



Skolkovo Institute of Science and Technology

Primed CRISPR-Cas adaptation in type I-E system of *Escherichia coli*: use of single-molecule and biochemical assays to verify models of the phenomenon at molecular level

Doctoral Thesis

by

Andrey Krivoy

DOCTORAL PROGRAM IN LIFE SCIENCES

Supervisor

Professor Konstantin Severinov

Moscow – 2018

© Andrey Krivoy 2018

Abstract

CRISPR-Cas are adaptive immune systems of prokaryotes that destroy genetic parasites such as viruses or plasmids. CRISPR-Cas systems, widely spread in both archaea and bacteria, have diverse genetic structure and molecular mechanisms for viral and/or plasmid nucleic acids (NA) degradation in a process called CRISPR interference. At the same time, they all share a common ability to acquire short fragments of viral or plasmid DNA, followed by integration of these fragments into special CRISPR loci in cellular genome. This process, CRISPR adaptation, is carried out by two most conserved CRISPR-associated (Cas) proteins Cas1 and Cas2. A CRISPR locus is transcribed and resulting RNA is processed into units harboring individual viral fragments (crRNA). crRNAs assemble with proteins encoded by the *cas* genes into RNA-protein effector complexes. These complexes are able to recognize foreign NAs complementary to crRNA sequence and cause their degradation. Mutations introducing mismatches between foreign NAs and crRNA weaken the effectiveness of CRISPR interference. Some of type I CRISPR-Cas systems exhibit a specific mode of adaptation dependent on recognition of partially matching targets, called “primed adaptation” or “priming”. Primed adaptation requires both the effector complex and the Cas1, Cas2 proteins and is orders of magnitude more efficient than “*de novo*” adaptation that requires the Cas1-Cas2 complex only. The mechanisms of primed adaptation are not understood yet and several competing models of the process have been proposed. We here focus on type I-E CRISPR-Cas system of *Escherichia coli* as the most studied one among those supporting primed adaptation. We consider two models: one that regards conformational changes in the Cascade effector complex as the cause of priming versus interference with the target and another where kinetics of interaction with fully or partially matching targets is thought to determine the outcome (interference or priming). To probe the first, conformational, model, we used magnetic tweezers that allow one to manipulate force and torque applied to single DNA molecules. Further, we use bulk biochemical assays with Cas3 nuclease that is recruited upon Cascade binding to the target and that destroys the target. Finally, we test if conformational state detected on different targets affect priming. The main conclusion of our work is that priming happens independently of altered Cascade conformations on different targets, thus lending support to the kinetic model of priming. Our results also include some method development for magnetic tweezers and bulk assays that should help move the field forward in the future.

Publications

Results of this thesis were presented on scientific conferences and are published in peer-reviewed journals.

List of publications:

1. Rutkauskas M., Krivoy A., Szczelkun M.D., Rouillon C. and Seidel R. (2017) Single-Molecule Insight Into Target Recognition by CRISPR–Cas Complexes. *Methods in Enzymology*. Vol. 582, pp. 239–273.
2. Krivoy, A., Rutkauskas, M., Kuznedelov, K., Musharova, O., Rouillon, C., Severinov, K., & Seidel, R. (2018). Primed CRISPR adaptation in *Escherichia coli* cells does not depend on conformational changes in the Cascade effector complex detected *in Vitro*. *Nucleic Acids Research*, 46(8), 4087–4098.

List of posters presented on scientific conferences:

1. A. Krivoy, C. Rouillon, K. Severinov and R. Seidel. Poster: "Target binding and processing of protospacers that support primed adaptation", CRISPR 2016, May 23-25, 2016, Weizmann Institute of Science, Rehovot, Israel.
2. A. Krivoy, K. Severinov and R. Seidel. Poster: "Insights into mechanisms of primed adaptation in type I-E CRISPR-Cas system of *E. coli*", Skoltech-MIT conference "Shaping the future", April 25-26, 2017, Skolkovo Institute of Science and Technology, Moscow, Russia.
3. A. Krivoy, K. Severinov and R. Seidel. Poster: "CRISPR-Cas system of *E. coli* supports primed adaptation regardless of conformational changes of effector complex Cascade bound to mismatched targets", CRISPR 2017, June 8-10, 2017, Big Sky, Montana, USA

Acknowledgements

The author is much grateful to Prof. Konstantin Severinov, Prof. Ralf Seidel, Dr. Olga Musharova, Marius Rutkauskas, Dr. Christophe Rouillon, Dr. Ekaterina Savitskaya, Anatoly Krivoy, Fergus Fettes, Dr. Konstantin Kuznedelov, Prof. Virginijus Siksnyus and Prof. Scott Bailey.

Table of Contents

Abstract.....	2
Publications	3
Acknowledgements	4
Table of Contents	5
List of Symbols, Abbreviations	8
List of Figures	10
List of Tables.....	11
Novelty and practical use	12
Personal contribution.....	12
1 Literature review	13
1.1 Antiviral defense systems of prokaryotes	13
1.2 Molecular mechanisms of CRISPR-Cas systems	16
1.3 Classification and diversity of CRISPR-Cas systems.....	19
1.4 Models of CRISPR-Cas primed adaptation	20
1.5 Single molecule techniques to study DNA targeting protein enzymes	21
1.6 Technology and value that CRISPR-Cas systems have brought up to date ..	23
2 Project objectives.....	24
3 Materials and methods	25
3.1 Magnetic tweezers	25
3.1.1 Principles of the magnetic tweezers setup	25
3.1.2 Flow-cell preparation	26
3.1.3 Tracking of the bead position in three dimensions.....	27
3.1.4 Preparation of DNA constructs for MT assays	28
3.1.5 DNA molecule behavior. Force calibration. Rotation curves	29
3.1.6 Magnetic tweezers experiments to measure R-loop properties	31
3.1.7 Data analysis: constant probability density events fit exponential distribution	31
3.1.8 Method development. Installing the temperature control system	32
3.2 Cloning methods	33
3.2.1 Preparation of competent <i>E. coli</i> cells	33
3.2.2 pUC19 targets cloning	34
3.2.3 Synthetic genes cloning	34
3.3 Protein purification and handling	35
3.4 Bulk protein assays.....	37
3.4.1 Cascade binding assays	37
3.4.2 Cas3-mediated DNA cleavage experiments	38
3.4.3 Cas1-Cas2 integration assays	38
3.4.4 <i>In vitro</i> reconstitution of priming	38
3.4.4.1 DNA extraction with phenol-chloroform-isoamyl alcohol. Ethanol precipitation.....	39

3.4.4.2 Urea PAGE.....	39
3.5 Detection of primed adaptation <i>in vivo</i>	39
3.5.1 Priming assay and CRISPR array elongation analysis using agarose gels	39
3.5.2 qPCR approach to develop and measure the adaptation score	40
3.6 High throughput sequencing of CRISPR array.....	41
3.7 Synthetic genes design	41
3.8 SDS PAGE	42
4 Results	42
4.1 Cascade specifically binds plasmids containing targeted sequences	42
4.2 Removal of nuclease contamination in Cascade by gel filtration chromatography.....	43
4.3 Selected target variants cover main possible outcomes of CRISPR immunity response	44
4.4 Cascade binds all potential priming targets within half hour	46
4.5 MT assays.....	47
4.5.1 Study of R-loop formation by Cascade	47
4.5.1.1 Mutated targets support R-loop formation	47
4.5.1.2 Locking ability of R-loop on mutated targets is reduced proportional to proximity of mismatch to PAM	49
4.5.2 Kinetics of Cascade binding to various targets	50
4.5.2.1 Modification of MT assay to allow comparative measurements of target variant R-loop formation kinetics	50
4.5.2.2 Mutated targets support R-loop formation at slower rates than wild-type target.....	54
4.5.3 RNP complexes with changed crRNA lengths and altered Cascade subunits	56
4.6 Altered Cse1 conformation model is supported by Cas3 recruitment ability upon R-loop formation	62
4.7 Primed adaptation occurs independently of conformational changes in Cascade.....	63
4.7.1 High throughput sequencing of adopted spacers shows the same spacer selection for every priming mutant.....	65
4.7.2 Quantitative adaptation measurement supports kinetics model of primed adaptation	66
4.8 Bulk assays development to reproduce priming conditions <i>in vitro</i>	68
4.8.1 Cas1-Cas2 effect on Cas3 degradation rate	68
4.8.2 Short fragments investigation	69
4.8.3 Complete reaction of primed adaptation. Attempt of reconstitution	71
4.9 Summary of results.....	75
5 Discussion	77

References	82
Appendix A. Results obtained by collaborators in this project.	96
Appendix B. Genetic plasmids and constructs used	98
Appendix C. Self-made code and algorithms	102
Appendix D. Supplementary data	109
Appendix E. Derivation of Poisson process to exponential distribution of the events times	116

List of Symbols, Abbreviations

Abi – Abortive infection

AFS – Acoustic Force Spectroscopy

ATP - Adenosine Triphosphate

BSA – Bovine Serum Albumin

CCD – charge coupled device

CMOS – complementary metal oxide semiconductor

CRISPR – Clustered Regularly Interspaced Short Palindromic Repeats

CRISPR-Cas – CRISPR associated

Dig – digoxigenin

DNA – Deoxyribonucleic Acid

DSB – Double-Stranded Breaks (i.e. breaks in both strands of double-stranded DNA)

dsDNA – double-stranded DNA

dUTP – Deoxyuridine Triphosphate

EDTA – Ethylenediaminetetraacetic acid

FRET – Förster Resonance Energy Transfer

GPU – Graphics Processing Unit

IPTG – Isopropyl β -D-1-thiogalactopyranoside

LB medium – Luria-Bertani medium

MT – Magnetic Tweezers

NA – Nucleic Acid(s)

NGS – Next Generation Sequencing

OD₆₀₀ – Optical Density at 600 nm wavelength

PAGE – Polyacrylamide gel electrophoresis

PAM – Protospacer Adjacent Motif

PMSF – Phenylmethylsulfonyl fluoride

pN – piconewton

RNA – Ribonucleic Acid

ROI – Region of Interest

SDS PAGE – Sodium dodecyl sulfate polyacrylamide gel electrophoresis

SEC – Size-exclusion chromatography

Sie – Superinfection exclusion

smFRET – single-molecule FRET

ssDNA – single-stranded DNA

TCEP – tris(2-carboxyethyl)phosphine

TIRF – total internal reflection fluorescence (microscopy)

α Dig – anti-digoxigenin

List of Figures

Figure 1. Type I-E CRISPR effector complex bound to the targeted DNA	18
Figure 2. Magnetic tweezers setup and principle of operation	26
Figure 3. The principle of tracking of magnetic bead in the tweezers	28
Figure 4. Supercoiling of single molecules of DNA by magnet rotations on magnetic tweezers and use of it for R-loop detection.....	30
Figure 5. Purification of <i>E. coli</i> Cascade with g8 spacer crRNA	43
Figure 6. Cascade binding ability, specificity and purity tested by bulk assays	44
Figure 7. Cascade binding to supercoiled plasmid DNA. Bulk assay resolved on agarose gel.....	47
Figure 8. Example trajectories of R-loop formation and dissociation by <i>E. coli</i> Cascade measured with magnetic tweezers	49
Figure 9. Mean dissociation times of Cascade bound target variants (with seed mutations)	50
Figure 10. Development of MT assay to allow repetitive Cascade binding-dissociation cycles.....	51
Figure 11. Shifted force calibration curves of MT assay measurements.....	53
Figure 12. Cascade binding kinetics measured on shortened targets.....	54
Figure 13. Comparison of kinetics for full-length and shortened (PAM-distal 6nt mismatched) targets.....	56
Figure 14. Altered crRNA length containing Cascades tested for R-loop formation using MT assays	57
Figure 15. Altered crRNA R-loop sizes	58
Figure 16. Altered crRNA R-loops tested for stability at high positive torque	60
Figure 17. Synthetic template backbone shows stability issues compared to pUC19 backbone	62
Figure 18. Kinetics of Cas3-mediated cleavage of plasmid DNA bound by Cascade	63
Figure 19. Primed adaptation tested in <i>E. coli</i> KD263	65
Figure 20. High-throughput sequencing (HTS) data of selected spacers via fast adaptation presented on Figure 19	66
Figure 21. qPCR assay to quantify adaptation levels	67
Figure 22. Cas proteins effect on priming mutant target degradation	69
Figure 23. Investigation of short DNA fragments generated in the reactions with Cascade, Cas3, and Cas1-Cas2 proteins	71
Figure 24. Oligoduplex integration by Cas1-Cas2 into plasmid containing CRISPR leader and repeats	72

Figure 25. Cas1-Cas2 and Cas3 plus Cascade activity under titration of KCl concentration 73

Figure 26. First attempt for complete *in vitro* primed adaptation reconstitution.... 74

Figure 27. A model of conformation states of Cascade at different target variants 78

List of Tables

Table 1. Sequences of the non-target strand of the target variants tested in this study (project)..... 46

Table 2. A summary of observations from MT, biochemical and *in vivo* experiments for the target variants tested 76

Novelty and practical use

Novelty of this work is in the use of single-molecule approach to study CRISPR-Cas effector target recognition for a system that supports primed adaptation. Together with complementary assays, this approach allowed to provide evidence that particular changes in the effector complex do not participate in primed adaptation. This key finding encourages development of kinetics-based priming model. As a general science project, this work does not imply any immediate biotechnological application. However, development of single-molecule methods in the field could allow to approach off-targeting issue that is very important for gene editing.

Personal contribution

The greater part of the work reported in results was performed by the author. In particular: protein purification; genetic constructs preparation, including cloning and MT constructs; MT assays; biochemical and *in vivo* assays; preparation of the samples for HTS sequencing; code writing for synthetic genes design; data analysis; majority of the experiments shown on agarose and SDS PAGE gels. At the same time, the author is deeply grateful to collaborators and peers who helped with or taught: how to perform MT assay; gel filtration chromatography workflow for protein purification; advanced Urea PAGE with single-nucleotide resolution; repeats of Cas3 biochemical assays; high-throughput sequencing run and scripts for data analysis; protocols and guidance for *in vivo* experiments; scripts for MT data analysis.

1 Literature review

1.1 Antiviral defense systems of prokaryotes

Every cellular species faces constant pressure to overcome genetic parasites, such as viruses, plasmids and transposons that attack it (1, 2). In order to protect themselves from invasions exhausting resources of cells and populations, bacteria and archaea have developed defense systems that vary in molecular mechanisms and ways they affect the fate of invaded cell (3). To overview these defense systems, it is convenient to focus on viruses, since they are the most common and sophisticated genetic invaders of prokaryotes.

We will start from the very beginning of viral replication cycle, when viral particle recognizes the targeted cell. In order to inject its genetic material (DNA or RNA) into the cell, a virus must recognize specific receptors on cell surface. These are typically either proteins or glycolipids. Broad specificity transport proteins such as porins or selective transporters, pili and flagella, as well as specific lipopolysaccharides are all known to be used by viruses for entry into cells (4). For gram-negative bacteria these can also be structural proteins interacting with peptidoclycan layer. For gram-positives, viruses can directly bind to glucose, galactose or rhamnose residues of polysaccharides presented on the cell wall. Each virus usually displays narrow specificity, infecting only some strains of the species, and much of this specificity is due to the presence of appropriate receptors on sensitive cells.

To prevent viral infection, prokaryotes have developed mechanisms that disrupt receptor recognition. One way is to block receptors. This is achieved by altering conformations of cell receptors, or by physically covering them by specific proteins or lipoproteins (5-7). Interestingly, phages themselves use this method to avoid superinfection by multiple viral particles (called virions) and to prevent new viral progeny from binding lysed cell fragments containing targeted receptors (8). Another way to make receptors unavailable for viruses is conditional expression of receptors' genes (so called phase variation). For instance, *Bordetella* spp. cells stop expression of pertactin autotransporter in phase of "no colonization", so that phages cannot infect cells via this receptor (9). Blocking receptors can bring up to 10^6 -fold resistance to virus infection, but some viruses are still able to infect, adapting to this resistance mechanism (10).

Another strategy prokaryotes use to make their receptors unavailable is to produce extracellular matrix. Carbohydrate polymers cover the cell membrane and protect from phage recognition (11). As an adaptive response, virions of some phages contain enzymes that degrade the polycarbohydrate layer. Sometimes such phage enzymes are expressed into media by infected cell, acting in close proximity of

lysed cells, which may “help” phage progeny (12). Finally, another way for bacteria to prevent receptor binding is to produce competitor molecules that compete with viral particles by binding at their receptor binding sites. In this way, *Escherichia coli* produced microcin J25, part of a toxin-antitoxin system, clogs T5 phage receptor binding sites preventing phage binding (13).

Next level of fighting off a viral infection occurs after a phage particle successfully bound to its receptor and viral genetic material is injected into the cell. To this end, many phages or prokaryotic cells employ the so-called superinfection exclusion (Sie) systems (3, 14). Mechanisms of most of such systems are not yet explained completely understood. All Sie systems contain proteins that are specifically binding viral proteins at or inside the cell membrane and disrupt their function at the step of genetic material injection. For example, blocking of tape measure protein of certain phages prevents normal function of DNA passage channels (15, 16). For gram-negative bacteria, Sie systems can also block the function of phage lysozyme, thus preventing peptidoglycan layer degradation (17). The Sie systems are believed to have originated as a mechanism of preventing superinfection by closely related phages (18). Indeed, in many bacteria Sie systems localize in prophages, supporting this hypothesis (19, 20).

If viral (or plasmid) genetic material is inserted into the cell, prokaryotes still can defend themselves by degrading invading DNA or RNA. Restriction-modification (RM) systems are well-studied friend-or-foe systems that degrade foreign DNA. Classically encoding two main enzymes: restriction endonuclease and methyltransferase, these systems degrade DNA unmethylated at specific sites. The sites are typically of no more than 8 nucleotides in length and are often palindromic. The sites in cellular genome and resident plasmids are methylated and thus protected, preventing self-damage (21). Since methylation is uncontrollable in a sense that it does not specifically distinguish cellular DNA from viral DNA, the latter can be modified as well before it is cleaved, leading to escape from immunity. However, type I R-M systems may exhibit higher probability of strand methylation of hemimethylated sites, which could be regarded as a mechanism favoring host over viral DNA through methyltransferase activity (22). Phages evolved to avoid RM-systems by excluding the recognition sites from their genome (23), changing nucleotides in their DNA (for example, using uracils instead of thymines), or expressing their own methylation enzymes (3, 24).

Restriction-modification systems are diverse and include at least four types in their taxonomy (25). Due to ability to cut DNA at specific places, purified restriction enzymes of type II RM-systems are ubiquitously used in cloning. Some systems are degrading only modified DNA and some have much more complex molecular organization than the classical type II systems (26, 27).

Other systems that degrade parasites' genetic material are CRISPR-Cas systems. These systems have been discovered relatively recently and attracted high interest (28–31). Number of scientific publications dedicated to CRISPR is rapidly growing and exceeded 3000 in 2017. CRISPR-Cas systems are in the focus of research in this thesis, thus, in-depth description of their molecular mechanisms is provided in the following sections. For this chapter, we will say, first of all, that these are adaptive prokaryotic immunity systems (32, 33). CRISPR–Cas systems are present in most archaea and in around 50% of bacteria (34, 35). The common feature of these systems is their ability to acquire short (typically around 20-35 nucleotides long) fragments of viral genomes and introduce them into a special chromosomal locus called CRISPR (or CRISPR array) (36, 37). The introduced sequences are called spacers and are identical to protospacers – sequences of viral DNA they originate from. Each integrated sequence (spacer) in CRISPR locus (38) is separated from the next spacer by a short repeat that often is palindromic (39). The first CRISPR locus was detected in late eighties in *E. coli* (40). In all systems, CRISPR arrays are transcribed and the transcripts, pre-crRNAs, are processed into individual spacer-harboring CRISPR RNA molecules (called crRNAs). The crRNAs are assembled with CRISPR-associated (Cas) protein(s) (encoded by *cas* genes often located close to CRISPR arrays) into ribonucleoprotein (RNP) effector complexes. These complexes recognize viral protospacer sequences, bind them by base-pairing and then induce viral NA degradation by either cleaving it directly or attracting dedicated nucleases (41, 42). This process of targeted genetic material degradation is called CRISPR interference. Adaptation machinery is the most conserved part of CRISPR-Cas systems while genetic and functional organization of interference machinery varies. Most CRISPR-Cas systems target DNA, while some are active towards RNA (and sometimes both) (43). To distinguish between viral protospacer and CRISPR array spacer in cellular genome, some CRISPR-Cas systems rely on the presence of short (2-6 bp) specific sequences next to protospacers named protospacer adjacent motifs or PAMs. Despite their frequent occurrence, CRISPR-Cas systems have a drawback of low mutation tolerance. Indeed, even a point mutation in protospacer or PAM sequence was shown to be enough for a virus to escape CRISPR immunity (44). A particular low tolerance for mismatches is observed at the PAM and PAM-proximal “seed” sequence of the protospacer. In contrast, even multiple crRNA-protospacer mismatches are tolerated at PAM-distant positions without apparent effects on CRISPR interference *in vivo* (45, 46).

Phages avoid CRISPR-Cas not only by acquiring mutations in their genomes, but also by small anti-CRISPR proteins sometimes encoded within viral early genes clusters. Diverse anti-CRISPR proteins with different mechanisms of action have been discovered: they interact directly with various parts of the effector complex and inhibit binding and/or degradation of targeted NA (47–49).

Even if a prokaryotic cell does not manage to prevent viral invasion and becomes a factory for phage production (virocell or ribovirocell), it still can preclude spreading of infection. Often leading to controllable self-death, abortive infection (Abi) systems serve to stop production and spreading of functional virions (50). Abi systems may disrupt viral cycle at any in-cell stage: DNA injection (51), DNA replication (52, 53), transcription at different stages (54), capsid protein synthesis (55), and even DNA packaging into viral particles (56). For many Abi systems the molecular mechanism is not completely understood (57, 58). As one of few studied examples, Rex system of *Escherichia coli* cells depletes ATP concentration via membrane channel expression, which reduces membrane potential and blocks ATP-dependent viral genes expression and viral particles assembly (59). This system activates by sensing phage DNA-protein involved in phage DNA replication-recombination (60). Another Abi system of the same bacterium blocks translation via cleavage of elongation factor EF-Tu, responsible for aa-tRNA binding to ribosome (59). This system is activated by the presence of capsid peptides in the cytoplasm (61). Notably, this system is located in prophage ϕ 14 of *E. coli* K12 genome. This fact may point to viral origin of some Abi systems. Some toxin-antitoxin system in bacteria have been shown to act like abortive systems during phage infection (62, 63). Some of CRISPR-Cas systems may also work as abortive infection systems, degrading all RNA or ssDNA in cell cytoplasm after being activated by binding to the targeted sequence (64, 65).

The many measures by prokaryotes against viruses and counter-measures by viruses to avoid cell immunity allow the continuing race between genetic parasites and cellular defense systems. This ongoing process is believed to significantly influence the evolution of both sides. It is interesting to understand all aspects of this reciprocal battle, and estimate the outcome of this coexistence for evolution of life (66). Great effort made by others allow us to understand mechanisms and classify protective systems. In this thesis, our contribution to the studies of molecular mechanisms of not yet completely understood CRISPR-Cas systems is described. Thus, from now on CRISPR-Cas systems will be our main focus.

1.2 Molecular mechanisms of CRISPR-Cas systems

CRISPR-Cas adaptation is a process of acquiring new spacers. A heteromultimer of Cas1 and Cas2 proteins is responsible for acquisition and insertion of short DNA fragments into the CRISPR array (67, 68). These proteins are the most conserved among all CRISPR systems. Recent structures have revealed a $2_{\text{Cas1}}:2_{\text{Cas2}}:2_{\text{Cas1}}$ dumbbell-like complex (69, 70). So far, both structural analysis and biochemical assays suggest that the Cas1-Cas2 complex inserts spacers as double-stranded DNA with forked (unwound and split) ends (71). Recent studies

demonstrated that Cas1-Cas2 acts similar to viral integrases and a model of integration was proposed, though not yet solidly supported. A protein complex with future spacer bound is believed to catalyze nucleophilic attack on chromosomal DNA at either end of the first repeat of CRISPR array, though, the CRISPR leader to first repeat border is a preferred site of attack (72). After the first attack, at least some of CRISPR-Cas systems do a “quality check” for appropriate length of new spacer, and only after that perform the second attack on the opposite strand at the opposite end of the CRISPR repeat (73). For that, the -OH groups are required at 3'-ends of DNA that is being inserted (72, 74). Thus, right after integration resulting structure has new spacer introduced in double-stranded form between two single-stranded regions of CRISPR repeat. These are likely filled in by cell repair systems (73). Directionality of spacer insertion is crucial for future interference, as PAM in targeted NA (virus) is located at a particular side of protospacer. Directionality is ensured by incorporating one (the most proximal) nucleotide of PAM with the spacer sequence into CRISPR array. This residue of PAM thus becomes a part of repeat rather than the first nucleotide of spacer (75).

Despite being conserved, Cas1-Cas2 complexes have system-to-system variation. Thus, type I-E systems use integration host factor (IHF) that recognizes AT-rich leader part in front of CRISPR array and facilitates Cas1-Cas2 navigation by bending DNA molecule at this region (76, 77). The Cas1-Cas2 complex of type II-A system of *Streptococcus pyogenes*, in contrast, is able to identify leader-repeat sequence in the absence of additional factors (73).

While integration of spacers has been studied to the point of proposing a detailed model, the understanding of the process of spacer selection and capture by Cas1-Cas2 complex remains vague (35). The *de novo* spacer capturing is believed to be a stochastic process that does not distinguish between viral and own cellular genome and, thus, is potentially toxic to the cell. Adaptation of own-genome-derived spacers may be fueled by DNA fragments generated by the reparation and recombination systems nucleases. The RecBCD complex (and its gram-positive analog AddAB) processively degrades DNA ends generated at double-strand breaks (DSBs) until specific Chi sites are reached where degradation stops and reparative recombination initiates (78). This degradation activity is believed to be an important source of short single-stranded DNA fragments that, when reannealed, can be captured by the Cas1-Cas2 adaptation complex. Indeed, more frequently acquired spacers originate from areas around replication fork stalling sites – sources of DSBs – bound by nearest Chi sites (79). Some bias of *de novo* acquisition of spacers from foreign DNA compared to self-originated spacers may be due to lower frequency of Chi sites in foreign DNA thus leading to its more extensive degradation per one RecBCD “run” (78). Consistent with this model, for

linear phage genomes, spacers are selected more frequently from a region between genome end and the first Chi site (80).

For some types of CRISPR systems the Cas2 protein is fused with nuclease-helicase protein Cas3 involved in target degradation (43). Indirect experimental evidence suggests that the *E. coli* type I-E system Cas1-Cas2 complex interacts with target-bound effector *and* Cas3 (81). These observations have brought speculation that interference components may alter the function of adaptation machinery, increasing its efficiency or the other way round (82). As mentioned earlier, spacers are likely integrated into CRISPR array as double-stranded forked DNA fragments that may originate from reannealed degradation products. However, a recent study reported that during primed adaptation spacer-size ssDNA intermediates can be pulled from cells lysates through immunoprecipitation with Cas1 (83). This supports a model of interference-adaptation machinery cooperation during priming and demonstrates that these fragments are excised from only one strand of DNA in targeted plasmid, which could explain the strand bias in spacer acquisition observed during priming.

CRISPR effector complexes serve as devices for targeting viral or other foreign DNA/RNA. Spacer sequence in crRNA guides the effector complex to complementary sequence (protospacer) in targeted NA with a “consensus” (strongly supporting interference) PAM next to it. The search process is believed to be a 3D diffusion during which the effector complex is attracted to strongest PAM sequences (81, 84). It is believed that each time the effector complex binds a strong PAM, the process of matching of crRNA spacer part to DNA is initiated (Figure 1). This nucleation is believed to be directional, starting from DNA unwinding of seed (next to PAM) nucleotides and probing base-pairing between crRNA and the target in this area first (85). If protospacer seed matches crRNA, further base pairing along the target is achieved in a reversible zipper-like fashion by displacing the non-target DNA strand, resulting in a triple-strand R-loop structure (86).

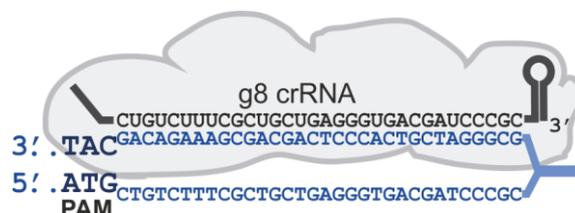


Figure 1. Type I-E CRISPR effector complex (Cascade) bound to the targeted DNA (blue). Cascade requires a particular PAM sequence to be present next to protospacer target. Binding occurs via base-pairing with crRNA complementary to target. Resulting R-loop structure (one DNA strand paired to crRNA, another is in ssDNA form) is stabilized by downstream locking

It is worth mentioning, that effector binding does not require ATP and R-loop structure is formed in the course of extensive conformational changes in the effector, in addition to RNA-DNA pairing. Upon zipping of the RNA-DNA hybrid to the end of the protospacer and formation of full-sized R-loop structure, conformational changes occur in the effector complex that energetically “lock” and stabilize the R-loop (87–89). For type I-E system of *E. coli*, the conformational change that leads to locking mainly involves the Cse1 and Cse2 subunits. R-loop locking is thought to be a signal for recruitment of the helicase-nuclease Cas3 (85), which cleaves DNA at and around the protospacer and, in the presence of ATP, proceeds to degrade foreign DNA in both directions from the bound Cascade (90–94).

1.3 Classification and diversity of CRISPR-Cas systems

The *cas* genes and CRISPR arrays show high diversity. Bioinformatics analysis of Cas proteins allowed the classification of the CRISPR-Cas systems into two classes: with multimeric and monomeric protein subunit structure of effector complexes. The last published classification distinguishes six types (from type I to VI) of CRISPR-Cas systems (43) among which the types I, II, and III represent the best studied systems (31). The different types of CRISPR–Cas systems use different effector complexes and are distinguished by the presence of specific “signature proteins” that are responsible for initial DNA degradation which are, respectively, Cas3, Cas9, and Cas10 for types I, II, and III. type I systems employ a large multisubunit effector complex called Cascade (stoichiometry Cse1₁Cse2₂Cas7₆Cas5₁Cas6₁ with 61-nt crRNA containing 32-nt spacer for type I-E system of *E. coli*) that recognizes double-stranded DNA (dsDNA) targets (86) (95). After target recognition and verification, Cascade recruits the signature protein Cas3—a helicase–nuclease—to degrade DNA (90, 91). In type II systems, the monomeric signature Cas9 protein is both the effector for dsDNA target recognition and the nuclease for target degradation. Using its two nuclease domains, it generates a double-strand break in bound targets (96, 97). Type V effectors are similar with type II and likely evolved from the same family of transposon-originated nucleases (98). The main differences for subtype V-A are the absence of activating RNA (tracrRNA), present in type II and subtype V-B. While type II effectors are known to produce blunt-ended DSBs, type V effectors make staggered DSBs (99, 100). Being compact and simple, type II and V effectors became the preferred tool in CRISPR–Cas-based genome engineering applications (101–103).

In type III systems, the effector complex is multimeric with a helicoid structure similar to that formed by Cas7 subunits oligomer of type I Cascade (104, 105). The

type III effector complex is recognizing not dsDNA but RNA sequences complementary to crRNA spacer (106–108). RNA recognition stimulates nonspecific DNA cleavage activity of the Cas10 signature nuclease that is part of the effector complex (109–111). This DNA cleavage occurs co-transcriptionally (111, 112), allowing to target actively transcribed genes of infecting phages or induced prophages.

1.4 Models of CRISPR-Cas primed adaptation

Mismatches between crRNA and DNA target represent kinetic barriers for R-loop formation that are difficult and sometimes impossible to overcome. Particularly, PAM-proximal mismatches in seed region lead to stronger hindrance and are thought to inhibit the R-loop nucleation (45, 85, 113). In case of mismatches that still lead to R-loop formation, the conformational states of the resulting complex are altered compared to the fully-matched complex (114).

In addition to mediating target recognition during interference, Cascade can promote the acquisition of new spacers from invader DNA, a response called “primed adaptation” or “priming” that allows to update the “invader memory” of the CRISPR-Cas system (94, 113, 115). So far, priming has been observed for type I-B (94, 115–117), I-C (118), I-E (45, 113, 119–121) and I-F systems (115, 122). In type I-E, priming requires all elements of the system’s machinery, i.e., the effector complex, Cas3, and the Cas1-Cas2 adaptation complex (45, 113, 120). The rapid and efficient adaptation typical for priming is stimulated by the recognition of protospacers that form mismatches with the crRNA spacer or at fully matching protospacers that contain suboptimal (not supporting efficient interference, but still supporting some CRISPR-Cas response) PAM sequences (113, 121). Majority of spacers that are acquired in the course of primed adaptation are located *in cis* with priming protospacers. In *E. coli*, the protospacers from which new spacers are selected have almost invariably a consensus interference-proficient AAG PAM, which should increase the ability of CRISPR-Cas system to fight off a genetic invader. While distances between the priming site and the protospacer from which a new spacer is acquired can be substantial (tens of thousands of nucleotides), the efficiency at which new spacers are acquired drops with increasing distance from the priming site (113, 115, 119, 120, 122, 123). In addition to the distance from the priming site, other factors such as sequence, and local context affect selection efficiency of spacers (119, 124). Some “hot” protospacers are selected as spacers with thousands-fold higher probability than other, “cold” protospacers.

Two main alternative mechanistic models of priming have been proposed. In the conformational-control model Cascade is believed to be able to adopt a conformation that supports priming and that is distinct from conformation that

supports interference. The model is based on an observation that for some targets that support priming but strongly attenuate interference, Cas3 recruitment is decreased, but can be restored and altered in the presence of the Cas1-Cas2 complex (81). On such targets, the Cse1 subunit of Cascade adopts predominantly an open conformation in contrast to a closed conformation found in complexes with targets that, once recognized, promote interference (114). Thus, priming could be a consequence of specific recognition of the open-form Cascade-target complex by Cas3 and, subsequently, by the Cas1-Cas2 complex. In the extreme case of this model, acquisition of new spacers could occur without interference initiated at the priming site.

Within the conceptually simpler interference-based model, both interference and priming are consequences of the same process of target DNA degradation. The model is based on observations that (i) during the short time window before their destruction, matching targets with consensus PAMs support primed acquisition of spacers that is considerably more robust than acquisition from mismatched targets that are poorly interfered with (124) and (ii) target DNA fragments generated by Cas3 fuel priming (125). During attenuated but not completely suppressed interference, targeted plasmids and phage genomes can replicate and persist for longer periods of time inside cells despite ongoing CRISPR interference. As a result, Cas3-generated fragments of foreign DNA, which are substrates for adaptation, will also be present for longer time, allowing spacer acquisition events to occur over longer periods. In contrast, a rapid interference reaction quickly depletes invader DNA providing insufficient time for adaptation (126).

1.5 Single molecule techniques to study DNA targeting protein enzymes

A broad range of techniques has been applied to decipher the molecular mechanisms of CRISPR–Cas systems. The combination of *in vivo* and *in vitro* studies, associated with structural snapshots, allowed understanding the pathways of CRISPR systems as briefly described above. Among *in vitro* approaches, single-molecule tools have uniquely revealed the dynamics of effector complexes during target protospacer recognition. Single-molecule fluorescence experiments, such as Förster (fluorescence) resonance energy transfer (FRET) and DNA curtain assays were able to monitor protospacer binding by the type I–E Cascade (81, 127). Furthermore, the dynamic search of DNA targets by *E. coli* Cascade and the type II Cas9 effector taking place by a three-dimensional diffusion mechanism were followed (81, 84). A force-based technique called magnetic tweezers (MT) was also used to study target recognition of CRISPR effectors. Compared to fluorescence approaches, magnetic tweezers are uniquely able to monitor in real time the formation and the extent of the R-loop structure. Additionally, the

dependence of R-loop formation on the applied mechanical stress (torque from DNA supercoiling) can be studied (87). This MT technique provided unique insights into the directionality of R-loop formation, the R-loop stability, and the conditions necessary for target cleavage (85).

Magnetic tweezers became possible to build with the development of sub-nanometer focusing objective stages, thus allowing to track less than micrometer-long DNA molecules with precision (128). It is an efficient method to study any DNA targeting protein that has an ability to unwind (melt) DNA (129). Further, combination of MT with fluorescence detection via total internal reflection fluorescence (TIRF) microscopy allows single-molecule studies of DNA binding enzymes that do not locally melt DNA (130). A combination of MT with FRET optical systems, and, sometimes, with TIRF, allows one to dissect specific functions of mono- and oligomers of proteins being studied (131), and detect conformational changes of either protein or targeted DNA under different torque (i.e., energy density values) (132).

Optical tweezers is another tool that can work with dsDNA, ssDNA, and sometimes other fiber-forming macromolecules (133). Optical tweezers have similar to magnetic tweezers force range (from 0.1 to few pN) but its major drawback is low throughput: while optical tweezers can work with one DNA molecule, MT assay can be designed to track 50 or even more molecules simultaneously. Multiplication of parallel measurements is crucial for getting statistics on rare events. The limit on molecules studied in MT assays is caused by optical system parameters, magnetic field horizontal uniformity as well as computational limitations: high-performance GPUs are used to track few dozens of DNA molecules with frequencies from hundred to thousand hertz. Attempts to allow more parallel measurements have resulted in development of assay variations. Acoustic force spectroscopy (AFS) has a principle very similar to MT but the pulling force is created by acoustic resonance rather than magnetic field gradient (134). The advantage of this approach is its simplicity and better force range, while its major drawback is the inability to generate and control torque, as is easily done with magnetic tweezers (135). Development of this method as a commercial platform allowed simultaneous tracking of few hundred DNA molecules. Another parallelization development exploits an NGS platform to track DNA-protein interactions for up to 10^7 individual DNA molecules (136). The platform initially serves to track millions of signals from fluorescently labeled nucleotides during DNA sequencing. Re-engineering assay to detect signals from fluorescently labeled proteins that target DNA allowed to probe CRISPR-Cas effector complex interactions with millions of target variants, thereby, studying its off-target activity.

