ACCESS DENIED

A closer look at anti-phage defense mechanisms

Marnix Vlot

A strong feeling of adventure is animating those who are working on bacterial viruses, a feeling that they have a small part in the great drive towards a fundamental problem in biology.

— Max Ludwig Henning Delbrück
1946

Access Denied

A closer look at anti-phage defense mechanisms

Marnix Vlot

Thesis committee

Promotor

Prof. Dr John van der Oost
Personal chair at the Laboratory of Microbiology
Wageningen University & Research

Co-promotor

Dr Stan J.J. Brouns
Associate Professor, Department of Bionanoscience
Delft University of Technology

Other members

Prof. Dr Eugene V. Koonin, National Center for Biotechnology Information (NCBI), USA
Prof. Dr Michiel Kleerebezem, Wageningen University & Research
Dr Joyce H.G. Lebbink, Erasmus Medical Centre Rotterdam
Dr Gorben P. Pijlman, Wageningen University & Research

Access Denied

A closer look at anti-phage defense mechanisms

Marnix Vlot

Table of Contents

Chapter 1

General introduction and thesis outline

Bacteriophage life cycle

Bacteriophages (phages) were discovered by Frederick Twort and Félix d'Hérelle in the beginning of the 20th century. d'Hérelle coined the term bacteriophage (derived from "bacteria" and the Greek: φαγεῖν (phagein), "to eat") for these 'invisible antagonists' (D'Herelle, 1917). Phages are the most abundant biological entities in earth, outnumbering their hosts in most ecosystems, up to 150 fold (Wigington et al., 2016). Phages consist of nucleic acid genomes encapsulated by a proteinaceous capsid and can have a wide range of morphologies. Most known phages have a polyhedral morphology, but others are filamentous, or pleomorphic (Ackermann, 2003). Most phage genomes consist of dsDNA, but others contain ssDNA, ssRNA, or dsRNA. Bacteriophages are metabolically inert and require a host for replication. The process of phage infections starts with adsorption of the virus to the cell surface. Phage adsorption is usually accomplished via binding of the phage particle to cell-surface receptors, membrane proteins, or lipopolysaccharides (LPS). Subsequent to successful adsorption, the viral DNA is injected into the host cell. Penetration of the bacterial cell wall is sometimes aided by phage-encoded lysozymes bound to the phage tail that partially degrades the peptidoglycan layer (Arisaka et al., 2003). The genomic DNA of the phage is then replicated and copy numbers can reach up to 200 (Demerec and Fano, 1945). Transcription and translation of viral DNA is followed by protein assembly and viral particle formation. The last step of the viral life cycle is lysis of the host cells, usually performed by viral enzymes, such as porins, that perforate the cell wall. Generally, there are two ways a viral infection can spread, namely lytic and lysogenic infection. In a lytic infection, which is considered the main lifestyle of phages, the phage directly takes over the metabolic activities of the host in order to replicate and release new phages. Unlike lytic infection, lysogenic infection is followed by integration of the viral genome into the bacterial genome. The integrated viral genome (prophage) turns the bacterial cell into a lysogen. Prophages can be activated upon cellular stress such as DNA damage which initiates the SOS response of the cell (Nanda et al., 2014). Interestingly, some phages show yet another lifestyle, that of a chronic infection. In this life cycle, there is a continues production of phage particles that are released from the cell without lysis of the cell (Clokie et al., 2011).

Biotechnological applications of phages

Phages are present in all ecosystems where bacteria are found, which can include industrial fermentation reactors. Although phages can cause problems in bacterial fermentation processes in food, feed, chemical, and pharmaceutical industries, phage related problems probably are best documented in dairy related fermentations (Garneau and Moineau, 2011). In a dairy fermentation process, lactic acid bacteria (LAB) produce lactic acid via lactose fermentation, which acidifies the milk. The decreased pH is a crucial factor for avoiding contamination by pathogenic and spoilage microorganisms. Phage infection of LAB starter cultures is the primary cause of slow fermentation or even culture collapse in industrial fermentation processes. The relevance and implications of phage infections are enormous, considering that the dairy industry is a multi-

