complex, allowing for the read out of the GG-PAM by interactions with the major groove of the target DNA. Therefore, this addition to Cas5fv compensates the lack of a large subunit (Pausch et al., 2017).

This type of PAM recognition differentiates from the Type I-F and Type I-E systems, where crystal structures point towards minor groove interactions (Hayes et al., 2016; Rollins et al., 2019). Major groove interactions have been described for Cas9 from *Streptococcus pyogenes* through conserved arginine residues, in addition to minor groove stabilization of the non-target strand (Anders, Niewoehner, Duerst, & Jinek, 2014). Recognition of the major groove has been reported as more stable, since they involve the formation of hydrogen bonds, while minor groove-interacting proteins rely mostly on DNA shape (Rohs et al., 2009). This difference explains the promiscuity seen for the Type I-E Cascade, a complex that can recognize up to 5 different PAMs with similar efficiency (Hayes et al., 2016). Therefore, it is expected that minor groove-PAM recognition is more stringent, preventing self-targeting. On the downside, this mechanism will depend more on efficient adaptation, as PAM mutations would easily evade CRISPR-Cas targeting.

To study the specific residues involved in the interaction, alanine scanning was performed. It was not possible to abolish PAM recognition by single mutations. As a first approach, we looked for conservation of residues described as relevant in Type I-E Cascade. Here, a lysine finger, a glutamine wedge and a glycine loop identify the AAG PAM and open the first position of the protospacer (Hayes et al., 2016). Mutation of the only glutamine present in the vicinity of the GG PAM, Q113, had no effect on the interference activity by the complex. On the other hand, three residues, T251, K253 and

D254, all located in the alpha helical domain, seem to act in a concerted manner to recognize the PAM. The threonine residue is in close proximity to one cytosine residue of the motif, suggesting that both strands are relevant in the process. The effect of K253 coincides with the description for Type I-F Cascade, where a lysine residue is responsible for dsDNA opening (Rollins et al., 2019). Further tests could involve the mutation of R248, as arginines have been shown to be important for major groove interaction (Luscombe, Laskowski, & Thornton, 2001). In addition, interference assays with Cas5fv mutants targeting different PAMs could be performed to see if the introduced changes in the pocket could lead to the recognition of other motifs.

The complete removal of the alpha helical domain still allows for Cascade formation, but completely abolishes the interference capacity of the complex. Furthermore, as seen by single-particle tracking, the removal of the AH domain results in faster Cascades, resembling free-diffusing proteins. This might indicate that Cascade depends on the protruding domain to grab to the DNA and reel through it in the search for targets. This mechanism would be comparable to DNA scanning by the hook domain of Cas8 of the Type I-F complex (Chowdhury et al., 2017; Rollins et al., 2019). The structural and dynamical data then point towards a one-dimensional DNA reading, although three dimensional jumps should not be discarded. In this regard, *in vitro* studies could help elucidate the scanning method of Cascade with higher precision. Biolayer interferometry assays with a PAM mutant and non-complementary dsDNA showed non-ideal binding. This could be interpreted as transient interactions with Cascades constantly reeling through and off the oligonucleotide. Nevertheless, further experiments are needed to be able to interpret these results properly. As an example, it could be possible to use BLI to test the binding of the alpha-helical mutant complex, but other methods, such as single-molecule FRET or DNA curtain assays, would give clearer insight into the DNA readout by the minimal Cascade (Redding et al., 2015; Xue, Zhu, Zhang, Shin, & Sashital, 2017).

Upon PAM recognition, CRISPR-Cas complexes open the dsDNA, relying on sequence complementarity, protein interactions and dsDNA supercoiling to generate a stable R-loop (Chowdhury et al., 2017; Hayes et al., 2016; Rollins et al., 2019; Rutkauskas et al., 2015; Westra et al., 2012; Xiao et al., 2017). In the Type I-F system, Cas8f, Cas5f and the adjacent Cas7f subunits contribute to the process. Upon dsDNA binding, Cas8f, Cas5f and the two PAM-proximal Cas7f subunits form a positively charged channel that helps maintain separation of the non-target strand. In addition, the Cas7f backbone elongates by approximately 18 Å. Both changes allow for an alpha helical bundle of Cas8f to rotate 180º, locking the complex to its target. This conformational change is

expected to recruit the Cas2-3 nuclease (Rollins et al., 2019). The rotation of Cas8f has only been described in the presence of a fully complementary dsDNA target, as structures of the complex bound to a partial R-loop, as the one used for Type I-Fv crystallization, exhibit only the elongation of the backbone and a smaller (10 to 16 Å) movement of the hook domain of Cas8f to close the PAM recognition pocket (Chowdhury et al., 2017).

If Cas5fv would act in a similar manner to Cas8f, it would be necessary that its AH domain would act both as the hook domain (for PAM recognition) and as the alpha helical Cas8f bundle. A more extreme movement of the AH domain, that already shifts 12.5 Å upon DNA binding, would require either the release of the PAM or high DNA bending. Both cases seem unlikely, therefore the conformational change of the Type I-Fv Cascade that leads to the lock state must be more modest, or rely on other components and interactions. In order to elucidate this, we are currently performing HDX experiments with Cascade and Cas2-3 to gain insights into the recruitment of the nuclease and to map the interaction surface.

Taking the strong differences between I-F complexes bound to a fully complementary dsDNA or a semi-open R-loop into consideration, it would be interesting to determine the crystal structure of the minimal complex bound to a canonical target. This option was initially discarded, as it has not been possible to observe this kind of binding through EMSAs, except in the presence of short mismatches between the strands of DNA. On the other hand, BLI shows that the minimal Cascade is able to interact with fully complementary dsDNA, albeit with lower than expected affinity, and that mismatches help the binding. Therefore, it should be possible to obtain dsDNA-bound Cascades for structural analyses.

Despite the differences on PAM recognition and target-induced locking, the stabilization of the non-target strand happens on a similar manner for Type I-Fv and other Type I complexes. The minimal complex has a charged vise formed by Cas5fv and Cas7fv residues, which guide the non-target strand away from its complement strand, aiding the formation of the R-loop. Structural specificities of Cas7fv make this possible, giving further insight into the replacement of the large subunit. Although the RNA recognition motif (RRM) and the thumb domain of Cas7fv, used for target strand recognition, are structurally conserved, the protein has also a wrist helix that locates to the concave face when assembled into the complex. The helices of each Cas7fv interact with each other, forming a trench route that can interact with the non-target strand. Upon mutation of the

wrist helix, Cascade is not able to form. It is possible that the removal of the trench route, leaving an empty space on the “belly” of the complex, leads to the collapse of the structure. Therefore, this variation of the Cas7 structure also fulfils several roles in order to allow for the minimal complex to be fully functional.