1.6 Technology and value that CRISPR-Cas systems have brought up to date

Due to their simplicity and ability to target specific, programmable sequences, type II and V systems Cas9 and Cpf1 effectors, have been used to develop new tools for genome engineering and gene regulation that are easier to generate than ZFNs and TALENs based tools (137). The introduction of CRISPR-Cas based tools boosted genome editing and gene therapy, though have brought many safety-related issues that are subject of hot discussions (138–140). CRISPR-Cas originated enzymes are used for gene “knockouts” by promoter or whole gene removal, gene “knockins”, editing of single nucleotides or short fragments at specific sites, gene regulation, i.e., silencing or activation – typically done by fusion of CRISPR-Cas effector complex to transcription factors –, as well as programmable histone modifications and DNA methylation (141). Hundreds of potential applications for CRISPR-Cas effectors are now being developed. The first organism modified with CRISPR-Cas9, *Agaricus bisporus* (common mushroom), with induced knockout of one of the six *polyphenol oxidase* genes was approved for sales in 2016 (142). Speaking about clinical applications, many concerns are raised about off-targeting effects that can lead to introduction of unintentional and uncontrolled genome damages (143). Bioinformatics, together with high-throughput screening, may help to minimize this risk (136). Even if CRISPR-Cas enzymes cut precisely at the intended target, there is a risk of undesired genome rearrangements (140). *Ex vivo* modification of limited pool of differentiated cells seems, thus, a good strategy of gene therapy development for the nearest future (144). In this case, control of correct modification is possible. Besides, *ex vivo* requires much simpler delivery systems than *in vivo* applications, where all somatic cells in organism might become a target. Currently, several clinical trials of anti-cancer therapies are held in China that rely on knockout of Programmed cell death protein 1 (PD-1) receptor in *ex vivo* derived T-lymphocytes (144). Removal of this immune checkpoint abolishes cancer cells evasion from immune response in many cancer types (145, 146). PD-1 knockout therapy is an example of one of the first gene therapies on humans that are undergoing phase I or II of clinical trials. The number of cancer types that are covered by such trials is going to grow rapidly in the near future.

2 Project objectives

As described in review of current knowledge in the field, CRISPR adaptation in general, and primed adaptation (or priming), in particular, remain insufficiently studied phenomena. It is crucial to understand how the mechanism of target recognition and DNA degradation is connected to primed adaptation. In order to push understanding in this field, we aim to develop experiments, which would discriminate between two main models of primed adaptation.

We have chosen CRISPR-Cas system of *Escherichia coli* since it is the most studied system that supports primed adaptation. In this project, we use magnetic tweezers to study CRISPR-Cas targeting (part of interference). With this method, we aimed to investigate whether all target mutants that support priming display conformational changes reported by other research groups. Besides, having all protein components of *E. coli* CRISPR-Cas system purified, we dissect system's function at each step during primed adaptation. And, finally, we aimed to test our insights in *in vivo* model experiments making sure they are relevant to CRISPR-Cas system function.

3 Materials and methods

3.1 Magnetic tweezers

3.1.1 Principles of the magnetic tweezers setup

Magnetic tweezers (MT) are a single-molecule manipulation and detection instrument extensively used in biophysics. It consists of a microscope onto which motor-controlled magnets are mounted right above (below) a specimen. Specimen is typically a flow-cell, allowing to repeatedly flush desired (e.g., DNA or protein solution) liquid into the field of view. The flow cells also serves as a support for immobilized molecules of study – in our case DNA – that are attached to the surface of cover glass. The key principle of molecule manipulation with magnetic tweezers is the use of superparamagnetic spherical particles (further called beads or magnetic beads) that attach to the tethered DNA molecules via high affinity ligand-receptor (or sometimes covalent) bonds (147, 148). The motor-controlled magnets generate a strong magnetic field gradient in the direction perpendicular to the surface, such that the bead is pulled away from the surface and a stretching force is applied to the tethered biomolecule under study.

A scheme of the setup can be seen in Figure 2. The measurements described below were done using the set up described in (149). Light produced by a light emitting diode (LED, 660 nm emission wavelength) passes a lens telescope (Thorlabs, USA), which focuses the light on the sample. Transmitted light from the sample passes an oil immersion objective (Olympus Ach 100x/ 1.25 oil, Japan) and is deflected by a mirror (Thorlabs, USA) and tube lenses to produce a high-resolution image of the sample on the chip of a CCD camera (TM-6710CL, JAI Pulnix Inc, USA). Images of individual beads in the sample are analyzed in real-time by a custom-developed software on a PC that operates the whole setup. Permanent NdFeB magnets (Supermagnete, Uster, Switzerland) are located just above the flow-cell and can be moved vertically with a precision of 10 μm and rotated using motors with servo-control (M126.PD1 and C150.PD, PI, Germany). The range of forces produced by these magnets on the superparamagnetic spherical particles with 1 μm diameter (MyOne, Invitrogen, USA) is in the range of 0.01 to 6 pN. The focal position of the objective is controlled by a piezoelectric nanopositioning stage (PIFOC P-721, PI, Germany) and allows to introduce precise focus steps with sub-nanometer precision. A second magnetic tweezers setup was used for some experiments described in this thesis. Its main differences compared to the described setup was the use of a faster CMOS camera with higher pixel resolution (EoSenc CL MC1360-63, Germany) and an objective with lower magnification (UplanSApo 60x/1.35 Oil, Olympus, Japan). CMOS cameras are typically used for rapid parallel (multiple objects) imaging with lower noise, while CCD cameras

display much higher sensitivity and are perfect for low contrast and signal applications.

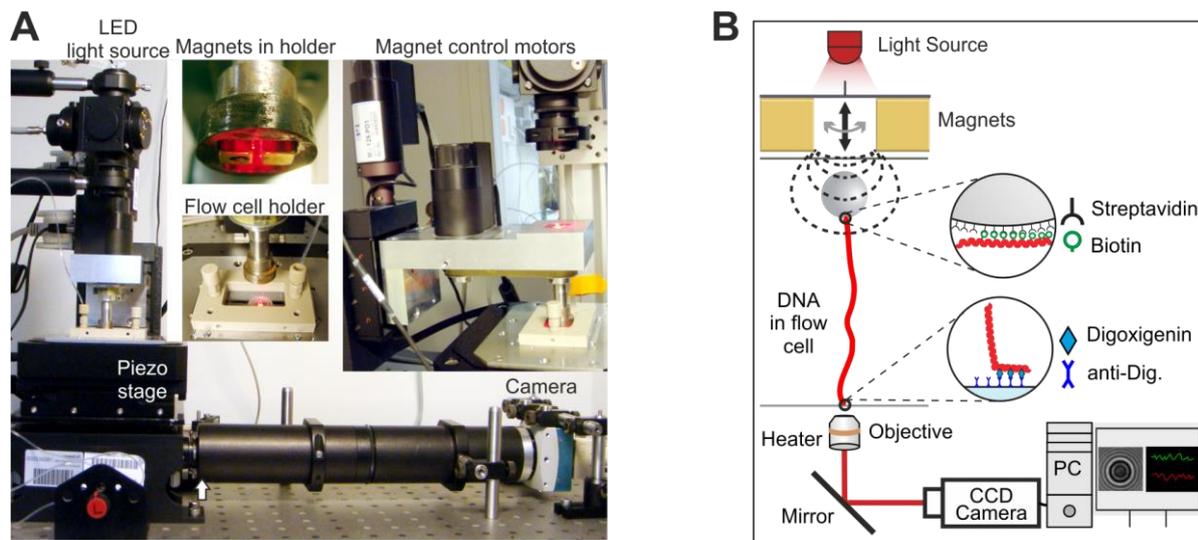


Figure 2. Magnetic tweezers setup and principle of operation (figures were adapted from (129, 150, 151)). (A) Photograph of the setup and individual components. The optical path can be seen in the overview image (left and bottom). Light from the LED goes through lenses, passes between a pair of magnets, illuminates the flow-cell and is collected by the objective beneath the flow cell. Subsequently it is reflected to the right (with respect to the photograph) by a mirror and detected by a camera. (B) Principle of DNA stretching experiments in magnetic tweezers. DNA is attached to the bottom of the flow-cell and to the magnetic bead by high affinity ligand-receptor interactions. The magnetic bead is pulled up by a magnetic field gradient generated by a pair of cubic magnets that are mounted on a translation stage. The bead image is collected by camera and analyzed by a custom-developed software that calculates its position in real time.

3.1.2 Flow-cell preparation

The flow-cell is a special microscope specimen chamber: two glass cover slips 20 x 60 mm in size (Menzel-Gläser, Germany) and a layer of Parafilm “M” (Bemis NA, USA) are assembled as a sandwich. The parafilm layer is cut so that a central channel is formed. Two holes in the upper cover slip provide access to the channel. This way any desired solutions can be exchanged in a convenient manner. The bottom glass cover slip is covered with anti-digoxigenin antibody (α Dig) to graft the DNA molecules that are modified with digoxigenin (Dig) at one end. At another end, DNA is modified with biotin to which streptavidin-coated magnetic beads bind.

The preparation of the flow-cell includes cleaning of the cover slips by sonication in acetone and isopropanol. The top cover slip is cut with a laser cutter to create two holes. The bottom cover slip is coated with polystyrene dissolved in toluene with the help of a spin coater. This covering helps to attach antibodies to the

bottom cover slip, as well as to fix on surface reference beads used during measurements (see below). Parafilm is cut by a laser cutter to match the flow channel form (empty space in the middle with channel contraction on sides). Ends of flow channel in parafilm match to the holes of top cover slip. All three parts are assembled and tightly pressed against each other at the Parafilm melting temperature of about 120 °C. Before use, the flow-cell is incubated with 50 µg/ml αDig (Roche, Germany) for at least one hour in order to cover the surface of the flow-cell with αDig. Then the flow-cell surface is passivated with Bovine Serum Albumin (BSA, New England Biolabs, USA) at room temperature for at least 4 hours (in most cases overnight). BSA is used to reduce the nonspecific adsorption of the magnetic beads to the flow-cell surface. The typical volume of the flow-cell is about 30-40 µl.

Special 3.2 µm sized polystyrene reference beads (Uniform microspheres, Invitrogen DYNAL AS, Norway) are flushed into the flow-cell after washing the remaining BSA. Flushing out BSA after its incubation is necessary to allow proper reference beads fixation on the surface of the bottom cover slip. These reference beads are used to detect relative paramagnetic bead position and subtract the drift of the whole sample holder. Typically, 20 minutes incubation time was used to let the reference beads settle down and fix on the surface of the flow-cell.

3.1.3 Tracking of the bead position in three dimensions

The parameter that is obtained during magnetic tweezers measurements is the extension of DNA. The DNA end-to-end length is calculated from the axial position of the attached magnetic bead. To this end, the bead is imaged in overfocus such that diffraction rings are formed by light scattering on the micrometer-sized bead (Figure 3). Custom-developed software is used to process the image. The lateral positions of the bead (i.e., along x and y) are determined as the center of the diffraction pattern (152). The axial position (z-position, along the direction of the applied force) is determined from the diffraction pattern that changes with the position of the bead with respect to the overfocus. To this end, a so-called look-up table – i.e., a set of diffraction profiles obtained from bead images taken at step-wise increased overfocus - is generated (Figure 3D). To achieve precise overfocus steps, the whole flow-cell is moved with nanometer precision by a piezoelectric nanopositioning stage. The typical range of a look-up table is 6 µm. During the measurement, the radial diffraction profile of the bead is compared to the profiles of the look-up table, which provides the position of the magnetic bead. Simultaneously, a non-magnetic 3.2 µm polystyrene bead, which is immobilized on the flow-cell surface, is tracked using a separate look-up table to collect the different interference profile of the reference bead. This bead serves to

remove the microscope drift. The position difference between magnetic and reference beads is then used to determine the DNA extension value. DNA extension is Z-direction projection of the distance between the two ends of the DNA molecule – from one attached to the bottom of the flow-cell to the one attached to the paramagnetic bead. This value can approach the total length of the DNA molecule itself at high pulling force (low magnets-pair position) and decreases at lower forces or/and supercoiled state of DNA molecule (see below).

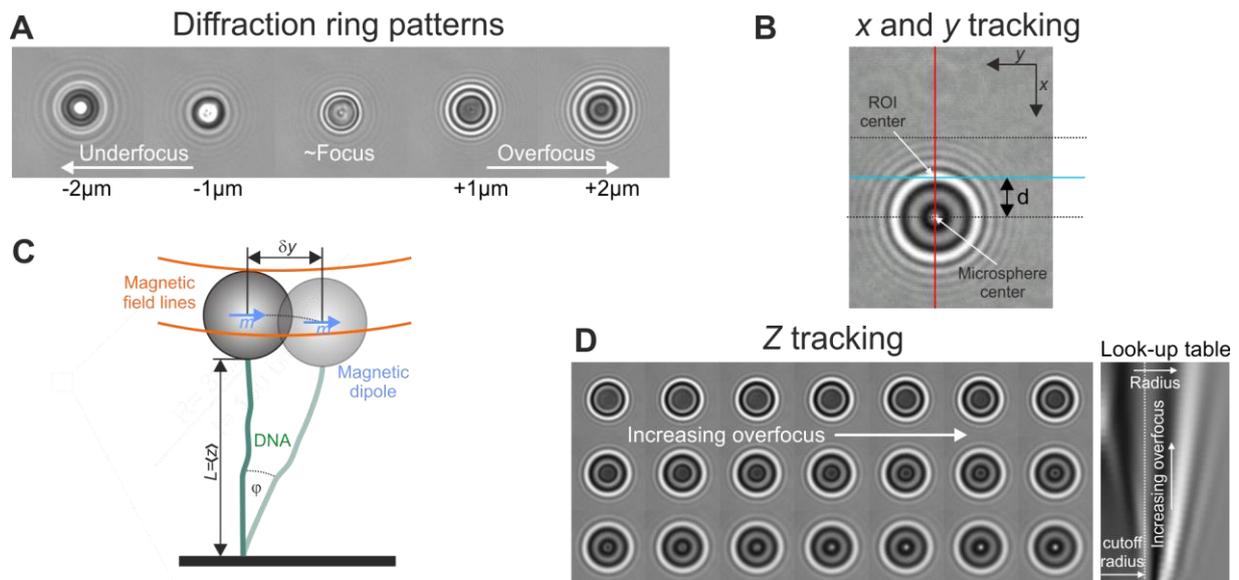


Figure 3. The principle of tracking of magnetic bead in the tweezers, adopted from (151). (A) Images of a magnetic bead taken at different positions with respect to the overfocus. The obtained diffraction ring pattern is analyzed by the tweezers software. (B) X- and Y- position tracking is used to determine the bead position as the center of the rings. (C) Movement of the magnetic bead attached to the DNA under the vertical force imposed by the magnetic field gradient. Fluctuations in X and Y directions are different since the bead dipole orients itself along the magnetic field in the Y direction (D) Z-position tracking. Collection of bead images over increasing overfocus from which a look-up table of radial intensity profiles (with respect to the bead center) is obtained (see right). During the measurement the radial intensity profile of the actual bead image is compared with the look-up table and the Z position of the bead is determined. For the DNA extension length value, the differential position of the magnetic bead and a polystyrene reference bead is calculated.

The rate of bead position tracking during the measurements is limited by the camera frequency characteristics. Here 120 Hz were used for single-molecule tracking and 180 Hz for multiple-molecules tracking.

3.1.4 Preparation of DNA constructs for MT assays

DNA constructs for magnetic tweezers were made of three parts: main body of desired sequence, Digoxigenin (Dig) handle able to bind antiDigoxigenin (α -Dig)

at the bottom of the flow-cell, and Biotin handle able to bind Streptavidin on magnetic bead (Figure 2B). Main body, pUC19 originated, (with g8 protospacer variants cloned in the middle) was about 2200 bp long and prepared by Phusion Polymerase (New England Biolabs, USA) PCR. The PCR was done using primers complementary to pUC19 plasmid at their 3' ends, with 5' overhangs containing SpeI or NotI restriction sites. Handles were produced from the pBluescript II SK+ template by Taq Polymerase PCR using special Biotin-16-dUTPs or Digoxigenin-11-dUTPs (Jena Bioscience, Germany). PCR product of pBluescript II SK+ contains one SpeI and one NotI site in the middle part. PCR products were purified using Nucleospin gel and PCR clean-up Kit (Macherey-Nagel, Germany). Dig handle was cut with NotI-HF enzyme (New England Biolabs, USA), Biotin handle with SpeI (New England Biolabs, USA). Main body was treated with both enzymes to form “sticky ends”. Restriction products were purified using the same purification kit. All three parts were ligated together with the T4 DNA Ligase (New England Biolabs, USA). In case of synthetic genes (see chapter 4.5.4), similar approach was used to produce parts of constructs by PCR. The difference was only in sequences of templates and primers used.

Ligation product was run on 1% agarose gel and the band of the right size (3200 bp) was extracted from it, while limiting EtBr (or other intercalators) contact or ultraviolet exposing which could damage DNA. Nucleospin gel and PCR clean-up Kit (Macherey-Nagel, Germany) was used for extraction of DNA from the gel. Resulting construct was ready to be used in MT assays. Typically the concentration was around from 5 to dozens ng/ μ l in 20 μ l volume. Constructs were aliquoted and frozen at -20 °C, as biotin and Dig are sensitive to multiple thawing-freezing cycles. Thawed aliquots were stored at $+4$ °C for a month.

3.1.5 DNA molecule behavior. Force calibration. Rotation curves.

The DNA molecule attached to the magnetic bead and the surface of the flow cell is stretched by the magnetic field gradient and fluctuates around its equilibrium position due to thermal fluctuations. Its mean height depends on the force applied by the magnet and on the length of the DNA molecule. In order to determine the force imposed by the magnets, force calibration is done for every magnetic bead. To this end, the force is estimated from thermal fluctuations in lateral direction considering the pendulum geometry of the bead-DNA system according to (153):

$$F = \frac{k_B T}{\langle \Delta x^2 \rangle} L \quad (3.1)$$

where F is the force produced by the magnet, k_B is the Boltzmann constant, T is temperature, $\langle \Delta x^2 \rangle$ is the mean square displacement of the lateral bead position x and L is the length of the pendulum (i.e. L is thus not the length of DNA molecule

itself, but the DNA extension value or, in other words, end-to-end distance). The formula is simply derived by calculating the mean pendulum energy and equating it to $\frac{1}{2} kBT$ according to the equipartition theorem.

A dsDNA molecule that is attached with both strands to α Dig at the bottom of the flow-cell and to streptavidin of the magnetic bead can be twisted, i.e., supercoiled (Figure 4). Rotation of the magnet lets the magnetic bead rotate synchronously due to its intrinsic dipole moment conservation. Thus, every magnet rotation introduces (or removes) one turn into DNA.

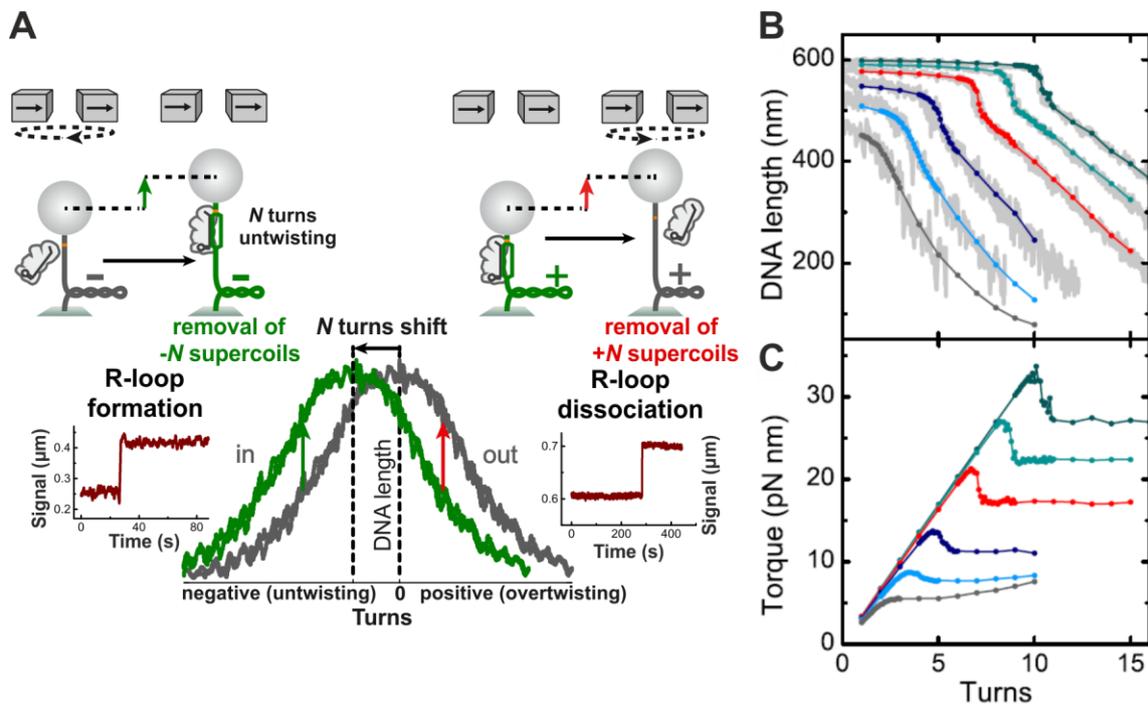


Figure 4. Supercoiling of single molecules of DNA by magnet rotations on magnetic tweezers leads to plectoneme formation (DNA supercoils or superhelices). After plectoneme formation the torque remains constant. This condition is used to measure Cascade binding and R-loop formation. After target location, Cascade complex unwinds the DNA helix and shifts the DNA supercoiled state. (A) Changes of DNA supercoiling during R-loop formation and dissociation (87). Magnet rotation leads to bead rotation and thus to supercoiling of the DNA molecule. DNA “stores” applied turns in a plectoneme form – a structure where a DNA molecule forms a superhelix with itself. DNA length here stands for DNA end-to-end distance (also called DNA extension). When Cascade binds to DNA, it unwinds the targeted region, and therefore, induces positive turns into the DNA molecule. This is detected as a shift of the rotation curve as well as a sudden jump in smoothed time traced signal. The opposite event of R-loop dissociation is seen as back shift to original rotation curve and a jump towards lower supercoiling state (seen as jump up since R-loop dissociation happens). (B) and (C) The rotation curve is the dependence of DNA length on the rotations introduced by magnet. Data shown were obtained for forces ranging from 0.5 to 4 pN (0.25, 0.5, 1.0, 2.0, 3.0 and 4.0 pN, correspondingly). Monte-Carlo simulations show

that the torque reaches a constant level as soon as the first plectoneme forms at B-form of DNA (154).

Initially rotations lead to a linear increase of the torque within the molecule. Then, at a certain torque value, the DNA molecule suddenly buckles and a plectoneme starts to form. This is seen as a linear decrease of the bead height. Plectoneme expansion/growth requires constant amount of work applied to the molecule by the magnet per induced turn. Therefore, the rest of the molecule experiences the same energy density and the same torque value no matter how many rotations will be added. This property, a so-called “torque wrench”, is very convenient for measuring binding of the Cascade complex, since R-loop formation leads to unwinding and localized melting of DNA in the region of protospacer sequence. This introduces additional turns in the whole DNA molecule. If the molecule is in the negatively supercoiled state, an abrupt step in the DNA extension signal will be observed, since some plectonemic turns will be released, leading to a permanent rotational shift, that is easy to detect (Figure 4A). R-loop dissociation, if it happens, is seen as molecule shift back to the original state showed as a grey rotation curve. Some R-loops are highly energetically favorable and dissociate only at high positive torque. In this case, R-loop dissociation will be seen as sudden jump towards lower value of DNA supercoiling (right part of Figure 4A).

3.1.6 Magnetic tweezers experiments to measure R-loop properties

R-loop formation/dissociation and extent of DNA opening (unwinding) experiments on full-length targets were carried out as previously described in 20 mM Tris-HCl pH 8.0, 150 mM NaCl and 0.1 mg/mL BSA at 37 °C. Experiments using the target with CCG and CCC PAM were also performed in the presence of 5 mM MgCl₂. Mg²⁺ facilitates R-loop formation, thus providing a more stringent negative control for Cascade target specificity. R-loop formation was detected at about -7 (7 negative) turns of negative supercoiling and a force of 0.4 pN, corresponding to a torque of -6.7 ± 0.5 pN·nm (148). Rotational shifts were estimated from the linear part on the left side of the rotation curve (87). R-loop dissociation experiments were performed at about +12 turns of supercoiling and a force of 5 pN, corresponding to a torque of $+32 \pm 3$ pN·nm. Each target variant was characterized on at least two individual DNA molecules.

3.1.7 Data analysis: constant probability density events fit exponential distribution

Cascade interaction with DNA is thought of as random Poisson process, with constant binding rate. In other words, the process has the constant probability of the protein to interact over a small time period. This assumed to be true along all

the time of measurements. As shown in the Appendix E, in this case random events of interaction will be exponentially distributed over time. The average time for the events distribution will be

$$T_{\text{exp}} = \frac{1}{\lambda} \quad (3.2)$$

where λ is the exponential rate of the process. Then, as derived in Appendix E, collected in experiments data can be fitted with function:

$$F_N(t) = N * (1 - e^{-\lambda*t}) \quad (3.3)$$

Where N is number of events collected, λ is rate of the process and t is time. By fitting the experimental data with this function, actual rate of the process observed in the experiment can be extracted. This, for example, is applicable to Cascade binding rate and Cas3 cleavage in the single-molecule experiments. Extracted rates are easily compared between each other and allow conclusions about conditions and variations investigated in the experiment. This approach was used to analyze all the data for magnetic tweezers experiments where considerable statistics was collected.

To calculate the error of the fitted rate, and therefore, of the average time, the following expression was used (see Appendix E):

$$\sigma(T_{\text{exp}}) = \frac{1}{\sqrt{N} * \lambda} \text{ or } \frac{T_{\text{exp}}}{\sqrt{N}} \quad (3.4)$$

Where N is number of events, λ is rate, σ is standard deviation.

Data analysis was carried out using customized software code written in Labview and Matlab as well as Origin 9 (87). The author expresses special thanks to Fergus Fettes for development of Matlab code for precise R-loop extent determination via averaging and fitting the rotation curves. Mean R-loop formation and dissociation times were determined from exponential fits to cumulative distributions of the data (Figures 9 and 12). Each mean-time value was calculated from ~25 events.

3.1.8 Method development. Installing the temperature control system

To effectively measure *E. coli* Cascade binding to target variants, temperature of 37 °C had been maintained. The original magnetic tweezers setup did not contain any temperature control. To allow measurements at 37 °C, heating systems were installed on two setups as part of this thesis (see indicated heater in Figure 2B). The systems were obtained from Biotech, USA (150803 and 150819-13) and Okolab, USA, Italy (H401-T-SINGLE-BL and OBJ-COLLAR-2532). For each of these, the central part of system was a heating loop that is installed on t objective;

it transmits heat through the objective and oil to the flow-cell, where the object of study is located (Figure 2B). Since heat transition through several material borders creates a complex heat flow distribution and causes a temperature gradient, the calibration of the device was done exactly at the position where the flow-cell with the sample was located. A thermistor sensor was included into the control device. Additionally a thermocouple sensor of an Agilent U1272A multimeter was used. The sensors were placed into a water or oil droplets placed on the top of cover slip mounted above the objective, where the flow-cell is normally located. As a result of calibration, a correct temperature value of the objective heating loop was established to obtain 37 °C on the top of the glass cover slip, where the interior space of the flow-cell is normally located. Small thermistor sensors were mounted on the top part of the objective, right underneath the flow-cell, for both calibration and further use in assays, to ensure temperature recording (as one cannot mount sensor inside each flow cell. For Olympus Ach 100x objective, the outer sleeve was removed to ensure better heat conductivity. For Olympus UplanSApo 60x objective the outer sleeve was not removed, which led to higher heating loop temperature, i.e., difference (offset) between heating loop temperature and calibrated 37 °C temperature in the flow-cell. Overall, both systems showed around the same temperature fluctuations (± 0.2 °C) from steady state of 37 °C. Both systems can be connected to computer to control and record the temperature, which was tested for the Okolab system. Overall, the resulting system is an efficient solution, which requires minimum resources and is quick to be installed and calibrated.

The effect of temperature on optical system and measurements was investigated during tests done at room temperature and at 37 °C. The piezo stage step movements of 1 μm were detected by software and compared with absolute values. As was expected, small changes in optical density of the immersion oil of the objective, the buffer solution in the flow-cell and temperature expansion of the optical parts negligibly affected measurement: only small deflection of less than 3% were seen. Much bigger effects on the measurement were produced by a change of viscosity of the buffer solution. As described in chapter 4.5.2.1, the viscosity of water reduces by approximately 23% when temperature is increased from 25 °C to 35 °C (155). Since viscosity of water is used for the force calibration, this effect had to be considered to ensure correct force calibration.

3.2 Cloning methods

3.2.1 Preparation of competent *E. coli* cells

Electro-competent NEB 5 α (DH5 α derivative, New England Biolabs, USA) and BL21 (DE3) derivative KD418 (156) *E. coli* cells were prepared for plasmid

cloning and protein expression, respectively, following the protocol provided at Skoltech course MA03046 “Advanced molecular biology techniques”. LB medium was inoculated with overnight culture in 1:100 ratio and grown until optical density at 600 nm wavelength (OD_{600}) of 0.6-0.8. Then the culture was left on ice for 15-30 minutes to cool. Further, the cells were washed with 10% autoclaved ice cold glycerol three times followed by centrifugation at 3000 g for 10 minutes at 4 °C, removing supernatant and resuspending in glycerol solution. The fourth time the cells were resuspended in about 1 ml of volume to reach the desired concentration 2×10^{10} to 3×10^{10} cells per ml. They were snap frozen in liquid nitrogen and transformed by electroporation when needed.

3.2.2 pUC19 targets cloning

Constructs containing the desired targets, including g8 protospacer variants (see Table 1), were cloned into plasmid pUC19 (New England Biolabs, USA) at a unique single SmaI site by blunt end ligation. The 73-bp insert DNA carried the target sequence (PAM and protospacer variant) in its center with “shoulders” to ensure target inviolability. Complementary oligonucleotide pairs used for cloning were ordered from Sigma (Germany) or Eurofins Genomics (Appendix B). Oligonucleotides were phosphorylated by T4 Polynucleotide Kinase (New England Biolabs, USA) in buffer containing ATP and hybridized. The resulting oligoduplex solution was used in a subsequent ligation reaction with SmaI linearized and dephosphorylated pUC19. Ligation products were transformed into NEB 5 α *E. coli* cells, and cells were plated on LB-Agar plates containing 50 μ g/ml Ampicillin. Colony PCR over targeted sequence and surrounding plasmid sequences was used to select recombinant plasmids. Plasmids were purified and the presence of correct protospacer variant in single copy and proper direction was confirmed by sequencing. Plasmids were used as templates for PCR for MT constructs preparation. In addition, plasmids were used in bulk Cascade binding and Cas3 assays and for *in vivo* priming tests.

3.2.3 Synthetic genes cloning

Synthetic genes is a special DNA construct – 2-kb length sequence cloned into a plasmid vector. It was intended to be used as a substitute for pUC19 backbone sequence for the magnetic tweezers constructs. It was designed specifically to have minimum potential interactions with Cascade complex containing certain crRNA sequence. During its design, five consensus PAM sequences for the CRISPR-Cas system of study were avoided in the sequence, as were sequences matching to the first five nucleotides of the targeted protospacer (see Appendix C). Because the

construct was designed from scratch intentionally and ordered for the synthesis in Eurofins Genomics (Luxembourg) as two 1-kb regions, name “synthetic genes” stuck for this construct. This backbone for the magnetic tweezers is intended to have minimum interactions with Cascade.

Synthetic genes were cloned using almost the same methods as for pUC19. Two synthetic genes were cloned together: they were produced as two 1-kb fragments “SG-1” and “SG-2” on commercial plasmid vectors pEX-K4 (Km^R) by Eurofins (sequence is provided in Appendix B). The SG-1 fragment was cut out of the plasmid using two Sall sites. It was purified and extracted from agarose gel, then cloned into linearized and dephosphorylated pEX-K4 SG-2 backbone (linearized with Sall). Correct clones were selected by PCR and confirmed by sequencing.

The resulting construct was changed by Quick Change II Site-Directed Mutagenesis Kit (Agilent Technologies, USA) to eliminate one of the Sall sites. The final version is suitable to clone into it any short fragment of desire, including protospacer sequence. Restriction sites Sall and PspOMI in direct proximity allow to follow the same strategy as for pUC19, but with more efficient sticky-end cloning. Restriction sites were specifically selected to avoid any potentially active PAM for *E. coli* Cascade (see Chapter 3.7).

3.3 Protein purification and handling

Cascade containing the g8 spacer crRNA was overexpressed in BL21 Star (DE3) *E. coli* derivative strain KD418 lacking CRISPR locus and *cas* genes in its genome (156) co-transformed with plasmids pCDF-casABCDE, a derivative of pWUR400 (157) encoding the Cascade complex with N-terminally Strep-tagged Cse2 subunit (156) and pWUR615 containing seven g8 spacers in a CRISPR array (46). Briefly, Cascade complex was purified by affinity chromatography on a Strep-trap column (157) followed by size-exclusion chromatography using a Superose 6 (GE) gel filtration column.

Expression was done in 0.5-1.5 L of LB media with 100 µg/ml Streptomycin (Sm) and 35 µg/ml Chloramphenicol (Cm), inoculated at 1:100 with saturated overnight culture and grown at 37 °C, induced at OD₆₀₀ = 0.5 with 1 mM IPTG and shaken at 16 °C overnight. Cells were harvested by centrifugation at 10 000 g, 4 °C, for 15 minutes and stored at –80 °C. For Cascade purification, cells were lysed by sonication in 100 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, pH 8 (Wash buffer for Strep-Trap column) containing 5 mM β-mercaptoethanol and 0.1 mM PMSF. Lysate was centrifuged at 45 000 g, 4 °C, for 45 minutes. Supernatant was carefully removed and filtered through 0.22 µm membrane filter.

Filtered lysate supernatant was applied to StrepTrap™ HP column (GE Healthcare Bio-Sciences, Germany). The column was washed with several column volumes of buffer and Cascade was eluted by a gradient of d-Desthiobiotin (Sigma-Aldrich, USA). All buffers were used according to the manufacturer protocol. Protein containing gradient fractions were pulled together and concentrated using 100 kDa cutoff Amicon® Ultra-0.5 Centrifugal Filter Devices (Merck, Germany). The material was next checked by 12% SDS PAGE for the presence and expected proportions of all individual subunits of the Cascade complex (Figure 5A).

First experiments with one-step purified Cascade protein, conducted in the presence of Mg^{2+} , showed that additional purification was needed to get rid of nuclease contaminant in the Cascade stock. Thus, Size Exclusion Chromatography (SEC) was applied using Gel Filtration column Superose™ 6 10/300 GL (GE Healthcare Bio-Sciences, Sweden). Running buffer was 20 mM Tris-HCl pH 8, 200 mM NaCl, 1 mM TCEP (same as storage buffer for Cascade). Fractions were checked by SDS PAGE (Figure 5B) and elution peak of Cascade monomer with all subunits was collected and concentrated with Amicon® 100 kDa cutoff concentration devices. Concentration of protein was measured by light absorbance at 280 nm using extinction coefficient value and adjusted to 8 μ M. 8 μ M stock was split into aliquots (usually 5 μ l) and snap-frozen in liquid nitrogen, stored at minus 80 °C. Cascade storage buffer was 20 mM Tris-HCl pH 8, 200 mM NaCl, 1 mM TCEP. For use, one aliquot was thawed, diluted to 2 μ M and kept at -20 °C in the same buffer with 50 % glycerol for no longer than 21 day.

Extinction coefficient value for the *E. coli* Cascade complex was kindly provided by Prof. Scott Bailey (Johns Hopkins University, USA). It can be calculated manually as a superposition of the extinction coefficients for each separate Cascade subunit and crRNA. Protein subunits coefficients can be calculated using ProtParam tool available on the online ExPASy server (158). Nucleic acid absorbance at 280 nm wavelength can be estimated using online service “UV spectrum calculator” by the Integrated DNA Technologies (USA). Resulting extinction coefficient in both cases was 725000 $M^{-1}cm^{-1}$, making it a reliable value.

The Cas3 protein was stored at 20 μ M concentration at - 80 °C in 20 mM Tris-HCl pH 8, 200 mM NaCl, 1 mM TCEP. An aliquot was thawed and diluted to 4 μ M and kept at -20 °C with 50% glycerol for no longer than 7 days. All comparative kinetics measurements were done in one day to avoid errors due Cas3 loss of function. Cas3 protein was kindly provided by the group of our collaborator Prof. Scott Bailey (Johns Hopkins University, USA).

Cas1-Cas2 complex was received and stored in 20 mM Tris-HCl pH 8, 500 mM NaCl, 1 mM DTT, 50 % glycerol and was kept at -20 °C at 35 μ M concentration.

Proper amount was diluted to lower concentration right before usage. The Cas1-Cas2 protein was kindly provided by the group of our collaborator Prof. Virginijus Siksnys (Vilnius University, Lithuania).

Cascade with altered crRNA lengths was provided by Dr. Konstantin Kuznedelov in the form of protein-crRNA complexes lacking the Cse1 (CasA) subunit (159). The Cse1 subunit was purified from the same strain as Cascade (KD418) but transformed with the pET30 (Km^R)-based plasmid containing the *cse1* gene (see Appendix B for the map) under T7 RNAP promoter control. The plasmid was generously provided by Dr. Konstantin Kuznedelov (Rutgers, The State University of New Jersey, USA). The recombinant Cse1 protein has 6xHis-tag at the N-terminus and was purified similar to Cascade, but with Ni²⁺ loaded HisTrap HP 5 ml column (GE Healthcare Bio-Sciences, Germany) in 50 mM HEPES pH 7.5, 300 mM NaCl, 5% (v/v) glycerol buffer. Binding to the column proceeded without imidazole, wash was done with 10 mM imidazole, and elution was done in gradient for up to 200 mM imidazole in the buffer. SEC was performed on the SuperoseTM 6 10/300 GL (GE Healthcare Bio-Sciences, Sweden) column to remove imidazole and possible protein contaminants. The SEC running buffer was 20 mM Tris, pH 7.5, 200 mM NaCl, 5% glycerol, 1 mM TCEP (see Appendix B for elution profiles). Protein was concentrated on Amicon[®] centrifugal filters (Merck, Germany) with 10 kDa cutoff, snap-frozen in liquid nitrogen and stored at -80 °C. Cse1 function was tested for reconstitution ability with other protein in bulk plasmid shift assays (Appendix B).

3.4 Bulk protein assays

3.4.1. Cascade binding assays

Cascade binding ability was tested in bulk assays, where in a tube, plasmid containing targeted protospacer variant was mixed with Cascade, incubated, and then run on agarose gel (1% typically). The shift of targeted plasmid was detected upon Cascade binding. Various buffers were used for Cascade binding assay as not only Cascade binding, but further Cas3, Cas1-Cas2 or both were used in some experiments. MT buffer 20 mM Tris-HCl pH 8, 200 mM NaCl, 0.1 mg/ml Bovine Serum Albumin (BSA) was used for Cascade binding experiments (Figure 6A). Cas3 reaction buffer (20 mM Hepes-KOH pH 7.5 supplemented with 35 mM KCl, 10 mM MgCl₂, 10 μM CoCl₂, 1.5 mM ATP and 1 mM TCEP) was used for Figure 6B. For priming reconstitution, KCl gradient (from 25 mM to 80 mM) in Cas3 reaction buffer was used. Room temperature or 37 °C was used depending on assay purpose. From 5 to 10 nM concentration of targeted plasmid was used. Cascade concentration was 100 nM unless specified differently.