billion dollar industry. A better understanding of phage-host interactions is expected to contribute to the development of more effective anti-phage measures (Horvath and Barrangou, 2010).
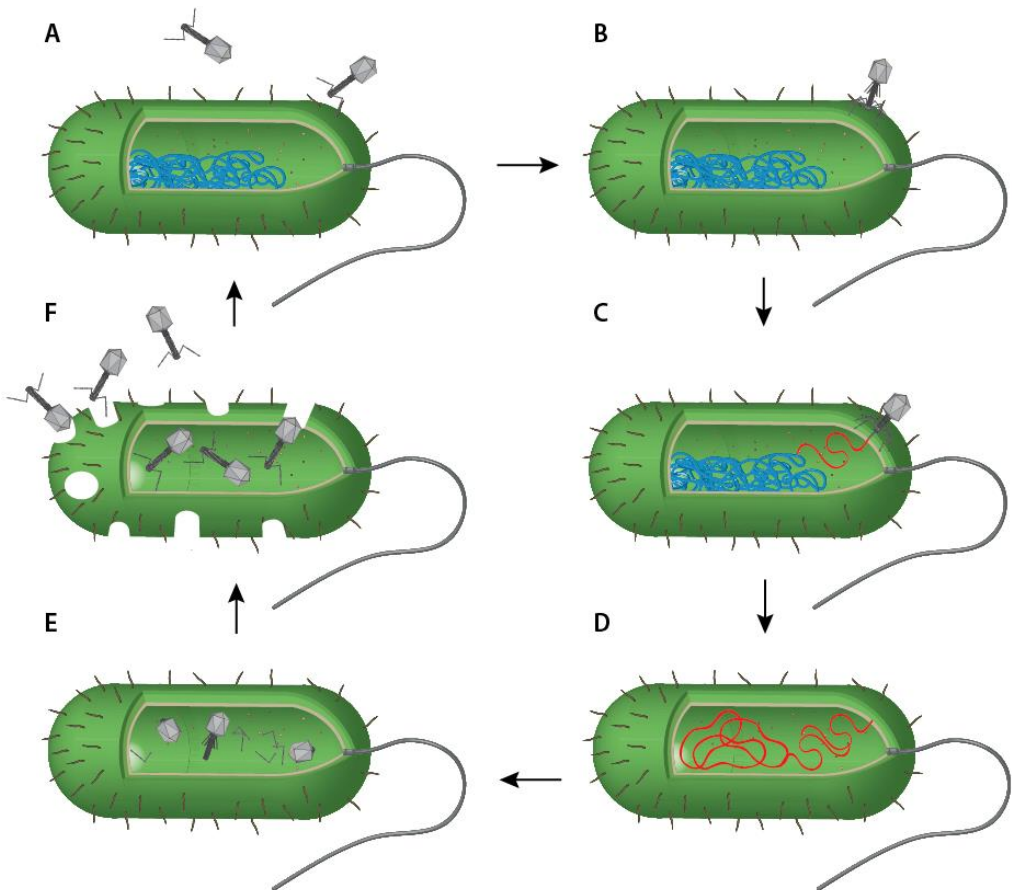
Figure 1. Schematic overview of the life cycle of *Escherichia coli* phage T4. (**A**) A single tail fiber binds to the outer membrane receptors OmpC or LPS. (**B**) Multiple long tail fibers unwrap from the tail sheath and bind to the outer membrane. Also the baseplate binds the membrane. (**C**) The tail sheath contracts and internal tube penetrates the outer membrane and DNA is injected into the cytoplasm. (**D**) Degradation of the host DNA and replication of viral DNA. (**E**) Synthesis of phage proteins followed by assembly of complete phage particles. (**F**) Holin proteins form pores in the bacterial cell membrane and the cell bursts to release phage progeny.

Box1. The quintessence of bacteriophages / War against the machines

Bacteriophages are by common definition not alive nor organisms. The lack of a universal definition of life indicates that the question whether something is alive might be a more philosophical than a physical question. Life definitions shift towards 'context-dependent' definitions to include entities that are argued to be alive but do not satisfy technical definitions (Schrödinger, 1967). Some biologists argue that the living part of a

bacteriophage is not the viral particle, but the viral factory within the infected cell with all its associated biosynthetic activity (Brüssow, 2009). The manner in which inanimate bacteriophages are being transmitted could ground one's argument for considering them as 'memes'. A meme, as introduced in Richard Dawkins' book *The Selfish Gene* (Dawkins, 1976), is a discrete unit that is replicated, mutated and responsive to selective pressure (Stotz, 2004). These philosophical ideas might not give a satisfactory distinction between life or inanimacy for scientists who are used to thinking in well-defined categories. The question whether bacteriophages are alive might forever stay controversial.