While the crystal structure of the complex helped elucidating the main characteristics of the interaction with a target, there are still open questions about the overall interference process. In the first place, a precise determination of the seed sequence is missing. Here, we used SPT-PALM to determine the binding times of Cascade to targets with different complementarities. We show that crRNA-DNA complementarity plays an important role in DNA-complex formation, since higher binding times correlated to longer stretches of crRNA-DNA hybridisation. The fact that this relation is not linear suggests that there might be a threshold for establishing a steady R-loop. In the future, we expect to perform SPT-PALM to also determine the bound times to seed mutants and without Cas2-3. These experiments, in addition to efficiency of transformation (EOT) assays and BLI, would be key to elucidate the stringency of the complex. Altogether, they would allow to propose a mechanism for abortive engagement, as it is not known how the minimal complex deals with the counterproductive binding of wrong sequences.

Furthermore, the regulation of the expression of the system has only been partially uncovered. Previous work on Type I-Fv system showed a negative regulation of the Cas operon by the histone-like nucleoid structuring (H-NS) protein (Dwarakanath et al., 2015). Furthermore, it was shown for Type I-E systems that this repression can be counteracted by the leucine response transcription factor LeuO, although the specific signal behind LeuO induction or H-NS down-regulation are unknown (Medina-Aparicio et al., 2011; Pul et al., 2010; Westra et al., 2010). For other systems, immunity has been shown to be regulated by environmental queues. Signals like quorum sensing, temperature, glucose levels and membrane stress have been described as part of the regulatory network of several CRISPR-Cas systems (Hoyland-Kroghsbo, Munoz, & Bassler, 2018; Hoyland-Kroghsbo et al., 2017; Patterson, Chang, Taylor, & Fineran, 2015; Patterson et al., 2016; Yosef, Moran, Kiro, Edgarm, & Qimron, 2011; Young et al., 2012). As these signals have pleiotropic effects, the exact way by which they interact with CRISPR loci is unknown.

Keeping an active system is a double-edged sword. On one side, hosts should be ready to interfere with fast-replicating elements, as any delay in expression or complex formation would be lethal. On the other hand, overly active CRISPR-Cas systems have

a higher chance of acquiring self-targeting spacers or eliciting non-specific interference. Therefore, it is expected that each organism tailors the regulation of immunity based on the threats it is exposed to. Further studies are needed to uncover the whole regulatory pathway of the minimal system in *Shewanella putrefaciens* CN-32. In the future, we expect to obtain phages capable of targeting this strain, as this would allow us to study the infection and defence process in a biologically relevant context.

Finally, the timing of the targeting has not been studied for the minimal system. In this study, SPT-PALM revealed that effector complexes do not have a specific sub-cellular localization. Therefore, a faster target encounter due to a strategic positioning may be ruled out. Furthermore, our data describes how Cascades with targets in the cell exhibit two main populations: one target-bound and the other scanning for targets. The resolution of the technique did not discern complexes pausing for PAM recognition.

Efficient and fast PAM discrimination improves the chances of Cascades to successfully perform interference. But other factors, such as the amount of targeting complexes and the sequence and location of the protospacer also play a role in the outcome of the interaction. It has been described that Cas9 proteins carrying spacers targeting the early-injected genomic region of a phage are better at eliciting a robust immune response. Accordingly, there is a higher rate of acquisition from these sections (Modell, Jiang, & Marraffini, 2017). This suggests that interference relies on target degradation upon viral entry, pointing towards the need for fast scanning and recognition.

Interestingly, recent studies propose that CRISPR-Cas immunity does not necessarily occur by total degradation of the invading element, but rather by eliciting cell dormancy or programmed cell death (Meeske, Nakandakari-Higa, & Marraffini, 2019; Watson et al., 2019). In these cases, the timing of recognition would not be a problem as long as it happens faster than the replication or insertion cycle of the invader. These mechanisms unveil a new level of population defence, where individuals prevent the spreading of the infection. It is still uncertain how widespread these mechanisms are, and what led to their selection. Nevertheless, it would be interesting to see how the Type I-Fv system fits into these models, for which the study on its natural host and environment would be needed.

What is the evolutionary pressure behind Cascade minimization?

It was possible to establish that the Type I-Fv complex is capable of performing interference, despite the diversification of some of its components (Dwarakanath et al., 2015; Gleditzsch et al., 2016; Pausch et al., 2017). Nevertheless, the reasons behind

the diversification are still unclear. This is of importance, as understanding what led to the stark changes of this and other minimal systems could help mapping the evolutionary route of CRISPR-Cas systems and unveiling the main evolutionary pressures that this and other defence mechanisms are subjected to.

In biotic populations, the main drive behind molecular evolution is antagonistic co-evolution (Stenseth & Smith, 1984). For microorganisms, the antagonism is represented by continuous viral predation, which leads to an accelerated selection of traits that improve fitness. Yet, as hosts adapt and find ways to avoid infection, phages counter-adapt in order to overcome the bacterial efforts (Paterson et al., 2010). These cycles result in the diversification of the defence mechanisms, as seen for CRISPR-Cas, restriction-modification and abortive infection systems (Koonin & Makarova, 2019; Stern & Sorek, 2011). Therefore, it is expected that the minimal CRISPR-Cas system emerged as a response to the obstruction of its ancestor by viral counter-adaptations.

As CRISPR-Cas systems are adaptive, viral escape by mutations on the target DNA can be easily blocked by acquisition of new spacers. Therefore, it is expected that an effective counter-adaptive measure follows a different route to deal with CRISPR-Cas defence. Recently, it was observed that prophages in *P. aeruginosa* were not targeted by the endogenous CRISPR-Cas system, despite carrying a matching spacer. Small proteins encoded by the prophage were pointed as responsible for this defect in interference, as they physically interacted with components of the CRISPR-Cas immune system. To date, several of these small proteins, termed Anti-CRISPRs (Acrs), have been discovered, with a wide range of mechanisms that match CRISPR-Cas diversity.

So far, 14 Acrs (AcrF1-14) that block the Type I-F Cascade have been described. For AcrF1, 2 and 3, the hindering mechanism against the *P. aeruginosa* complex has been elucidated. AcrF1 interacts with a lysine residue in the thumb domain of Cas7f, plus two other lysines in the web domain of the same and adjacent subunits (Chowdhury et al., 2017). As these residues can be found several times within a Cascade, three different binding modes involving one to three AcrF1 subunits are possible (Peng et al., 2017). Similarly, AcrF2 interacts with lysine residues within Cas8f and the adjacent Cas7f subunit through an acidic-charged surface, resulting in specific binding. The binding of these proteins to Cascade prevents the formation of an R-loop by blocking the access of the target to the crRNA.