3.4.2 Cas3-mediated DNA cleavage experiments

The assays were carried out in 20 mM HEPES-KOH pH 7.5 supplemented with 35 mM KCl, 10 mM MgCl₂, 10 μM CoCl₂, 1.5 mM ATP and 1 mM TCEP. First, Cascade binding was ensured by incubating 10 nM of pUC19 plasmids containing targeted sequences with 100 nM Cascade (g8 crRNA) at 37° C for 30 minutes. Cascade binding was seen as a small but noticeable shift towards lower mobility (Figure 6). To measure DNA degradation, 100 nM Cas3 was added to the Cascade-bound plasmid. The reaction was allowed to proceed at 37 °C for variable times and was stopped by adding 30 mM EDTA and rapid cooling on ice. Reaction products were separated on 1% agarose gel and visualized by ethidium bromide staining using a Bio-Rad gel imaging system. For each target variant, the intensity of Cascade-bound plasmid in absence of Cas3 was taken as zero-time reference. The processed fraction of plasmid was calculated from the intensity decrease of the supercoiled plasmid species normalized by the zero-time reference.

3.4.3 Cas1-Cas2 integration assays

Cas1-Cas2 reactions appear to be very sensitive to KCl concentration, and the final buffer used for them is Reaction Buffer mentioned before. Cas1-Cas2 concentration used was 75 nM following (72). Higher concentrations of protein are not recommended due to stickiness issues. Integration assays were used to test the Cas1-Cas2 activity. As an integration substrate a 33 bp DNA oligoduplex was used in 200 nM concentration. The oligoduplex was left with OH- group at 3' end of each oligonucleotide which was shown to be essential for integration (72). Cas1-Cas2 showed integrase activity with this duplex on the plasmid containing CRISPR leader region with repeats (pCRISPRmini) derived from pUC19 with *E. coli* CRISPR leader and repeat cloned into it. Reaction product was seen as relaxed open circle plasmid and topoisomers accumulation, believed to be the products of integration-disintegration reaction.

3.4.4 *In vitro* reconstitution of priming

Targeted plasmid pT7blue containing g8 protospacer variants was mixed with Cascade in Reaction buffer and incubated for 30 minutes to form R-loops. Cas3 was added for target degradation. Cas1-Cas2 and plasmid pCRISPRmini were added immediately after, or after one-hour incubation at 37 °C. The products were analyzed on agarose or denaturing (Urea) PAGE gels to reveal target degradation, insertions to the pCRISPRmini plasmid, and for the presence of potential ~33 nucleotide long fragments (possible source for primed adaptation).

3.4.4.1 DNA extraction with phenol-chloroform-isoamyl alcohol. Ethanol precipitation

For experiments studying Cascade, Cas3, and Cas1-Cas2 action on plasmid DNA containing target protospacer, DNA extraction after the reaction was done in some cases. Firstly, it allowed to purify and concentrate short fragments (less than 100 nt long) of DNA generated in the reaction. Secondly, it allowed to obtain concentrated products of CRISPR interference and primed adaptation for further use as integration substrates in Cas1-Cas2 integration assays, or for downstream deep sequencing analysis.

Phenol-chloroform-isoamyl alcohol (25:24:1, Sigma, Germany) was used for post-reaction DNA extraction done according to manufacturer's recommended protocol. The reaction product was mixed by vortex shaking with organic solvent mixture in 1 to 1 ratio. The phases were separated by centrifugation for 5 minutes at 12000 g. Water phase containing nucleic acid was carefully removed.

Nucleic acids were precipitated by addition of NaOAc (sodium acetate) to 300 mM final concentration and 2 volumes of absolute ethanol, incubation at minus 80 °C for 20 minutes, and centrifuging for 1 hour at 16000 g, 4 °C, removing the supernatant and drying the tubes under the vacuum. For denaturing Urea PAGE, samples nucleic acids pellets were resuspended in 10% pure formamide (Applichem, Germany).

3.4.4.2 Urea PAGE

Denaturing Urea Polyacrylamide Gel Electrophoresis (Urea PAGE) was used to resolve short DNA and RNA fragments, typically in 10-100 nucleotides length range. Urea PAGE recipe is: 7 M Urea (Applichem, Germany), 20% Acrylamide:Bis-acrylamide (19:1, Carl Roth GmbH, Germany), 1x Tris-borate-EDTA buffer (Carl Roth GmbH, Germany). The gel was prerun before loading to reach the temperature of 45 °C, samples loaded and run at 45 °C, with power no more than 30 W for about two hours. To stain single-stranded DNA and RNA, SYBR® Gold nucleic acid gel stain (Invitrogen, USA) was used. Stained nucleic acids were imaged in ultraviolet spectrum using Gel-Doc Bio-Rad imaging system.

3.5 Detection of primed adaptation *in vivo*

3.5.1 Priming assay and CRISPR array elongation analysis using agarose gels

Primed adaptation *in vivo* was studied using *E. coli* KD263 cells (K-12 F+, *lacUV5-cas3 araBp8-cse1*, CRISPR I: repeat-spacer g8-repeat) as described in (83). Cells were transformed with pUC19-based plasmids containing g8

protospacer variant (Table 1, Figure 19). Single colonies were picked, inoculated in LB medium containing 100 µg/ml of ampicillin and grown overnight. The cultures were then used to inoculate fresh LB without antibiotic and cells were grown for few hours until an OD₆₀₀ of 0.4 was reached. Expression of *cas* genes was induced by the addition of 1 mM IPTG (induction of the *cas3* gene) and 1 mM arabinose (induction of operon containing genes encoding Cascade subunits and Cas1-Cas2). At various times, 10-µl culture aliquots were withdrawn and diluted 1:10 in deionized water. 1 µl of diluted cultures was used in a 20-µl PCR reaction with Taq polymerase using the 5'-AAGGTTGGTGTCTTTTTTAC-3' and 5'-GTCGCTGCCGTGACGTTATG-3' primers to amplify CRISPR array (including part of the leader and all repeats and spacers). The PCR product was 308 bp long without newly incorporated spacers and 369 bp long with one newly incorporated spacer. The PCR products were analyzed on 2% agarose gels. Gel images were quantified using Image Lab 5.0 (Bio-Rad, USA) software. At least two independent experiments for each time point and each target variant were performed to calculate priming efficiency.

3.5.2 qPCR approach to develop and measure the adaptation score

The efficiency of primed adaptation was also measured using qPCR. The amount of CRISPR arrays that acquired a particular plasmid-derived spacer (hotspot 1, HS1) was quantified and normalized by the amount of the *gyrA* gene on the bacterial genome (See chapter 3.6 and Figure 21). For qPCR amplification of extended CRISPR arrays primer 5'-CATGAGTGATAAACTGCGGCC-3' complementary to HS1 and 5'-AAGGTTGGTGGGTTGTTTTTATGG-3' complementary to the CRISPR array leader were used. For *gyrA* amplification primers 5'-CGGTCAACATTGAGGAAGAGC-3' and 5'-TACGTCACCAACGACACGG-3' were used. DNA amounts were obtained from the qPCR cycle threshold using calibration curves from diluted DNA samples (Figure 21A and B). The adaptation score was calculated as percentage of CRISPR arrays that adapted the HS1 spacer over all CRISPR arrays in the sample (see test measurement in Figure 21C). In all primed adaptation experiments the adaptation score is significantly below 100% since spacers from HS1 represent only a subset of all acquired spacers. The adaptation score is nonetheless proportional to the overall level of adaptation, since all targets supported comparable patterns of spacer acquisition (see high throughput sequencing experiments below).

3.6. High throughput sequencing of CRISPR array

KD263 *E. coli* cells harboring pUC19-based plasmids with g8 protospacer variants (Table 1) were collected 6 hours after induction of *cas* genes expression. 1 μ l of ten-fold diluted culture was used in a 20- μ l PCR reaction with 5'-AAGGTTGGTGGGTTGTTTTTATGG-3' and 5'-GGATCGTCACCCTCAGCAGCG-3' primers to amplify sequences between the leader region to the priming g8 spacer (118 bp in KD263 with unexpanded array). PCR products were separated on 2% agarose gels and a 179 bp band that corresponds to a single acquired spacer was excised and purified using a GeneJet Gel Extraction and DNA Cleanup kit (Thermo Fisher Scientific, USA). ~100 ng of purified DNA amplicons from each target were sequenced using Illumina MiniSeq system according to the recommended protocol of the manufacturer. Each amplicon was read for 150 nucleotides from each side. Results were trimmed and paired using the CLCgenomics software (Qiagen, Germany). Further analysis was done using in-house developed R scripts. Acquired spacer sequences were identified, counted for frequency and mapped onto the pUC19 plasmid backbone. Graphical visualization of results was done using in-house developed EasyVisio1500 software. The author is grateful to Dr. Ekaterina Savitskaya for run of Illumina system, development of R scripts and teaching how to analyze data.

3.7 Synthetic genes design

Synthetic genes were designed in order to eliminate any possible Cascade off-target activity and ensure binding only to the target protospacer. The sequence of synthetic genes was designed to eliminate 5 PAM known at the time to be able to cause interference in the cell (45, 136) - AAG, ATG, AGG, AAA, GAG. The second requirement for the sequence was absence of sequences matching the seed sequence of g8 protospacer 5'-CTGTC-3' (not-targeted strand); and another protospacer used for *in vitro* studies in a parallel project: 5'-CCAGT-3'. The rule was set to have at least two mismatches with these sequences along the entire sequence of synthetic genes to ensure low efficiency to the Cascade crRNA base pairing and, therefore, limit attempts of the Cascade to initiate R-loop formation anywhere but on the target protospacer. To generate the sequence, a program in Matlab was written, which implemented simplest Monte Carlo simulations (code provided in Appendix C.1). About 30 kb of such sequence was checked (1 kb after another) in Mfold service for minimal secondary structures like hairpins (160). Then they were ordered and cloned as described above.

3.8 SDS PAGE

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS PAGE) was used to analyze protein composition and purity. The recipe of the gel is: for “stacking” gel part, 4.875% Acrylamide:Bis-acrylamide (AA:BA, 37.5:1 ratio, “Rotiphores Gel 30”, Carl Roth GmbH, Germany), 187.5 mM Tris-HCl pH 6.8, 0.1% SDS, 0.1% Ammonium persulfate (APS), 1.25 μ l/ml N, N, N', N'-tetramethylethylenediamine (TEMED). For “separating” gel part: 12% AA:BA (37.5:1), 375 mM Tris-HCl pH 8.8, 0.1% SDS, 0.1% APS, 0.5 μ l/ml TEMED. Gel was run on 25 mM Tris base, 192 mM Glycine and 0.1% SDS buffer at 100 V for stacking and 200 V for separation. After the run, gels were stained with Coomassie Brilliant Blue staining solution or ready-to-use Roti®-Blue quick staining solution (Carl Roth GmbH, Germany) and imaged. Usually, Protein PAGERuler™ Unstained Protein Ladder (Thermo Fisher Scientific, USA) was used as a molecular weight marker.

4. Results

4.1 Cascade specifically binds plasmids containing targeted sequences

As mentioned in chapter 3.3, Cascade was purified by affinity chromatography followed by gel filtration (size exclusion) chromatography (SEC) step (Figure 5). Subunit composition was in accordance with published Cascade stoichiometry (86). In order to test activity and function, bulk assays were performed after each new Cascade batch purification. Plasmids containing targeted protospacer were mixed with Cascade, incubated, and then resolved by native agarose gel electrophoresis. Since Cascade molecular weight (~ 405 kDa) constitutes ~24% of the ~2.8 kb pUC19 plasmid with cloned short target sequence (1705 kDa), Cascade binding results in a significantly slower migrating complex and can be detected visually (Figure 6A). Well-studied crRNA with g8 spacer and corresponding plasmid protospacer pair was used in most experiments in this project (156). The buffer used was based on Cascade purification buffer and supported robust activity of *Streptococcus thermophilus* Cascade (St-Cascade) (85). Purified Cascade-g8 crRNA complex bound plasmids containing wild-type g8 protospacer with consensus ATG PAM (non-targeted strand nomenclature) (45). In contrast, plasmids with no protospacer or with CCG in place of PAM (a sequence of CRISPR array repeat found in the position of PAM) showed no binding after a one-hour incubation. Thus, target-specific binding and correct composition of protein and crRNA (further chapter 4.8.3) allows us to conclude that our Cascade preparation is functioning in proper way. Interestingly, complete plasmid binding after just few minutes at both 37 °C, room temperature and, possibly, even on ice (with a caveat that samples were transferred to room-temperature agarose gel

wells; Figure 6C) was observed. The complex of Cascade bound to plasmid also survived a freezing-thawing cycle. Thus, the complex is very robust.

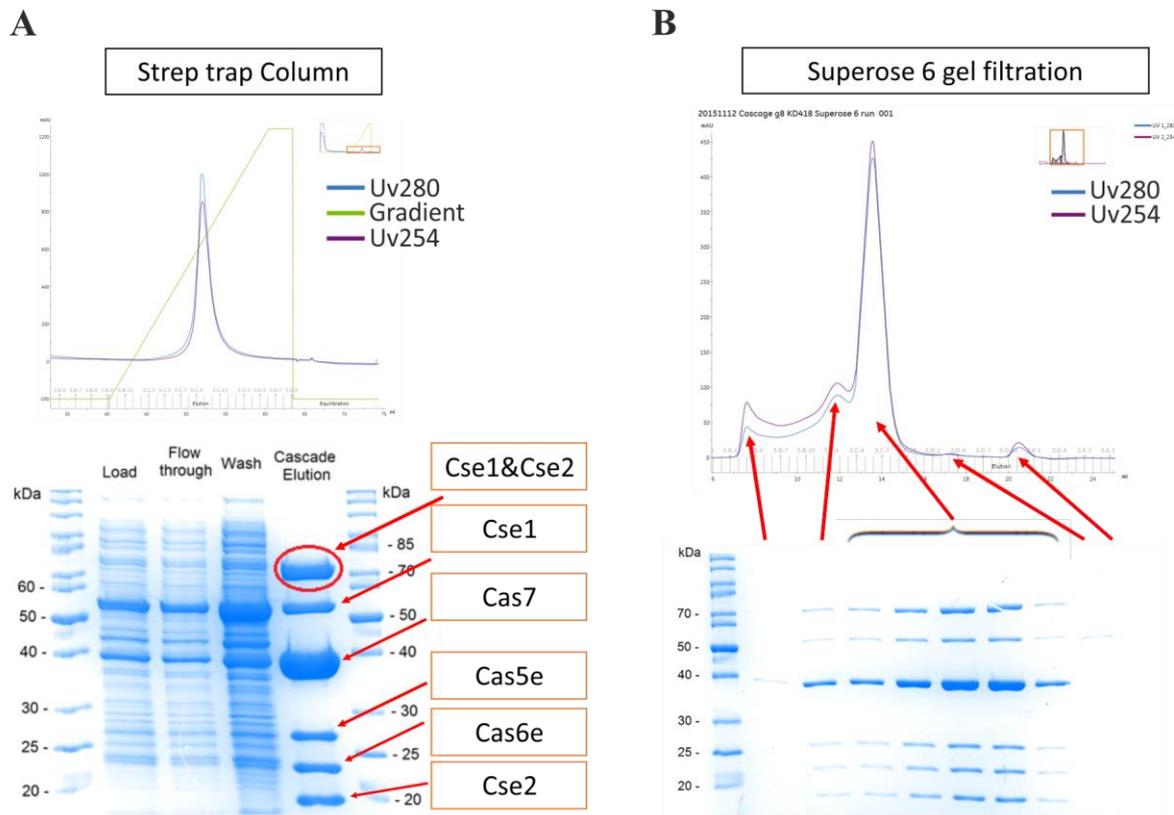


Figure 5. Purification of *E. coli* Cascade with g8 spacer crRNA. (A) First step purification with Strep-trap affinity column. On the top there is a profile of elution with gradient of dethiobiotin (Sigma, Germany). Peaks of absorption at 280 nm and 254 nm at the equal intensity show elution of protein containing nucleic acids (in our case RNA). SDS PAGE confirms that eluted complex contains all Cascade subunits (161). (B) Top: gel filtration column elution profile. The highest peak contains active Cascade with no nuclease contamination. No other proteins apart from Cascade subunits are detected on SDS PAGE.

4.2 Removal of nuclease contamination in Cascade by gel filtration chromatography

SEC was performed not only to get rid of incorrectly folded RNP complexes, but also to remove contamination. Having a homogeneous population of properly folded Cascade also allowed to measure functional Cascade concentration more precisely, removing batch-to-batch variance. Since Cascade was intended to be used not only for target binding studies but also for reconstruction of the entire interference process (includes Cas3 recruitment and DNA degradation) and primed adaptation, it was important to have nuclease-free Cascade. It appeared that after affinity purification on Strep trap column Cascade displayed Me^{2+} (Mg^{2+})-dependent nuclease activity. Additional step of SEC removed this contamination (Figure 6B, 6C and 6D).

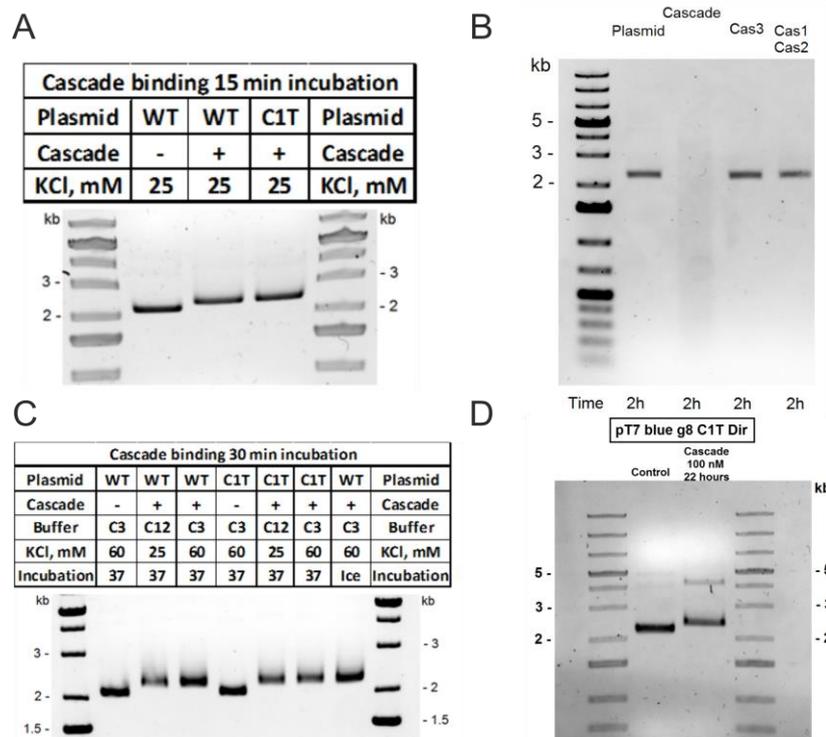


Figure 6. Cascade binding ability, specificity and purity tested by bulk assays. (A) Cascade binding ability tested on wild-type (WT) and mutated target C1T (see next chapter). Plasmids bound to Cascade migrated slower due to Cascade significant mass. Even mutated target appears to be fully (all plasmid in sample) bound by Cascade in a matter of 15 minutes. (B) Cascade purified with only affinity chromatography showed nuclease contamination which is Mg^{2+} dependent (not shown on the gel). Cas3 and Cas1-Cas2 are free from nonspecific nuclease contaminants. (C). Cascade after gel filtration shows no nuclease contamination either in presence of Mg^{2+} (buffers C3 and C12) or in the presence of Mg^{2+} and ATP (buffer C3). (D) Incubation of Cascade with mutated target for 22 hours at 37 °C in presence of Mg^{2+} . Only effect seen is target binding, with small conversion of plasmid into relaxed form. It demonstrated that contaminating nuclease was removed. Assay also demonstrates stability of Cascade being bound on mutated target for a long time.

4.3 Selected target variants cover main possible outcomes of CRISPR immunity response

In order to study Cascade binding to g8 protospacer variants that result in different interference and adaptation efficiencies, several mutations were selected (Table 1). These included two fully matching protospacers with interference-proficient ATG and AAG consensus PAM sequences. The ATG PAM is initially present in the M13 phage genome, where g8 protospacer originates from (156). A g8 crRNA reduces M13 infection efficiency by 10^7 -fold (46). For g8 protospacer with ATG PAM it is known that substitution of the first cytosine for thymine in not-targeted (not base-pairing with crRNA) strand supports strong primed adaptation *in vivo* (113). This mutation, further named “C1T”, was thoroughly studied and was also shown to reduce the interference rate. Another variant, fully matching g8

protospacer with ATT PAM, further called “G-1T”, also was shown to support priming (113). Thus, G-1T and C1T targets were chosen to represent priming substrates for further experiments. Further, several consecutive mutations in the seed region were chosen, as they were earlier shown to reduce interference, leading to viral escape from CRISPR defense similarly to the G-1T and C1T variants (46). When this project was initiated, no seed mutations beyond the second position were tested for ability to support priming, thus, the priming properties of T2A, G3T, and T4G variants were unknown. Finally, as negative controls that should not support any R-loop formation we used CCG and CCC trinucleotides in place of PAM. CCG is a sequence of repeat in front of spacers in CRISPR array. CCC at the PAM position was shown to be as inert as CCG in the work of Xue C. et al. (121). As was shown in Semenova et al. (46), single substitutions beyond protospacer position 8 were unable to efficiently escape from CRISPR interference. This makes mutations further than the 8th nucleotide not interesting for our study, as fast interference and no prominent priming are expected. While first 8 nucleotides after PAM, which when mutated lead to phage escape, are all promising to study priming and R-loop properties (except for 6th nucleotide which does not base pair to crRNA), we here focus on the first four positions closest to PAM. The first four nucleotides, in our opinion, should make the most impact on interference when mutated. As revealed in this work and shown further, their range was enough to reveal R-loop properties important for scientific questions under consideration.

Targets with mutations in PAM-distal protospacer parts were also created to study the effect of this target variation on R-loop. It was previously shown for St-Cascade of I-E type that lack of complementarity between crRNA and protospacer in the PAM-distal end of the protospacer abolish locking of Cascade on the target upon R-loop formation (85). Since priming was proposed to be promoted by an altered conformational state of Cse1 subunit of the Cascade (114), which is believed to participate in Cascade complex structural rearrangements upon locking on the targeted DNA (88), a study of effects of PAM-distal mismatches on priming is essential. PAM-proximal mismatched targets in conjunction with wild-type were also used to study R-loop formation ability, and R-loop properties, in particular, extend of their locking ability.

Table 1. Sequences of the non-target strand of the target variants tested in this study. The protospacer sequence and the PAM are shown in blue and dark blue, respectively. Base substitutions are represented in red. Black are overhangs that were used for cloning the constructs.

Not-target strand insert full sequence 5' to 3'		
Wild type (WT) full size		
GACCACCCTTTT	ATG CTGTCTTTCGCTGCTGAGGGTGACGATCCCGC	GCAGATACGTTCTGAGGGAA
Wild type (WT) shortened		
GACCACCCTTTT	ATG CTGTCTTTCGCTGCTGAGGGTGACGA CGTTAT	GCAGATACGTTCTGAGGGAA
AAG PAM full size		
GACCACCCTTTT	AAG CTGTCTTTCGCTGCTGAGGGTGACGATCCCGC	GCAGATACGTTCTGAGGGAA
AAG PAM shortened		
GACCACCCTTTT	AAG CTGTCTTTCGCTGCTGAGGGTGACGA CGTTAT	GCAGATACGTTCTGAGGGAA
G-1T full size		
GACCACCCTTTT	ATT CTGTCTTTCGCTGCTGAGGGTGACGATCCCGC	GCAGATACGTTCTGAGGGAA
G-1T shortened		
GACCACCCTTTT	ATT CTGTCTTTCGCTGCTGAGGGTGACGA CGTTAT	GCAGATACGTTCTGAGGGAA
C1T full size		
GACCACCCTTTT	ATG TTGTCTTTCGCTGCTGAGGGTGACGATCCCGC	GCAGATACGTTCTGAGGGAA
C1T shortened		
GACCACCCTTTT	ATG TTGTCTTTCGCTGCTGAGGGTGACGA CGTTAT	GCAGATACGTTCTGAGGGAA
T2A full size		
GACCACCCTTTT	ATG CAGTCTTTCGCTGCTGAGGGTGACGATCCCGC	GCAGATACGTTCTGAGGGAA
T2A shortened		
GACCACCCTTTT	ATG CAGTCTTTCGCTGCTGAGGGTGACGA CGTTAT	GCAGATACGTTCTGAGGGAA
G3T full size		
GACCACCCTTTT	ATG CTTTCTTTCGCTGCTGAGGGTGACGATCCCGC	GCAGATACGTTCTGAGGGAA
G3T shortened		
GACCACCCTTTT	ATG CTTTCTTTCGCTGCTGAGGGTGACGA CGTTAT	GCAGATACGTTCTGAGGGAA
T4G full size		
GACCACCCTTTT	ATG CTGCTTTCGCTGCTGAGGGTGACGATCCCGC	GCAGATACGTTCTGAGGGAA
T4G shortened		
GACCACCCTTTT	ATG CTGCTTTCGCTGCTGAGGGTGACGA CGTTAT	GCAGATACGTTCTGAGGGAA
CCG PAM full size		
GACCACCCTTTT	CCG CTGTCTTTCGCTGCTGAGGGTGACGATCCCGC	GCAGATACGTTCTGAGGGAA
CCC PAM full size		
GACCACCCTTTT	CCC CTGTCTTTCGCTGCTGAGGGTGACGATCCCGC	GCAGATACGTTCTGAGGGAA

4.4. Cascade binds all potential priming targets within half hour

To test the ability of Cascade to bind various targets, bulk assays were first conducted (Figure 7). Plasmids containing targeted sequences were incubated with Cascade and resolved on agarose gels by native electrophoresis. Results indicated that all known “interfering” and “priming” target variants, as well as the seed

mutants were bound by Cascade. In contrast, g8 protospacer targets with CCG and CCC PAM sequences and control plasmid without the g8 protospacer were not bound. This indicated that our Cascade preparation is functional, and, that priming targets are bound at our conditions. Since binding itself does not mean R-loop formation, DNA opening (melting) in the protospacer region was tested via permanganate footprinting by Dr. Kontantin Kuznedelov (Rutgers, The State University of New Jersey, USA) and described in Appendix A.2. Opening of all binding-proficient targets was revealed. Thus, R-loops can be formed on targets that behave as escape and cause priming *in vivo*. Kinetic analysis of R-loop formation on various targets measured with MT assays are described further.

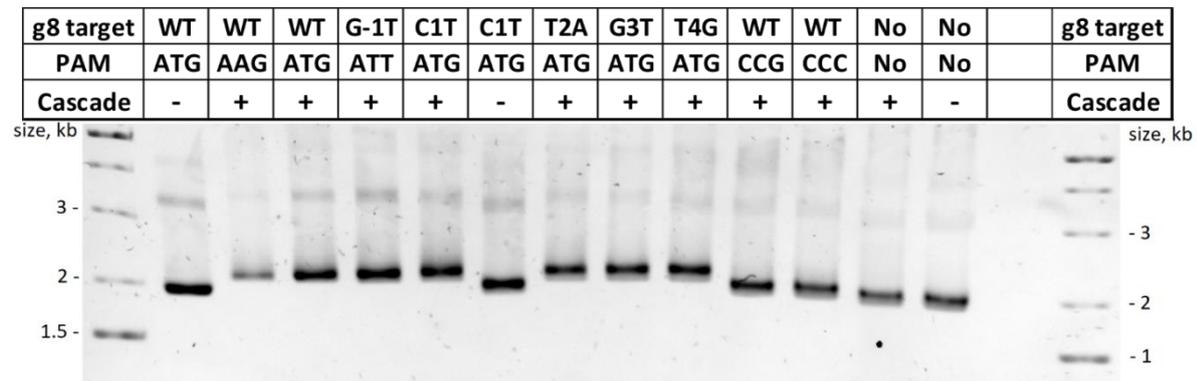


Figure 7. Cascade binding to supercoiled plasmid DNA. 100 nM Cascade was mixed with 10 nM plasmid carrying the indicated target variant in cleavage buffer (with Mg^{2+}). The mixture was incubated at 37 °C for 30 minutes and products separated by electrophoresis in 1% agarose gel. Cascade binding is seen as a slower migration of plasmid DNA. The WT target and all targets with single substitutions in PAM or seed are bound by Cascade. No binding is seen for CCG and CCC PAM variants, as well as for plasmid without a protospacer.

4.5 MT assays

4.5.1 Study of R-loop formation by Cascade

4.5.1.1 Mutated targets support R-loop formation

When investigating R-loop formation on the wild-type target carrying a fixed number of negative turns, a sudden DNA length increase in the time-trajectory as well as a shift of the whole rotation curve (when subsequently rotating the magnets) are detected and indicate DNA unwinding as a result of R-loop formation (Figure 8). DNA unwinding was not detected when the substrate contained a fully matching protospacer with a CCG trinucleotide in the place of PAM (Figure 8). Surprisingly and in strong contrast to *St-Cascade* (87), R-loops of *E. coli* Cascade formed on the WT target could not be dissociated (rotation curve did not back-shift) even at maximal applicable positive supercoiling/torque at which the B-form

DNA structure collapses (see Figure 3 right side and the trajectory in Figure 8, top row, right panel). Thus, conformational locking (88, 90, 162) by the *E. coli* Cascade appears to be practically irreversible. Besides being evidence of diversity even among same type (I-E) CRISPR-Cas effectors, the result presented an obstacle to study repeated R-loop formation-dissociation cycles in the MT assay. Thus, the assay had to be modified in order to collect statistics of R-loop formation events (i.e., kinetics) as reported in chapter 4.5.2.1. Another difference from St-Cascade was that *E. coli* Cascade did not show any torque dependence for the R-loop binding rates (Appendix A.1, supplementary Figure D8). These two facts allow to conclude that energy gain of R-loop formation by *E. coli* Cascade is much higher than for St-Cascade. This fact is convenient for measurements since using just one torque for Cascade binding and R-loop formation (-6.7 ± 0.5 pN) was enough to accurately determine rates of binding: the rate will not significantly change if range of torque values is applied, remaining in narrow range of values. Thus, rate at a certain torque defines the rate at any feasible torque value.

When investigating target variants bearing single point mutations, R-loop formation was observed on every target tested as a single-step DNA length increase (Figure 8). Possible R-loop intermediates, if they exist, were too short-lived to be detected. Measured rotational shifts were between 2 to 3 turns, suggesting that R-loops on all targets covered the same or very similar regions of unwound DNA. Thus, the ability of *E. coli* Cascade to form R-loops on the set of escape point-mutant targets used was directly demonstrated. However, in contrast to the WT target, most of the mutant variants exhibited R-loop dissociation, i.e., were less stably locked (Figure 8, see also below).

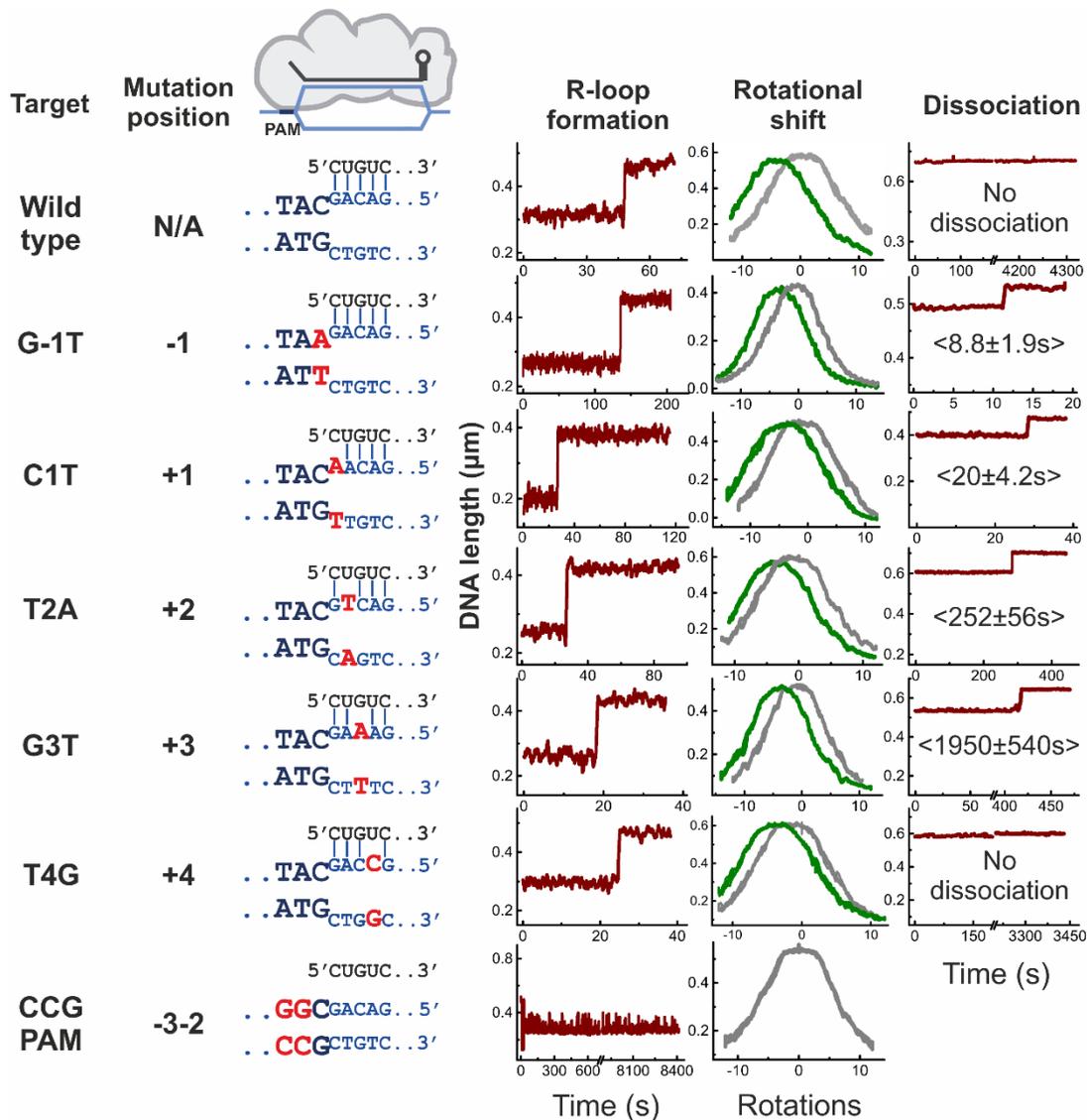


Figure 8. Example trajectories of R-loop formation and dissociation by *E. coli* Cascade measured with magnetic tweezers. Data is shown for all investigated target variants (nomenclature, mutation positions, and base-pairing schemes, with mutations indicated in red, are shown on the left side). R-loop formation is seen as a sudden DNA length jump at negative supercoiling (left trajectories recorded at 0.4 pN force and about -7 turns) and as a shift of the supercoiling curve (middle trajectories, grey and green curves refer to unbound and bound DNA). R-loop dissociation is seen as a length jump at positive supercoiling (right trajectories including mean dissociation times taken at 5 pN force and about +12 turns). Shown trajectories were smoothed with a sliding average filter of 1 s.

4.5.1.2. Locking ability of R-loop on mutated targets is reduced proportional to proximity of mismatch to PAM

While R-loop locking on the WT target was irreversible even at highest positive torque, R-loop dissociation could be observed for most of the target variants, in particular the G-1T PAM mutation as well as seed mutations at positions +1, +2, and +3 (Figure 8, seen as sudden DNA length increase at conditions of positive

supercoiling). To quantify R-loop dissociation we measured the times required for R-loop dissociation after introduction of positive supercoiling (torque of $+32 \pm 3$ pN·nm). Fitting of cumulative time distributions provided mean dissociation times of 9 ± 1.9 s for the G-1T PAM variant and 20 ± 4.2 s, 252 ± 56 s, and 1950 ± 540 s for targets with single seed mutations at positions +1, +2, and +3, respectively (Figure 9). R-loops on the WT target as well on the target with a seed mutation at position +4 did not dissociate under the conditions of the experiment over the course of at least one hour (Figure 8). Overall, these results clearly indicate that the presence of mutations in the PAM or in the seed region very close to the PAM (positions +1 and +2) strongly attenuates R-loop locking by *E. coli* Cascade without altering the extent of DNA unwinding. The presence of more distant seed mutations (position +4 and to the extent position +3) supports, however, strong WT-like R-loop locking.

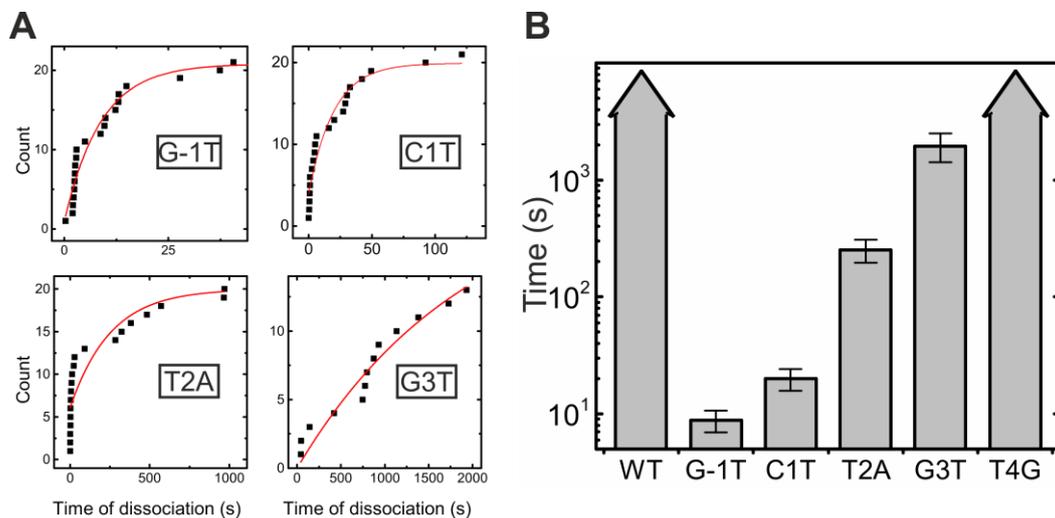


Figure 9. Mean dissociation times of Cascade bound target variants (with seed mutations). (A) Single events of Cascade (R-loop) dissociation times under high positive torque for target variants where it occurred. Time over count is fitted to exponential distribution and mean times are calculated. Tracks of no dissociation for wild-type and T4G are provided on Figure 8 as representative traces. (B) Mean times required to dissociate single R-loops at high positive torque. Arrowheads on bars indicate that no R-loop dissociation events were observed in the course of 1-hour observation.