Anti-viral defense mechanisms

The parasitic nature of bacteriophages poses an incredible selective pressure on micro-organisms to develop anti-viral defense mechanisms. The competitive interactions between hosts and parasites forces a continuous arms-race to prevent extinction. This evolutionary law was postulated as the Red Queen Hypothesis (van Valen, 1973). Originally describing extinction biology of eukaryotes, this hypothesis is most applicable on the co-evolution of prokaryotes and their viruses largely due to the fast mutation rates of phage genomes (Paez-Espino et al., 2015).

Recently, new types of defense mechanisms have been discovered, of which the CRISPR-Cas mechanisms have had a large impact on the field of genome engineering (Doudna and Charpentier, 2014). Anti-viral defense mechanisms can be divided into two categories: prevention of phage infection and prevention of phage replication. In the following section examples are described for several anti-viral defense mechanisms and escape mechanisms used by bacteriophages are explained.

Inhibition of phage adsorption

Bacteriophage adsorption is initiated by interaction between cell surface receptors and bacteriophage components. With a distinct composition of cell wall and surface structures, gram-negative and gram-positive bacteria have different receptors available for phage adsorption (Figure 2). Adsorption inhibition strategies include blocking of phage receptors, production of extracellular matrix, and production of competitive inhibitors.

In gram-negative bacteria, lipopolysaccharides (LPS) represent the most common receptor for phage attachment. LPS is composed of fatty acids and monosaccharides typically organized into three sub-structures: lipid A, core polysaccharide and O-chain (i.e. O-antigen). LPS with such organization is known as smooth LPS; rough LPS lacking the O-antigen can also be found in many gram-negative bacteria (Pupo et al., 2013). Phages targeting the O-antigen (smooth LPS) are expected to reveal a narrower host range due to the high variability of the O-antigen chain among species and strains (Alicja Niewiadomska et al., 2005). Conversely, phages

recognizing elements of the LPS core may demonstrate a broader host range as the core structure (exposed in rough LPS and less accessible in smooth LPS) is highly conserved in a number of species and genera of gram-negative bacteria. Production of a structured extracellular matrix such as LPS can provide bacteria with a physical barrier between phages and their receptors. In response these extracellular matrices have become receptors themselves, as described above. Antithetically, phages have evolved polysaccharide degrading enzymes, such as lysozymes (reviewed by Sutherland, 1995). These viral proteins are either bound to the phage particle or released in the environment when cells lyse.
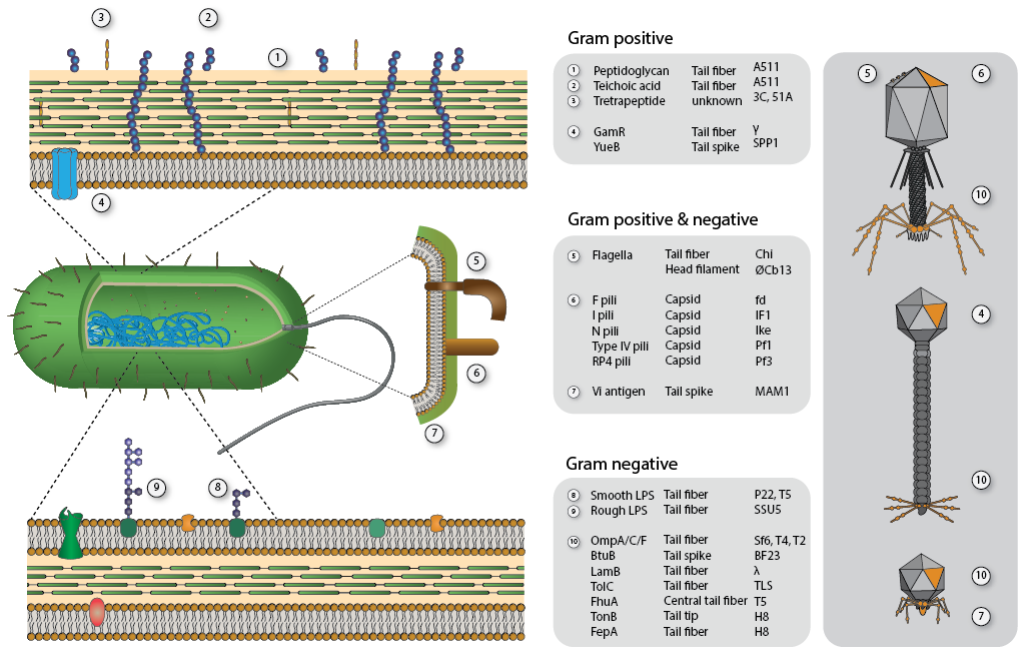
Figure 2. Representation of Gram-positive and Gram-negative bacterial receptors used by bacteriophages for adsorption.