AcrF3, on the other hand, forms a homodimer that interacts with Cas2-3, blocking the path of the non-target strand to the HD nuclease domain. Because of this, the nuclease cannot be recruited to the Cascade, affecting both interference and primed adaptation (J. Wang et al., 2016; X. Wang et al., 2016). Structurally, AcrF3 mimics the Cas2-3 recruitment helix in Cas8f, suggesting that phages can co-opt domains of CRISPR-Cas systems and exploit them for their defence (Rollins et al., 2019).

For AcrF1 and 2, the attachment sites to Type I-F Cascade are altered in the I-Fv complex structure. The wrist helix formed by Cas7fv would prevent the binding of AcrF1 to the backbone, and the absence of the large subunit makes the attachment of AcrF2 highly unlikely, as the AH domain significanlty differs from Cas8f. Furthermore, the low sequence identity between the nuclease portion of the Cas2-3 protein of Type I-F and I-Fv also implies the incapacity of AcrF3 to block the recruitment to the minimised complex. Overall, the specific sites of modification suggest that the minimal Cascade might be an evolutionary response to the blocking activity of Acr proteins on a postulated Type I-F ancestor.

In order to test this hypothesis, we studied the effect of several AcrF proteins on Type I-Fv interference. Acrs have been classified according to their targeting range, with AcrF1, 2 and 3 being very stringent, binding only the *P. aeruginosa* Cascade. AcrF9 and 10 show a broader degree of inhibition, blocking Type I-F complexes from various species. AcrF6 was shown to attack both Type I-E and I-F complexes (Bondy-Denomy, Pawluk, Maxwell, & Davidson, 2013; Pawluk et al., 2016). Despite the wide diversity, the individual co-expression of 13 different I-F inhibitors with the minimal Cascade did not lead to a significant decrease in transformation efficiency (Figure VII.1). Therefore, the divergence of the minimal complex provides immunity against anti-CRISPR measures that block the closely related Type I-F system.

Furthermore, we also tested the interaction of the minimal complex with AcrF9 Vpa through BLI, using the Type I-F system from *Shewanella baltica* OS195 as a positive control. In addition to corroborating the lack of binding of the inhibitor to the Type I-Fv Cascade, we were able to report a novel mechanism of Acr interaction. In this case, a 1:1 complex-AcrF9 ratio promoted the binding of the Type I-F Cascade to both complementary and non-complementary dsDNA, with increasing concentrations of Acr abolishing the binding to correct targets.

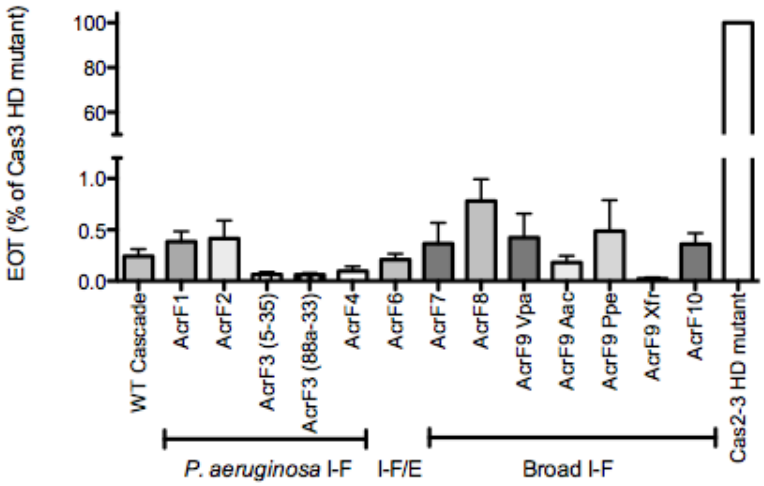


Figure VII.1. Efficiency of Transformation of *E. coli* BL21-AI carrying Type I-Fv Cascade and Anti-CRISPR (Acr) proteins. Quantification of three independent EOT assays of Type I-Fv WT Cascade carrying a plasmid-targeting crRNA; alone or co-expressed with the Acr proteins described in Table II. EOT is expressed as the colony ratio between the strain of interest and the Cas2/3 HD mutant strain. Error bars represent the standard error of the mean (SEM). Materials and methods for this experiment can be found in the Appendix section.

Although the Type I-Fv effector complex is not affected by the assayed Acrs, there might be others that do block it. The tested inhibitors were found mostly in prophages, which were integrated in hosts that do not have the Type I-F variant. Therefore, an adaptive counter-pressure is absent. Furthermore, it is possible that their mechanisms affect the highly diverged components of the system, as shown for AcrF1, 2 and 3, while Acrs targeting the conserved proteins, as Cas6f or the Cas2 portion of Cas2-3, could have a higher chance of blocking a wider range of systems.

As we lack knowledge on phages or mobile genetic elements (MGEs) that specifically target organisms that carry Type I-Fv systems, currently it is not possible to study with higher detail the dynamics behind this arms-race. Recent studies have identified 7 lytic phages against *Shewanella putrefaciens* strains, with 3 of them sequenced (Han et al., 2014; Yang et al., 2019). Yet, none of the host strains has been sequenced. Furthermore, no spacers from the I-Fv CRISPR array have sequence complementarity to the genome of the phages. Therefore, it is not possible to know whether these *S. putrefaciens* strains carry CRISPR-Cas systems and if the viral activity is dependent on anti-CRISPR proteins. In the future, sequencing these Shewanella strains or testing the activity of the phages against *S. putrefaciens* CN-32 might open the possibility of studying the evolution of a warfare where the minimal CRISPR-Cas system is involved.

Previous studies have relied on the inverse search of spacers to identify viral interactors of the CRISPR-Cas host. When doing so for the 81 spacers in the Type I-Fv CRISPR array, 4 spacers can be found in the viral genome of Inoviridae sp. isolate ctce6 (MH617040.1). This sequence was assembled from metagenomics samples, therefore testing the activity of the phage in laboratory conditions is not possible. Nevertheless, the lack of matches for the rest of the spacers suggests that there are more invading agents for *S. putrefaciens* CN-32, but they are yet to be found and sequenced.

Although there are plausible hypotheses for the minimization of the Type I-Fv Cascade; it is still unclear where the new components, Cas5fv and Cas7fv, originated from and how they spread through populations. Therefore, looking into the other systems and organisms closely related to *S. putrefaciens* CN-32 might help elucidating the evolutionary pathway of the minimization.