4.5.2 Kinetics of Cascade binding to various targets

4.5.2.1 Modification of MT assay to allow comparative measurements of target variant R-loop formation kinetics

Since measurements of R-loop formation rates by MT assay require analysis of multiple individual binding/dissociation events, it was practically impossible to obtain such data for substrates on which highly stable R-loops formed. We,

therefore, sought conditions that prevent locking but not R-loop formation. To this end, we introduced six consecutive mismatches between crRNA spacer part at the PAM-distal end of the protospacer with ATG PAM (Table 1 and Figure 10A). The same number of mismatches abolished locking of St-Cascade by preventing zipping of the R-loop to the end of the protospacer (87). As anticipated, R-loops formed by *E. coli* Cascade on the target with six mismatches dissociated instantaneously upon the application of mild positive supercoiling. This indicates the absence of locking (Figure 10B), making multiple R-loop formation-dissociation cycles possible.

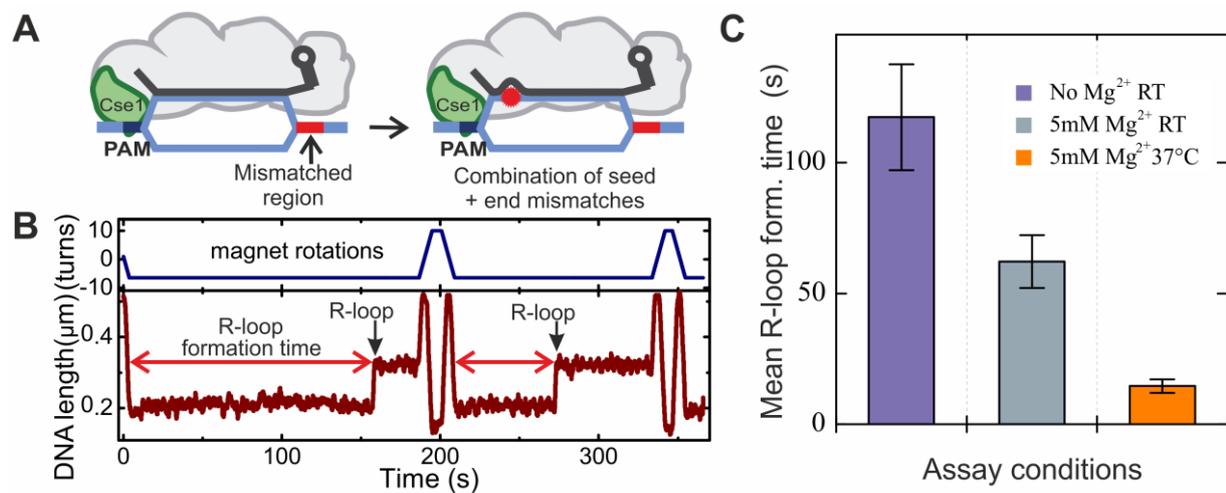


Figure 10. Development of MT assay to allow repetitive Cascade binding-dissociation cycles. (A) A sketch of Cascade bound to DNA targets containing 6 PAM distal mismatches that prevent locking of the shortened 26-bp long R-loop. (B) Representative trajectory from magnetic tweezers measurement of repetitive R-loop formation-dissociation events obtained with *E. coli* Cascade on the shortened target. Jumps corresponding to R-loop formation are indicated. Dissociation of R-loops is observed as soon as positive supercoiling is applied. The R-loop formation time is measured from the moment the desired negative supercoiling was reached. (C) Mg²⁺ ions and 37 °C temperature effect on acceleration of R-loop formation. Average times of R-loop formation measured on wild-type target are less in the presence of 5 mM Mg²⁺ allowing to speed up R-loop formation for mutated targets. Increasing temperature to 37 °C results in even more significant acceleration. The effect of only temperature increase is reported as separate measurement (Figure D12).

Another obstacle for comparative measurements of R-loop formation kinetics was extremely slow Cascade binding rate to targets with mutations in the seed. One binding event happening over a few hours of observation was impractical and insufficient to collect proper events for fitting to constant probability density distribution. To solve this problem we investigated ways to speed up the R-loop

formation process. The most obvious measure, increasing Cascade concentration, was not an option since *E. coli* Cascade required a high (100 nM) working concentration even with the wild-type targets, in contrast to St-Cascade (requires up to 100-fold lower concentrations). Thus, it was not possible to raise the Cascade concentration further due to concomitant increase in viscosity (the RNP complex was stored in 50% glycerol at 2 μ M stock concentration), which made magnetic particles stick to the surface. Other measure to solve the problem would be to add compounds that decrease the energy barrier for DNA melting during R-loop formation. Increasing Mg^{2+} concentration was an obvious choice: by lowering the energy penalty for local opening of DNA structure Mg^{2+} stabilizes non-standard DNA structures like hairpins and cruciform (163). Indeed, 5 mM of Mg^{2+} ions in the solution resulted in 1.8-fold increase of R-loop formation rate (Figure 10C). The main drawback of using Mg^{2+} was the overall lower stability of DNA: it was “breathing” more, opening for short time under negative supercoiling, which was seen as spikes in time-traces (see Figure 8 bottom row, CCG control was done in the presence of Mg^{2+}). This is why Mg^{2+} was only included in reactions to measure kinetics of R-loop formation and in negative controls to make them more stringent.

The second measure tested to facilitate R-loop formation was temperature increase from ambient 24 $^{\circ}$ C to 37 $^{\circ}$ C. The rationale here was two-fold. On the one hand, 37 $^{\circ}$ C is a physiological temperature for *E. coli* and, on the other hand, increased temperature generally facilitates enzymatic reactions via kinetics energy raise of the substrates and products, thus decreasing the activation energy barrier. Technically, controlled 37 $^{\circ}$ C conditions were created by implementing a special heater-controller system on oil immersion objective of inverted microscope of MT setup (Figure 2). Since heat transferred via several material borders (objective to oil to bottom cover slip to flow-cell interior), thorough calibration was required. Three points of temperature measurements were tracked simultaneously for calibration: heating ring, top of objective, and top of one of cover slip. The last one was used as a desired point for set temperature of 37 $^{\circ}$ C, while the first two were used to track temperature during measurements, since it is not physically possible to put a sensor inside flow-cell. This approach is sensitive to surrounding conditions, thus, room temperature must be kept constant. Overall, the effect of increased temperature was even more significant and together with Mg^{2+} resulted in approximately 8-fold increase of R-loop formation rate. This significantly improved the assay and allowed to collect data for every target variant that supported R-loop formation.

The drawback of temperature increase was the disturbance (significant shift) of force calibration curve towards apparently lower values (Figure 11). The calibration curve is used to determine the force of magnet pulling for every DNA molecule with magnetic bead on the top. Correction for the k_bT value (from 4.1

pN·nm at 24 °C to 4.28 pN·nm at 37 °C) did not correct the shift of force calibration. Therefore, the effect of temperature on the optical system and the measurements was investigated during test measurements done at room temperature and 37 °C. The piezo stage step movements of 1 μm were detected by the software and compared with absolute values. As was expected, small changes in optical density of the immersion oil of the objective, the buffer solution in the flow-cell and temperature expansion of the optical parts had little effect on the measurements: only small errors of less than 3% were seen. As it turned out, much bigger effect was produced by a change of the viscosity of the buffer solution. It is known, that the viscosity of water solutions similar to the binding buffer used reduces by approximately 23% as temperature is raised from 25 °C to 35 °C (155). Since the viscosity of water was used for magnetic bead fluctuations modeling during force calibration, incorrect values were leading to errors. Correction of viscosity value could not be performed since the precise value of solution viscosity could not be measured directly. Thus, the force calibration was measured at room temperature prior to the system heat up and actual measurements with Cascade.

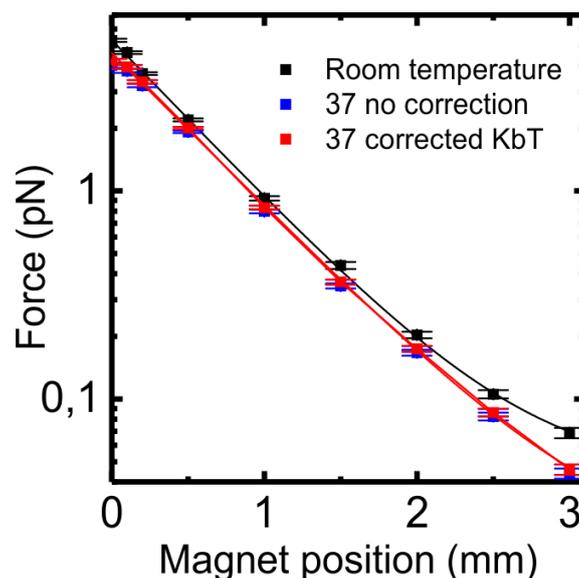


Figure 11. Shifted force calibration curves of MT assay measurements. In the absence of Cascade, DNA molecule tethered to the surface and a magnetic bead is pulled by a pair of magnets. Force values (as function of magnet position) result in an individual curve for every molecule and therefore has to be measured for each molecule. Raising temperature to 37 °C makes the actual force calibration (blue) shift towards lower forces for the same magnet positions (black). Distortion is linear in the range of binding measurements (0.4 pN). Changing $k_B T$ from 4.1 to 4.28 pN·nm did not help correct the error (red).

4.5.2.2 Mutated targets support R-loop formation at slower rates than wild-type target

Shortening of targets along with modifications of the MT assay allowed us to investigate the R-loop formation kinetics differences between wild-type and mutant targets. To this end, a set of substrates was produced where single PAM or seed substitutions were combined with the PAM-distal mutation introducing 6 mismatches (See Figure 10 and Table 1). Mean R-loop formation times were obtained from the analysis of consecutive single R-loop formation events (see Figure 12). R-loop formation on the target with 26 fully matching base pairs from the PAM (i.e., a shortened version of the wild-type target) was the fastest (6.27 ± 1.7 s at defined supercoiling conditions with a torque of -6.7 ± 0.5 pN·nm). The G-1T PAM mutation and the C1T seed mismatch retarded R-loop formation strongly, by 35- to 50-fold. The T2A, G3T and T4G mutants exhibited a moderate 5- to 15-fold retardation of R-loop formation.

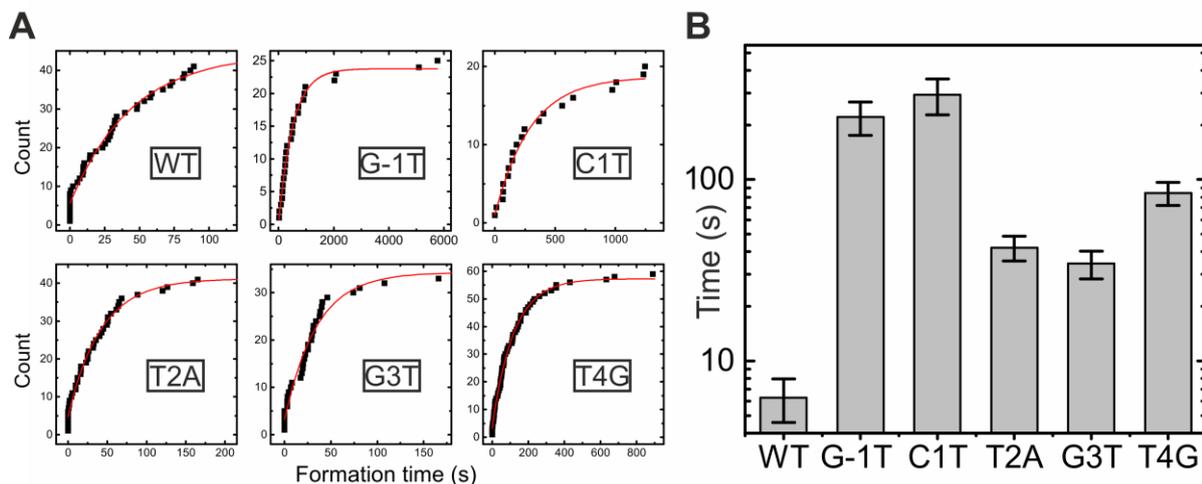


Figure 12. Cascade binding kinetics measured on shortened targets. (A) R-loop formation kinetics (black squares) measured at negative supercoiling and 0.4 pN force (-6.7 ± 0.5 pN·nm). Plotted is the cumulative event number as function of time between introducing the negative torque till R-loop formation time. From exponential fits (red lines) mean formation times were determined. All variants but the WT were measured using 100 nM Cascade. For measurements with the WT target, 20 nM Cascade was used in order to decrease the formation rate and allow reliable measurements. The resulting average time was then rescaled accordingly. (B) Average R-loop formation times for different protospacer/PAM variants obtained from exponential fits of recorded kinetics of R-loop formation for seed and PAM mutants is at least 5 times slower than that for the WT target.

Since earlier data support that R-loop formation is a directional process that nucleates at the seed region and propagates till PAM-distal protospacer end (85), the time of R-loop formation (nucleation at the PAM sequence and propagation of R-loop) should not be affected by PAM-distal mismatches. Indeed, nothing should be different for the first 26 nucleotides either in target to crRNA pairing or in Cascade conformation (as locking is believed to happen only upon the pairing of

target at the very end of 32-nt protospace. In other words Cascade does not “sense” any mismatch until it tries to pair its crRNA to with mismatched nucleotide. At the same time, the probability of R-loop to collapse and Cascade to dissociate from a target with 26-bp duplex is minimal, based on assumption that crRNA interaction with targeted DNA is energetically favorable, due to RNA-DNA pairing and, possibly, protein stabilization of the resulting complex. Energetics of this intermediate state is the same as in structures considered by the model of toehold-mediated strand displacement (164). In this model, a competing oligonucleotide that displaces the other one by having more complementarity to the “target”, has almost zero chance to dissociate (detach by collapsing paired region to free DNA strands) when 10 or more nucleotides of the “attacker” and “target” are paired. Thus, we assume that the times of R-loop formation measured by above assays should be the same as for full-length targets. Luckily, there is a direct way to test this hypothesis: not all of full-length target variants lock strongly and, therefore, can be dissociated. Specifically, G-1T, C1T, and T2A variants supported R-loop formation-dissociation cycles even in the context of full-length protospacer targets. To prove that kinetics of binding is the same for both full-length and shortened targets, comparison was done on these mutants (Figure 13): R-loop formation rates for shortened targets are compared to R-loop formation rates for full-length targets. It is seen that same mutants show similar R-loop formation rates. The only observed difference towards slower binding for full-length targets is attributed 5 mM MgCl₂ present for shortened targets but absent for full-length ones. If one multiplies by a 1.8 correction factor, all values become equal within the error. Thus, measured kinetics represents actual rates for Cascade R-loop formation. Results thus indicate that R-loops form much slower with mutants tested here.

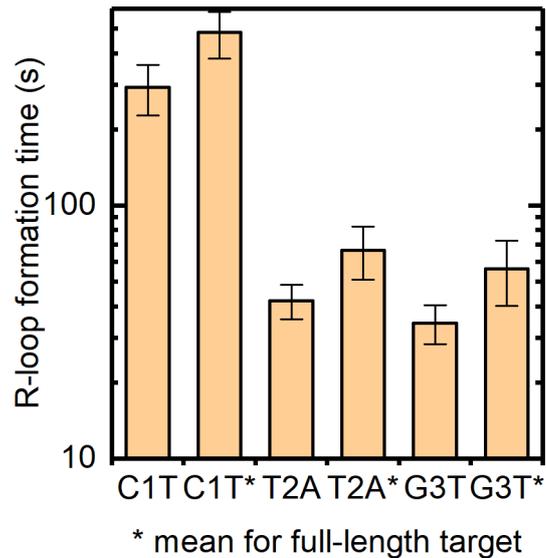


Figure 13. Comparison of kinetics for full-length and shortened (PAM-distal 6nt mismatched) targets. Asterisk stands for full-length, while its absence stands for PAM-distal mismatched targets. It is clearly seen that values for every identical seed mismatches are in high correlation. The difference in absolute values comes from different MT assays conditions (Mg^{2+} included in full-sized targets), which affects the overall rate of R-loop formation by a manner similar to constant ~ 1.8 multiplier of rates (see chapter 4.5.2.1).

4.5.3 RNP complexes with changed crRNA lengths and altered Cascade subunits

Part of the project of Cascade targeting study on single-molecule level was dedicated to analysis of Cascades with altered crRNA spacer length and consequent change in Cascade subunits stoichiometry (159). If one intentionally adds or removes parts of a spacer sequence in CRISPR array, CRISPR transcript will still be processed into individual crRNAs at repeat region by the Cas6e Cascade subunit. The resulting non-standard length crRNA will be incorporated into a complex that comprises Cascade subunits but not in original stoichiometry (and, therefore, overall structure). We here focused on particular crRNA variants studied by Kuznedelov et al. (159): a set of six crRNA variants with spacer lengths ranging from 18-nt shorter (14-nt spacer) than natural 32-nt spacer to 38-nt spacer. Specifically, -18, -12, -6 and -3 shortened versions (from PAM-distal end) of g8 spacer and +6 elongated version with PAM-distal 5' '-ATGTAT-3' extension of the spacer were studied along with standard g8 crRNA Cascade. Changes in crRNA length lead to changes in Cas7 subunits stoichiometry, since they compose the backbone of Cascade, whose major function is to hold the spacer part of crRNA. As expected, there was less Cas7 subunits in complexes with shorter crRNA, and more (7 instead of standard 6) in Cascade containing +6 crRNA. In addition, only one (instead of standard 2) Cse2 subunit was present in Cascades with shortened crRNAs, while 3 Cse2 monomers were present in the +6 crRNA complex.

In vivo, all complexes supported primed adaptation, and all but -18 and -12 variants supported efficient interference, while -12 variant supported interference to some extent. *In vitro*, all altered complexes have shown target binding in bulk assays, but stable R-loop formation on permanganate footprinting was not detected for -18 and -12 Cascades. We decided to study the variants for R-loop properties on MT assays to directly measure the extent of unwinding.

At first, we probed all complexes for R-loop formation, and observed shifts on negative supercoiling states for every crRNA variants (Figure 14). This proves the ability of RNP complexes to open DNA for R-loop formation and explains why they all show target bindings and support primed adaptation. Complexes with two shortest crRNAs, -18 and -12, did not withstand positive supercoiling at constant force (moderate torque of $+6.7 \pm 0.5$ pN·nm). The -6 and -3 spacer variants supported more pronounced locking, with around a third of R-loops surviving $+6.7 \pm 0.5$ pN·nm positive torque. Still, the majority of R-loops for these variants dissociated immediately after positive rotations. For wild-type and +6 variants, all R-loops were stable and survived positive rotations.

From Figure 14 one can see that extent of DNA opening, at least on negative supercoiling (left slope), corresponds to spacer length. We quantified the shifts on at least 30 rotations of several R-loops per each spacer variant (Figure 15).

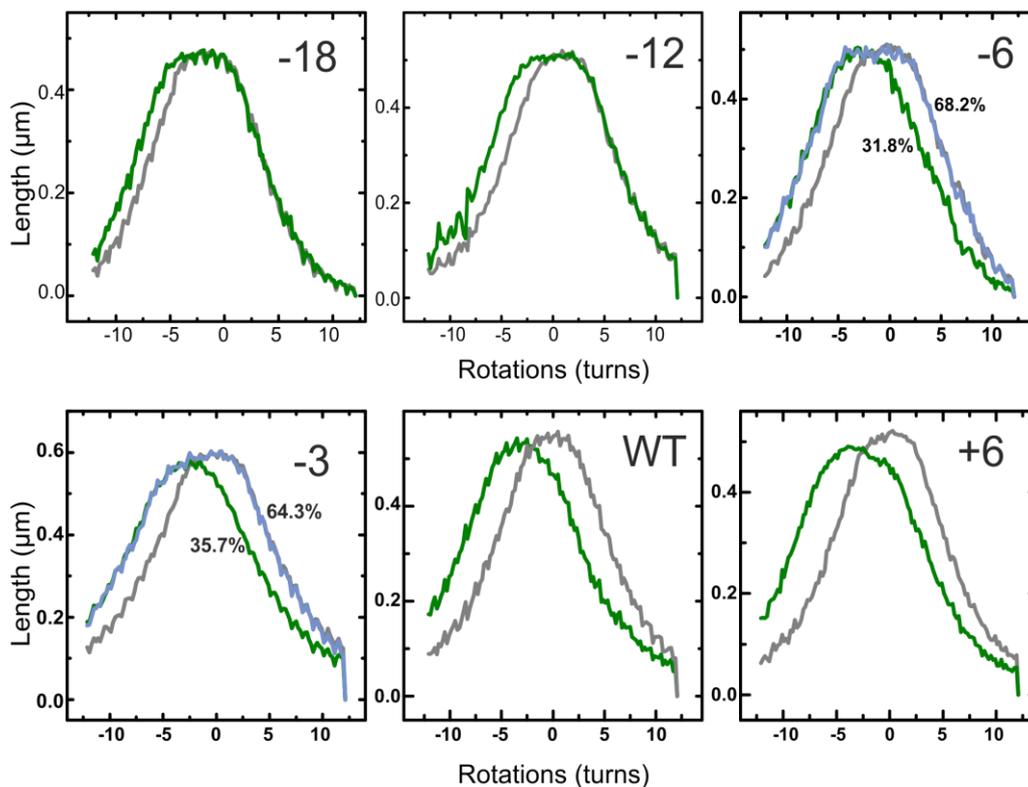


Figure 14. Altered crRNA length containing Cascades tested for R-loop formation using MT assays. Averaged rotation curves representing shifts of DNA opening (green) in at least 10 R-

loop formation events for each trajectory. It is seen that -18 and -12 variants support R-loop only when negative torque is applied to DNA molecule, meaning lack of locking for such complexes. The -6 and -3 variants support some locking with a fraction of R-loops surviving positive torque (blue compared to green). Wild-type and +6 variants demonstrate stable locking. Shifts extent is proportional to spacer length. Initial molecules rotation curves (without R-loop) are shown in grey.

The trend of rotation curve shifts is in full agreement with expected R-loop sizes. Two features here were observed. First is that the right-side shift (positive supercoiling) for +6 crRNA variant displays value closer to wild-type or even to -3 R-loop size, much smaller than expected and observed for the left-side shift (negative supercoiling). An explanation of such phenomenon can be that despite of 38-nt matching region that support 38-nt R-loop formation at negative torque, locking of the altered Cascade complex still happens at native 32-nt R-loop size. In order to test this idea, the +6 crRNA Cascade was tested with wild-type g8 target, where there was no match of the last 6 crRNA nucleotides to DNA. In such assay, stable R-loop with same as for original g8 crRNA rotational shifts was observed (Figure 15). Thus, the ability of +6 (38-nt) spacer to lock at 32-nt target was proved. Another interesting phenomenon was observed for the -6 target: it is seen that both negative and positive supercoiling shifts for -6 crRNA are a bit lower than expected by the common trend values of all other crRNA variants. The same effect, even more prominent, was observed for St-Cascade (unpublished data). Explanation should be based on altered RNP complex stoichiometry, but so far, there is not specific model to explain this phenomenon.

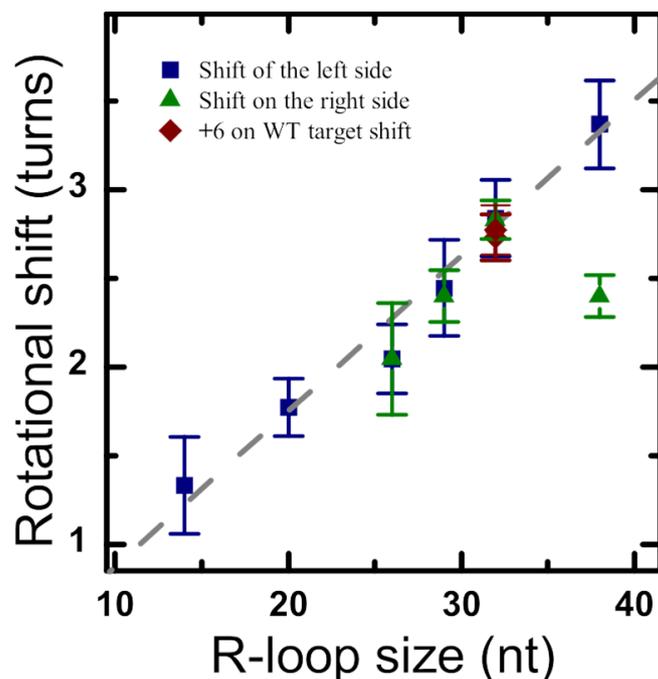


Figure 15. Altered crRNA R-loop sizes. Quantification of rotational curves shifts plotted as function of expected R-loop lengths. Left side (negative supercoiling) rotational shifts (blue)

perfectly match expected trend. Right side (positive supercoiling) shifts (green) for -18 and -12 variants were not observed. For +6 variant, right side shift is less than expected, but matches to either 29-32-nt R-loop size. The +6 variant on 32-nt (wild-type g8 protospacer) target forms R-loop identical to wild-type g8 crRNA. The -6 variant shows lower than expected shifts, which, however, still fit the trend within error values. Errors are standard deviations of measurements for at least three R-loops with at least 30 rotations in total.

Since all of crRNA variants starting from -6 supported R-loops that survived moderate positive torque, locking ability of such complexes was tested at high $+ 22.7 \pm 2.1$ pN·nm positive torque (Figure 16). Wild-type and +6 variants showed strong locking and did not dissociate during the time of measurements – at least 45 minutes (Figure 16A and B). The -6 and -3 variants showed complex behavior: in majority of cases they dissociated on the positive side of the rotation curve (Figure 14). But in some of the cases they survived rotations to the positive side and high positive torque forces were required for the complex to dissociate (Figure 16C-F). Interestingly, different behavior of -6 and -3 R-loops was observed under high positive torque. While -6 crRNA R-loops dissociated in all cases after seconds (average of 1.78 sec), the -3 complexes showed both fast, within few seconds dissociation, and much longer, few dozens of minutes, survival on DNA. Thus, locking ability of shortened crRNA complexes is increasing with spacer length and is stronger for spacers closest to the natural 32-nt length. The -3 complex displayed three distinct locking modes: weak locking not surviving $+6.7 \pm 0.5$ pN·nm torque, intermediate locking surviving $+6.7 \pm 0.5$ pN·nm but dissociating in seconds after $+ 22.7 \pm 2.1$ pN·nm torque application, and strong, closer to wild-type locking, with stable for half-hour R-loop at $+ 22.7 \pm 2.1$ pN·nm torque. Different stoichiometry of RNP complexes for -3 crRNA is very unlikely, based on mass spectrometry data (159), implying that identical -3 complexes are able to support these distinct locking states.

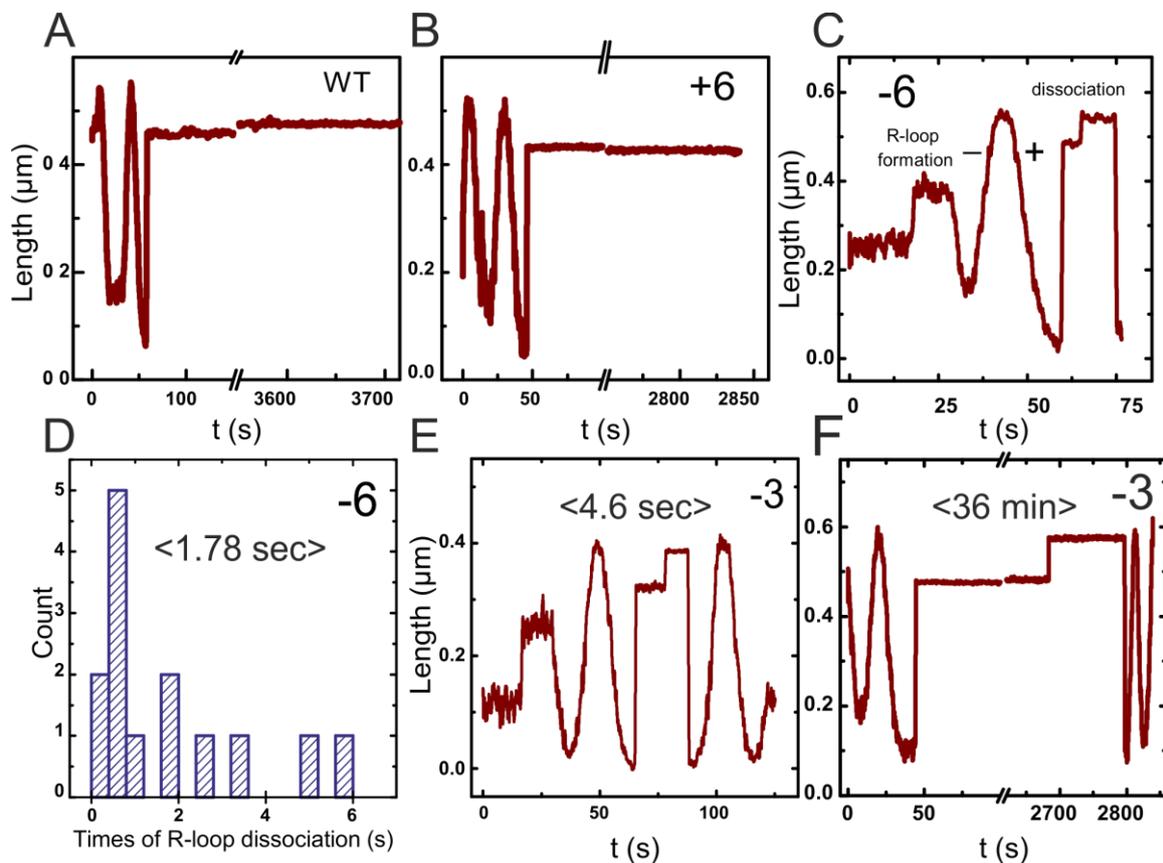


Figure 16. Altered crRNA R-loops tested for stability at high positive torque. (A) and (B) wild-type and +6 crRNA variants do not dissociate under $+ 22.7 \pm 2.1$ pN·nm torque for at least 45 minutes. (C) Representative trace of the -6 variant dissociating fast after high positive torque application (shift after positive rotations). (D) Histogram of the -6 variant R-loop dissociation times under high positive torque. No R-loop sustained for longer than 6 seconds; average is 1.78 seconds. (E) and (F) representative traces of fast and long -3 variant dissociation events, correspondingly. While fast dissociation events are observed in most cases, with average of 4.6 seconds, few long dissociation events happen with typical “average” time of 36 minutes, indicating strong locking ability.

4.5.4 MT assay developments for future use: synthetic genes

Developments of Magnetic tweezers assays is an essential part of workflow optimization to get new insights into molecular mechanisms of DNA targeting enzymes. This chapter describes the developments that were done but not yet used for systematic studies. The first development is focused on sequence of DNA backbone where targeted sequences were cloned (see Materials and Methods). Off-targeting by CRISPR effectors is a big issue that is discussed mainly when biotechnological applications are considered (165). Biophysical experiments can become a very useful tool for detection of off-targeting (166, 167). Regardless of applications, off-targeting can lead to misinterpretations of experimental results and must be minimized or controlled. For example, research that proposed two distinct Cascade binding modes based on single-molecule FRET (smFRET) data (127) did not consider that the target protospacer they used contained at least one internal consensus AAG PAM and two ATG PAM sequences inside it. Thus, their

interpretation of short-lived events as distinct binding modes of Cascade might be due to Cascade interactions with these PAM sequences. Considering this case, we worried whether our experiments may suffer from similar problems.

For all experiments presented above the pUC19 plasmid was used as a backbone for cloning in protospacers. This plasmid sequence contains lots of consensus PAM sequences in it that should be recognized by Cascade and may be used to nucleate R-loops. Though no off-targeting signal was detected in negative controls, we worried that increase of the RNP complex concentration (as one of the means to speed up R-loop formation measurements) could cause off-targeting Cascade activity to become more prominent. As a preventive solution that could be useful for bulk assays as well, we designed synthetic 2 kbp templates (see Materials and Methods and Appendix C) that do not contain any of PAM that support high levels of interference (AAG, ATG, AGG, AAA, GAG, not-targeted strand nomenclature). Besides, these templates did not contain seed region of g8 protospacer (5'-CTGTC-3') and another protospacer used for torque-dependence study (Appendix A) with seed region of 5'-CCAGT-3'. Synthesized construct was cloned into commercial vector and used for protospacer cloning in the middle of it.

Tests of synthetic genes backbone on MT measurements revealed less stability of the complex under negative torque: it apparently produced significantly more noise than the pUC19 backbone, at around -6.7 pN·nm torque, used for R-loop formation (Figure 17). Free energy gain of any not B form DNA structures was minimized by Mfold analysis on dozens of candidates (160), thus, preventing as possible from potential base-pairing of detached parts of designed construct. Nevertheless, Ugene software repeat finder, used for post-problem discovery analysis, showed increased number of 8-10 nt repeats in synthetic templates compared to the pUC19 backbone (168). This or other factors caused “breathing” of DNA. The high number of repeats could be the result of too strict limits imposed on synthetic genes sequences: absence of 5 trinucleotides and of 2 five-nucleotides regions at both strands. There may be a chance to improve the backbone, by redesigning synthetic templates to lack only the g8 protospacer seed. In this case, as well, more attention should be dedicated to avoid small matching (repeated) regions in mutual proximity, rather than overall potential energy of secondary structures (as was done for the tested version of synthetic templates). In order to progress with experimental results, we decided to stick with the pUC19 backbone, considering the fact that synthesis and cloning of new synthetic genes is matter of months. It was a correct decision, since new MT assay developments allowed to progress with R-loop formation kinetics measurement (see chapter 4.5.2.1), and no Cascade off-targeting was observed for these experiments. Still, synthetic template concept is an interesting avenue to explore for future MT and bulk assays.

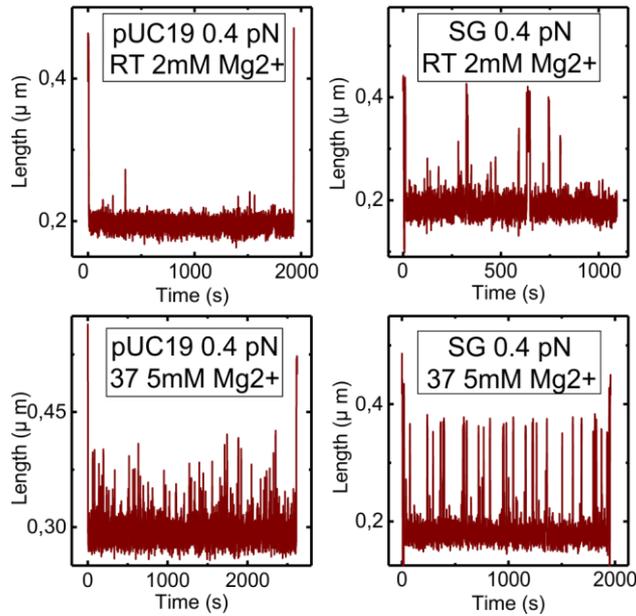


Figure 17. Synthetic template backbone shows stability issues compared to pUC19 backbone. Measured at constant moderate -6.7 ± 0.5 pN·nm torque, in the absence of Cascade or any DNA binding proteins peaks of DNA “breathing” – spontaneous DNA melting can be detected. Synthetic template backbone (right column) shows more noise than pUC19 (left column). The effect is increased at R-loop facilitating conditions (37 °C, 5 mM Mg^{2+}). The increase of noise is seen for pUC19 as well at these conditions (bottom row), but for synthetic template it is much more pronounced making assay too noisy for proper measurements.

4.6 Altered Cse1 conformation model is supported by Cas3 recruitment ability upon R-loop formation

The results presented above indicate that with the exception of negative controls (CCG PAM), all tested mutated targets that escape CRISPR interference *in vivo* are bound by *E. coli* Cascade *in vitro* and form full-length R-loops. As proposed earlier, the locking differences for these targets must be caused by different Cse1 conformation, more “open” when the mismatch is close to PAM (114). In the same work, it was proposed that Cas3 recruitment ability is compromised by the conformational changes in the Cse1 subunit, compared to wild-type target. If this were so, then Cas3 nuclease activity on targets we study should be in direct correspondence with locking observed in MT assays. To test this idea, we investigated Cas3-mediated degradation of plasmids containing g8 protospacer variants that showed R-loop formation (Figure 18). Cascade-bound plasmids were combined with an excess of purified Cas3 protein at conditions optimized for target degradation (114, 121). DNA cleavage was seen as conversion of supercoiled Cascade-bound plasmid DNA into relaxed and/or linearized forms as well as the appearance of a smear of shorter degradation products at longer reaction times (Figure 18) (90, 114). For the WT target, most of the supercoiled

plasmid species disappeared (corresponding to a processed fraction of 0.84, Figure 18B and C) with a mean reaction time of ~8 min at 100 nM Cas3. Cas3-catalyzed DNA degradation of targets with seed mutations at positions +3 and +4 was similarly fast. In contrast, DNA degradation of targets bearing the G-1T PAM mutation or seed mutations at position +1 and +2 was greatly decreased. The strong attenuation of DNA cleavage in these cases was not caused by dissociation of less stable R-loop complexes, since all supercoiled (and, therefore, uncut) DNA remained bound to Cascade during the entire 120-min time course of the reaction (Figure 18A). Cas3 was added to no Cascade samples as control of specific R-loop mediated DNA degradation and no degradation was revealed. We conclude that the rate and efficiency of Cas3 cleavage strongly correlates with the R-loop dissociation time, i.e., locking strength, on a particular target (compare Figure 9 with Figures 18C and D, Pearson correlation coefficients r of 0.88 and 0.94 with probability values p of 0.021 and 0.006 for rate and efficiency, respectively).