Membrane proteins, such transport proteins and efflux pumps, can also serve as phage receptors. Outer membrane proteins OmpC and OmpF in gram-negative bacteria are used by bacteriophages as receptor for tail-fiber attachment (Hantke, 1978; Ho and Slauch, 2001; Marti et al., 2013). Maltose transport protein LamB, vitamin B12 transporter BtuB, and efflux pump TolC are used as receptors by phage λ, BF23, and ST29 respectively (Meyer et al., 2012; Mondigler et al., 2006; Ricci and Piddock, 2010). A common mechanism of bacteria to block phage infection is mutation of receptors. In response phages can evolve the ability to recognize new receptors. Phage T4 has developed an OmpC-dependent and OmpC-independent/LPS-dependent adsorption mode to circumvent adsorption block by OmpC mutation (Washizaki et al., 2016).

Another example is the adaptation of phage λ to recognize a new receptor upon reduced expression of outer-membrane receptor LamB (Meyer et al., 2012). While LamB is its preferred receptor, phage λ was able to change the receptor specificity by minor mutations of its tail attachment protein J.

In gram-positive bacteria the main component of the cell wall is peptidoglycan, a heteropolymer composed of N-acetylglucosamine and N-acetylmuramic acid. A tetrapeptide is covalently attached to the N-acetylmuramic acid. The other vital component of gram-positive bacteria cell wall are teichoic acids, polyol phosphate polymers with either ribitol or glycerol linked by phosphodiester bonds. The substituent groups on the polyol chains (D-alanine, N-acetylglucosamine, N- acetylgalactosamine, and glucose) are characteristic of a particular bacterial species and thus can act as a specific antigenic determinant. For example, *Staphylococcus aureus* and *Bacillus subtilis* have similar peptidoglycan and teichoic acids structure, so phages specific for *S. aureus* may also adsorb on the surface of *B. subtilis* (Rakieten and Rakieten, 1937). Conversely, the specific distinction of the teichoic acid substituent between these species (D-glucose for Bacillus and N-acetylglucosamine for S. aureus) is key for the specific adsorption of phages to B. subtilis (e.g. phages Ø1, SP3 and Sp02)(Glacer, L., Lonesco, H. & Shaefer, 1966; Young, 1967). Proteins of gram-positive bacteria may also serve as receptors for phage adsorption. Examples include GamR of the *Bacillus anthracis* cell wall which is involved in the adsorption of phage γ 63, and the extra-membrane domain YueB of *Bacillus subtilis* to which phage SPP1 irreversibly binds (São-José et al., 2006). As some of the receptors used by phages are not essential for the host, phage resistance frequently occurs by random mutation.

Super infection exclusion / DNA injection blocking

Shortly after phage adsorption the viral DNA is injected into the cell via a syringe-like mechanism and within a few minutes the first phage-encoded proteins are synthesized (Molineux, 2008). In the case of T-even phages, these include superinfection exclusion (Sie) proteins. Sie systems were discovered by Dubecco in the 1950s, who observed that infection with a mix of two different phages was inefficient unless infection occurred simultaneously. Sie systems are typically encoded by phages or prophages and protect bacteria from secondary infections by the same or closely related phages. Membrane-associated proteins *imm* and *sp* were later characterized as responsible for Sie in T4 (Dulbecco, 1952; Lu and Henning, 1994). Imm inhibits DNA transfer across the cytoplasmic membrane and Sp inhibits local degradation of the murein layer by phage lysozyme (Lu and Henning, 1994).

Restriction-Modification

Bacterial Restriction-Modification (R-M) systems can provide protection from mobile genetic elements by degradation of foreign DNA. R-M systems are active against phage DNA but also strongly involved in the regulation of plasmid immigration, illustrated by the fact that R-M systems are over-represented in naturally competent organisms (Oliveira et al., 2014). This phenomenon was first discovered when Arber and Dussoix

found that phage λ isolated from E. coli B strains could not infect E. coli K strains and vice versa (Kühnlein and Arber, 1972). This inability to infect was due to DNA restriction by EcoKI and EcoBI restriction systems respectively. Typically, R-M systems have enzymes responsible for recognition and cleavage of specific DNA sequences and a cognate methyltransferase that confers protection from cleavage by methylation of the same DNA sequence. Typically, specific DNA sequences of the host are methylated whereas invading DNA usually does not contain these methylation patterns. A multitude of DNA base modifications is known that allows hosts to distinguish between their own and phage DNA. The most prominent DNA modifications are nucleobase methylations of either adenine or cytosine.