The minimal system is classified as a variant of Type I-F due to the sequence similarity to Cas6f, a conserved endoribonuclease with a RAMP domain (Makarova et al., 2015). Cas6f from *S. putrefaciens* CN-32 shares 36% of amino acid sequence identity with its counterpart from *P. aeruginosa* PA14, with high conservation of the functional domains. Furthermore, Cas1 and the Cas2 portion of the Cas2-3 fusion also have high sequence conservation, of 59% and 56%, respectively.

Stronger similarities are found when the minimal system is compared to the I-F type of closer related organisms, such as *Shewanella baltica* OS195 or *Shewanella sp.* W3-18-1. Again, Cas6f, Cas1 and the Cas2 section of Cas2-3 are conserved, but in addition, the repeat sequences of the CRISPR array are identical. This could be explained by the higher sequence identity of the acquisition complex, Cas1 and Cas2-3, although Cas6f, the protein responsible for repeat recognition, keeps a sequence similarity below 35%.

Different kinds of minimisation of Type I-F systems have been found *in silico.* Some Vibrio species have loci with conserved Cas7f and Cas5f proteins, but miss Cas8f. They also lack the acquisition module and an associated CRISPR array with more than 1 spacer, suggesting that they do not play an active role in interference (Makarova et al., 2015). Recently, one of these minimal systems was shown to help the DNA integration of associated transposons (Klompe, Vo, Halpin-Healy, & Sternberg, 2019). This suggests that the inclusion of Cas5fv and Cas7fv proteins was required to cope with the loss of the large subunit and keep the immune system active, as other minimisation steps lead to complexes that do not perform interference.

CRISPR-Cas systems are located within mobile genetic elements, and they are easily spread through populations by horizontal gene transfer (Baliga, Shekar, & Venugopal, 2019; Chakraborty et al., 2010; McDonald, Regmi, Morreale, Borowski, & Boyd, 2019; Varble, Meaden, Barrangou, Westra, & Marraffini, 2019). In spite of the expected advantage, the minimal variety is not as widespread as other Type I systems (Makarova et al., 2015), possibly due to detrimental crosstalk or overlapping functions with other CRISPR-Cas systems, leading to fitness costs, as it has been shown for other Type I systems (Touchon & Rocha, 2010).

S. putrefaciens CN-32 and *Shewanella benthica* DB21MT-2 only carry a Type I-Fv locus. The existence of stand-alone Type I-Fv CRISPR-Cas systems suggests that they are able to elicit a robust immune response. Consistent with this, the CRISPR arrays of *S. putrefaciens* CN-32 and *S. benthica* DB21MT-2 carry 81 and 126 spacers, respectively. *S. putrefaciens* 200 has an additional Type III-B system, while *Photobacterium profundum* SS9 has a plasmid-encoded Type I-E system in addition to the genomic minimal system. Interestingly, all of these organisms carry orphan CRISPR arrays, with repeats that do not correspond to the reported systems. This suggests the loss of other systems, in benefit of more efficient ones. Furthermore, it has been shown that these orphan arrays can target Cas proteins, limiting the transfer of other systems (Almendros, Guzman, Garcia-Martinez, & Mojica, 2016).

Cas5fv and Cas7fv are highly conserved among Type I-Fv systems. Therefore, the limited diversification of the homologues does not allow to predict an ancestor among other Cas protein. Previous evolutionary theories have pointed at Cas10, the nuclease from Type III systems, as the common ancestor of Type I Cascade proteins, as they all belong to the RAMP family (Koonin & Makarova, 2019; Makarova, Wolf, & Koonin, 2013). For Type I-Fv, both novel proteins could not be classified to any previously described families, suggesting that they were not acquired from other CRISPR-Cas systems, rather than from proteins fulfilling other tasks in the cell, or even in viruses. In addition, these proteins must have undergone extreme mutation or re-arrangement in order to be able to participate in an active effector complex. One of this independent re-arrangements might have led to the addition of the AH domain of Cas5fv, as its removal does not affect complex formation.

Overall, it is not possible to map the evolution of Cas5fv and Cas7fv with the available data. As more strains are sequenced, we expect to find the missing links to bridge the I-Fv proteins with their homologues. Although we do not know their origin, we can

hypothethise on the causes behind Cas protein diversification. The work here presented supports the theory where the evolutionary pressure exerted by Acr proteins acts as the driving force of diversification or replacement of the targeted Cas proteins. Therefore, the new mechanisms for PAM recognition and non-target stabilization give an advantage to the organisms carrying the Type I-Fv system, in contrast to Type I-F systems. The discovery of phages and plasmids that target strains carrying this minimal system will further help studying the evolution of adaptive and counter-adaptive mechanisms, taking interactions that take place in the environment into account.

What is the cost of complex minimization?

The replacement of the large subunit by specific features of Cas5fv and Cas7fv is a modification that has advantages and disadvantages. It is expected that the diversification protects the system from anti-CRISPR proteins that attack Cas8f, Cas5f, Cas7f and Cas2-3. Furthermore, the need for less proteins might aid in complex formation, making them available faster. This would be extremely relevant in the scenario where CRISPR-Cas immunity is up-regulated upon viral entry.

A recent study showed that, *in vivo*, Type I-E Cascades can lose the large subunit upon encounter of the transcriptional machinery or highly structured DNA (as in CRISPR arrays) (Vink et al., 2019). It is not known whether Cascades can recover Cas8e, or if they are degraded after this disassembly, yet this phenomenon might pose a big setback in the establishment of immunity. For Type I-Fv, effector complexes are highly stable, although they require an excess of crRNA to form efficiently and with the right stoichiometry. Nevertheless, as the crRNA acts as a linker for all Cas proteins in the complex, the loss of components is highly unlikely.

Despite these benefits, the study of the kinetics of the complex revealed a dissociation constant (K_D) of 46.65 nM, which is higher than for the Type I-F system from *P. aeruginosa* (1 nM) (Rollins, Schuman, Paulus, Bukhari, & Wiedenheft, 2015). Although the difference could be explained by technical discrepancies, as Type I-F interactions were studied by EMSAs, we also see an attenuated targeting *in vivo.* The minimal system shows interference levels 10-fold lower than *S. baltica* OS195 and *P. aeruginosa* PA14 (Steube, 2018). This could be explained at the Cascade level as a result of the major groove PAM recognition mechanism, leading to higher target recognition times, or as a lack of stability of the R-loop complex due to the structural minimisation. This last scenario is reinforced by BLI data showing that mismatches between the strands of the

DNA target improve binding of the minimal complex, suggesting that Cas7fv could not efficiently stabilise the non-target strand.

The decreased interference could also be explained by different recruitment and activity of Cas2-3. The nuclease portion of the protein has conserved nuclease and helicase domains, but overall low sequence identity to other Type I-F Cas nucleases. Previous experiments aiming at elucidating the structure of this protein and the interaction mode with Cascade have been unsuccessful due to the low stability of the purified nuclease. In the future, we expect to study the timing of the recruitment of the nuclease by sptPALM, as well as the kinetics of the interaction by BLI.