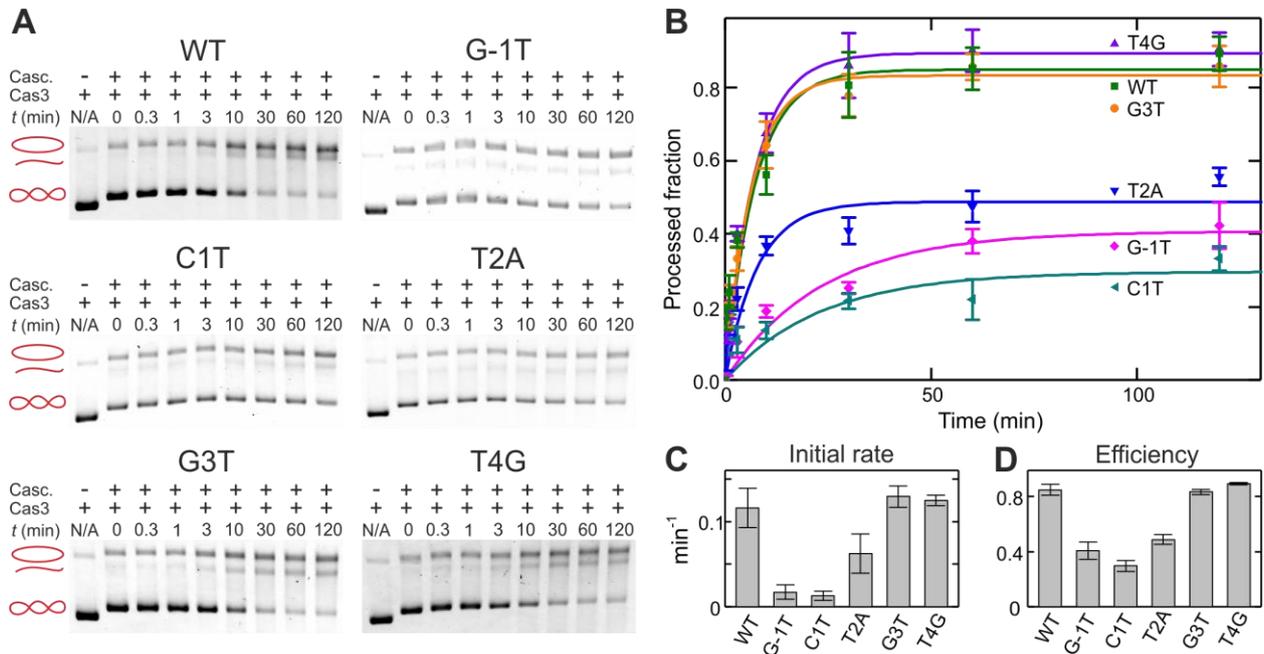


Figure 18. Kinetics of Cas3-mediated cleavage of plasmid DNA bound by Cascade. (A) Target plasmids containing indicated g8 protospacer variants were preincubated with Cascade-g8-crRNA. DNA cleavage reaction was initiated by the addition of Cas3. At indicated times after Cas3 addition reactions were terminated. Products were separated by agarose gel electrophoresis. Sketches next to gel images illustrate positions of the DNA topoisomers (supercoiled, nicked and linear plasmid). (B) Kinetics of DNA cleavage quantified from the disappearance of the supercoiled DNA species in the agarose gels. Error bars represent standard deviations from 2 to 3 repeat measurements. (C) and (D) Initial cleavage rates and cleavage efficiencies obtained from the fits of the data shown in B.

4.7 Primed adaptation occurs independently of conformational changes in Cascade

Seed mutants of g8 protospacer all displayed different locking abilities and Cas3 recruitment efficiency that correlates with proximity of mismatch to PAM. Thus, if priming is conformationally triggered, they should support priming to different extents. I.e., priming should correlate with changes observed on MT assays and almost no priming should be seen for the T4G variant, as its conformation, upon R-loop formation, is undistinguishable by our methods from R-loop on wild-type target. While G-1T and C1T were already shown to support priming (113), T2A and G3T mutants were expected to support priming at lower rates, with G3T closer to T4G level than to C1T. To test ability of target variants to cause primed adaptation, *E. coli* KD263 cells harboring *cas* genes under control of inducible promoters and a CRISPR array with g8 spacer were transformed with plasmids harboring the WT g8 target or its variants (Figure 19A). The *cas* gene expression was induced in plasmid-bearing cell cultures, and adaptation was followed by PCR amplification of CRISPR array at various times post-induction (Figure 19B). In agreement with previous work (119, 169), no adaptation was detected in induced cultures of cells harboring plasmids without protospacer, with the WT target or with the CCG PAM variant, while robust adaptation was observed in cultures transformed with plasmids carrying targets with a G-1T PAM and the +1 seed mutation (C1T). Notably, targets with seed mutations at positions +2, +3, and +4 (T2A, G3T, and T4G) also supported adaptation at comparable rates and apparently comparable levels (Figure 19C). No correlation between mismatch position and priming rates was detected. In fact, T4G and C1T target variants were the ones most actively supporting fast adaptation.

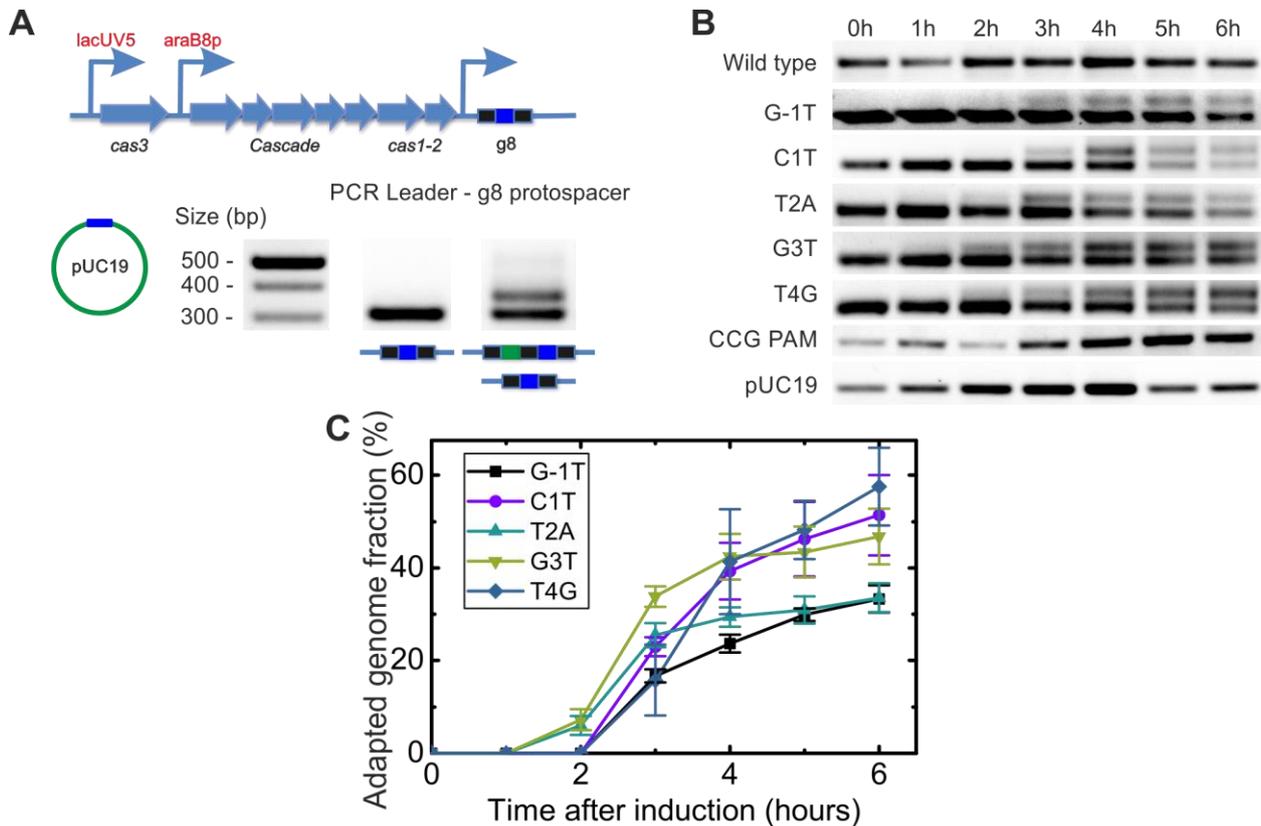


Figure 19. Primed adaptation tested in *E. coli* KD263. (A) Genetic structure of KD263 CRISPR locus. CRISPR array contains only g8 spacer. The *cas* genes are under inducible promoters and were induced after plasmids containing g8 protospacer were transformed into cells. PCR over CRISPR array detects array expansion after new spacers are integrated. (B) Results of g8 target variants tested for primed adaptation. All seed mutants that are escapes *in vivo* and support R-loop formation demonstrate high levels of adaptation. Wild-type target does not support adaptation. Negative controls of CCG PAM and lack of g8 in pUC19 do not show detectable levels adaptation. (C) Quantification of gel images shows very similar rates and levels of adaptation for all mutants. PCR part of genomes with elongated CRISPR array over total PCR product is plotted as function of time. C1T and T4G variants support rapid adaptation, showing no correlation with R-loop locking ability. Errors are standard deviations of at least 2 repeats.

4.7.1 High throughput sequencing of adopted spacers shows the same spacer selection for every priming mutant

We further tested whether the locking state of Cascade affects the specificity to select particular spacers during priming. For all target variants that supported priming, PCR fragments corresponding to expanded CRISPR arrays were subjected to high-throughput sequencing. After filtering, the acquired spacer sequences were extracted and mapped onto the donor plasmid backbone (Figure 20 and supplementary Figure D1). No significant differences among the target variants could be detected: all of them displayed the previously reported strand-biased hot-spot pattern of protospacers from which newly acquired spacers originated (113, 115, 119, 120, 122, 123). These patterns were for all targets highly

correlated to each other (Figure 20C) with Pearson correlation coefficients >0.97 (supplementary Figure D1).

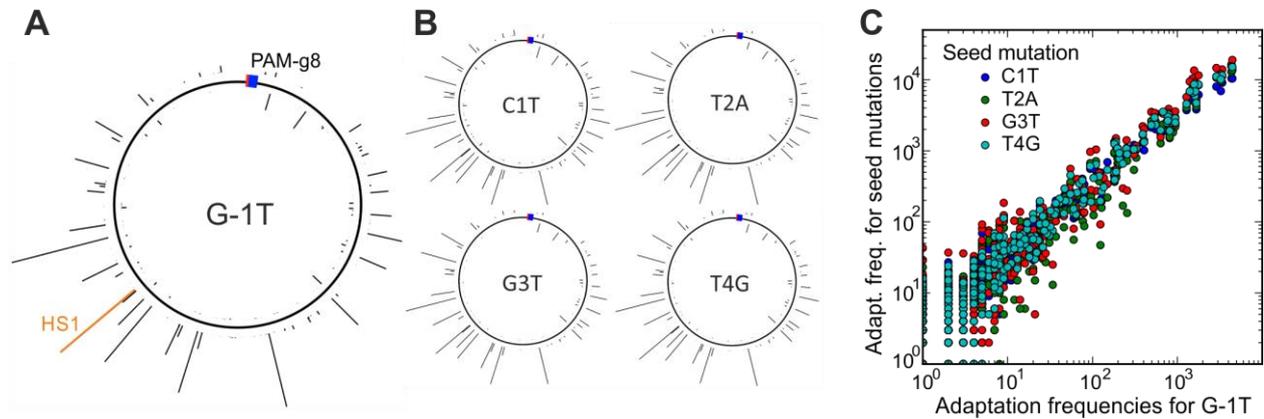


Figure 20. High-throughput sequencing (HTS) data of selected spacers via fast adaptation presented on Figure 19. (A) Mapping of spacers acquired from the g8 G-1T variant target protospacer plasmid to the pUC19 backbone. The height of the histogram bars corresponds to the number of HTS reads found for a particular position. The location of the priming protospacer and the PAM is shown as a blue-red box. A bar marked in orange marks hotspot HS1 protospacer, which was used for semi-quantitative measurements of the primed adaptation efficiency (see chapter 4.7.2). (B) The same histograms made for every target variants that supported priming. Every histogram looks identical/similar. (C) Position-dependent acquisition frequency for targets with seed mutation plotted over the acquisition frequency for the G-1T PAM mutation target. High correlation between spacer acquisition patterns of all tested target variants (see supplementary Figure D1 for correlation coefficients) is apparent.

4.7.2 Quantitative adaptation measurement supports kinetics model of primed adaptation

The identical hot-spot selection of spacers during primed adaptation for the different targets allowed us to design a semi-quantitative assay to measure adaptation efficiency. The assay is more accurate than gel images analysis (Figure 19B and C). The assay involves qPCR reactions where one of the primers is specific to a frequently acquired spacer (orange colored hotspot HS1, see Figure 20A), and another specific to CRISPR array leader. By normalizing the qPCR signal from HS1 by qPCR signal from *gyrA* gene from bacterial genome, an adaptation score could be calculated: it measures the percentage of genomes containing HS1 in CRISPR array over all the *E. coli* KD263 genomes present in the sample (Figure 21, see Materials and Methods for details). The HS1 is one of the most frequently acquired spacers and is acquired with same relative efficiency for every target variant tested. All necessary calibrations were done to ensure reliability of the assay.

Implementation of the assay revealed that all targets that supported priming had a similar adaptation score, i.e., priming occurred for all of them at a comparable level (the scores are not going above 10% since it is only HS1 adaptation that we follow, while the real adaptation score that takes all acquired spacers in account is several times higher). We conclude that the locking state of Cascade, while well correlated with the *in vitro* ability to recruit Cas3 for target degradation, does not influence the extent of primed adaptation (insignificant correlation with $r = -0.21$ and $p = 0.7$) or sequence preferences during spacer selection *in vivo*. In other words, *in vivo* priming occurs independently of the particular locking state of target-bound Cascade.

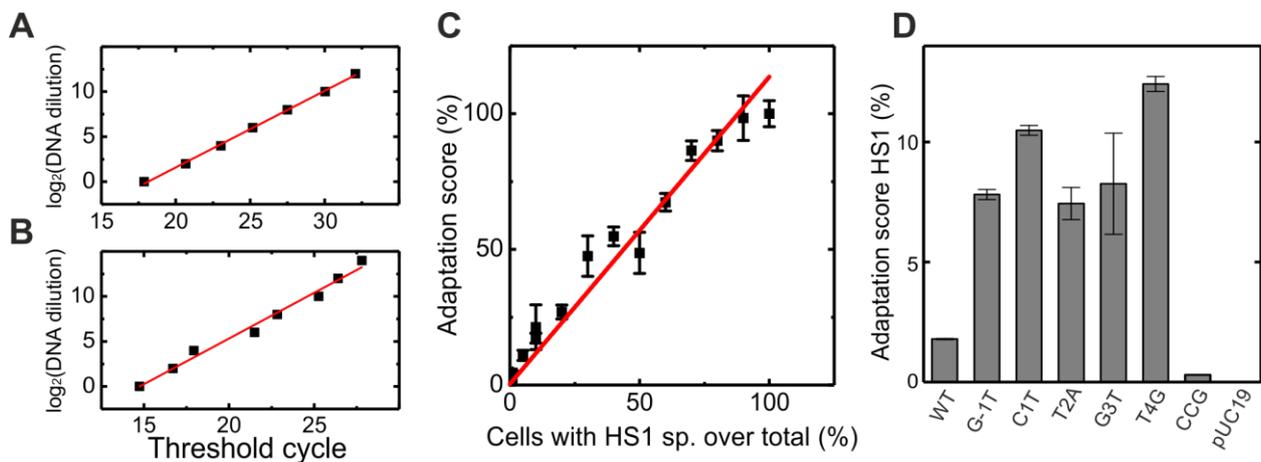


Figure 21. qPCR assay to quantify adaptation levels. (A) Calibration of qPCR measurements on hotspot (Figure 20A). Bacterial genome DNA containing an elongated CRISPR array with HS1 was diluted between 1- and 4096-fold and subjected to qPCR to determine the threshold cycle. Plotting the binary logarithm of the DNA dilution over the threshold cycle provides the expected linear relation with a slope of 0.85 (red line, $R^2=0.9987$). This corresponds to a 1.8-fold increase of the DNA amount per PCR cycle being close to the expected value of 2. (B) Calibration of qPCR measurements on the *gyrA* gene used for normalization. The calibrations used the same DNA as in A. The slope of a linear fit to the data is 0.978 (red line, $R^2 = 0.9856$) corresponding to a 1.97-fold DNA increase per PCR cycle. (C) Testing the qPCR-based quantification of primed adaptation. Cells containing either a non-elongated CRISPR array or an elongated CRISPR array with HS1 were mixed in known ratios and the genomic DNA was extracted. The obtained DNA was subjected to qPCR analysis. The adaptation score was obtained from the ratio of the amount of DNA containing elongated CRISPR arrays and the amount of DNA containing the *gyrA* gene. The DNA amounts were calculated from the measured qPCR threshold cycles using the calibration curves in A and B. The measured adaptation score returns within error the input percentage of cells containing HS1 thus verifying the calibration. (D) Relative frequency of priming (i.e., CRISPR array extension; over HS1 spacer) probed by qPCR for the different target variants at time point of 6 hours after CRISPR-Cas system induction. Error bars represent the standard deviation of at least three technical (qPCR) of two biological (adaptation experiments) repeats.

Kinetics model of priming, thus, is much more compatible with our observations: primed adaptation does not depend on R-loop stability (locking strength), neither for seed mutants nor for crRNA length variants. In the same time, all the targets tested that attenuate interference while still supporting R-loop formation, support primed adaptation. It is hard to derive a clear formula for primed adaptation, since kinetics model depends on virus (in our case plasmid) copy number, replication rate and the time of CRISPR-Cas system action and level of its induction (126). What it seen on Figure 21C it that even wild-type target appears to support primed adaptation to some extent, something that was not detected by agarose gel analysis of CRISPR array PCR products (Figure 19B). What we can estimate is that for every seed mutant where Cascade binding is slower than for wild-type, the scores of adaptation are few times higher. The best mutant to focus on is T4G: its Cas3 recruitment ability is the same as for wild-type protospacer (g8 with ATG PAM), and R-loop has the same undistinguishable locking strength for both of these targets. Thus, the only difference in CRISPR-Cas system action upon the wild-type and T4G variant detected so far is Cascade binding (R-loop formation) rate: 6.27 ± 1.7 sec and 84.4 ± 12.3 sec, respectively. The adaptation scores for these targets are 1.45 ± 0.05 and 10 ± 0.51 %, respectively. While the rate of Cascade binding is constant value, the adaptation score is cumulative. If the kinetics model of primed adaptation is valid (126), then adaptation score should be proportional to the time of growth after induction. According to this model, the difference (ratio) between T4G and wild-type should become larger at longer times, as wild-type protospacer plasmid should be depleted, while the T4G variant should be constantly degraded for much longer times, maintaining some amount of DNA in the cell.

4.8 Bulk assays development to reproduce priming conditions *in vitro*

4.8.1 Cas1-Cas2 effect on Cas3 degradation rate

Further development of bulk assays involved the Cas1-Cas2 complex. As mentioned, some evidence supports that Cas3 activity on mutated targets is altered by Cas1-Cas2 (81). This distinct mode of Cas3 function is believed to be connected to primed adaptation. Thus, next experiments were dedicated to analyze the effect of Cas1-Cas2 on Cas3 degradation. If the model is true, the presence of Cas1-Cas2 with Cascade and Cas3 should speed up plasmid degradation, causing faster accumulation of relaxed and linear fraction than at interference conditions with Cascade and Cas3 only. Experiments were done to check this, and the results are presented in Figure 22.

A plasmid containing priming target protospacer with C1T mutation was mixed with: only Cascade for binding to the substrate; Cascade and Cas3 to model interference, all three Cas protein complexes to model primed adaptation

conditions. The reaction was left at 37 °C for quite a long time – 18 hours, since Cas1-Cas2 activity rate is slow. Results reveal, first of all, that Cascade remains bound even after such a long time of incubation. Cas3 successfully degrades target plasmid, as expected. Unfortunately, the presence of Cas1-Cas2 did not appear to have any effect.

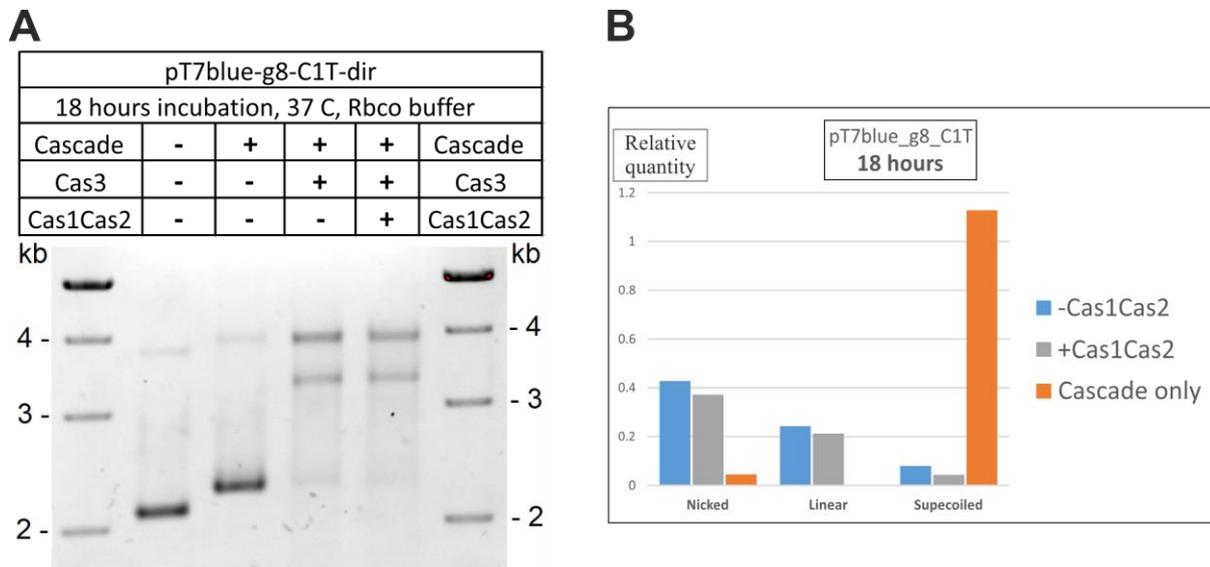


Figure 22. Cas proteins effect on priming mutant target degradation. Cascade, Cascade with Cas3, and Cascade, Cas3, Cas1-Cas2 proteins were mixed with the target plasmid and incubated for 18 hours at 37 °C. Plasmid without any Cas proteins was used as a control. (A) Reaction products were analysed on agarose gel. (B) Quantification of gel bands is presented. No significant difference induced by Cas1-Cas2 is observed. In opposite of expected outcome, the Cas1-Cas2 presence leads to a slight loss of linear and nicked forms – 3% difference.

Thus, Cas1-Cas2 appear to have no significant effect on the process of interference products accumulation at our conditions.

4.8.2 Short fragments investigation

Despite the fact that Cas1-Cas2 does not drastically speed up degradation of target plasmid by Cas3, it may still generate fragments necessary for insertion into the array. Data from Musharova et al (83) indicate that in the cell 33 nt DNA fragments can be cut out of target DNA by Cas proteins activity from templates containing priming protospacers. If this process creates exactly 33 nt length parts of DNA, they should be detectable. Even if the rate of Cas3 only (not triggered by Cas1-Cas2) target degradation is significantly higher than at priming conditions, the former should generate DNA fragments of random lengths distributed as a smear on gels, while Cas1-Cas2 activity should be seen as an enhanced band of defined length. Thereby, the reaction shown in Figure 22 was repeated but on a

larger scale. Reaction products were purified by phenol-chloroform DNA extraction, such that short fragments generated or grabbed by Cas1-Cas2 were released. After ethanol DNA precipitation, the samples were separated by denaturing Urea PAGE (Figure 23), which allows to analyze short fragments at single-nucleotide resolution. Oligonucleotides of different sizes were used as size markers. However, no difference in 33-nt area was detected in the presence or absence of Cas1-Cas2. The conclusion from these results is that some co-factors are missing in our reactions, preventing the generation of spacer-sized fragments for insertion (or that there is no specific 33-nt fragments generation during CRISPR-Cas response under priming conditions).

One of the possibilities discussed was that additional nucleases degrade material partially damaged by Cas3 in the cell, leaving only 33 nt fragments bound by Cas1-Cas2 protected. To test this theory, the same reaction set was repeated, but this time the products were treated with DNase I, which could degrade all DNA, except for that protected by a bound protein. Again, no enhancement of specific bands in the presence of Cas1-Cas2 was seen after Urea PAGE (supplementary Figure D9). While the assay has still to be optimized in terms of DNase concentration and incubation time, there could be a possibility that the binding of Cas1-Cas2 is not strong enough *in vitro*. Alternatively, some new insertion-substrate generating process not understood properly has to be revealed in future research.

One interesting fact that appeared during this experiment was shortening of crRNA after the incubation with target plasmid. It is hard to say why it is happening, but it is unlikely to be related with target binding, since Cascade is present in the solution in 100 nM concentration, when target plasmid is only at 5 nM.

It should be mentioned, that this assay was repeated multiple times during optimization of buffer conditions and proving activity of Cas1-Cas2 protein complex. Results presented on the figures were obtained with optimal conditions. It was done in the same buffer, where spacer integrase activity of Cas1-Cas2 was observed (see next chapter).

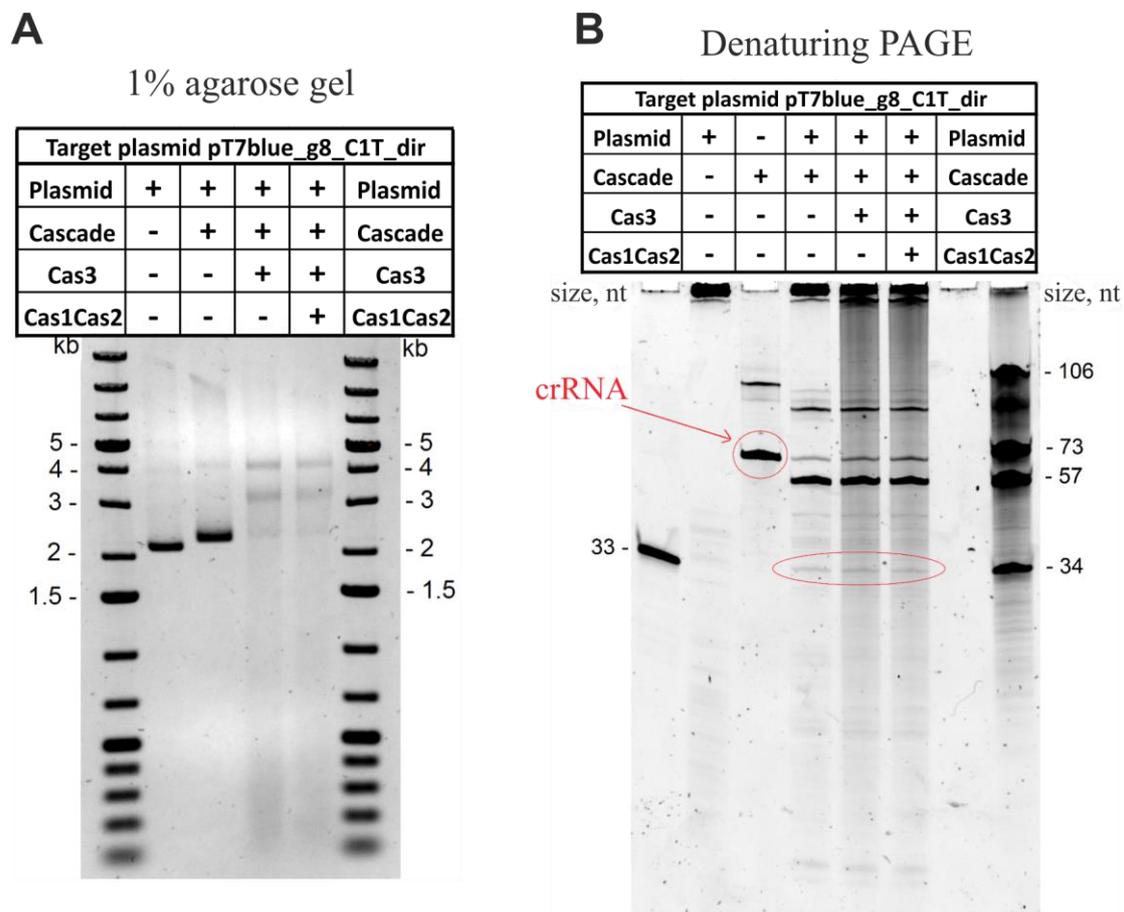


Figure 23. Investigation of short DNA fragments generated in the reactions with Cascade, Cas3, and Cas1-Cas2 proteins. On (A) is repeat of experiment presented in Figure 22 (B) represents denaturing Urea PAGE showing short DNA and RNA fragments in reactions. It can be seen, that non-cleaved plasmid contains almost no short fragments. 61-nt crRNA of Cascade complex can be clearly seen on the gel, although, so far not clear why, it gets shortened after the incubation, and a band corresponding to around double size of crRNA appears. Bands corresponding to ~33-nt size fragments do not show any difference in the presence of Cascade, Cascade plus Cas3, and Cascade plus Cas3 plus Cas1-Cas2.

4.8.3 Complete reaction of primed adaptation. Attempt of reconstitution

The main goal during this project development was to reproduce the full cycle of primed adaptation. Even despite the fact that insights and ideas into how the process can happen did not bring results in Cas1-Cas2 activity experiments, it does not mean that the full cycle is not reproducible *in vitro*. In this chapter summary of experiments describing such attempts is presented.

As was already seen, interference can be reproduced *in vitro*, showing Cascade induced Cas3 nuclease activity, which is required for primed adaptation (113). The best way to check the adaptation activity in any conditions is to perform Cas1-Cas2 dependent integration test. In (72) it was shown that Cas1-Cas2 can integrate 33-bp non-phosphorylated oligoduplex into the plasmid containing CRISPR leader

region and several repeats with spacers. Indeed, mixing 200 nM of 33-nt oligoduplex with plasmid and Cas1-Cas2 in the presence of Mg^{2+} initiates integration, changing the plasmid state (Figure 24). Due to Cas1 nuclease activity, initiation of the integration events leads to accumulation of the circular relaxed plasmid form, signaling accumulation of single-stranded breaks in DNA. But the most noticeable distinctive effect is the secondary product of integration reaction, which is believed to be integration-disintegration result leading to accumulation of plasmid topoisomers with different supercoiling index. Those topoisomers run in the gel between original supercoiled plasmid and the relaxed circle form and can be seen as multiple faint bands.

It is seen on the figure, that Cas1-Cas2 is inactive on integration target without oligoduplex substrate. Two batches of Cas1-Cas2 were tested and both show expected activity. The need of preincubation of Cas1-Cas2 with oligoduplex before contact with plasmid was studied. The reason was high stickiness of Cas1-Cas2, which could lead to Cas1-Cas2 binding to the plasmid with activity loss. The results indicate that preincubation step can be skipped.

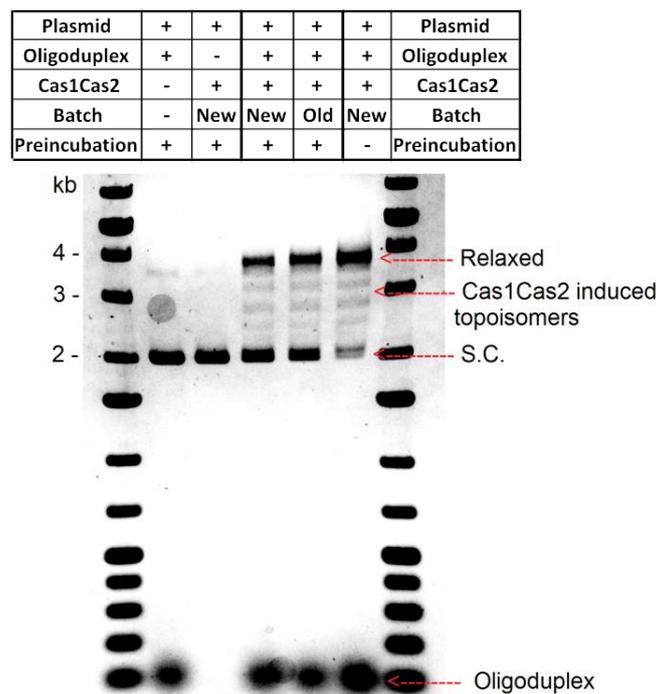


Figure 24. Oligoduplex integration by Cas1-Cas2 into plasmid containing CRISPR leader and repeats. Products of Cas1-Cas2 activity are clearly seen as the relaxed plasmid form and topoisomers.

First assays on Cas1-Cas2 integration in the conditions used for Cascade and Cas3 activity assays, showed, surprisingly, that the Cas1-Cas2 complex is inactive in the

buffers used. Additional research revealed Cas1-Cas2 salt concentration dependence (Figure 25). The integration activity was not detected at higher than 100 mM NaCl or KCl in the solution. Taking this into account, the buffer base was switched from Tris to HEPES, and salt concentrations compatible with both Cas1-Cas2 and Cas3-Cascade reactions were established. The optimized buffer was 20 mM HEPES pH 7.5, 35 mM KCl, 10 mM MgCl₂, 10 μM CoCl₂, 1.5 mM ATP, 1 mM TCEP. Cobalt and magnesium are cofactors for Cas3 ATP-dependent activity, magnesium as well is a cofactor for Cas1-Cas2 activity, reducing agent tris (2-carboxyethyl)phosphine (TCEP, Sigma, Germany) was added to minimize possible protein aggregation due to disulfide bonds formation.

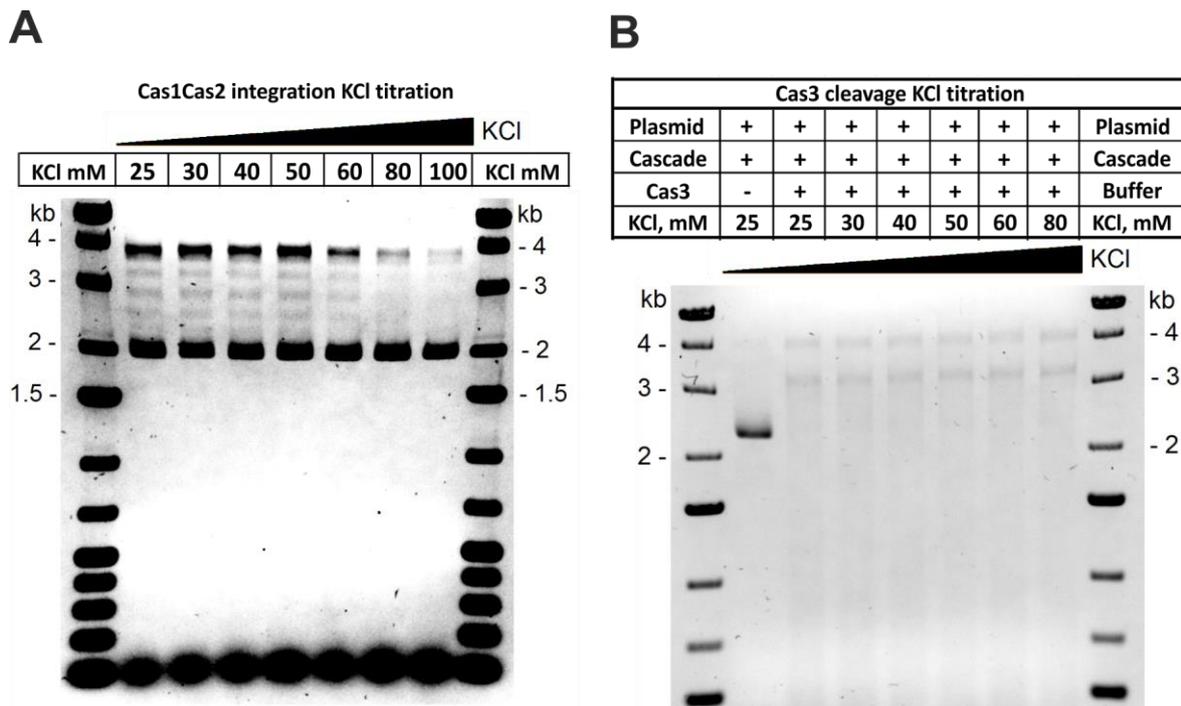


Figure 25. Cas1-Cas2 and Cas3 plus Cascade activity under titration of KCl concentration. (A) Cas1-Cas2 is limited by salt concentration and working better in the range of 25-50 mM KCl. (B) Cas3 cleavage on Cascade bound target is happening efficiently at any salt concentration tested.

Once the Cas1-Cas2 integration ability was proved, we tested if Cas1-Cas2 can use the products of *in vitro* Cas3 interference as a substrate for spacer insertion. Accordingly, a combined experiment was done: target plasmid with g8 protospacer was added to Cascade and Cas3 mixture, and then, immediately or after incubation for 1 hour, a Cas1-Cas2 solution containing “pUC_mini_CRISPR (*E. coli*)” – plasmid for insertion was added (Figure 26). Unfortunately, the results obtained so far did not show any processing of plasmid containing CRISPR part.