R-M systems can be divided into four different classes (Roberts et al., 2003). Type I R-M systems are characterized by a multisubunit protein complex that usually contains two restriction endonuclease (REase) subunits, two methyltransferase (MTase) subunits, and one specificity (S) subunit. The S subunit determines the DNA sequence that is recognized. When the recognized DNA sequence does not contain methylated adenosine bases (m6A) the REase subunits of the protein complex cleave the DNA in an ATP-dependent manner. When the recognized DNA sequence is hemimethylated, in which one of the strands is methylated, the MTase subunit uses S-adenosylmethionine (SAM) as methyl donor for methylation of the unmethylated strand. The best known example of a Type I R-M system is the EcoKI system in *E. coli* K-strains that recognizes the ACC(N_6)GTGC sequence (Murray, 2000; Sain and Murray, 1980)

Type II R-M systems are a widely diverse class subdivided into 11 subclasses (Roberts et al., 2003). REase subunits and MTase subunits function separately and recognize short palindromic sequences. A monomeric or homodimeric REase cleaves in or close to that sequence producing 5'-phosphates and 3'-hydroxyls. Type II R-M enzymes, such as EcoRI, are commonly used in recombinant DNA technologies and more than 3500 REases have been characterized (Pingoud and Jeltsch, 2001; Roberts et al., 2015).

Type III R-M systems are similar to Type I systems as they form a protein complex containing two REase subunits and two MTase subunits. Cleavage occurs when two inversely oriented copies of unmethylated recognition sequences are bound to the R-M complex (Meisel et al., 1992). Recognition sequences are typically 5-6 bp asymmetric, exemplified by the EcoP1 enzyme that recognizes AGACC, and cleavage only requires Mg^{2+} as a cofactor (Bickle and Kruger, 1993)

Unlike other R-M systems that recognize methylated DNA as host DNA, Type IV R-M systems will only cleave bases that have been modified (Loenen and Raleigh, 2014). The first Type IV restriction enzymes characterized were rglA and rglB (restrict glucoseless) which were later renamed to McrA and McrBC (modified cytosine restriction). The structural organization of modification dependent systems (MDS) suggests that they have emerged multiple times independently of each other (Loenen and Raleigh, 2014). Sequence specificity is usually very low for Type IV restriction enzymes.

Anti-restriction systems

In response to the selective pressure of R-M systems bacteriophages have developed several different mechanisms to subvert these defense systems (Tock and Dryden, 2005). One classic example of R-M diversion is the production of Ocr (overcome restriction) protein encoded by gene 0.3 of bacteriophage T7 (Walkinshaw et al., 2002). This anti-restriction protein contains a characteristic sequence of negatively charged Asp and Glu residues. Dimers of Ocr mimic approximately 24 base pairs of bent B-form DNA making it an efficient inhibitor of all known Type I R-M systems (Walkinshaw et al., 2002). The inhibitory activity of Ocr has been exploited to improve transformation efficiencies by blocking Type I R-M systems in various organisms by electroporation of Ocr proteins along with transforming DNA (Hoffman et al., 2002). Upon injection of the bacteriophage DNA it is directly vulnerable to attack by restriction enzymes. Therefore, in order to provide protection from restriction it is cardinal significance that anti-restriction proteins are synthesized directly after DNA injection. In the case of T7, the ocr gene is the first to be transcribed and translated (Atanasiu et al., 2001; Walkinshaw et al., 2002). To prevent belated protection by phage-encoded anti-restriction proteins that have to be transcribed and translated, anti-restriction proteins can also be co-injected with the phage DNA. Anti-restriction proteins DarA and DarB from bacteriophage P1 are packaged in the phage head and injected into the recipient cell along with the DNA and bind to restriction sites in the phage genome, protecting from attack by Type I R-M systems (Iida et al., 1987).