The described results give insights into the existence of a range of interference and affinity levels, depending on the effector complex and associated nuclease. For the Type I-Fv Cascade, the adjustment of complex components in order to gain Acr resistance might have come at a cost of affinity loss. Further studies into the target search of the minimal complex and the binding times of the I-F Cascade will help further describe the gains and losses that came with the diversification of the Type I-Fv system.

The current state of minimal CRISPR-Cas systems is a result of a long process of natural selection. Therefore, the described interference levels must suffice to elicit a proper immune response in the organisms carrying Type I-Fv systems, without affecting the overall physiology of the hosts. The establishment of an equilibrium within defence systems is relevant for the generation of diversity, as mobile genetic elements are not always detrimental. Horizontal gene transfer has been described as the main drive behind the fast spread of antibiotic resistance, as well as the gain of virulence traits and new metabolic pathways that allow microorganisms to adapt and colonize new environments (Frost, Leplae, Summers, & Toussaint, 2005; Ochman, Lawrence, & Groisman, 2000). Consistent with this, bacterial genomes can carry up to 21% of mobile element genes (Newton & Bordenstein, 2011). Consequently, it is expected that CRISPR-Cas systems act more as gatekeepers for mobile genetic elements rather than executioners, allowing some degree of integration in order to maintain evolution and diversification of the hosts. Further comparative studies on the interference activity of CRISPR-Cas systems would be useful to study this interplay, as correlations might be found between defense and MGE integration levels. Moreover, these comparisons could help understand the presence of more than one CRISPR-Cas system on some organisms, providing further insights into the evolution of this adaptive immune system.

References

Almendros, C., Guzman, N. M., Garcia-Martinez, J., & Mojica, F. J. (2016). Anti-cas spacers in orphan CRISPR4 arrays prevent uptake of active CRISPR-Cas I-F systems. *Nat Microbiol, 1*(8).

Anders, C., Niewoehner, O., Duerst, A., & Jinek, M. (2014). Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature, 513*(7519), 569-573.

Baliga, P., Shekar, M., & Venugopal, M. N. (2019). Investigation of direct repeats, spacers and proteins associated with clustered regularly interspaced short palindromic repeat (CRISPR) system of Vibrio parahaemolyticus. *Mol Genet Genomics, 294*(1), 253-262.

Bondy-Denomy, J., Pawluk, A., Maxwell, K. L., & Davidson, A. R. (2013). Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature, 493*(7432), 429-432.

Brouns, S. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J., Snijders, A. P., et al. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science, 321*(5891), 960-964.

Chakraborty, S., Snijders, A. P., Chakravorty, R., Ahmed, M., Tarek, A. M., & Hossain, M. A. (2010). Comparative network clustering of direct repeats (DRs) and cas genes confirms the possibility of the horizontal transfer of CRISPR locus among bacteria. *Mol Phylogenet Evol, 56*(3), 878-887.

Chowdhury, S., Carter, J., Rollins, M. F., Golden, S. M., Jackson, R. N., Hoffmann, C., et al. (2017). Structure Reveals Mechanisms of Viral Suppressors that Intercept a CRISPR RNA-Guided Surveillance Complex. *Cell, 169*(1), 47-57 e11.

Dwarakanath, S., Brenzinger, S., Gleditzsch, D., Plagens, A., Klingl, A., Thormann, K., et al. (2015). Interference activity of a minimal Type I CRISPR-Cas system from Shewanella putrefaciens. *Nucleic Acids Res, 43*(18), 8913-8923.

Frost, L. S., Leplae, R., Summers, A. O., & Toussaint, A. (2005). Mobile genetic elements: the agents of open source evolution. *Nat Rev Microbiol, 3*(9), 722-732.

Gleditzsch, D., Muller-Esparza, H., Pausch, P., Sharma, K., Dwarakanath, S., Urlaub, H., et al. (2016). Modulating the Cascade architecture of a minimal Type I-F CRISPR-Cas system. *Nucleic Acids Res, 44*(12), 5872-5882.

Gleditzsch, D., Pausch, P., Muller-Esparza, H., Ozcan, A., Guo, X., Bange, G., et al. (2019). PAM identification by CRISPR-Cas effector complexes: diversified mechanisms and structures. *RNA Biol, 16*(4), 504-517.

Han, F., Li, M., Lin, H., Wang, J., Cao, L., & Khan, M. N. (2014). The novel Shewanella putrefaciens-infecting bacteriophage Spp001: genome sequence and lytic enzymes. *J Ind Microbiol Biotechnol, 41*(6), 1017-1026.

Hayes, R. P., Xiao, Y., Ding, F., van Erp, P. B., Rajashankar, K., Bailey, S., et al. (2016). Structural basis for promiscuous PAM recognition in type I-E Cascade from E. coli. *Nature, 530*(7591), 499-503.

Hoyland-Kroghsbo, N. M., Munoz, K. A., & Bassler, B. L. (2018). Temperature, by Controlling Growth Rate, Regulates CRISPR-Cas Activity in Pseudomonas aeruginosa. *MBio, 9*(6), e02184-18.

Hoyland-Kroghsbo, N. M., Paczkowski, J., Mukherjee, S., Broniewski, J., Westra, E., Bondy-Denomy, J., et al. (2017). Quorum sensing controls the Pseudomonas aeruginosa CRISPR-Cas adaptive immune system. *Proc Natl Acad Sci U S A, 114*(1), 131-135.

Klompe, S. E., Vo, P. L. H., Halpin-Healy, T. S., & Sternberg, S. H. (2019). Transposon-encoded CRISPR-Cas systems direct RNA-guided DNA integration. *Nature* doi: 10.1038/s41586-019-1323-z (Epub ahead of print).

Koonin, E. V., & Makarova, K. S. (2019). Origins and evolution of CRISPR-Cas systems. *Philos Trans R Soc Lond B Biol Sci, 374*(1772), 20180087.

Kuznedelov, K., Mekler, V., Lemak, S., Tokmina-Lukaszewska, M., Datsenko, K. A., Jain, I., et al. (2016). Altered stoichiometry Escherichia coli Cascade complexes with shortened CRISPR RNA spacers are capable of interference and primed adaptation. *Nucleic Acids Res, 44*(22), 10849-10861.

Luo, M. L., Jackson, R. N., Denny, S. R., Tokmina-Lukaszewska, M., Maksimchuk, K. R., Lin, W., et al. (2016). The CRISPR RNA-guided surveillance complex in Escherichia coli accommodates extended RNA spacers. *Nucleic Acids Res, 44*(15), 7385-7394.