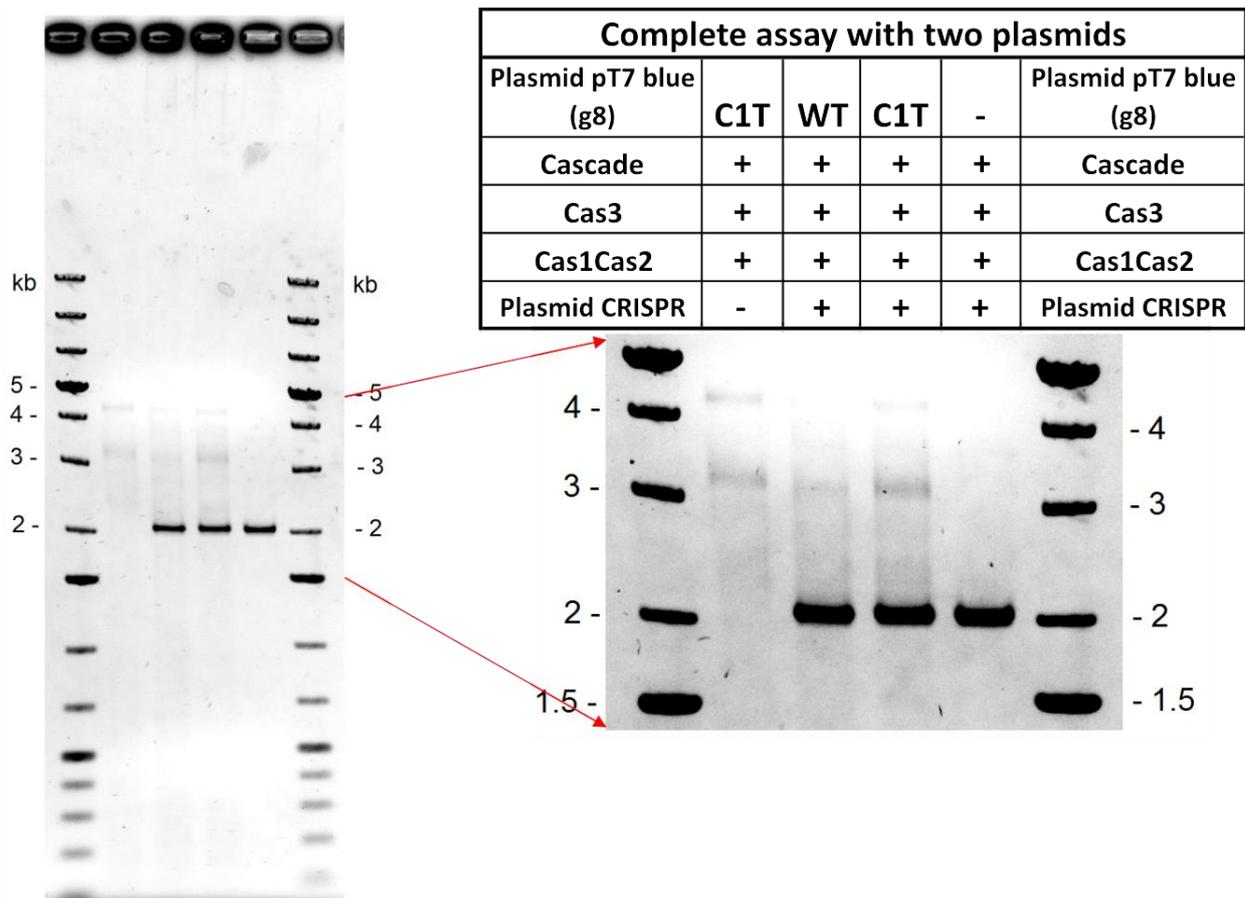


Figure 26. First attempt for complete *in vitro* primed adaptation reconstitution; experiment in the presence of all Cas proteins, target degradation plasmid and target plasmid for insertion. The target plasmid pT7blue containing target g8 protospacer with wild type or priming C1T mutation was mixed with Cascade (g8 crRNA) and Cas3. Then, immediately, Cas1-Cas2 protein with plasmid containing *E. coli* derived CRISPR leader sequence was added. The mixture was incubated at 37 °C for 1 hour. No sign of insertion is seen. Missing processing machinery or lack of substrate for integration are the most probable issues.

The reason for the lack of integration activity could be that some other machinery in the cell could be needed for processing degradation products generated by Cas3. In that case, as was discussed above, some nucleases can degrade DNA overhangs around the double-stranded part bound by Cascade. One possible way to solve this problem would be to use available nucleases in add-back experiments. Other candidates could be phosphatases that remove phosphate groups from 3' ends of oligoduplex, or even polymerases that can take part in the processing of the insertion substrate fragments. In this cases, more data on *in vivo* process and/or extract fractionation coupled with activity assays is required to identify such missing factors.

Another reason can be the lack of sufficient amounts of integration substrate generated. Integration tests were performed at 200 nM concentration of oligoduplex, while target plasmid with g8 protospacer was used usually at 5 nM. If from one DNA molecule only one fragment for insertion is generated, then the

reaction would be 40 times slower. However, the exact number of substrates that can be taken from one target molecule is not known. Trying to solve this issue, the product of g8 target plasmid incubated with all Cas proteins was carried through phenol-chloroform DNA extraction and was resuspended in small volume, concentrating degradation products 20 times. This concentrated product was incubated with Cas1-Cas2 and pCRISPR plasmid for insertion (supplementary Figure D11). The design of experiment should be elaborated further. Since both plasmids have about the same size, concentrated degradation products mentioned above makes indistinguishable any changes happening to pCRISPR. In that sense, there is clearly room to develop the procedures and get more results out of these experiments.

4.9 Summary of results

We would now like to summarize the insights obtained during the project. The most important (in our opinion) results for understanding the priming phenomenon are reflected in Table 2, with asterisk indicating the novel results obtained in this project. First, the MT assays showed that all point seed mutants tested supported unwinding of DNA in the presence of Cascade. Target protospacer melting was stable at both positive and negative rotations and was undistinguishable in its extent from that observed on wild-type g8 protospacer, meaning that, most likely, full-sized R-loops were formed for each seed mutant. This conclusion was supported by bulk assays: agarose gels with plasmids and permanganate footprinting experiments made by our collaborator. Negative control with CCG trinucleotide in the place of PAM, as expected, did not show any R-loop formation in either of bulk, MT, or permanganate footprinting experiments.

Kinetics of R-loop formation for seed mutants was directly studied and shown to be much slower (by 5 to 50 fold) than for the wild-type target. Locking, tested by measuring R-loop dissociation times under positive torque applied was shown to be dependent on the position of mismatch: the closer it was to PAM, the less stable the complex was. PAM mutant (at -1 position) demonstrated weak locking similar to mismatch in the +1 position (first position after PAM in protospacer). At the same time, mismatch in the fourth (+4) position showed the locking ability that was as high as that for wild-type target. The second and third positions showed intermediate locking strengths. Overall, *E. coli* Cascade locking was, surprisingly, much stronger compared to the previously studied *Streptococcus thermophilus* Cascade.

Cas3 cleavage kinetics observed in bulk assays correlated well with the strength of R-loop locking measured in the magnetic tweezers assay, i.e., targets with longer

dissociation times of R-loops under high positive torque were cleaved faster by Cas3.

Finally, priming experiments *in vivo* revealed that every tested seed mutant supported primed adaptation independent of its locking or Cas3 recruitment ability. HTS data revealed same profiles of acquired spacers for every seed mutant.

The qPCR approach allowed to evaluate adaptation efficiency quantitatively, and to detect some adaptation for wild-type target. The result supports genetic experiments of (124) and shows that qPCR approach is much more sensitive in detecting low-level adaptation than standard amplification methods.

Table 2. A summary of observations from MT, biochemical and *in vivo* experiments for the target variants tested. This data was obtained in the current work (marked with asterisk) or taken from literature (45, 113, 114, 156).

Results					
Target variant	Escape in vivo	R-loop formation	R-loop locking	Cas3 cleavage	Priming
CCG PAM	Yes	No	N/A	N/A	No
WT	No	Fast*	Strong*	Fast	No
G-1T	Yes	Very slow*	Weak*	Slow*	Yes
C1T	Yes	Very slow*	Weak*	Slow	Yes
T2A	Yes	Slow*	Moderate*	Moderate*	Yes*
G3T	Yes	Slow*	Moderate*	Fast*	Yes*
T4G	Yes	Slow*	Strong*	Fast*	Yes*

5 Discussion

It has been previously shown that a +1 seed mutation causes escape from CRISPR interference and stimulates primed adaptation in *E. coli* cultures (113). The Cse1 subunit of Cascade for an R-loop formed on +1 seed mutation target (as well as on targets with single PAM mutations) adopted an open conformation, while it was found at closed conformation on wild-type targets (114). A +1 seed mutation was also reported to support Cas3 DNA degradation at significantly reduced rate. The magnetic tweezers assay measures a global effect of conformational changes that contribute to locking. Given that we observe extremely strong locking on the WT target and weak locking for a +1 seed mismatch target, we may conclude that full locking requires closed Cse1 conformation. Thus, the attenuated locking and DNA degradation that we observe for targets with the G-1T PAM and with the +1 as well as the +2 seed mutations are consistent with a model of predominantly open Cse1 conformation on these substrates. Locking involves a large movement of Cse1 and the Cse2 dimer – the latter establishing PAM-distal DNA contacts (88). The crystal structure of Cascade with bound single-stranded DNA (170), as well as Cryo-EM data and molecular dynamics simulations (170), suggest that on weakly locked targets the Cse2 dimer adopts a locked position that stabilizes the R-loop on the PAM-distal side, while Cse1 remains in the open conformation. The open conformation of Cse1 fails to support full R-loop locking, leaving the R-loop in a “semi-locked” state (Figure 27). The differences in R-loop stabilities between the weakly locked targets may be due to Cse1 being in dynamic equilibrium between the predominantly open and closed conformations. Targets with seed mismatches more distal to PAM (from position +3 onwards) support wild-type like locking and thus Cse1 should adopt here a closed conformation, which is additionally supported by the wild-type like DNA degradation rates. The fact that locking is practically irreversible on these substrates suggests that the closed state is almost exclusively occupied. Important to note that there may be slight differences in the occupancy of the closed Cse1 conformation for the wild-type and the +4 mismatch target, since we cannot quantitatively evaluate the differences in locking strength between these two targets with our methods.

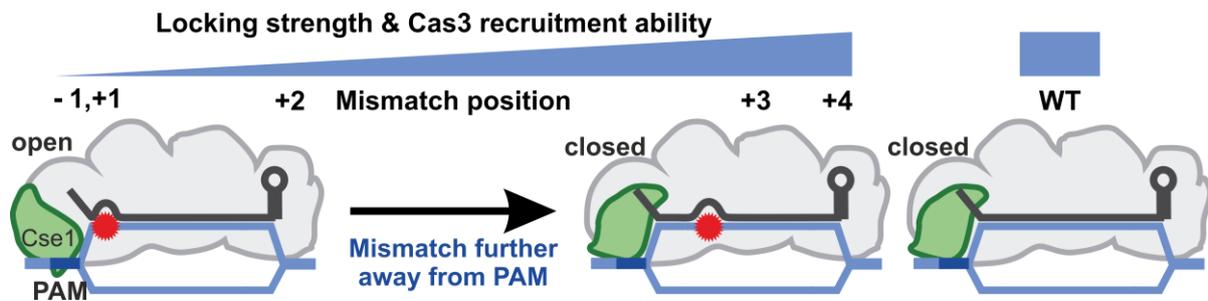


Figure 27. A model of conformation states of Cascade at different target variants. The Cse1 subunit is predominantly in the open state for the -1 and +1 substitutions resulting in a weak (semi-locked) R-loop that is not able to support efficient Cas3 recruitment. For other seed substitutions the locking strength and the ability to recruit Cas3 increase with increasing distance from the PAM, reaching the wild-type level for +4 substitution (suggesting a closed wild-type like Cse1 conformation that readily recruits Cas3).

Dual control of DNA degradation by (i) triggering locking upon R-loop expansion until the PAM-distal end of the target and (ii) additional verification of PAM by Cse1 seems to be a shared mechanism of type I-E CRISPR-Cas systems Cascade complexes. It has been shown for *S. thermophilus* Cascade (85) that R-loop degradation is impeded for PAM mutants, while R-loops with a +2 seed mismatch (corresponding to the +1 position in *E. coli* Cascade) were cleaved at wild-type rate (85). For *T. fusca* Cascade, Cas3-recruitment is impeded by PAM and +1 seed mutations but not by more PAM-distal mutations (136). The relative involvement of the first base pairs of the seed in this additional verification step seems, however, to vary between these species.

The strong differences in R-loop degradation for target variants tested in our work did not lead to changes in primed adaptation, which occurred at comparable levels for all mutant targets. This result is difficult to reconcile with a model in which priming is triggered by the open conformation of Cse1 (114). The open-form Cascade is thought to represent a specific priming signal for a Cas3-Cas1-Cas2 complex, inducing a distinct mode of Cas3 movement along the DNA molecule with less DNA degradation and concomitant spacer acquisition (81). While this model can explain the behavior of protospacers with mutations at positions -1, +1 and +2, the priming behavior of +3 and +4 mismatches is not explained.

An alternative kinetic model (126) posits that the persistent presence of target plasmid DNA at conditions of reduced interference allows bulk levels of spacer acquisition in cultures that by far exceed the acquisition levels that can be attained during a restricted time window in the case of rapid interference. According to this model, high yields of primed adaptation are a consequence of a steady slow Cas3-based production of target DNA fragments at low interference rates such that the loss of invader DNA can be compensated by its ongoing replication. Since R-loop complexes with mismatches at positions +3 and +4 readily recruit Cas3 and support rapid DNA degradation, their ability to promote priming should arise only

from the slower R-loop formation kinetics detected *in vitro*. Likewise the kinetic model can explain priming for the targets with mutations at positions -1, +1, and +2 since they also exhibit a low overall rate of target degradation. Thus, for all tested target variants, invader DNA should, at conditions of ongoing replication, persist over longer periods of time. Together with the ongoing degradation of the foreign DNA a constant production of substrates for spacer acquisition by Cas1-Cas2 should be ensured. Thus, our data - while clearly supporting the existence of multiple conformations of the Cascade complex on target protospacers - are more consistent with a minimalistic kinetic model for primed adaptation.

Approaches developed in this work allow to design experiments to further test the models of priming. First, if proposed by Severinov et al. kinetic model is true (supplementary Figure D10), targets that are subject for strong interference, should deplete out after a reasonable time (from minutes to hours), and should produce no fuel for adaptation thereafter. If so, experiments presented in chapter 4.7.2 performed for a series of time points, especially in few hours to dozens of hours time range, should be of interest.

The nature of the spacer selection process still remains unknown: what causes *in cis* strand preference and hot spot profile for selected spacers (113, 115, 119, 120, 122, 123). Several ideas were proposed, but no solid proofs were made so far. If Cas1-Cas2 in type I-E system of *E. coli* is capable to cooperate with Cas3 during DNA fragments (future spacers) selection, single-molecule studies could be used to reveal this cooperation (81). Fluorescently labeled functional Cas1-Cas2 would allow to demonstrate directly whether Cas1-Cas2-Cas3-Cascade interaction happens. Another approach is to capture the Cas1-Cas2 interaction with Cas3 *in vivo*: by co-purification of proteins, hopefully, with DNA adaptation intermediates. Another approach to understand the process is to compare molecular machinery of *E. coli* system with others that support primed adaptation but demonstrate different spacer selection preferences, such as inverted strand bias polarity as is the case of I-E and I-F systems (94, 120, 123, 171) .

Another development that can help to better understand priming is FRET studies of Cse1 conformation of Cascade when bound to the PAM-distal seed mutated targets (point mutations at positions +3,+4 and further). We expect that Cse1 conformation will become closed starting from +3 and +4, with small occupancy of open conformation for +3 seed mutant (114). However, we cannot undoubtedly claim from our experiments that the conformation of Cse1 subunit at +4 seed mutant is identical to the wild-type. There might be changes that are not detected at MT assays: since we applied close to highest possible torque, it is unknown if the +4 mutant would show R-loop dissociation under the higher torque and if there will be any difference between it and the wild-type target.

With regards to MT measurements of Cascade complexes with altered length crRNA, first of all, they provide an additional proof for MT assay validity: the shifts of DNA molecules' supercoiled states were in linear correlation with expected R-loop lengths. Observed alterations in crRNA complexes locking properties could help understand the conformational changes in Cascade required for locking. Even though the -6 and -3 crRNA variants were shown to contain two Cse2 and five Cas7 subunits (159), their locking ability was weak in ~66% of cases, while stronger locking was observed in remaining cases. This "stronger" locking still did not survive high positive torque for longer than dozens of second. The -3 variant has shown even stronger locking level, dissociating only after dozens of minutes. Since mass-spectrometry demonstrated homogeneity of -6 and -3 variants (no variation in subunit composition of complexes) (159), such difference may be explained by different, discrete occupancy levels of locked (presumably closed) to not-locked (presumably open) conformation of Cse1 subunit or, alternatively, by the presence of several, more than two stable conformational states of Cse1 (all affecting the locking strength). The mechanism that defines what mode of Cse1 is assumed to remain unknown. This part of the project resulted in preparation of scripts that are able to analyze rotational shifts with independent averaging and fitting of parts of rotational curve ensuring better precision of measurement. These scripts shall be helpful in future investigations.

As for the improvements and adjustments of the MT assays for *E. coli* Cascade measurements, we believe that majority of the developments were useful and efficient. The most questionable development is synthetic genes concept. In addition to issues with backbone "breathing" uncovered during MT measurements, new data indicate that many more trinucleotides can function as functional PAMs for *E. coli* Cascade, making it difficult or even impossible to create the needed templates (45, 136).

As for bulk assays reported in chapter 4.8, the main achievement of those is defining conditions where every protein component of the CRISPR-Cas system is functional: Cascade, Cas3, and Cas1-Cas2. Attempts to reconstitute the full priming cycle *in vitro* were unsuccessful, however. Previous studies suggested that an additional cell machinery other than the Cas proteins may be required for primed adaptation (172). In this case, more *in vivo* studies would be performed before reconstituting primed adaptation *in vitro*. One model proposes that Cas3 is not the main nuclease that degrades targeted DNA molecule, but that it just starts the degradation process (79). It is also possible that *in vitro* Cas3 just does not produce enough necessary "fuel" for Cas1-Cas2 to capture. Another potential reason for lack of success in reconstitution of primed adaptation may be low concentrations of generated substrates for spacer insertion.

Overall, our project has brought valuable insights to the model of priming, and laid ground for further development and studies of CRISPR-Cas systems function by combining biophysical, biochemical and genetic approaches.

References

1. Koonin, E. V and Dolja, V. V (2013) A virocentric perspective on the evolution of life. *Curr. Opin. Virol.*, **3**, 546–557.
2. Brüssow, H. and Hendrix, R.W. (2002) Phage genomics: small is beautiful. *Cell*, **108**, 13–6.
3. Labrie, S.J., Samson, J.E. and Moineau, S. (2010) Bacteriophage resistance mechanisms. *Nat. Rev. Microbiol.*, **8**, 317–327.
4. Rakhuba, D. V, Kolomiets, E.I., Dey, E.S. and Novik, G.I. (2010) Bacteriophage receptors, mechanisms of phage adsorption and penetration into host cell. *Polish J. Microbiol.*, **59**, 145–55.
5. Foster, T.J. (2005) Immune evasion by staphylococci. *Nat. Rev. Microbiol.*, **3**, 948–958.
6. Nordström, K. and Forsgren, A. (1974) Effect of protein A on adsorption of bacteriophages to *Staphylococcus aureus*. *J. Virol.*, **14**, 198–202.
7. Riede, I. and Eschbach, M.L. (1986) Evidence that TraT interacts with OmpA of *Escherichia coli*. *FEBS Lett.*, **205**, 241–5.
8. Pedruzzi, I., Rosenbusch, J.P. and Locher, K.P. (1998) Inactivation in vitro of the *Escherichia coli* outer membrane protein FhuA by a phage T5-encoded lipoprotein. *FEMS Microbiol. Lett.*, **168**, 119–25.
9. Utsumi, R. (2008) Bacterial signal transduction: networks and drug targets. Preface. *Adv. Exp. Med. Biol.*, **631**, v.
10. Liu, M., Deora, R., Doulatov, S.R., Gingery, M., Eiserling, F.A., Preston, A., Maskell, D.J., Simons, R.W., Cotter, P.A., Parkhill, J., *et al.* (2002) Reverse Transcriptase-Mediated Tropism Switching in *Bordetella* Bacteriophage. *Science.*, **295**, 2091–2094.
11. Stummeyer, K., Schwarzer, D., Claus, H., Vogel, U., Gerardy-Schahn, R. and Muhlenhoff, M. (2006) Evolution of bacteriophages infecting encapsulated bacteria: lessons from *Escherichia coli* K1-specific phages. *Mol. Microbiol.*, **60**, 1123–1135.
12. Sutherland, I.W. (1995) Polysaccharide lyases. *FEMS Microbiol. Rev.*, **16**, 323–47.
13. Destoumieux-Garzón, D., Duquesne, S., Peduzzi, J., Goulard, C., Desmadril, M., Letellier, L., Rebuffat, S. and Boulanger, P. (2005) The iron-siderophore transporter FhuA is the receptor for the antimicrobial peptide microcin J25: role of the microcin Val11-Pro16 beta-hairpin region in the recognition mechanism. *Biochem. J.*, **389**, 869–76.
14. Akçelik, M., Sanlibaba, P. and Tükel, C. (2000) Phage resistance in *Lactococcus*

- lactis subsp. lactis strains isolated from traditional fermented milk products in Turkey. *Int. J. Food Sci. Technol.*, **35**, 473–481.
15. Cumby,N., Edwards,A.M., Davidson,A.R. and Maxwell,K.L. (2012) The Bacteriophage HK97 gp15 Moron Element Encodes a Novel Superinfection Exclusion Protein. *J. Bacteriol.*, **194**, 5012–5019.
 16. Bebeacua,C., Lorenzo Fajardo,J.C., Blangy,S., Spinelli,S., Bollmann,S., Neve,H., Cambillau,C. and Heller,K.J. (2013) X-ray structure of a superinfection exclusion lipoprotein from phage TP-J34 and identification of the tape measure protein as its target. *Mol. Microbiol.*, **89**, 152–165.
 17. Lu,M.J. and Henning,U. (1994) Superinfection exclusion by T-even-type coliphages. *Trends Microbiol.*, **2**, 137–9.
 18. Seed,K.D. (2015) Battling Phages: How Bacteria Defend against Viral Attack. *PLOS Pathog.*, **11**, e1004847.
 19. Garvey,P., Hill,C. and Fitzgerald,G.F. (1996) The lactococcal plasmid pNP40 encodes a third bacteriophage resistance mechanism, one which affects phage DNA penetration. *Appl. Environ. Microbiol.*, **62**, 676–679.
 20. Mahony,J., McGrath,S., Fitzgerald,G.F. and van Sinderen,D. (2008) Identification and characterization of lactococcal-prophage-carried superinfection exclusion genes. *Appl. Environ. Microbiol.*, **74**, 6206–15.
 21. Tock,M.R. and Dryden,D.T. (2005) The biology of restriction and anti-restriction. *Curr. Opin. Microbiol.*, **8**, 466–472.
 22. Krüger,D.H. and Bickle,T.A. (1983) Bacteriophage survival: multiple mechanisms for avoiding the deoxyribonucleic acid restriction systems of their hosts. *Microbiol. Rev.*, **47**, 345–60.
 23. Krüger,D.H., Barcak,G.J. and Smith,H.O. (1988) Abolition of DNA recognition site resistance to the restriction endonuclease EcoRII. *Biomed. Biochim. Acta*, **47**, K1-5.
 24. McGrath,S., Seegers,J.F., Fitzgerald,G.F. and van Sinderen,D. (1999) Molecular characterization of a phage-encoded resistance system in *Lactococcus lactis*. *Appl. Environ. Microbiol.*, **65**, 1891–9.
 25. Roberts,R.J., Belfort,M., Bestor,T., Bhagwat,A.S., Bickle,T.A., Bitinaite,J., Blumenthal,R.M., Degtyarev,S.K., Dryden,D.T.F., Dybvig,K., *et al.* (2003) A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res.*, **31**, 1805–12.
 26. Stewart,F.J., Panne,D., Bickle,T.A. and Raleigh,E.A. (2000) Methyl-specific DNA binding by McrBC, a modification-dependent restriction enzyme. *J. Mol. Biol.*, **298**, 611–622.
 27. Ofir,G., Melamed,S., Sberro,H., Mukamel,Z., Silverman,S., Yaakov,G.,

- Doron,S. and Sorek,R. (2018) DISARM is a widespread bacterial defence system with broad anti-phage activities. *Nat. Microbiol.*, **3**, 90–98.
28. 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.
29. Mojica,F.J.M., 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.
30. Sorek,R., Lawrence,C.M. and Wiedenheft,B. (2013) CRISPR-Mediated Adaptive Immune Systems in Bacteria and Archaea. *Annu. Rev. Biochem.*, **82**, 237–266.
31. van der Oost,J., Westra,E.R., Jackson,R.N. and Wiedenheft,B. (2014) Unravelling the structural and mechanistic basis of CRISPR–Cas systems. *Nat. Rev. Microbiol.*, **12**, 479–492.
32. Bolotin,A., Quinquis,B., Sorokin,A. and Ehrlich,S.D. (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology*, **151**, 2551–2561.
33. Mojica,F.J.M., Díez-Villasenor,C., Garcia-Martinez,J. and Soria,E. (2005) Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements. *J. Mol. Evol.*, **60**, 174–182.
34. Burstein,D., Sun,C.L., Brown,C.T., Sharon,I., Anantharaman,K., Probst,A.J., Thomas,B.C. and Banfield,J.F. (2016) Major bacterial lineages are essentially devoid of CRISPR-Cas viral defence systems. *Nat. Commun.*, **7**, 10613.
35. Hille,F., Richter,H., Wong,S.P., Bratovič,M., Ressel,S. and Charpentier,E. (2018) The Biology of CRISPR-Cas: Backward and Forward. *Cell*, **172**, 1239–1259.
36. Sternberg,S.H., Richter,H., Charpentier,E. and Qimron,U. (2016) Adaptation in CRISPR-Cas Systems. *Mol. Cell*, **61**, 797–808.
37. Wright,A.V., Nuñez,J.K. and Doudna,J.A. (2016) Biology and Applications of CRISPR Systems: Harnessing Nature’s Toolbox for Genome Engineering. *Cell*, **164**, 29–44.
38. Grissa,I., Vergnaud,G. and Pourcel,C. (2007) CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.*, **35**, W52–W57.
39. Lawrence,C.M. and White,M.F. (2011) Recognition of Archaeal CRISPR RNA: No P in the alindromic Repeat? *Structure*, **19**, 142–144.
40. Ishino,Y., Shinagawa,H., Makino,K., Amemura,M. and 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**, 5429–33.
41. Marraffini, L.A. and Sontheimer, E.J. (2010) CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat. Rev. Genet.*, **11**, 181–90.
 42. Rath, D., Amlinger, L., Rath, A. and Lundgren, M. (2015) The CRISPR-Cas immune system: Biology, mechanisms and applications. *Biochimie*, **117**, 119–128.
 43. Koonin, E. V., Makarova, K.S. and Zhang, F. (2017) Diversity, classification and evolution of CRISPR-Cas systems. *Curr. Opin. Microbiol.*, **37**, 67–78.
 44. Deveau, H., Barrangou, R., Garneau, J.E., Labonte, J., Fremaux, C., Boyaval, P., Romero, D.A., Horvath, P. and Moineau, S. (2008) Phage Response to CRISPR-Encoded Resistance in *Streptococcus thermophilus*. *J. Bacteriol.*, **190**, 1390–1400.
 45. Fineran, P.C., Gerritzen, M.J.H., Suarez-Diez, M., Kunne, T., Boekhorst, J., van Hijum, S.A.F.T., Staals, R.H.J. and Brouns, S.J.J. (2014) Degenerate target sites mediate rapid primed CRISPR adaptation. *Proc. Natl. Acad. Sci.*, **111**, E1629–E1638.
 46. Semenova, E., Jore, M.M., Datsenko, K.A., Semenova, A., Westra, E.R., Wanner, B., van der Oost, J., Brouns, S.J.J. and Severinov, K. (2011) Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl. Acad. Sci.*, **108**, 10098–10103.
 47. Maxwell, K.L., Garcia, B., Bondy-Denomy, J., Bona, D., Hidalgo-Reyes, Y. and Davidson, A.R. (2016) The solution structure of an anti-CRISPR protein. *Nat. Commun.*, **7**, 13134.
 48. Borges, A.L., Davidson, A.R. and Bondy-Denomy, J. (2017) The Discovery, Mechanisms, and Evolutionary Impact of Anti-CRISPRs. *Annu. Rev. Virol.*, **4**, 37–59.
 49. He, F., Bhoobalan-Chitty, Y., Van, L.B., Kjeldsen, A.L., Dedola, M., Makarova, K.S., Koonin, E. V., Brodersen, D.E. and Peng, X. (2018) Anti-CRISPR proteins encoded by archaeal lytic viruses inhibit subtype I-D immunity. *Nat. Microbiol.*, **3**, 461–469.
 50. Chopin, M.-C., Chopin, A. and Bidnenko, E. (2005) Phage abortive infection in lactococci: variations on a theme. *Curr. Opin. Microbiol.*, **8**, 473–479.
 51. García, L.R. and Molineux, I.J. (1995) Incomplete entry of bacteriophage T7 DNA into F plasmid-containing *Escherichia coli*. *J. Bacteriol.*, **177**, 4077–83.
 52. Emond, E., Holler, B.J., Boucher, I., Vandenberg, P.A., Vedamuthu, E.R.,

- Kondo, J.K. and Moineau, S. (1997) Phenotypic and genetic characterization of the bacteriophage abortive infection mechanism AbiK from *Lactococcus lactis*. *Appl. Environ. Microbiol.*, **63**, 1274–83.
53. Domingues, S., Chopin, A., Ehrlich, S.D. and Chopin, M.-C. (2004) The Lactococcal abortive phage infection system AbiP prevents both phage DNA replication and temporal transcription switch. *J. Bacteriol.*, **186**, 713–21.
54. Cluzel, P.J., Chopin, A., Ehrlich, S.D. and Chopin, M.C. (1991) Phage abortive infection mechanism from *Lactococcus lactis* subsp. *lactis*, expression of which is mediated by an Iso-ISS1 element. *Appl. Environ. Microbiol.*, **57**, 3547–51.
55. Durmaz, E., Higgins, D.L. and Klaenhammer, T.R. (1992) Molecular characterization of a second abortive phage resistance gene present in *Lactococcus lactis* subsp. *lactis* ME2. *J. Bacteriol.*, **174**, 7463–9.
56. Emond, E., Dion, E., Walker, S.A., Vedamuthu, E.R., Kondo, J.K. and Moineau, S. (1998) AbiQ, an abortive infection mechanism from *Lactococcus lactis*. *Appl. Environ. Microbiol.*, **64**, 4748–56.
57. Bouchard, J.D. and Moineau, S. (2004) Lactococcal phage genes involved in sensitivity to AbiK and their relation to single-strand annealing proteins. *J. Bacteriol.*, **186**, 3649–52.
58. Fortier, L.-C., Bouchard, J.D. and Moineau, S. (2005) Expression and Site-Directed Mutagenesis of the Lactococcal Abortive Phage Infection Protein AbiK. *J. Bacteriol.*, **187**, 3721–3730.
59. Molineux, I.J. (1991) Host-parasite interactions: recent developments in the genetics of abortive phage infections. *New Biol.*, **3**, 230–6.
60. Snyder, L. (1995) Phage-exclusion enzymes: a bonanza of biochemical and cell biology reagents? *Mol. Microbiol.*, **15**, 415–20.
61. Bingham, R., Ekunwe, S.I.N., Falk, S., Snyder, L. and Kleanthous, C. (2000) The Major Head Protein of Bacteriophage T4 Binds Specifically to Elongation Factor Tu. *J. Biol. Chem.*, **275**, 23219–23226.
62. Fineran, P.C., Blower, T.R., Foulds, I.J., Humphreys, D.P., Lilley, K.S. and Salmond, G.P.C. (2009) The phage abortive infection system, ToxIN, functions as a protein–RNA toxin–antitoxin pair. *Proc. Natl. Acad. Sci.*, **106**, 894–899.
63. Hazan, R. and Engelberg-Kulka, H. (2004) *Escherichia coli* mazEF-mediated cell death as a defense mechanism that inhibits the spread of phage P1. *Mol. Genet. Genomics*, **272**, 227–34.
64. Abudayyeh, O.O., Gootenberg, J.S., Konermann, S., Joung, J., Slaymaker, I.M., Cox, D.B.T., Shmakov, S., Makarova, K.S., Semenova, E., Minakhin, L., *et al.* (2016) C2c2 is a single-component programmable RNA-guided RNA-

- targeting CRISPR effector. *Science.*, **353**, aaf5573.
65. Chen,J.S., Ma,E., Harrington,L.B., Da Costa,M., Tian,X., Palefsky,J.M. and Doudna,J.A. (2018) CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science.*, **360**, 436–439.
 66. Koonin,E. V (2016) Viruses and mobile elements as drivers of evolutionary transitions. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, **371**, 20150442.
 67. Makarova,K.S., Haft,D.H., Barrangou,R., Brouns,S.J.J., Charpentier,E., Horvath,P., Moineau,S., Mojica,F.J.M., Wolf,Y.I., Yakunin,A.F., *et al.* (2011) Evolution and classification of the CRISPR–Cas systems. *Nat. Rev. Microbiol.*, **9**, 467–477.
 68. Yosef,I., Goren,M.G. and Qimron,U. (2012) Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. *Nucleic Acids Res.*, **40**, 5569–5576.
 69. Jackson,S.A., McKenzie,R.E., Fagerlund,R.D., Kieper,S.N., Fineran,P.C. and Brouns,S.J.J. (2017) CRISPR-Cas: Adapting to change. *Science.*, **356**, eaal5056.
 70. Nuñez,J.K., Kranzusch,P.J., Noeske,J., Wright,A. V, Davies,C.W. and Doudna,J.A. (2014) Cas1–Cas2 complex formation mediates spacer acquisition during CRISPR–Cas adaptive immunity. *Nat. Struct. Mol. Biol.*, **21**, 528–534.
 71. Wang,J., Li,J., Zhao,H., Sheng,G., Wang,M., Yin,M. and Wang,Y. (2015) Structural and Mechanistic Basis of PAM-Dependent Spacer Acquisition in CRISPR-Cas Systems. *Cell*, **163**, 840–853.
 72. Nuñez,J.K., Lee,A.S.Y., Engelman,A. and Doudna,J.A. (2015) Integrase-mediated spacer acquisition during CRISPR–Cas adaptive immunity. *Nature*, **519**, 193–198.
 73. Wright,A. V and Doudna,J.A. (2016) Protecting genome integrity during CRISPR immune adaptation. *Nat. Struct. Mol. Biol.*, **23**, 876–883.
 74. Nuñez,J.K., Harrington,L.B., Kranzusch,P.J., Engelman,A.N. and Doudna,J.A. (2015) Foreign DNA capture during CRISPR–Cas adaptive immunity. *Nature*, **527**, 535–538.
 75. Rollie,C., Schneider,S., Brinkmann,A.S., Bolt,E.L. and White,M.F. (2015) Intrinsic sequence specificity of the Cas1 integrase directs new spacer acquisition. *Elife*, **4:e08716**, 1-19.
 76. Yoganand,K.N.R., Sivathanu,R., Nimkar,S. and Anand,B. (2017) Asymmetric positioning of Cas1-2 complex and Integration Host Factor induced DNA bending guide the unidirectional homing of protospacer in CRISPR-Cas type I-E system. *Nucleic Acids Res.*, **45**, 367–381.

77. Wright,A. V, Liu,J.-J., Knott,G.J., Doxzen,K.W., Nogales,E. and Doudna,J.A. (2017) Structures of the CRISPR genome integration complex. *Science*, **357**, 1113–1118.
78. Dillingham,M.S. and Kowalczykowski,S.C. (2008) RecBCD Enzyme and the Repair of Double-Stranded DNA Breaks. *Microbiol. Mol. Biol. Rev.*, **72**, 642–671.
79. Levy,A., Goren,M.G., Yosef,I., Auster,O., Manor,M., Amitai,G., Edgar,R., Qimron,U. and Sorek,R. (2015) CRISPR adaptation biases explain preference for acquisition of foreign DNA. *Nature*, **520**, 505–510.
80. Modell,J.W., Jiang,W. and Marraffini,L.A. (2017) CRISPR–Cas systems exploit viral DNA injection to establish and maintain adaptive immunity. *Nature*, **544**, 101–104.
81. Redding,S., Sternberg,S.H., Marshall,M., Gibb,B., Bhat,P., Guegler,C.K., Wiedenheft,B., Doudna,J.A. and Greene,E.C. (2015) Surveillance and Processing of Foreign DNA by the Escherichia coli CRISPR-Cas System. *Cell*, **163**, 854–865.
82. Dillard,K.E., Brown,M.W., Johnson,N. V., Xiao,Y., Dolan,A., Hernandez,E., Dahlhauser,S.D., Kim,Y., Myler,L.R., Anslyn,E. V., *et al.* (2018) Assembly and Translocation of a CRISPR-Cas Primed Acquisition Complex. *Cell*, **175**, 934–946.e15.
83. Musharova,O., Klimuk,E., Datsenko,K.A., Metlitskaya,A., Logacheva,M., Semenova,E., Severinov,K. and Savitskaya,E. (2017) Spacer-length DNA intermediates are associated with Cas1 in cells undergoing primed CRISPR adaptation. *Nucleic Acids Res.*, **45**, 3297–3307.
84. Sternberg,S.H., Redding,S., Jinek,M., Greene,E.C. and Doudna,J.A. (2014) DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*, **507**, 62–7.
85. Rutkauskas,M., Sinkunas,T., Songailiene,I., Tikhomirova,M., Siksnys,V. and Seidel,R. (2015) Directional R-loop formation by the CRISPR-cas surveillance complex cascade provides efficient off-target site rejection. *Cell Rep.*, **10**, 1534–1543.
86. Jore,M.M., Lundgren,M., van Duijn,E., Bultema,J.B., Westra,E.R., Waghmare,S.P., Wiedenheft,B., Pul,Ü., Wurm,R., Wagner,R., *et al.* (2011) Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat. Struct. Mol. Biol.*, **18**, 529–536.
87. Szczelkun,M.D., Tikhomirova,M.S., Sinkunas,T., Gasiunas,G., Karvelis,T., Pschera,P., Siksnys,V. and Seidel,R. (2014) Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proc. Natl. Acad. Sci.*, **111**, 9798–9803.

88. Hayes,R.P., Xiao,Y., Ding,F., van Erp,P.B.G., 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.
89. Xiao,Y., Ng,S., Nam,K.H. and Ke,A. (2017) How type II CRISPR–Cas establish immunity through Cas1–Cas2-mediated spacer integration. *Nature*, 10.1038/nature24020.
90. Sinkunas,T., Gasiunas,G., Fremaux,C., Barrangou,R., Horvath,P. and Siksnys,V. (2011) Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. *EMBO J.*, **30**, 1335–1342.
91. Sinkunas,T., Gasiunas,G., Waghmare,S.P., Dickman,M.J., Barrangou,R., Horvath,P. and Siksnys,V. (2013) In vitro reconstitution of Cascade-mediated CRISPR immunity in *Streptococcus thermophilus*. *EMBO J.*, **32**, 385–394.
92. Westra,E.R., van Erp,P.B.G., Künne,T., Wong,S.P., Staals,R.H.J., Seegers,C.L.C., 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.
93. Gong,B., Shin,M., Sun,J., Jung,C.-H., Bolt,E.L., van der Oost,J. and Kim,J.-S. (2014) Molecular insights into DNA interference by CRISPR-associated nuclease-helicase Cas3. *Proc. Natl. Acad. Sci.*, **111**, 16359–16364.
94. Li,M., Wang,R., Zhao,D. and Xiang,H. (2014) Adaptation of the *Haloarcula hispanica* CRISPR-Cas system to a purified virus strictly requires a priming process. *Nucleic Acids Res.*, **42**, 2483–2492.
95. Wiedenheft,B., Lander,G.C., Zhou,K., Jore,M.M., Brouns,S.J.J., van der Oost,J., Doudna,J.A. and Nogales,E. (2011) Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature*, **477**, 486–489.
96. Gasiunas,G., Barrangou,R., Horvath,P. and Siksnys,V. (2012) Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci.*, **109**, E2579–E2586.
97. 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.
98. Shmakov,S., Abudayyeh,O.O., Makarova,K.S., Wolf,Y.I., Gootenberg,J.S., Semenova,E., Minakhin,L., Joung,J., Konermann,S., Severinov,K., *et al.* (2015) Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. *Mol. Cell*, **60**, 385–397.
99. Fonfara,I., Richter,H., Bratovič,M., Le Rhun,A. and Charpentier,E. (2016) The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor

CRISPR RNA. *Nature*, **532**, 517–521.