Similar to the Ocr protein from T7, phage T4 also encodes a DNA mimicking protein; Arn (anti-restriction nuclease) (Goldberg and Dharmalingam, 1976, 1979). Besides binding to restriction enzyme McrBC, Arn also binds to bacterial histone-like protein H-NS. H-NS plays a regulatory role in host defense and by disruption of the H-NS-DNA binding Arn neutralizes the gene silencing effect of H-NS (Ho et al., 2014). The identification of this dual function of Arn may suggest that anti-restriction proteins play more important role in overcoming host defense systems than previously anticipated.

Other mechanisms to overcome host defense systems include methylation of phage DNA, loss of restriction sites or by hiding restriction sites in a non-recognizable conformation. *E .coli* phage T3 evades the Type II-E restriction enzyme EcoRII by keeping copies of the pseudo-palindromic restriction sites genomically distant from each other (H.Kruger et al., 1979). Similarly, T7 evades EcoP15 by using inverse orientation of recognition sites (Meisel et al., 1992).

BREX and DISARM

Recently Sorek and colleagues have discovered two novel phage resistance systems that are widespread in microbial genomes (Goldfarb et al., 2015; Ofir et al., 2017). The Bacteriophage Exclusion (BREX) system is a novel system that allows phage adsorption but blocks phage DNA replication. Unlike R-M systems, BREX proteins do not cleave or degrade phage DNA. BREX systems are found in approximately 10% of all sequenced

microbial genomes. The BREX system characterized in *Bacillus cereus* includes a methylase that methylates the non-palindromic 5'-TAGGAG-3' sequence, providing bases of discrimination between self and non-self DNA. A putative Lon-like protease, a protein containing an alkaline phosphatase domain, a putative RNA- binding protein, and a protein containing an ATPase-domain, and a protein of unknown function together with the DNA methylase confer resistance to both virulent and temperate phages (Goldfarb et al., 2015). Interestingly, infection by several Lambda-like phages is not blocked by BREX. This could indicate that this phage family has evolved strategies to evade the BREX defense.

The Defense Islands System Associated with Restriction-Modification (DISARM) system is yet another variation on the theme of inhibition of phage replication. DISARM systems are found in approximately 3% of all sequenced genomes. Similar to R-M systems, the DISARM system also prevents phage DNA replication by restriction but distinguishes itself from R-M systems by the multi-gene restriction-modification module. The DISARM system in *Bacillus paralicheniformis* 9945A includes a methylase that methylates the palindromic CCWGG sequence. A protein with a helicase domain, a protein with a DUF1998 domain, a protein with a catalytic nuclease PLD domain, and a protein of unknown function (Ofir et al., 2017). The DISARM system in B. *paralicheniformis* 9945A provides protection against diverse phages.