Luscombe, N. M., Laskowski, R. A., & Thornton, J. M. (2001). Amino acid–base interactions: a three-dimensional analysis of protein–DNA interactions at an atomic level. *Nucleic Acids Res, 29*(13), 2860–2874.

Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J., Charpentier, E., Horvath, P., et al. (2011). Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol, 9*(6), 467-477.

Makarova, K. S., Wolf, Y. I., Alkhnbashi, O. S., Costa, F., Shah, S. A., Saunders, S. J., et al. (2015). An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol, 13*(11), 722-736.

Makarova, K. S., Wolf, Y. I., & Koonin, E. V. (2013). The basic building blocks and evolution of CRISPR-CAS systems. *Biochem Soc Trans, 41*(6), 1392-1400.

McDonald, N. D., Regmi, A., Morreale, D. P., Borowski, J. D., & Boyd, E. F. (2019). CRISPR-Cas systems are present predominantly on mobile genetic elements in Vibrio species. *BMC Genomics, 20*(1), 105.

Medina-Aparicio, L., Rebollar-Flores, J. E., Gallego-Hernandez, A. L., Vazquez, A., Olvera, L., Gutierrez-Rios, R. M., et al. (2011). The CRISPR/Cas immune system is an operon regulated by LeuO, H-NS, and leucine-responsive regulatory protein in Salmonella enterica serovar Typhi. *J Bacteriol, 193*(10), 2396-2407.

Meeske, A. J., Nakandakari-Higa, S., & Marraffini, L. A. (2019). Cas13-induced cellular dormancy prevents the rise of CRISPR-resistant bacteriophage. *Nature, 570*(7760), 241-245.

Modell, J. W., Jiang, W., & Marraffini, L. A. (2017). CRISPR-Cas systems exploit viral DNA injection to establish and maintain adaptive immunity. *Nature, 544*(7648), 101-104.

Mohanraju, P., Makarova, K. S., Zetsche, B., Zhang, F., Koonin, E. V., & van der Oost, J. (2016). Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science, 353*(6299), aad5147.

Newton, I. L. G., & Bordenstein, S. R. (2011). Correlations Between Bacterial Ecology and Mobile DNA. *Current Microbiology, 62*(1), 198-208.

Ochman, H., Lawrence, J. G., & Groisman, E. A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature, 405*(6784), 299-304.
Paterson, S., Vogwill, T., Buckling, A., Benmayor, R., Spiers, A. J., Thomson, N. R., et al. (2010). Antagonistic coevolution accelerates molecular evolution. *Nature, 464*(7286), 275-278.
Patterson, A. G., Chang, J. T., Taylor, C., & Fineran, P. C. (2015). Regulation of the Type I-F CRISPR-Cas system by CRP-cAMP and GalM controls spacer acquisition and interference. *Nucleic Acids Res, 43*(12), 6038-6048.
Patterson, A. G., Jackson, S. A., Taylor, C., Evans, G. B., Salmond, G. P. C., Przybilski, R., et al. (2016). Quorum Sensing Controls Adaptive Immunity through the Regulation of Multiple CRISPR-Cas Systems. *Mol Cell, 64*(6), 1102-1108.
Pausch, P., Muller-Esparza, H., Gleditzsch, D., Altegoer, F., Randau, L., & Bange, G. (2017). Structural Variation of Type I-F CRISPR RNA Guided DNA Surveillance. *Mol Cell, 67*(4), 622-632.
Pawluk, A., Staals, R. H., Taylor, C., Watson, B. N., Saha, S., Fineran, P. C., et al. (2016). Inactivation of CRISPR-Cas systems by anti-CRISPR proteins in diverse bacterial species. *Nat Microbiol, 1*(8), 16085.
Peng, R., Xu, Y., Zhu, T., Li, N., Qi, J., Chai, Y., et al. (2017). Alternate binding modes of anti-CRISPR viral suppressors AcrF1/2 to Csy surveillance complex revealed by cryo-EM structures. *Cell Res, 27*(7), 853-864.
Plagens, A., Richter, H., Charpentier, E., & Randau, L. (2015). DNA and RNA interference mechanisms by CRISPR-Cas surveillance complexes. *FEMS Microbiol Rev, 39*(3), 442-463.
Pul, U., Wurm, R., Arslan, Z., Geissen, R., Hofmann, N., & Wagner, R. (2010). Identification and characterization of E. coli CRISPR-cas promoters and their silencing by H-NS. *Mol Microbiol, 75*(6), 1495-1512.
Redding, S., Sternberg, S. H., Marshall, M., Gibb, B., Bhat, P., Guegler, C. K., et al. (2015). Surveillance and Processing of Foreign DNA by the Escherichia coli CRISPR-Cas System. *Cell, 163*(4), 854-865.
Rohs, R., West, S. M., Sosinsky, A., Liu, P., Mann, R. S., & Honig, B. (2009). The role of DNA shape in protein-DNA recognition. *Nature, 461*(7268), 1248-1253.
Rollins, M. F., Chowdhury, S., Carter, J., Golden, S. M., Miettinen, H. M., Santiago-Frangos, A., et al. (2019). Structure Reveals a Mechanism of CRISPR-RNA-Guided Nuclease Recruitment and Anti-CRISPR Viral Mimicry. *Mol Cell, 74*(1), 132-142 e135.
Rollins, M. F., Schuman, J. T., Paulus, K., Bukhari, H. S., & Wiedenheft, B. (2015). Mechanism of foreign DNA recognition by a CRISPR RNA-guided surveillance complex from Pseudomonas aeruginosa. *Nucleic Acids Res, 43*(4), 2216-2222.
Rutkauskas, M., Sinkunas, T., Songailiene, I., Tikhomirova, M. S., Siksnys, V., & Seidel, R. (2015). Directional R-Loop Formation by the CRISPR-Cas Surveillance Complex Cascade Provides Efficient Off-Target Site Rejection. *Cell Rep, 10*(9), 1534-1543.
Stenseth, N. C., & Smith, J. M. (1984). Coevolution in Ecosystems: Red Queen Evolution or Stasis? *Evolution, 38*(4), 870-880.
Stern, A., & Sorek, R. (2011). The phage-host arms race: shaping the evolution of microbes. *Bioessays, 33*(1), 43-51.
Steube, N. (2018). *Characterization of Anti-CRISPR proteins against three variants of subtype I-F CRISPR-Cas systems.* Master Thesis. Philipps-Universität Marburg.
Touchon, M., & Rocha, E. P. (2010). The small, slow and specialized CRISPR and anti-CRISPR of Escherichia and Salmonella. *PLoS One, 5*(6), e11126.
Varble, A., Meaden, S., Barrangou, R., Westra, E. R., & Marraffini, L. A. (2019). Recombination between phages and CRISPR-cas loci facilitates horizontal gene transfer in staphylococci. *Nat Microbiol, 4*(6), 956-963.
Vink, J. N. A., Martens, K. J. A., Vlot, M., McKenzie, R. E., Almendros, C., Estrada Bonilla, B., et al. (2019). Direct visualization of native CRISPR target search in live bacteria reveals Cascade DNA surveillance mechanism. *bioRxiv* doi :10.1101/589119.
Wang, J., Ma, J., Cheng, Z., Meng, X., You, L., Wang, M., et al. (2016). A CRISPR evolutionary arms race: structural insights into viral anti-CRISPR/Cas responses. *Cell Res, 26*(10), 1165-1168.
Wang, X., Yao, D., Xu, J. G., Li, A. R., Xu, J., Fu, P., et al. (2016). Structural basis of Cas3 inhibition by the bacteriophage protein AcrF3. *Nat Struct Mol Biol, 23*(9), 868-870.
Watson, B. N. J., Vercoe, R. B., Salmond, G. P. C., Westra, E. R., Staals, R. H. J., & Fineran, P. C. (2019). Type I-F CRISPR-Cas resistance against virulent phage infection triggers abortive infection and provides population-level immunity. *bioRxiv* doi: 10.1101/679308
Westra, E. R., Pul, U., Heidrich, N., Jore, M. M., Lundgren, M., Stratmann, T., et al. (2010). H-NS-mediated repression of CRISPR-based immunity in Escherichia coli K12 can be relieved by the transcription activator LeuO. *Mol Microbiol, 77*(6), 1380-1393.
Westra, E. R., van Erp, P. B., Kunne, T., Wong, S. P., Staals, R. H., Seegers, C. L., et al. (2012). CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Mol Cell, 46*(5), 595-605.
Xiao, Y., Luo, M., Hayes, R. P., Kim, J., Ng, S., Ding, F., et al. (2017). Structure Basis for Directional R-loop Formation and Substrate Handover Mechanisms in Type I CRISPR-Cas System. *Cell, 170*(1), 48-60 e11.
Xue, C., Zhu, Y., Zhang, X., Shin, Y. K., & Sashital, D. G. (2017). Real-Time Observation of Target Search by the CRISPR Surveillance Complex Cascade. *Cell Rep, 21*(13), 3717-3727.
Yang, Z. Q., Tao, X. Y., Zhang, H., Rao, S. Q., Gao, L., Pan, Z. M., et al. (2019). Isolation and characterization of virulent phages infecting Shewanella baltica and Shewanella putrefaciens, and their application for biopreservation of chilled channel catfish (Ictalurus punctatus). *Int J Food Microbiol, 292*, 107-117.
Yosef, I., Moran, G. G., Kiro, R., Edgarm, R., & Qimron, I. (2011). High-temperature protein G is essential for activity of the Escherichia coli clustered regularly interspaced short palindromic repeats (CRISPR)/Cas system. *Proc Natl Acad Sci U S A, 108*(50), 20136-20141.
Young, J. C., Dill, B. D., Pan, C., Hettich, R. L., Banfield, J. F., Shah, M., et al. (2012). Phage-induced expression of CRISPR-associated proteins is revealed by shotgun proteomics in Streptococcus thermophilus. *PLoS One, 7*(5), e38077.