100. Yang,H., Gao,P., Rajashankar,K.R. and Patel,D.J. (2016) PAM-Dependent Target DNA Recognition and Cleavage by C2c1 CRISPR-Cas Endonuclease. *Cell*, **167**, 1814–1828.e12.
101. Hsu,P.D., Lander,E.S. and Zhang,F. (2014) Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell*, **157**, 1262–1278.
102. Karvelis,T., Gasiunas,G., Miksys,A., Barrangou,R., Horvath,P. and Siksnys,V. (2013) crRNA and tracrRNA guide Cas9-mediated DNA interference in *Streptococcus thermophilus*. *RNA Biol.*, **10**, 841–51.
103. Sander,J.D. and Joung,J.K. (2014) CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.*, **32**, 347–355.
104. Benda,C., Ebert,J., Scheltema,R.A., Schiller,H.B., Baumgärtner,M., Bonneau,F., Mann,M. and Conti,E. (2014) Structural Model of a CRISPR RNA-Silencing Complex Reveals the RNA-Target Cleavage Activity in Cmr4. *Mol. Cell*, **56**, 43–54.
105. Rouillon,C., Zhou,M., Zhang,J., Politis,A., Beilsten-Edmands,V., Cannone,G., Graham,S., Robinson,C.V., Spagnolo,L. and White,M.F. (2013) Structure of the CRISPR Interference Complex CSM Reveals Key Similarities with Cascade. *Mol. Cell*, **52**, 124–134.
106. Hale,C.R., Zhao,P., Olson,S., Duff,M.O., Graveley,B.R., Wells,L., Terns,R.M. and Terns,M.P. (2009) RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell*, **139**, 945–56.
107. Tamulaitis,G., Kazlauskienė,M., Manakova,E., Venclovas,Č., Nwokeoji,A.O., Dickman,M.J., Horvath,P. and Siksnys,V. (2014) Programmable RNA Shredding by the Type III-A CRISPR-Cas System of *Streptococcus thermophilus*. *Mol. Cell*, **56**, 506–517.
108. Zhang,J., Rouillon,C., Kerou,M., Reeks,J., Brugger,K., Graham,S., Reimann,J., Cannone,G., Liu,H., Albers,S.-V., *et al.* (2012) Structure and Mechanism of the CMR Complex for CRISPR-Mediated Antiviral Immunity. *Mol. Cell*, **45**, 303–313.
109. Elmore,J.R., Sheppard,N.F., Ramia,N., Deighan,T., Li,H., Terns,R.M. and Terns,M.P. (2016) Bipartite recognition of target RNAs activates DNA cleavage by the Type III-B CRISPR-Cas system. *Genes Dev.*, **30**, 447–59.
110. Estrella,M.A., Kuo,F.-T. and Bailey,S. (2016) RNA-activated DNA cleavage by the Type III-B CRISPR-Cas effector complex. *Genes Dev.*, **30**, 460–70.
111. Kazlauskienė,M., Tamulaitis,G., Kostiuk,G., Venclovas,Č. and Siksnys,V. (2016) Spatiotemporal Control of Type III-A CRISPR-Cas Immunity: Coupling DNA Degradation with the Target RNA Recognition. *Mol. Cell*, **62**,

295–306.

112. Samai,P., Pyenson,N., Jiang,W., Goldberg,G.W., Hatoum-Aslan,A. and Marraffini,L.A. (2015) Co-transcriptional DNA and RNA Cleavage during Type III CRISPR-Cas Immunity. *Cell*, **161**, 1164–1174.
113. Datsenko,K.A., Pougach,K., Tikhonov,A., Wanner,B.L., Severinov,K. and Semenova,E. (2012) Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nat. Commun.*, **3**, 945.
114. Xue,C., Whitis,N.R. and Sashital,D.G. (2016) Conformational Control of Cascade Interference and Priming Activities in CRISPR Immunity. *Mol. Cell*, **64**, 826–834.
115. Richter,C., Dy,R.L., McKenzie,R.E., Watson,B.N.J., Taylor,C., Chang,J.T., McNeil,M.B., Staals,R.H.J. and Fineran,P.C. (2014) Priming in the Type I-F CRISPR-Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer. *Nucleic Acids Res.*, **42**, 8516–8526.
116. Li,M., Wang,R. and Xiang,H. (2014) *Haloarcula hispanica* CRISPR authenticates PAM of a target sequence to prime discriminative adaptation. *Nucleic Acids Res.*, **42**, 7226–7235.
117. Li,M., Gong,L., Zhao,D., Zhou,J. and Xiang,H. (2017) The spacer size of I-B CRISPR is modulated by the terminal sequence of the protospacer. *Nucleic Acids Res.*, **45**, 4642–4654.
118. Rao,C., Chin,D. and Ensminger,A.W. (2017) Priming in a permissive type I-C CRISPR-Cas system reveals distinct dynamics of spacer acquisition and loss. *RNA*, **23**, 1525–1538.
119. Savitskaya,E., Semenova,E., Dedkov,V., Metlitskaya,A. and Severinov,K. (2013) High-throughput analysis of type I-E CRISPR/Cas spacer acquisition in *E. coli*. *RNA Biol.*, **10**, 716–725.
120. Swarts,D.C., Mosterd,C., van Passel,M.W.J. and Brouns,S.J.J. (2012) CRISPR interference directs strand specific spacer acquisition. *PLoS One*, **7**.
121. Xue,C., Seetharam,A.S., Musharova,O., Severinov,K., Brouns,S.J.J., Severin,A.J. and Sashital,D.G. (2015) CRISPR interference and priming varies with individual spacer sequences. *Nucleic Acids Res.*, **43**, 10831–10847.
122. Staals,R.H.J., Jackson,S.A., Biswas,A., Brouns,S.J.J., Brown,C.M. and Fineran,P.C. (2016) Interference-driven spacer acquisition is dominant over naive and primed adaptation in a native CRISPR–Cas system. *Nat. Commun.*, **7**, 12853.
123. Shmakov,S., Savitskaya,E., Semenova,E., Logacheva,M.D., Datsenko,K.A. and Severinov,K. (2014) Pervasive generation of oppositely oriented spacers during CRISPR adaptation. *Nucleic Acids Res.*, **42**, 5907–5916.

124. Semenova,E., Savitskaya,E., Musharova,O., Strotskaya,A., Vorontsova,D., Datsenko,K.A., Logacheva,M.D. and Severinov,K. (2016) Highly efficient primed spacer acquisition from targets destroyed by the *Escherichia coli* type I-E CRISPR-Cas interfering complex. *Proc. Natl. Acad. Sci.*, **113**, 7626–7631.
125. Künne,T., Kieper,S.N., Bannenberg,J.W., Vogel,A.I.M., Miellet,W.R., Klein,M., Depken,M., Suarez-Diez,M. and Brouns,S.J.J. (2016) Cas3-Derived Target DNA Degradation Fragments Fuel Primed CRISPR Adaptation. *Mol. Cell*, **63**, 852–864.
126. Severinov,K., Ispolatov,I. and Semenova,E. (2016) The Influence of Copy-Number of Targeted Extrachromosomal Genetic Elements on the Outcome of CRISPR-Cas Defense. *Front. Mol. Biosci.*, **3**.
127. Blosser,T.R., Loeff,L., Westra,E.R., Vlot,M., Künne,T., Sobota,M., Dekker,C., Brouns,S.J.J. and Joo,C. (2015) Two distinct DNA binding modes guide dual roles of a CRISPR-Cas protein complex. *Mol. Cell*, **58**, 60–70.
128. Gu,G.Y., Zhu,L.M., Su,C.Y., Ding,H. and Fatikow,S. (2016) Modeling and control of piezo-actuated nanopositioning stages: A survey. *IEEE Trans. Autom. Sci. Eng.*, **13**, 313–332.
129. Kemmerich,F.E., Kasaciunaite,K. and Seidel,R. (2016) Modular magnetic tweezers for single-molecule characterizations of helicases. *Methods*, **108**, 4–13.
130. Dikić,J., Menges,C., Clarke,S., Kokkinidis,M., Pingoud,A., Wende,W. and Desbiolles,P. (2012) The rotation-coupled sliding of EcoRV. *Nucleic Acids Res.*, **40**, 4064–70.
131. Kostiuk,G., Dikić,J., Schwarz,F.W., Sasnauskas,G., Seidel,R. and Siksnys,V. (2017) The dynamics of the monomeric restriction endonuclease BcnI during its interaction with DNA. *Nucleic Acids Res.*, **45**, 5968–5979.
132. Kemmerich,F.E., Swoboda,M., Kauert,D.J., Grieb,M.S., Hahn,S., Schwarz,F.W., Seidel,R. and Schlierf,M. (2016) Simultaneous Single-Molecule Force and Fluorescence Sampling of DNA Nanostructure Conformations Using Magnetic Tweezers. *Nano Lett.*, **16**, 381–386.
133. Heller,I., Hoekstra,T.P., King,G.A., Peterman,E.J.G. and Wuite,G.J.L. (2014) Optical Tweezers Analysis of DNA–Protein Complexes. *Chem. Rev.*, **114**, 3087–3119.
134. Sitters,G., Kamsma,D., Thalhammer,G., Ritsch-Martens,M., Peterman,E.J.G. and Wuite,G.J.L. (2015) Acoustic force spectroscopy. *Nat. Methods*, **12**, 47–50.
135. Kamsma,D. and Wuite,G.J.L. (2018) Single-Molecule Measurements Using Acoustic Force Spectroscopy (AFS). In *Methods in molecular biology (Clifton, N.J.)*. Vol. 1665, pp. 341–351.

136. Jung,C., Hawkins,J.A., Jones,S.K., Xiao,Y., Rybarski,J.R., Dillard,K.E., Hussmann,J., Saifuddin,F.A., Savran,C.A., Ellington,A.D., *et al.* (2017) Massively Parallel Biophysical Analysis of CRISPR-Cas Complexes on Next Generation Sequencing Chips. *Cell*, **170**, 35–47.e13.
137. Gilles,A.F. and Averof,M. (2014) Functional genetics for all: engineered nucleases, CRISPR and the gene editing revolution. *Evodevo*, **5**, 43.
138. Wang,X., Wang,Y., Wu,X., Wang,J., Wang,Y., Qiu,Z., Chang,T., Huang,H., Lin,R.-J. and Yee,J.-K. (2015) Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors. *Nat. Biotechnol.*, **33**, 175–178.
139. Wu,X., Scott,D.A., Kriz,A.J., Chiu,A.C., Hsu,P.D., Dadon,D.B., Cheng,A.W., Trevino,A.E., Konermann,S., Chen,S., *et al.* (2014) Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat. Biotechnol.*, **32**, 670–676.
140. Kosicki,M., Tomberg,K. and Bradley,A. (2018) Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.*, **36**, 765–771.
141. Wang,H., La Russa,M. and Qi,L.S. (2016) CRISPR/Cas9 in Genome Editing and Beyond. *Annu. Rev. Biochem.*, **85**, 227–264.
142. Waltz,E. (2016) Gene-edited CRISPR mushroom escapes US regulation. *Nature*, **532**, 293.
143. Zhang,X.-H., Tee,L.Y., Wang,X.-G., Huang,Q.-S. and Yang,S.-H. (2015) Off-target Effects in CRISPR/Cas9-mediated Genome Engineering. *Mol. Ther. - Nucleic Acids*, **4**, e264.
144. Zhan,T., Rindtorff,N., Betge,J., Ebert,M.P. and Boutros,M. (2018) CRISPR/Cas9 for cancer research and therapy. *Semin. Cancer Biol.*, 10.1016/j.semcancer.2018.04.001.
145. Iwai,Y., Ishida,M., Tanaka,Y., Okazaki,T., Honjo,T. and Minato,N. (2002) Involvement of PD-L1 on tumor cells in the escape from host immune system and tumor immunotherapy by PD-L1 blockade. *Proc. Natl. Acad. Sci.*, **99**, 12293–12297.
146. Reck,M., Rodríguez-Abreu,D., Robinson,A.G., Hui,R., Csőszi,T., Fülöp,A., Gottfried,M., Peled,N., Tafreshi,A., Cuffe,S., *et al.* (2016) Pembrolizumab versus Chemotherapy for PD-L1–Positive Non–Small-Cell Lung Cancer. *N. Engl. J. Med.*, **375**, 1823–1833.
147. Schwarz,F.W., Toth,J., van Aelst,K., Cui,G., Clausing,S., Szczelkun,M.D. and Seidel,R. (2013) The Helicase-Like Domains of Type III Restriction Enzymes Trigger Long-Range Diffusion Along DNA. *Science.*, **340**, 353–356.

148. Maffeo,C., Schöpflin,R., Brutzer,H., Stehr,R., Aksimentiev,A., Wedemann,G. and Seidel,R. (2010) DNA-DNA interactions in tight supercoils are described by a small effective charge density. *Phys. Rev. Lett.*, **105**.
149. Klaue,D. and Seidel,R. (2009) Torsional stiffness of single superparamagnetic microspheres in an external magnetic field. *Phys. Rev. Lett.*, **102**.
150. F.W,S. (2012) The role of 1D diffusion for directional long-range communication of DNA.
151. D,K. (2012) DNA Unwinding by Helicases investigated on the Single Molecule Level.
152. Huhle,A., Klaue,D., Brutzer,H., Daldrop,P., Joo,S., Otto,O., Keyser,U.F. and Seidel,R. (2015) Camera-based three-dimensional real-time particle tracking at kHz rates and Ångström accuracy. *Nat. Commun.*, **6**, 5885.
153. Daldrop,P., Brutzer,H., Huhle,A., Kauert,D.J. and Seidel,R. (2015) Extending the range for force calibration in magnetic tweezers. *Biophys. J.*, **108**, 2550–2561.
154. Schöpflin,R., Brutzer,H., Müller,O., Seidel,R. and Wedemann,G. (2012) Probing the elasticity of DNA on short length scales by modeling supercoiling under tension. *Biophys. J.*, **103**, 323–30.
155. Qiblawey,H.A. and Abu-Jdayil,B. (2010) Viscosity and Density of the Ternary Solution of Magnesium Chloride + Sodium Chloride + Water from (298.15 to 318.15) K. *J. Chem. Eng. Data*, **55**, 3322–3326.
156. Beloglazova,N., Kuznedelov,K., Flick,R., Datsenko,K.A., Brown,G., Popovic,A., Lemak,S., Semenova,E., Severinov,K. and Yakunin,A.F. (2015) CRISPR RNA binding and DNA target recognition by purified Cascade complexes from *Escherichia coli*. *Nucleic Acids Res.*, **43**, 530–543.
157. Brouns,S.J.J., Jore,M.M., Lundgren,M., Westra,E.R., Slijkhuis,R.J.H., Snijders,A.P.L., Dickman,M.J., Makarova,K.S., Koonin,E. V and van der Oost,J. (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, **321**, 960–4.
158. Wilkins,M.R., Gasteiger,E., Bairoch,A., Sanchez,J.C., Williams,K.L., Appel,R.D. and Hochstrasser,D.F. (1999) Protein identification and analysis tools in the ExPASy server. *Methods Mol. Biol.*, **112**, 531–52.
159. Kuznedelov,K., Mekler,V., Lemak,S., Tokmina-Lukaszewska,M., Datsenko,K.A., Jain,I., Savitskaya,E., Mallon,J., Shmakov,S., Bothner,B., *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**, 10849–10861.
160. Zuker,M. (2003) Mfold web server for nucleic acid folding and hybridization

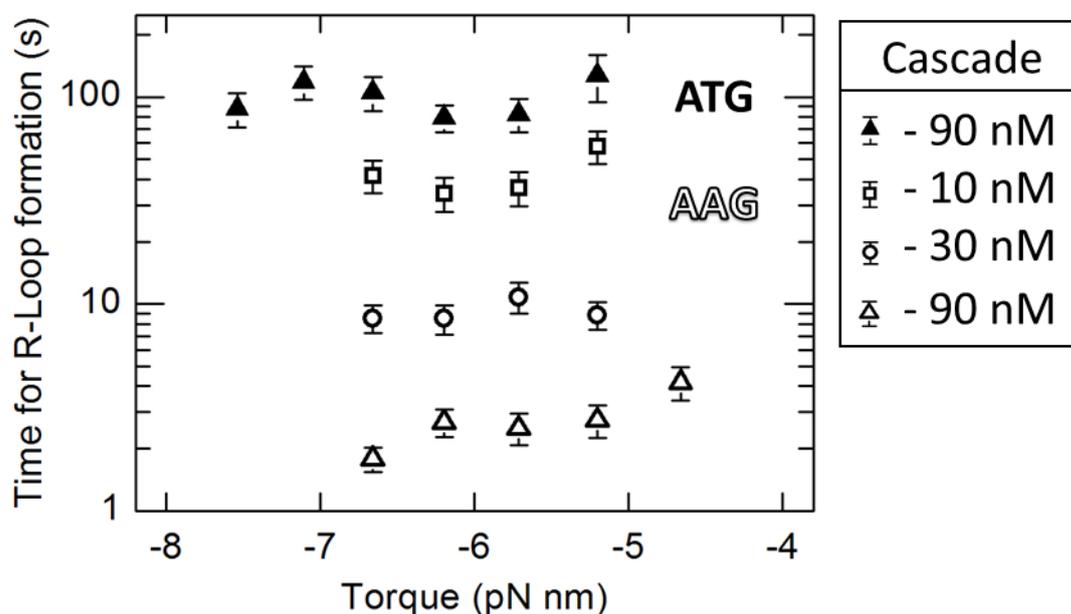
- prediction. *Nucleic Acids Res.*, **31**, 3406–15.
161. Van Erp, P.B.G., Jackson, R.N., Carter, J., Golden, S.M., Bailey, S. and Wiedenheft, B. (2015) Mechanism of CRISPR-RNA guided recognition of DNA targets in *Escherichia coli*. *Nucleic Acids Res.*, **43**, 8381–8391.
 162. Xiao, Y., Luo, M., Hayes, R.P., Kim, J., Ng, S., Ding, F., Liao, M. and Ke, A. (2017) Structure Basis for Directional R-loop Formation and Substrate Handover Mechanisms in Type I CRISPR-Cas System. *Cell*, **170**, 48–60.e11.
 163. Vologodskaya, M.Y. and Vologodskii, A. V (1999) Effect of magnesium on cruciform extrusion in supercoiled DNA. *J. Mol. Biol.*, **289**, 851–9.
 164. Machinek, R.R.F., Ouldrige, T.E., Haley, N.E.C., Bath, J. and Turberfield, A.J. (2014) Programmable energy landscapes for kinetic control of DNA strand displacement. *Nat. Commun.*, **5**, 5324.
 165. Schaefer, K.A., Wu, W.-H., Colgan, D.F., Tsang, S.H., Bassuk, A.G. and Mahajan, V.B. (2017) Unexpected mutations after CRISPR–Cas9 editing in vivo. *Nat. Methods*, **14**, 547–548.
 166. Jung, C., Hawkins, J.A., Jones, S.K., Xiao, Y., Rybarski, J.R., Dillard, K.E., Hussmann, J., Saifuddin, F.A., Savran, C.A., Ellington, A.D., *et al.* (2017) Massively Parallel Biophysical Analysis of CRISPR-Cas Complexes on Next Generation Sequencing Chips. *Cell*, **170**, 35–47.e13.
 167. Mekler, V., Minakhin, L. and Severinov, K. (2017) Mechanism of duplex DNA destabilization by RNA-guided Cas9 nuclease during target interrogation. *Proc. Natl. Acad. Sci.*, **114**, 5443–5448.
 168. Okonechnikov, K., Golosova, O., Fursov, M. and UGENE team (2012) Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics*, **28**, 1166–1167.
 169. Westra, E.R., Semenova, E., Datsenko, K.A., Jackson, R.N., Wiedenheft, B., Severinov, K. and Brouns, S.J.J. (2013) Type I-E CRISPR-Cas Systems Discriminate Target from Non-Target DNA through Base Pairing-Independent PAM Recognition. *PLoS Genet.*, **9**.
 170. Mulepati, S., Heroux, A. and Bailey, S. (2014) Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target. *Science.*, **345**, 1479–1484.
 171. Richter, H., Rompf, J., Wiegel, J., Rau, K. and Randau, L. (2017) Fragmentation of the CRISPR-Cas Type I-B signature protein Cas8b. *Biochim. Biophys. Acta - Gen. Subj.*, **1861**, 2993–3000.
 172. Amitai, G. and Sorek, R. (2016) CRISPR–Cas adaptation: insights into the mechanism of action. *Nat. Rev. Microbiol.*, **14**, 67–76.

Appendix A. Results obtained by collaborators in this project.

A.1. Absence of Cascade binding torque dependence.

This part was done by Dr. Christophe Rouillon (University of Leipzig, University of St Andrews).

Measured at different torques binding rate for two most active PAM sequences ATG and AAG, displayed, in contrast to St-Cascade, that *E. coli* Cascade displays no torque dependence. Thus, time of R-loop formation is the same within error limits, no matter what negative torque is applied to DNA molecule. This may indicate on large energetic barrier of Cascade to start R-loop propagation via PAM binding and DNA opening.



A.2. Permanganate footprinting results are in agreement with other assays on Cascade binding to mutated targets.

This part was done by Dr. Konstantin Kuznedelov (Rutgers, The State University of New Jersey, USA).

The results of MT and bulk assays of target binding are in full agreement with permanganate footprinting protocol provided below (Fig A2). The wild-type target, as well as the seed mutants in positions 1-4 and the PAM variant with single

substitution at position -1 all showed Cascade-dependent localized melting of DNA and the extent of this melting was identical for all targets (the differences in patterns of permanganate sensitivity between the wild-type and some of the mutant complexes are due to introduction/removal of thymine residues by substitutions). In reactions containing a DNA substrate with a fully matching g8 protospacer and a CCG PAM, the permanganate sensitivity was weak and the pattern that was detected was markedly different compared to other complexes indicating the absence of an R-loop.

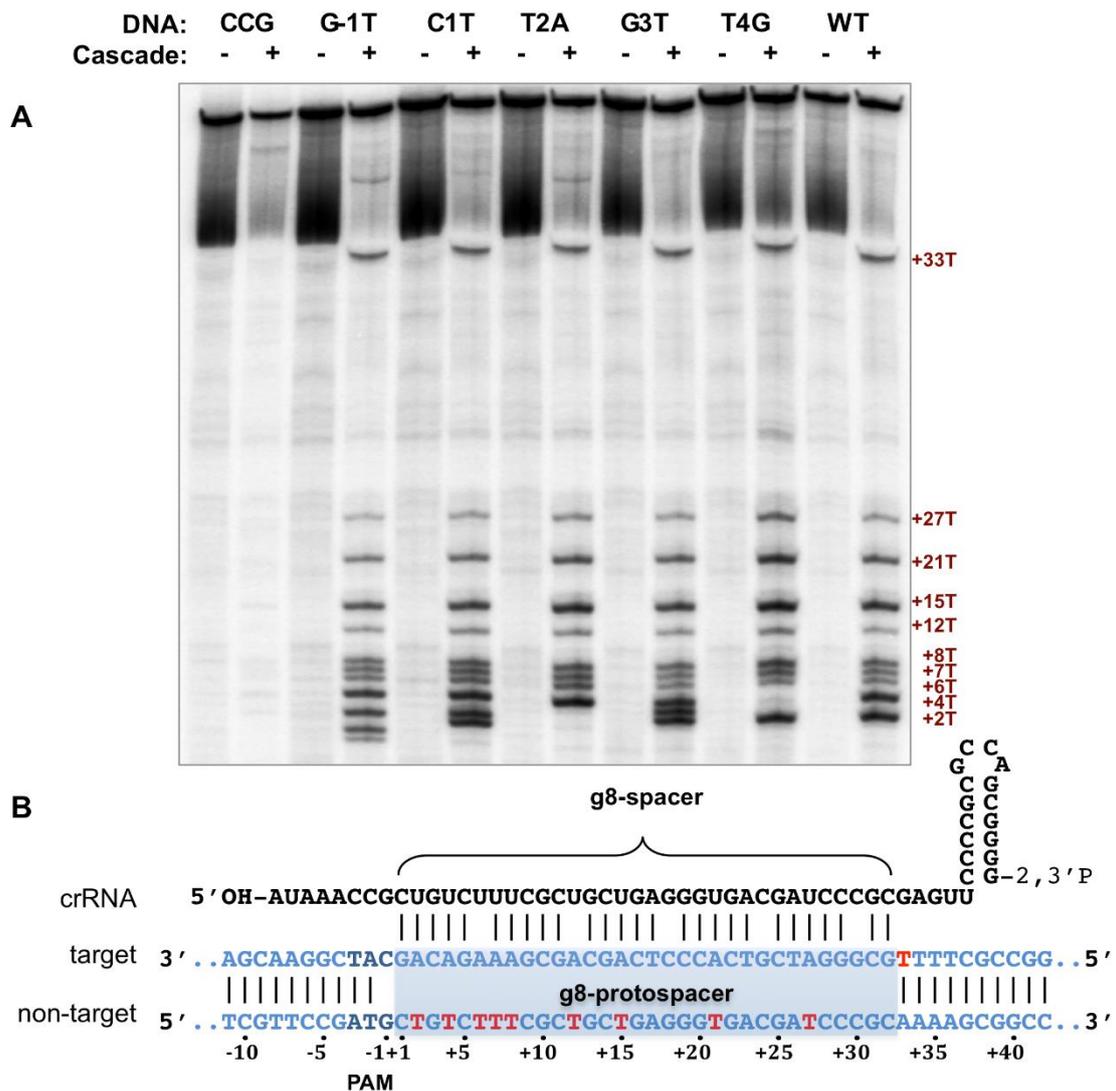


Figure A2. KMnO₄ probing of mutant g8 protospacers bound by Cascade-g8-crRNA. (A) Probing of complexes formed on WT, PAM (CCG, G-1T), and indicated protospacer mutant templates. Positions of permanganate-sensitive thymines are indicated by dark red numbers on the right side of the gel. (B) A schematic model of R-loop formed upon the recognition of g8 protospacer target. Thymines sensitive to permanganate oxidation are shown in red (see also panel A).

Protocol of permanganate probing.

The target g8 DNA fragment (213 bp) and its mutant variants were amplified by PCR of M13mp18 phage DNA (wild-type and engineered escape mutants (34)) using g8-dir 5'-AGTCTTTAGTCCTCAAAGCCTCTG-3' and g8-rev 5'-AGCTTGCTTTCGAGGTGAATTTC-3' primers. For radioactive labeling, 3-5 pmoles of the target DNA fragments were combined with 8 pmoles of [γ - 32 P]-ATP (3000 Ci/mmol) and 10 units of T4 polynucleotide kinase (NEB) in 20 μ l of the reaction buffer containing 70 mM Tris-HCl (pH 7.6), 10 mM MgCl₂, 5 mM DTT, and incubated for 30 min at 37 °C. 32 P-labeled DNA fragments were purified by micro Bio-Spin™ chromatography on columns packed with Bio-Gel P-30 (Bio-Rad) and used for permanganate probing reactions performed as described before (41). Target binding was performed in 10 μ l of binding buffer (40 mM Tris-HCl, pH 8.0, 50 mM NaCl, 10 mM MgCl₂, 0.5 mM TCEP, 50 μ g/ml BSA) using 15 nM labeled DNA fragment and 2 μ M Cascade. After 30 min incubation at 37 °C, the probing reaction was initiated by adding KMnO₄ to a final concentration of 2.5 mM. The reaction was quenched after 15 s by the addition of 10 μ l 1% 2-mercaptoethanol. The reaction products were extracted using a phenol–chloroform mixture, followed by an ethanol precipitation. The DNA pellets were dissolved in 100 μ l of freshly prepared 1 M piperidine solution and placed in a 90 °C water bath for 10 minutes. After chloroform extraction DNA was ethanol precipitated. The pellets were dissolved in 8 μ l of formamide loading buffer. The reaction products were separated using an 8% denaturing PAGE gel and visualized with a Typhoon 9400 phosphorimager.

Appendix B. Genetic plasmids and constructs used.

B.1. Protein expression (for further purification) vectors

CCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGA
GTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGG
TAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTT
GAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTAT
CTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCT
TGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTTGTTTGCA
AGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGA
TCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACCTCACGTAAAGG
GATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTA
AATTA AAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAAAA
TATTCCGGAATTGCCAGCTGGGGCGCCCTCTGGTAAGGTTGGGAAGCC
CTGCAAAGTAAACTGGATGGCTTTCTTGCCGCCAAGGATCTGATGGCGC
AGGGGATCAAGATCTGATCAAGAGACAGGATGAGGATCGTTTCGCATG
ATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAG
AGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGAT
GCCGCCGTGTTCCGGCTGTCAGCGCAGGGGGCGCCCGGTTCTTTTTGTCA
AGACCGACCTGTCCGGTGCCCTGAATGAACTGCAGGACGAGGCAGCGC
GGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGA
CGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCC
GGGGCAGGATCTCCTGTCATCCCACCTTGCTCCTGCCGAGAAAGTATCC
ATCATGGCTGATGCAATGCGGGCGGCTGCATACGCTTGATCCGGCTACCT
GCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTC
GGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATC
AGGGGCTCGCGCCAGCCGAACTGTTCCGCCAGGCTCAAGGCGCGCATGC
CCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGA
ATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCG
GCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGA
TATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTT
TACGGTATCGCCGCTCCCGATTTCGCAGCGCATCGCCTTCTATCGCCTTC
TTGACGAGTTCTTCTGAACCGGTAATATTATTGAAGCATTATCAGGGT
TATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAAC
AAATAGGGGTTCGCGCACATTTCCCGAAAAGTGCCACCTGACGTCT
AAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCA
CGAGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTG
ACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGC
CGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTG
TCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGT
GCACCAATTGGTCGACCTCGAGTTAATTAACGTA

Synthetic gene 1

GTCGCACACGGTGTGTTGAACACCCGATATTATTATCGTGCTGCAACACAA
TAGCCGTTGCACTACTAATTAATAGATCACGTTAATAGCGTATAGATTA
GTTGTACTATTATACCGACGTCGTAGTATTGCTATATACGGCGTTGGGT
ATCCGTGATCTAACACGGATTCTGCTGCGACGAACGCCGTACGTCTAGA
TCGATTATTCCCGGGCGGCTAATATCACTATAATTGATTCACAATTCGG
CTAGCCCGATACGATCGTTGGAACCTATTGTGAACAACCGTGTGTGATTC
TGCTAATTCAACGATTCAACGCTAGCGGGTGGGCGGGCTATATTATCGT
AGCTATTGAATAGCGTGTGTTATTCCACGTGTTCCGGCCGGATTCGATA
TAATCTGATCGATCAATACCGACGTATATTGTAGTTCAGCACCGCGTAC
TACGTGGATTGCGGATAATACTATACGCGATTTCGCACGAACGTCCCCGC
TGCACACCGATCGCCGTAACTAACACAACACCCGCGCGTTAATATTAG
ACGATTGTGGTAATAATATCCAACCGCGGCGCAGATTAGTGTGTTACTA
ACGCCCAACCGTGGGTACCGACTATTATCGGTTGTTGTAATAATCCGGA
TAATTCGTGGATTGTATATTCTACGACGACTATTCGACGATTTCGTTAGT
ACCAACTATTATTCTGCGTAGCTGCGCGCAATCAATTGATACAACCTATT
GTTTCGCGGATATCGGGAACTAATCCGCCGACGTAATTCTGCGATTCACG
CAGCGCTAATTCTACGACTATTGTGATCCACGCACCCGACGTACGTTTCG
CGAACTATATAGATTACGCACGAATATTGTGGCTAATAACCCGATCCGG
CCGATTAGCGGGGCCGACGAATTCAGAATAGTGCGGGGTGGTTCGGTGA
ACGCCGGGGTAGTGATCTGCGCGGGCGATACCGTGGAATTGTGGTTCC
GGCGATCCGGCAACGATCCGGG

Synthetic gene 2

CGCTAATCTATATATTAATTAGACCCACGATACTATAGTATTAATAGAC
GTCTATAATTGTGTAGCTACAACGCCGAATTCGTGCGTGACGCGTGCG
ATCGCACGAATTGAATCACACGTGAATTCGGTTGATTCGGGCCGTGGCC
CCACGTGTAACGGTGTGTTGCCAACAACCGTGCAATATCCACGCGGACGC
GTACTATTGCCGGAACGGTTCCGCGCGTTCCGTGTGCAATTGTATAGAT
CCCACACACTACAATTCAATAACGACGAATCAGCGCACACACTACGAT
TGTTGTAGTATATCAGAACGGCTATAGATTCTAGATTGTGCTACGTGCA
CAATAGTGAATTCAATCACGCGGGATTCAACCGTTGGTTGTGGTTCCGGC
CGCCGTTGACTAATTCGTTCGGTACACAACCCGATCAGCGGGTCCGGGG
ACGTGATATATATTACCGTGCCGCGATTGCCGTGGAACGCCGAATCTGC
GCCGCACGTGGTCCGTAGTTCGTTTCGATTTCGAATATTCGTTGGTTGGCA
ATTCATAATATACCGTGTGGCCGGGCGGTGCAATAACAACGGTTGGT
ACGTGGTCGTATCCACGCGCAATCGCACGACGTATCCCGCGTCACCGA
ACAATAGTTGTGTAGTTACCCAACGCTGCTAGAATAGCGATAATATCG
ATACGGCGTACGTGGTTAATAGATACACCGTAGTGTAGCTACGATAAT
CGCCCGGCGTTCAGAATCTATTCGACTACCGACTATTATTATCGTTGTG
GGACCCAACCTATTAGATCGGGTAATTATCGGATTTCGTTGTACTACGTTA
GTTGTAATTAGTGATCAGCCGTAATTCTAGCGTAATTAGCGGGGCCACC

GCGGCGTTAACGGTTCTATACGGATATCTATTGTGGGGTTGTTCTGAAT
 AGCCGTAATATCCACGTAACCGAACGAATATTGTGTGTTAGTAATTCAC
 ACGCGGCAATCTAATCGTCTGCACTA

Appendix C. Self-made code and algorithms

C.1 Synthetic genes design

Code for Monte-Carlo simulations:

% For now the program is making random DNA without any AAG; ATG; CCT;
 TTT; CTC and 5xN repeats

```

Nuc = ['A' 'T' 'C' 'G'];
N=10000; %length of the sequence
seq_1 = []; % blanks(N); % [];
seq_1=strcat(seq_1, Nuc(random('unid',4))); %first letter
seq_1=strcat(seq_1, datasample(['A' 'T' 'C' 'G'], 1)); %second letter

for i = 3:4 % without repeats cheks
    sub=Nuc;
    if (seq_1(i-2)=='A')&&(seq_1(i-1)=='A')           %No AAG I
        sub=strrep(sub, 'G', "");
    end
    if (seq_1(i-2)=='C')&&(seq_1(i-1)=='T')           %No CTT I
        sub=strrep(sub, 'T', "");
    end
    if (seq_1(i-2)=='A')&&(seq_1(i-1)=='T')           %No ATG II
        sub=strrep(sub, 'G', "");
    end
    if (seq_1(i-2)=='C')&&(seq_1(i-1)=='A')           %No CAT II
        sub=strrep(sub, 'T', "");
    end
    if (seq_1(i-2)=='A')&&(seq_1(i-1)=='G')           %No AGG III
        sub=strrep(sub, 'G', "");
    end
    if (seq_1(i-2)=='C')&&(seq_1(i-1)=='C')           %No CCT III
        sub=strrep(sub, 'T', "");
    end
    if (seq_1(i-2)=='A')&&(seq_1(i-1)=='A')           %No AAA IV
        sub=strrep(sub, 'A', "");

```

```

end
if (seq_1(i-2)=='T')&&(seq_1(i-1)=='T')           %No TTT IV
    sub=strrep(sub, 'T', '');
end
if (seq_1(i-2)=='C')&&(seq_1(i-1)=='T')           %No CTC V
    sub=strrep(sub, 'C', '');
end
if (seq_1(i-2)=='G')&&(seq_1(i-1)=='A')           %No GAG V
    sub=strrep(sub, 'G', '');
end
seq_1=strcat(seq_1, datasample(sub, 1));           %Adding new letter which satisfy
all conditions
end
for i = 5:N % with 5xN repeats cheks
    sub=Nuc;
    if (seq_1(i-2)=='A')&&(seq_1(i-1)=='A')         %No AAG I
        sub=strrep(sub, 'G', '');
    end
    if (seq_1(i-2)=='C')&&(seq_1(i-1)=='T')         %No CTT I
        sub=strrep(sub, 'T', '');
    end
    if (seq_1(i-2)=='A')&&(seq_1(i-1)=='T')         %No ATG II
        sub=strrep(sub, 'G', '');
    end
    if (seq_1(i-2)=='C')&&(seq_1(i-1)=='A')         %No CAT II
        sub=strrep(sub, 'T', '');
    end
    if (seq_1(i-2)=='A')&&(seq_1(i-1)=='G')         %No AGG III
        sub=strrep(sub, 'G', '');
    end
    if (seq_1(i-2)=='C')&&(seq_1(i-1)=='C')         %No CCT III
        sub=strrep(sub, 'T', '');
    end
    if (seq_1(i-2)=='A')&&(seq_1(i-1)=='A')         %No AAA IV
        sub=strrep(sub, 'A', '');
    end
    if (seq_1(i-2)=='T')&&(seq_1(i-1)=='T')         %No TTT IV
        sub=strrep(sub, 'T', '');
    end
    if (seq_1(i-2)=='C')&&(seq_1(i-1)=='T')         %No CTC V
        sub=strrep(sub, 'C', '');

```

```

end
if (seq_1(i-2)=='G')&&(seq_1(i-1)=='A')           %No GAG V
    sub=strrep(sub, 'G', "");
end
if seq_1(i-4)==seq_1(i-3)&&seq_1(i-3)==seq_1(i-2)&&seq_1(i-2)==seq_1(i-1)
%if 4xN repeat
    sub=strrep(sub, seq_1(i-1), "");
end
seq_1=strcat(seq_1, datasample(sub, 1));           %Adding new letter which satisfy
all conditions
end
fileID=fopen('Crispr Leipzig\seq_1.txt','w');
fprintf(fileID, seq_1);