CRISPR-Cas systems

The anti-viral defense mechanisms described above are robust and generic systems but not adaptive. Clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) systems were the first adaptive and heritable immunity systems found in prokaryotes. CRISPR-Cas systems provide prokaryotes with resistance against mobile genetic elements (MGEs), such as plasmids and bacteriophages (Barrangou et al., 2007; Brouns et al., 2008; Marraffini and Sontheimer, 2008). CRISPR-Cas systems are found in most bacterial and almost all archaeal genomes (Westra et al., 2012). CRISPR-Cas systems can be classified into two classes, six types, and multiple subtypes (Makarova et al., 2015; Shmakov et al., 2017). The mechanisms of CRISPR-Cas interference differs among different classes, types, and sub-types.
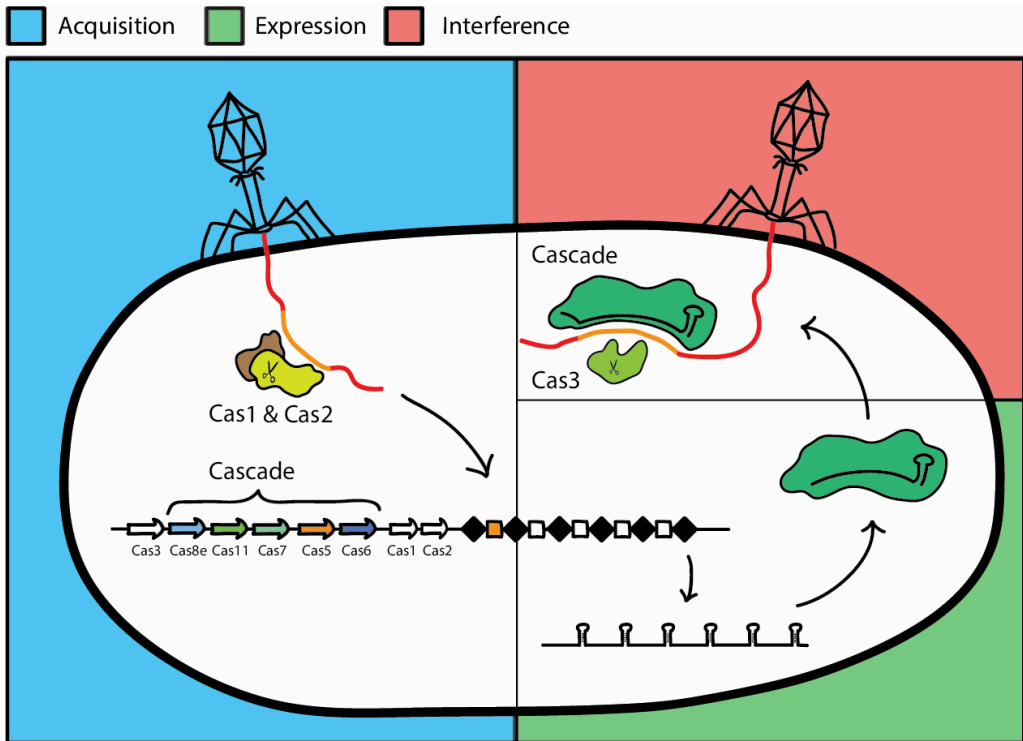
Figure 3. Schematic overview of the three stages of the CRISRP-Cas type I-E system in *E. coli*. In the acquisition phase Cas1 and Cas2 are involved in the expansion of the CRISPR by addition of short fragments of foreign DNA. In the expression stage Cas genes are transcribed and translated and form CRISPR-Cas complexes. The CRISPR array is transcribed into long pre-crRNAs. Processing and loading of crRNAs result in the formation of crRNA-guided CRISPR complexes. In the intereference stage previously encountered DNA is recognized and boudn by the crRNA-guided complexes. The nuclease Cas3 is recruited and degrades the foreign DNA.

CRISPR-Cas systems function in three stages; adaptation, expression, and interference. In the adaptation process invader DNA sequences are incorporated into a CRISPR array. Incorporation is mediated by Cas1 and Cas2 which are conserved in most CRISPR-Cas systems (Koonin and Makarova, 2013). In the expression stage CRISPR arrays are transcribed into pre-crRNA (pre-CRISPR RNA) and further processed into crRNAs (CRISPR RNAs). In the interference stage CRISPR-Cas effector complexes loaded with crRNA bind their cognate foreign DNA targets (termed "protospacers") when flanked by a protospacer-adjacent motif (PAM) (Barrangou; Marraffini and Sontheimer, 2010; Sorek et al., 2013; Wiedenheft et al., 2012) followed by degradation of target DNA.

Escherichia coli K12 contains a type I-E CRISPR-Cas hallmarked by the Cascade crRNA-effector complex encoded by Cas8e, Cas11, Cas7, Cas5, and Cas6. The crRNA guides Cascade to dsDNA target sequences by forming base pairs with the complementary DNA strand. Binding affinity between crRNA-Cascade and target DNA is within

the nanomolar range (Beloglazova et al., 2015). Target binding induces displacement of the noncomplementary strand to form an R-loop (Jore et al., 2011; Thomas et al., 1976). R-loop formation marks the target DNA for degradation by the CRISPR-associated helicase-nuclease Cas3. Cas3 degrades the DNA in a processive manner and hereby generating fragments of about 30-100 nucleotides (Kunne et al., 2016).

New spacers from not previously encountered MGEs can be acquired in a process called naïve acquisition. The naïve acquisition process solely requires Cas1 and Cas2 for successful incorporation of new spacers in the CRISPR array. New spacers are usually inserted a the leader proximal end (Yosef et al., 2012). This feature of leader proximal insertion of new spacers provides a molecular chronological record of previously encountered MGEs. Interestingly, this attribute of the CRISPR array was recently used to lay the foundations of a multimodal intracellular recording device (Shipman et al., 2016).