Appendix

Materials and Methods for Figure VII.1

Efficiency of Transformation (EOT) Assay

E. coli BL21-AI strains (F- ompT hsdSB (rB-mB-) gal dcm araB::T7RNAPtetA, Invitrogen) were used for efficiency of transformation (EOT) assays and grown in 2YTL media (16 g/l tryptone, 10 g/l yeast extract, 5 g/l NaCl, 10 mM MgSO4, 0.2% maltose) supplemented with appropriate antibiotics at 37°C. The pCDFDuet-1 plasmid encoding Cas7fv, Cas5fv and Cas6f and wild-type Cas2-3 or an inactive Cas2-3 HD domain mutant (C466G, A467C, A470C mutations in *cas2-3*) plus a minimal CRISPR with a spacer targeting the ampicillin resistance cassette of pETDuet-1 (spacer sequence: 5´-AGTCACAGAAAAGCATCTTACGGATGGCATGA-3´, Pausch et al., 2017) were transformed into BL21-AI with a pRSFDuet-1 plasmid encoding for one Acr protein (Table VIII.1). Plasmids carrying the *Acr* genes used for sub-cloning were kindly provided by Dr. Alan Davidson (Bondy-Denomy, Pawluk, Maxwell, & Davidson, 2013; Pawluk et al., 2016). Single colonies of the obtained strains were cultured overnight, diluted 1:500 on a main culture of 2YTL and grown until OD600 = 0.3. Subsequently, 0.1 mM of IPTG and 0.2% of arabinose (Sigma-Aldrich) were added for induction. After OD_{600} = 0.6 was reached, electrocompetent cells were prepared as previously described (Thomason et al., 2014). 50 µl aliquots of each strain were mixed with 1 µl of pETDuet-1 (5 ng/µl) and transferred to a pre-chilled 0.1 cm cuvette, for posterior electroporation at 1.8 kV (Micropulser, Biorad). Cells were then mixed with 550 µl of warm 2YTL media and transferred to a culture tube. After 1 hr of recovery at 37° C, serial dilutions were plated in 2YTL plates with ampicillin (50 µg/ml), kanamycin (25 µg/ml) and spectinomycin (25 µg/ml). After overnight incubation at 37°C, countable dilutions were selected and the amount of colonies per plate was determined. Efficiency of Transformation (EOT) was expressed as the ratio between colony count of the strains of interest carrying the wild-type Cas3 and the corresponding ones with the Cas3 HD mutation. Assays were performed in triplicate and error bars were calculated as standard error of the mean (SEM).

References

Bondy-Denomy, J., Pawluk, A., Maxwell, K. L., & Davidson, A. R. (2013). Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature, 493*(7432), 429-432.

Pausch, P., Muller-Esparza, H., Gleditzsch, D., Altegoer, F., Randau, L., & Bange, G. (2017). Structural Variation of Type I-F CRISPR RNA Guided DNA Surveillance. *Mol Cell, 67*(4), 622-632 e624.

Pawluk, A., Staals, R. H., Taylor, C., Watson, B. N., Saha, S., Fineran, P. C., et al. (2016). Inactivation of CRISPR-Cas systems by anti-CRISPR proteins in diverse bacterial species. *Nat Microbiol, 1*(8).

Thomason, L. C., Sawitzke, J. A., Li, X., Costantino, N., & Court, D. L. (2014). Recombineering: genetic engineering in bacteria using homologous recombination. *Curr Protoc Mol Biol, 106*, 1 16 11-39.