```

%This program aims to get rid of g_8 and another protospacer seed matching

```

file_seq_1=fopen('Crispr Leipzig\seq_1.txt', 'r');
seq_1=fscanf(file_seq_1,'%s');
seed_g8=['C' 'T' 'G' 'T' 'C']; %g8 seed region NTS;
seed_g8_r=['G' 'A' 'C' 'A' 'G']; %g8 seed region TS;
seed_Skott=['C' 'C' 'A' 'G' 'T']; %Skotts protospacer seed region NTS;
seed_Skott_r=['A' 'C' 'T' 'G' 'G']; %Skotts protospacer seed region TS;
Nuc = ['A' 'T' 'C' 'G'];
i=1;
N=0;
while i<(length(seq_1)-length(seed_g8)+1)
    if ((seq_1(i)==seed_g8(1))...           %matching >3 for seed g8 direct
        +(seq_1(i+1)==seed_g8(2))...
        +(seq_1(i+2)==seed_g8(3))...
        +(seq_1(i+3)==seed_g8(4))...
        +(seq_1(i+4)==seed_g8(5)))>3
        seq_1=clean_seed_func(seq_1,i);
        if i>4                               %going back for 4 nucleotides
            i
            i=i-4;
            N=N+1;
        else
            i
            i=1;
            N=N+1;
        end
    end
end

```

```

elseif ((seq_1(i)==seed_g8_r(1))... %matching for seed g8 reverse complement
+(seq_1(i+1)==seed_g8_r(2))...
+(seq_1(i+2)==seed_g8_r(3))...
+(seq_1(i+3)==seed_g8_r(4))...
+(seq_1(i+4)==seed_g8_r(5)))>3
seq_1=clean_seed_func(seq_1,i);
if i>4          %going back for 4 nucleotides
    i
    i=i-4;
    N=N+1;
else
    i
    i=1;
    N=N+1;
end
elseif ((seq_1(i)==seed_Skott(1))... %matching for Skotts seed direct
+(seq_1(i+1)==seed_Skott(2))...
+(seq_1(i+2)==seed_Skott(3))...
+(seq_1(i+3)==seed_Skott(4))...
+(seq_1(i+4)==seed_Skott(5)))>3
seq_1=clean_seed_func(seq_1,i);
if i>4          %going back for 4 nucleotides
    i
    i=i-4;
    N=N+1;
else
    i
    i=1;
    N=N+1;
end
elseif ((seq_1(i)==seed_Skott_r(1))... %matching for Skotts seed reverse
complement
+(seq_1(i+1)==seed_Skott_r(2))...
+(seq_1(i+2)==seed_Skott_r(3))...
+(seq_1(i+3)==seed_Skott_r(4))...
+(seq_1(i+4)==seed_Skott_r(5)))>3
seq_1=clean_seed_func(seq_1,i);
if i>4          %going back for 4 nucleotides
    i
    i=i-4;
    N=N+1;

```

```

else
    i
    i=i+1;
    N=N+1;
end
else
    i=i+1;
end
end

```

```

fileID=fopen('Crispr Leipzig\seq_2.txt','w');
fprintf(fileID, seq_1);

```

```

% This function checks for not creating "forbidden" PAMs and mutate the
% selected region

```

```

function seq_out = clean_seed(seq_in,i)

```

```

Nuc = ['A' 'T' 'C' 'G'];

```

```

seq_out=seq_in;

```

```

if i>2

```

```

    for j = 0:4

```

```

        sub=strrep(Nuc, seq_out(i+j), "");

```

```

        if (seq_out(i+j-2)=='A')&&(seq_out(i+j-1)=='A')    %No AAG before I

```

```

            sub=strrep(sub, 'G', "");

```

```

        end

```

```

        if (seq_out(i+j+1)=='A')&&(seq_out(i+j+2)=='G')    %No AAG after I

```

```

            sub=strrep(sub, 'A', "");

```

```

        end

```

```

        if (seq_out(i+j-1)=='A')&&(seq_out(i+j+1)=='G')    %No AAG Middle I

```

```

            sub=strrep(sub, 'A', "");

```

```

        end

```

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

```

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

```

```

        if (seq_out(i+j-2)=='A')&&(seq_out(i+j-1)=='T')    %No ATG before II

```

```

            sub=strrep(sub, 'G', "");

```

```

        end

```

```

        if (seq_out(i+j+1)=='T')&&(seq_out(i+j+2)=='G')    %No ATG after II

```

```

            sub=strrep(sub, 'A', "");

```

```

        end

```

```

        if (seq_out(i+j-1)=='A')&&(seq_out(i+j+1)=='G')    %No ATG Middle II

```

```

            sub=strrep(sub, 'T', "");

```

```

        end

```

```

%%%%%%%%%
%%%%%%%%%
if (seq_out(i+j-2)=='C')&&(seq_out(i+j-1)=='T')    %No CTT before I
    sub=strrep(sub, 'T', "");
end
if (seq_out(i+j+1)=='T')&&(seq_out(i+j+2)=='T')    %No CTT after I
    sub=strrep(sub, 'C', "");
end
if (seq_out(i+j-1)=='C')&&(seq_out(i+j+1)=='T')    %No CTT Middle I
    sub=strrep(sub, 'T', "");
end
%%%%%%%%%
%%%%%%%%%
if (seq_out(i+j-2)=='C')&&(seq_out(i+j-1)=='A')    %No CAT before II
    sub=strrep(sub, 'T', "");
end
if (seq_out(i+j+1)=='A')&&(seq_out(i+j+2)=='T')    %No CAT after II
    sub=strrep(sub, 'C', "");
end
if (seq_out(i+j-1)=='C')&&(seq_out(i+j+1)=='T')    %No CAT Middle II
    sub=strrep(sub, 'A', "");
end
%%%%%%%%%
%%%%%%%%%
if (seq_out(i+j-2)=='C')&&(seq_out(i+j-1)=='C')    %No CCT before III
    sub=strrep(sub, 'T', "");
end
if (seq_out(i+j+1)=='C')&&(seq_out(i+j+2)=='T')    %No CCT after III
    sub=strrep(sub, 'C', "");
end
if (seq_out(i+j-1)=='C')&&(seq_out(i+j+1)=='T')    %No CCT Middle III
    sub=strrep(sub, 'C', "");
end
%%%%%%%%%
%%%%%%%%%
if (seq_out(i+j-2)=='A')&&(seq_out(i+j-1)=='G')    %No AGG before III
    sub=strrep(sub, 'G', "");
end
if (seq_out(i+j+1)=='G')&&(seq_out(i+j+2)=='G')    %No AGG after III
    sub=strrep(sub, 'A', "");
end

```

```

if (seq_out(i+j-1)=='A')&&(seq_out(i+j+1)=='G')    %No AGG Middle III
    sub=strrep(sub, 'G', "");
end
%%%%%%%%%
%%%%%%%%%
if (seq_out(i+j-2)=='A')&&(seq_out(i+j-1)=='A')    %No AAA before IV
    sub=strrep(sub, 'A', "");
end
if (seq_out(i+j+1)=='A')&&(seq_out(i+j+2)=='A')    %No AAA after IV
    sub=strrep(sub, 'A', "");
end
if (seq_out(i+j-1)=='A')&&(seq_out(i+j+1)=='A')    %No AAA Middle IV
    sub=strrep(sub, 'A', "");
end
%%%%%%%%%
%%%%%%%%%
if (seq_out(i+j-2)=='T')&&(seq_out(i+j-1)=='T')    %No TTT before IV
    sub=strrep(sub, 'T', "");
end
if (seq_out(i+j+1)=='T')&&(seq_out(i+j+2)=='T')    %No TTT after IV
    sub=strrep(sub, 'T', "");
end
if (seq_out(i+j-1)=='T')&&(seq_out(i+j+1)=='T')    %No TTT Middle IV
    sub=strrep(sub, 'T', "");
end
%%%%%%%%%
%%%%%%%%%
if (seq_out(i+j-2)=='C')&&(seq_out(i+j-1)=='T')    %No CTC before V
    sub=strrep(sub, 'C', "");
end
if (seq_out(i+j+1)=='T')&&(seq_out(i+j+2)=='C')    %No CTC after V
    sub=strrep(sub, 'C', "");
end
if (seq_out(i+j-1)=='C')&&(seq_out(i+j+1)=='C')    %No CTC Middle V
    sub=strrep(sub, 'T', "");
end
%%%%%%%%%
%%%%%%%%%
if (seq_out(i+j-2)=='G')&&(seq_out(i+j-1)=='A')    %No GAG before IV
    sub=strrep(sub, 'G', "");
end

```

```

if (seq_out(i+j+1)=='A')&&(seq_out(i+j+2)=='G')    %No GAG after IV
    sub=strrep(sub, 'G', "");
end
if (seq_out(i+j-1)=='G')&&(seq_out(i+j+1)=='G')    %No GAG Middle IV
    sub=strrep(sub, 'A', "");
end
%%%%%%%%%%%%%%
%%%%%%%%%%%%%%
%   if seq_out(i+j-4)==seq_out(i+j-3)...          %No 5xN repeats
%       &&seq_out(i+j-3)==seq_out(i+j-2)...
%       &&seq_out(i+j-2)==seq_out(i+j-1)
%       sub=strrep(sub, seq_out(i+j-1), "");
%   end
    if ~isempty(sub)                                %substitute the letter if there are options
        seq_out(i+j)=datasample(sub, 1);
    end
end
else
    for j = 1:5
        seq_out(j)=datasample(Nuc, 1);
    end
end
end
end

```

Appendix D. Supplementary data

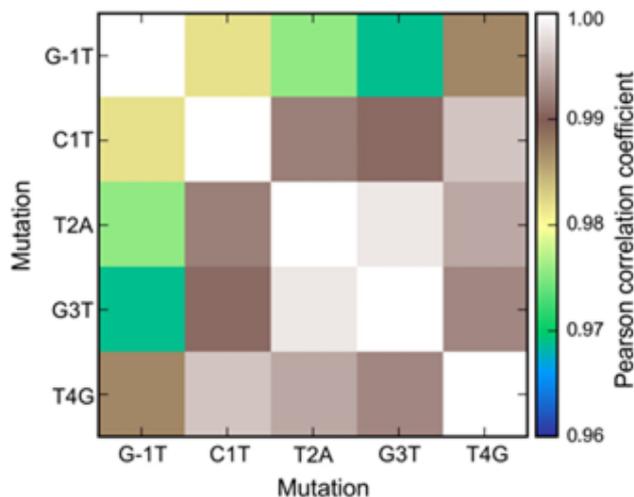


Figure D1. Pearson correlation coefficients of selector spacers for target seed mutants

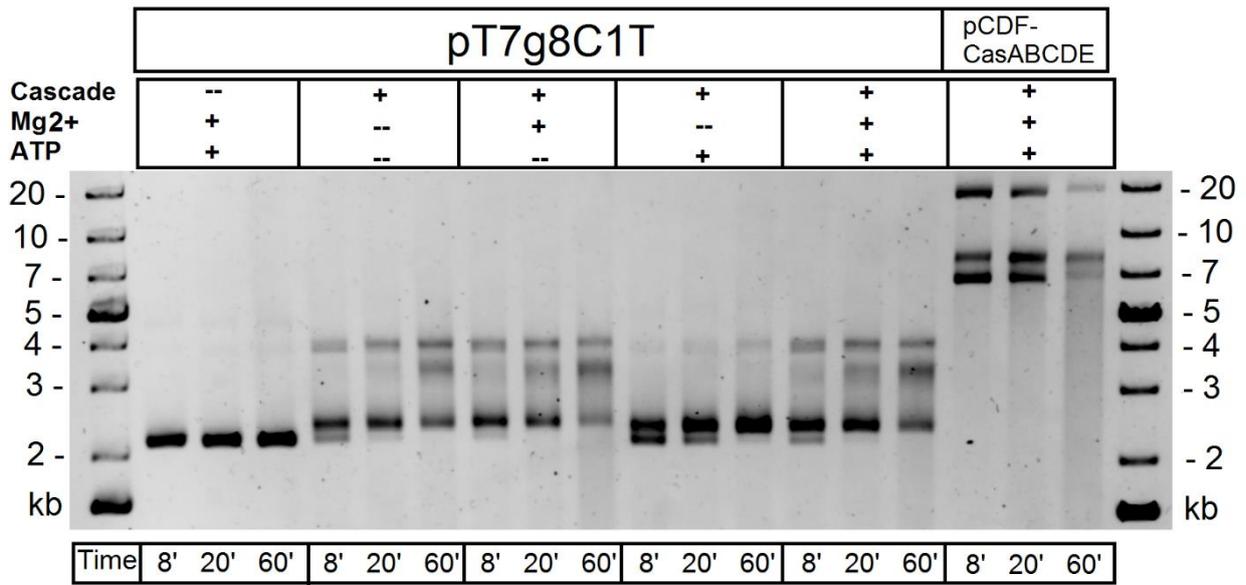


Figure D2. Non-specificity of Cascade Mg²⁺ nuclease contamination.

20160712 HiTrap Ni column 5ml E. coli CasA purification (636039297954157889)

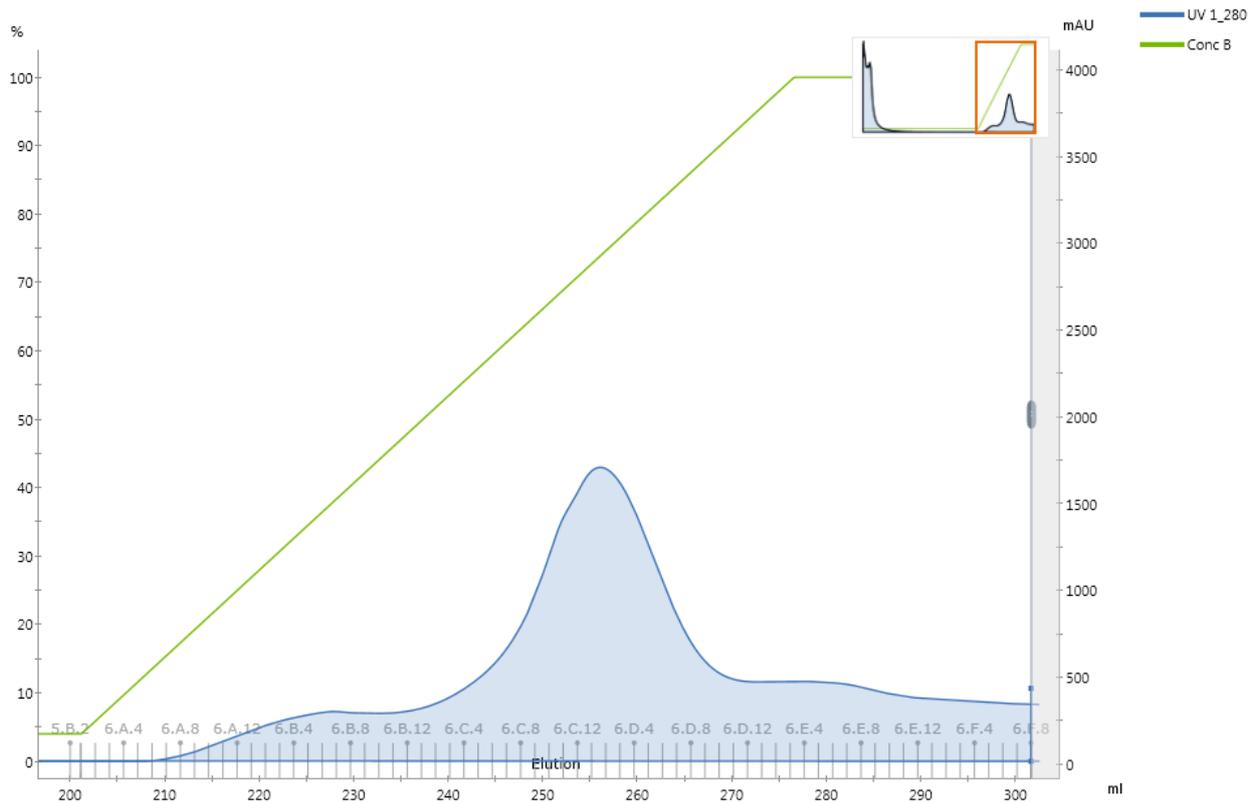


Figure D3. CasA purification (elution) profile from 280 nm Acta detector.

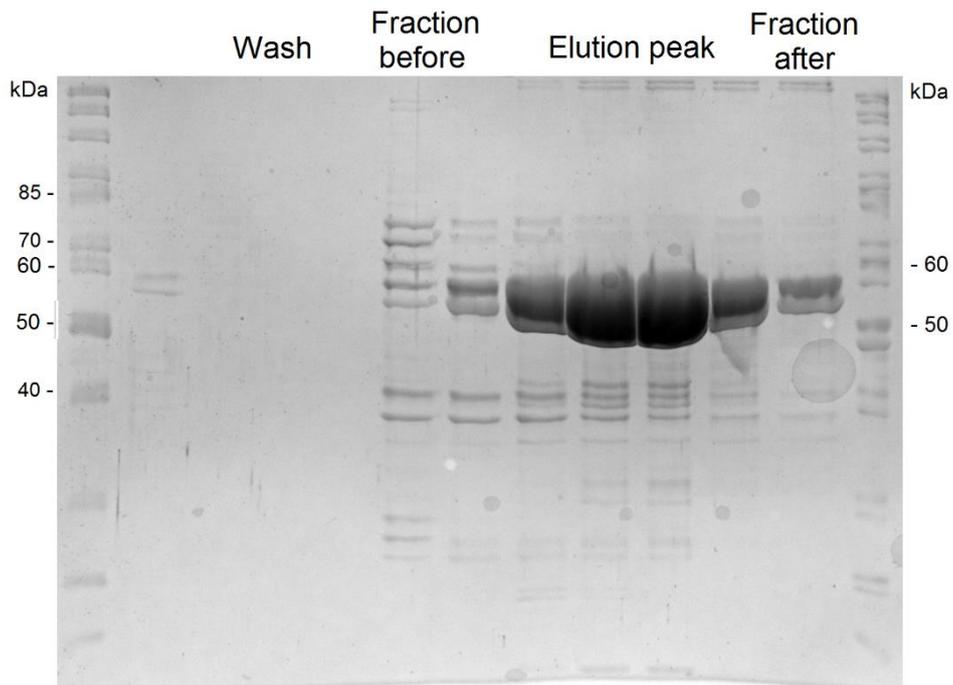


Figure D4. CasA purification SDS PAGE. Double bands are due to gel (not sample) issues.

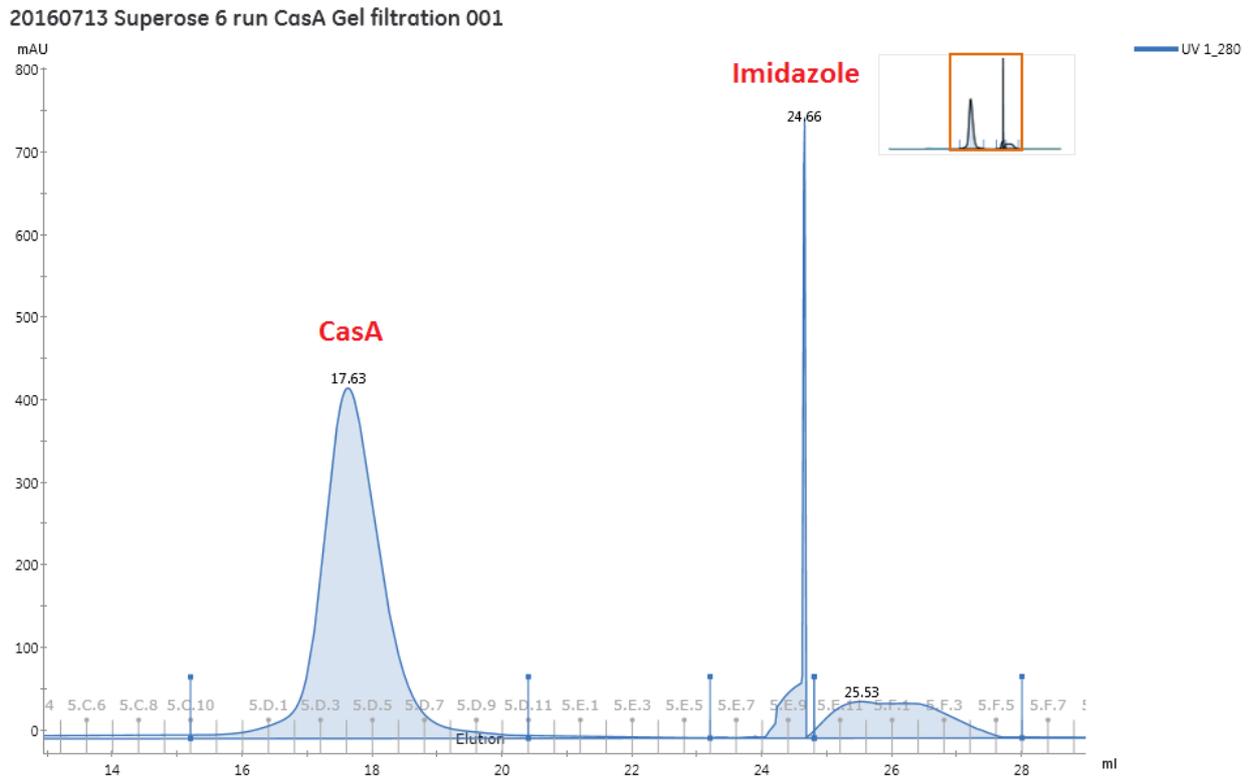


Figure D5. CasA gel filtration purification step. CasA and imidazole picks.

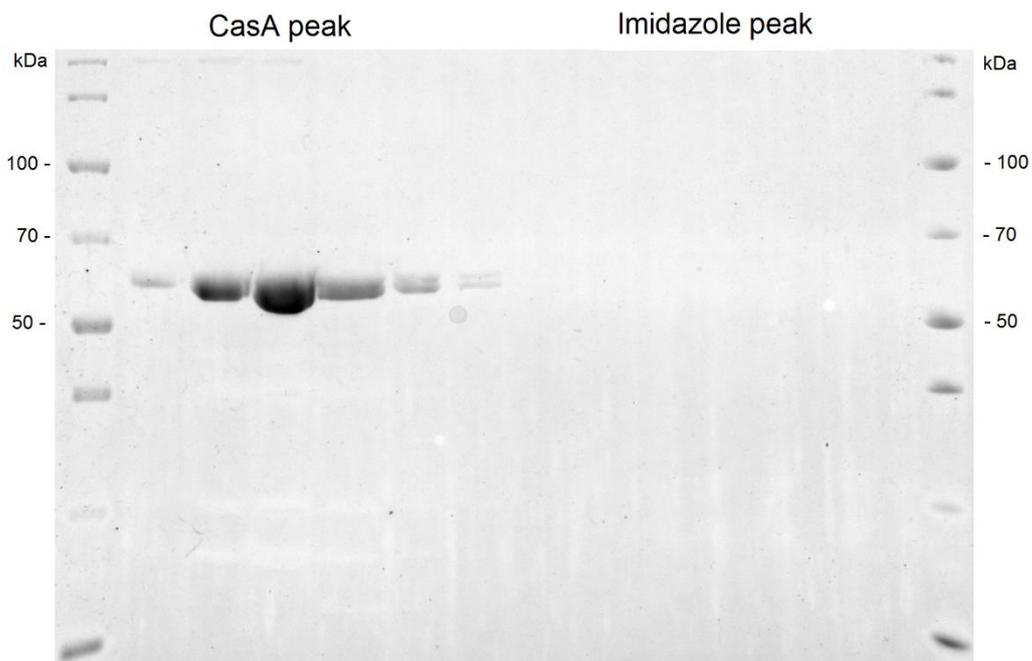


Figure D6. CasA purification SDS PAGE after gel filtration. Double bands are due to gel (not sample) issues. Pure protein is obtained.

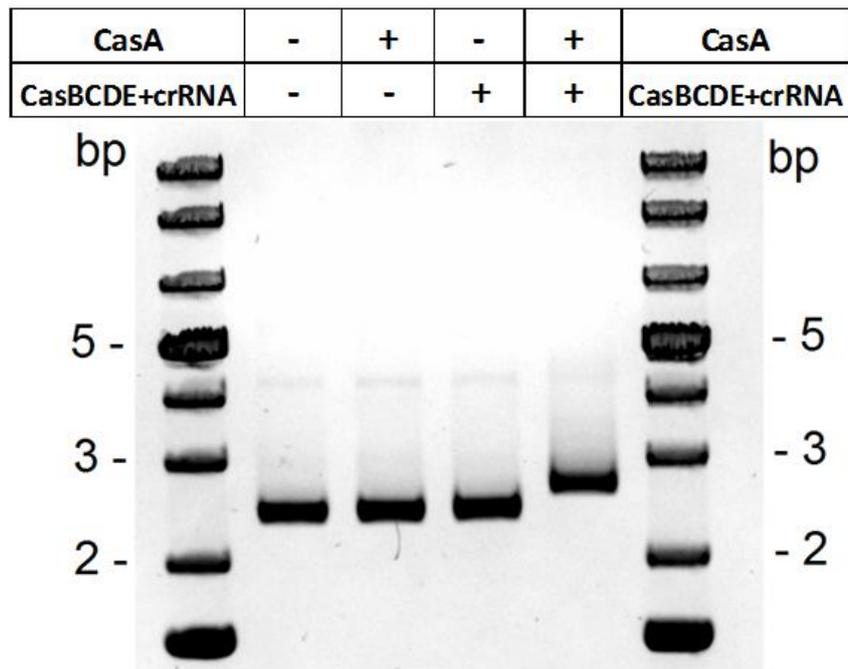


Figure D7. CasA function is tested with the rest of Cascade complex. After incubation together, reconstituted Cascade is capable to bind protospacer-containing plasmid.

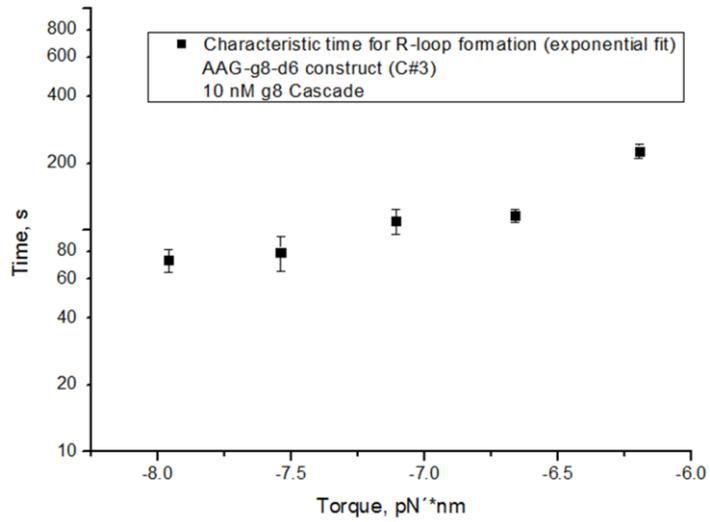
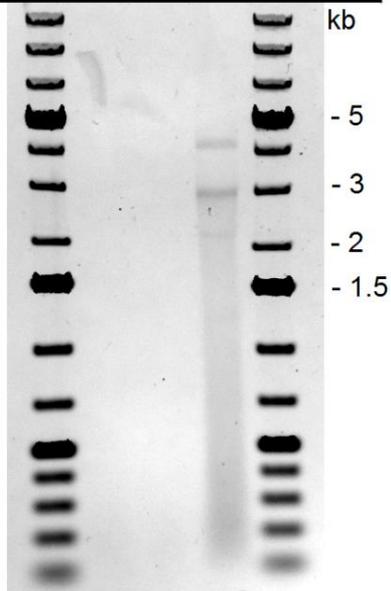


Figure D8. Lack of torque dependence tested for g8 protospacer with ATG PAM.

a

1% agarose gel

Sample ?	1	2	3	Sample ?
pT7g8wtdir	+	+	+	pT7g8wtdir
Cascade	+	+	+	Cascade
Cas3	-	+	+	Cas3
Cas1Cas2	-	+	+	Cas1Cas2
DNase treat	+	+	-	DNase treat



b

Denaturing PAGE

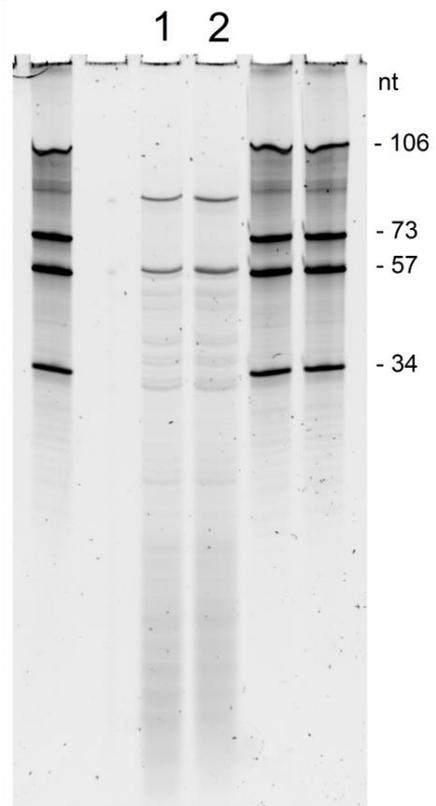


Figure D9. The same as Figure 23 but DNase I treated after incubation.

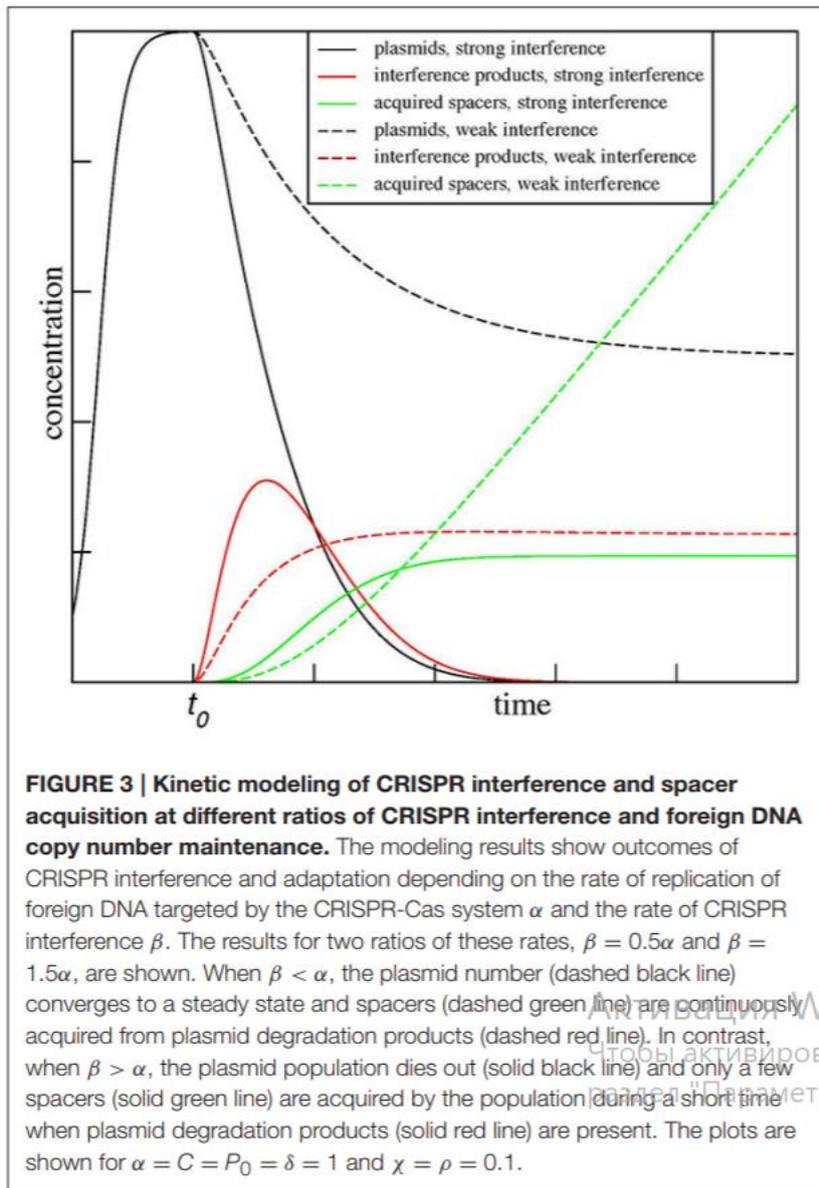


Figure D10. Kinetics model of priming fueled by degradation of targeted DNA proposed by (126)

pCRISPR	+	+	+	+	pCRISPR
Cas3 prod	-	+	-	-	Cas3 prod
Cas1Cas2	+	+	+	-	Cas1Cas2
Oligodup	-	-	+	-	Oligodup

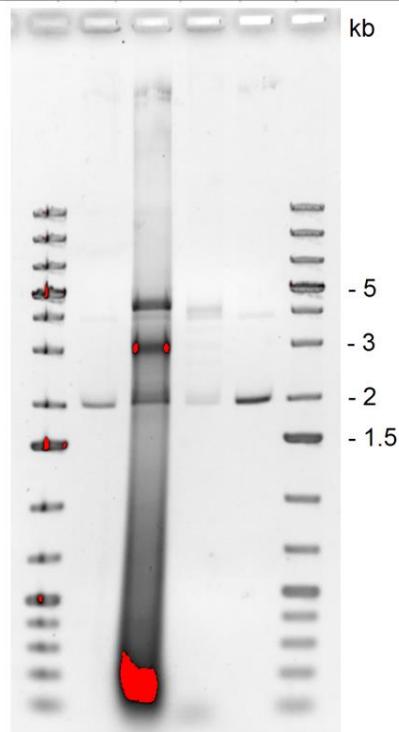


Figure D11. Tests for Cas1-Cas2 integration using concentrated Cas3 degradation product. The product of 5 nM target plasmid degradation was concentrated 20 times through phenol-chloroform extraction and incubated with Cas1-Cas2 and pCRISPR. Intense smear does not allow to look carefully on pCRISPR changes. The method can be developed using bigger plasmid for integration target.

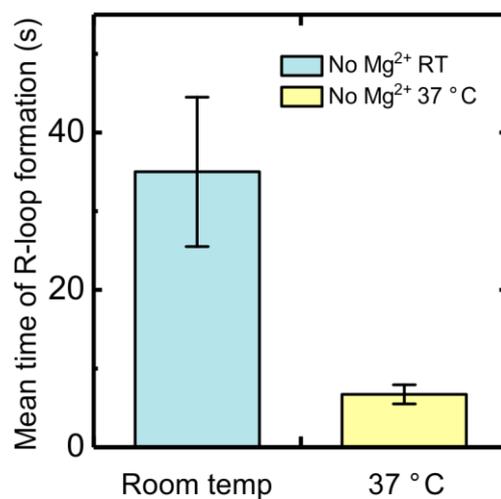


Figure D12. Set of measurements same as presented on Figure 10: Cascade binding mean times in absence of Mg²⁺, comparing conditions of room temperature and 37 °C. The effect of temperature is approximately 5-fold decrease the mean R-loop formation times (seed up R-loop

formation time). This is an intermediate between no Mg^{2+} , room temperature and 5 mM Mg^{2+} , 37 °C.

Appendix E. Derivation of Poisson process to exponential distribution of the events times

The following derivations were used extensively on data analysis for the results of measurements. Let's imagine the Poisson process with events happening independently and repeatedly at the constant rate. Definition of the rate is

$$\text{rate of event } \lambda = \frac{\text{expected number of events over time } t}{\text{time } t} \quad (\text{E.1})$$

That means that at every infinitely small period of time dt there is a probability of event happening is

$$P(N_{dt} = 1) = \lambda * dt \quad (\text{E.2})$$

And not happening is

$$P(N_{dt} = 0) = 1 - \lambda * dt \quad (\text{E.3})$$

Thus, by time t in this case one can expect integral of (E.2) from 0 to t which is λt events, with average. Now, let's think about the distribution of single events happening at the constant rate. The probability that no single event will happen over time T is superposition of (E.3) probabilities that nothing happened during every single infinitely small period of time dt

$$P(N_{t+dt} = 0) = P(N_t = 0) * (1 - \lambda * dt) \quad (\text{E.4})$$

What with the fact that $P(N_0 = 0)$ is equal 1 leads to the multiplication

$$P(N_T = 0) = \prod_0^T (1 - \lambda * dt) \quad (\text{E.4})$$

Which is the simple exponential limit:

$$P(N_T = 0) = \lim_{dt \rightarrow 0} (1 - \lambda * dt)^{\frac{T}{dt}} = \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda * T}{n}\right)^n = e^{-\lambda * T} \quad (\text{E.6})$$

Now, if we look at the random event happened, probability of it to happen after certain time t and before time t are correspondingly

$$S_1(t) = e^{-\lambda * t} \quad (\text{E.7})$$

$$F_1(t) = 1 - e^{-\lambda * t} \quad (\text{E.8})$$

That means that if we collected data of N events, they should be distributed exponentially as the normalized probability density function:

$$\text{pdf}(t) = \lambda * e^{-\lambda * t} \quad (\text{E.9})$$

And should lay on the fitting function:

$$F_N(t) = N * (1 - e^{-\lambda*t}) \quad (E. 10)$$

Fit by this function to the number of events data will allow to determine the rate of the reaction λ , or the characteristic time T_{exp} or $\frac{1}{\lambda}$, which is an Expected value of time and will be:

$$E(t) = \int_0^{\infty} t * pdf * dt \Rightarrow \frac{1}{\lambda} \quad (E. 11)$$

To estimate the error of the T_{exp} determination with N events sample, let's firstly calculate variance for the time:

$$V(t) = E(t^2) - (E(t))^2 \quad (E. 12)$$

Twice integrating by parts, find that

$$E(t^2) = \int_0^{\infty} t^2 * \lambda * e^{-\lambda*t} * dt \Rightarrow \frac{2}{\lambda^2} \quad (E. 13)$$

And the Variance is:

$$V(t) = \frac{1}{\lambda^2} \quad (E. 14)$$

Then, for N events T_1, T_2, \dots, T_N mean value

$$T_{exp} = \frac{\sum_{n=1}^N T_n}{N} \quad (E. 15)$$

Since, N is constant, variance of the T_{exp} will be sum of variances for every T_n over N^2 . And the variance of every T_n is the same as just calculated in (E.12). Then,

$$V(T_{exp}) = \frac{1}{N^2} \frac{N}{\lambda^2} = \frac{1}{N * \lambda^2} \quad (E. 16)$$

Standard deviation of T_{exp} then will be square root of variance by definition:

$$\sigma(T_{exp}) = \frac{1}{\sqrt{N} * \lambda} \text{ or } \frac{T_{exp}}{\sqrt{N}} \quad (E. 17)$$

The reader can find a lot of examples of such procedure applied to the experimental data further in the Results, so, it's very beneficial to understand why the constant rate leads to such kind of distribution.