Table VIII.1. Plasmids used for EOT assays of Figure VII.1

Plasmid name	Description
pCascadeWT	pCDFDuet-1 + Cas2-3 + Cas7fv + Cas5fv + Cas6f + repeat-spacer anti-amp-repeat sequence. Repeats and *Cas* genes from *S. putrefaciens* CN-32
pCascadeHD	pCDFDuet-1 + Cas2-3 HD mutant + Cas7fv + Cas5fv + Cas6f + repeat-spacer anti-amp-repeat sequence. Repeats and *Cas* genes from *S. putrefaciens* CN-32
pETDuet-1	Target Plasmid (Novagen)
pAcrF1	pRSFDuet-1 + AcrF1 from *Pseudomonas aeruginosa*
pAcrF2	pRSFDuet-1 + AcrF2 from *Pseudomonas aeruginosa*
pAcrF3 (5-35)	pRSFDuet-1 + AcrF3 from *Pseudomonas aeruginosa*
pAcrF3 (88a-33)	pRSFDuet-1 + AcrF3 from *Pseudomonas aeruginosa*
pAcrF4	pRSFDuet-1 + AcrF4 from *Pseudomonas aeruginosa*
pAcrF6	pRSFDuet-1 + AcrF6 from *Pseudomonas aeruginosa*
pAcrF7	pRSFDuet-1 + AcrF7 from *Pseudomonas aeruginosa*
pAcrF8	pRSFDuet-1 + AcrF8 from *Pseudomonas aeruginosa*
pAcrF9 Aac	pRSFDuet-1 + AcrF9 from *Aggregatibacter actinomycetemcomitans*
pAcrF9 Ppe	pRSFDuet-1 + AcrF9 from *Proteus penneri*
pAcrF9 Vpa	pRSFDuet-1 + AcrF9 from *Vibrio parahaemolyticus*
pAcrF9 Xfr	pRSFDuet-1 + AcrF9 from *Xanthomonas fragariae*
pAcrF10	pRSFDuet-1 + AcrF10 from *Shewanella xiamenensis*

Curriculum Vitae

Hanna Constanza Müller Esparza

DOB: April 16, 1991
POB: Antofagasta, Chile
Citizenship: Chilean/German

Education

09.2015-09.2019	**PhD.** Philipps-Universität Marburg and IMPRS for Environmental, Cellular and Molecular Microbiology. Thesis: "Characterization of DNA interference by a minimal Type I-F CRISPR-Cas system". Advisor: Dr. Lennart Randau. Max Planck Institute for Terrestrial Microbiology, Marburg, Germany.
08.2013-07.2015	**Biochemistry Degree.** Pontificia Universidad Católica de Chile. Professional Certificate Equivalent to German Diplom. Thesis: "Novel extra-circadian functions of the FREQUENCY protein in *B. cinerea:* implications in conidiation and sexual development". Advisor: Dr. Luis F. Larrondo. Department of Molecular Genetics and Microbiology, School of Biological Sciences, PUC, Santiago, Chile.
01.2009-07.2013	**BSc in Biochemistry.** Pontificia Universidad Católica de Chile.
03.2014-12.2008	**High School.** The Antofagasta British School.

Research Experience

09.2015- 09.2019	**PhD student.** Prokaryotic Small RNA Biology Group. Max Planck Institute for Terrestrial Microbiology.
08.2013 - 07.2015	**Undergraduate Research Assistant.** School of Biological Sciences. Pontificia Universidad Católica de Chile. Undergraduate thesis to obtain degree in Biochemistry. Advisor: Dr. Luis F. Larrondo.
07. 2012 - 07.2013	**Research Unit.** School of Biological Sciences. Pontificia Universidad Católica de Chile. "Molecular and phenotypic characterisation of *Botrytis cinerea* circadian clock mutants". Advisor: Dr. Luis F. Larrondo
01. 2012 - 06. 2012	**Research Unit.** School of Medicine. Pontificia Universidad Católica de Chile. "Genic and proteic expression of RPRM in gastric cancer cell lines". Advisor: Dr. Alejandro Corvalán.

Acknowledgements

The conclusion of this chapter would have not been possible without the help of many wonderful people. This past four years I have received an incredible amount of support, which was crucial both for my research and personal life. For this, I would like to thank everyone involved.

First and foremost, I would like to thank Dr. Lennart Randau for the opportunity of joining his lab, allowing me to explore a new topic within microbiology. His advise, support and encouragement have really contributed into shaping my present as a researcher. Thank you for always being available and open to my ideas.

I would also like to thank the members of my Thesis Advisory Committee, Prof. Dr. Kai Thormann and Prof. Dr. Peter Graumann, for the fruitful discussions and valuable input to my project. I have to further thank Dr. Ulrike Endesfelder for her help and mentorship, alongside introducing me to the world of photophysics and leading an important portion of this work. Also, I would like to express my gratitude to Prof. Dr. Martin Thanbichler, for his time and expertise, which significantly helped this work.

Infinite thanks go to Julia Wiegel, who accompanied me from day one of this adventure. Your help and friendship are really invaluable to me. Also, I would like to thank Dr. Álvaro Orell for guiding me through the German system, always having excellent advice and making me feel close to home in the lab. Furthermore, this work would have not been possible without Ruth Breuer, Edda Janssen and Niklas Steube, who helped me learn how to be a supervisor. Working side by side with you was very fulfilling, thank you for your commitment!

Additionaly, I would like to thank all the members of the AG Randau, past and present. You all contributed one way or another to this work, both scientifically and by providing a great atmosphere in the lab. I also have to thank the members of the AG Endesfelder for being great collaborators, especially to Bartosz Turkowyd. Without him it would have not been possible to decipher the misteries of Cascade dynamics. Having you as a colleague and friend made carrying out this project a real delight. I further extend my thanks to the members of the AG Thanbichler, especially to Manuel Osorio, for bearing up with my extended BLItz sessions. I appreciate your interest in my project and invaluable help.

Finally, I would like to thank my friends and family. To Manuel, Tarryn, Francisco, Marc, Kristen and specially Stefano, thank you for being a constant support and brightening my time in Marburg. You became my second family and it would have been a very different story if I hadn´t met you. To Carolina and Shankar, my thesis buddies, thank you for your friendship, the perfect nights out and all the help to get this done. To Rafa, thank you for always lending an ear and supporting me. To Vale, Pía, Natalia and all my friends from Chile, thank you for staying close despite the distance. Your comfort lifted my spirits throughout this four years. To my parents and brothers, infinite thanks for believing in me and encouraging me to follow this path away from home. Without you this would have not been possible.