Invaders are capable of avoiding immunity by introduction of mutations that prevent recognition by the CRISPR-Cas system. Mutated protospacers are often not capable of initiating direct interference; nevertheless, immunity can be restored due to a positive feedback mechanism termed priming (Datsenko et al., 2012; Savitskaya et al., 2013; Swarts et al., 2012). The priming process requires all CRISPR-Cas components for increased acquisition of spacers primed by a partially matching protospacer. Recent studies show that DNA degradation products produced by Cas3 are enriched for thymine-stretches in their 3' ends. These degradation fragments form the ideal substrate for the Cas1-2 complex and are efficiently integrated in the CRISPR array (Kunne et al., 2016).

Previous work by Fineran and colleagues revealed that direct interference tolerates mutations mostly at specific positions and can cope with a number of mutations in the non-seed region of the protospacer. The priming process is much more robust and can facilitate spacer acquisition of target containing up to 11 mismatches (Fineran et al., 2014). Major findings by Fineran and colleagues include identification of new PAMs, determination of protospacer locations that promote priming. Further studies by Xue and colleagues showed that differences in CRISPR activity depend on PAM and seed sequences (Xue et al., 2015).

Both the naïve and priming acquisition expand the spacer diversity within a population. As a result of this increased diversity, bacteriophages can no longer escape CRISPR-Cas interference by point mutations. Rapid acquisition of a highly diverse set of spacers constrains phage evolution and can even lead to phage extinction (Houte et al., 2016; Paez-Espino et al., 2015). This incredibly strong selection drives phages to adapt new strategies of CRISPR-Cas evasion. One of these strategies employed by phages infecting *Streptococcus thermophilus* is recombination-based formation of chimeric phages in which sequences targeted by spacers were replaced (Paez-Espino et al., 2015). The presence of multiple phages that infect the same host increases phage genome mosaicism and can prevent phage extinction. Increased rates of recombination events demonstrate that recombination is a mechanism that is specifically used by phages to escape CRISPR interference (Paez-Espino et al., 2015).

Recently, several families of phage-encoded proteins that inhibit CRISPR-Cas systems have been identified (Bondy-Denomy et al., 2013; Pawluk et al., 2016a). These viral anti-CRISPR (Acr) proteins were discovered in lysogens of *Pseudomonas aeroginosa* strains containing a CRISPR-Cas Type I-F system when these systems appeared to be inactivated by small phage-encoded proteins. It was recently demonstrated that different families of Acrs exhibit distinct modes of inhibition (Bondy-Denomy et al., 2015). AcrF1 and AcrF2 bind to the Type I-F Csy complex thereby preventing crRNA binding or crRNA-DNA binding respectively. AcrF3 binds specifically to the Cas3 helicase-nuclease preventing DNA degradation, essentially converting AcrF3 into a gene regulator (Bondy-Denomy et al., 2015). Analogous to AcrF3 proteins, AcrIIC1-3 proteins encoded by mobile elements in *Neisseria meningitidis* bind to Cas9 thereby inhibiting target recognition of Cas9. These Acrs can potentially be used as 'off-switch' for genome editing by conditional control of Cas9 activity (Pawluk et al., 2016b) ACR proteins AcrIIA1-4 encoded by *Listeria monocytogenes* prophages prevent Cas9 binding and can be used to regulate the genome engineering activity of Cas9 (Rauch et al., 2017).

Abortive infection and Toxin-Antitoxin systems

Abortive infection (Abi) systems block phage replication by altruistic death of the host cell. Most Abis have been found in *Lactococcus lactis*. Large scale milk fermentation by *L. lactis* is highly susceptible to phage infection involving economic consequences (Forde and Fitzgerald, 1999; Whitehead and Cox, 1935). Although the mechanism of many Abis remains unknown, the general theme involves activation of dormant enzymes which cleave essential components of the cellular translational apparatus. Protein synthesis is arrested by inactivation of the translational machinery and phage infections is aborted (Chopin et al., 2005). One of the best characterized examples of abortive infection is the cleavage of translation elongation factor Tu (EF-Tu) in *Escherichia coli* K12 strains. Transcription and translation of *gol* region of T4 infecting phage DNA activates proteolysis of EF-Tu by e14-encoded Lit protein. Cleavage of EF-Tu blocks translation and multiplication of the infecting phage (Georgiou et al., 1998) Mutations of the *gol* peptide partially overcome EF-Tu proteolysis.

Lactococcus lactis strains that contain the abortive infection gene *abiD1* prevent growth of phage bIL66 (Bidnenko et al., 1995). Expression of abiD1 is tightly controlled but induced by phage-encoded Orf1 (Bidnenko et al., 2002). AbiD1 is proposed to decrease the activity of phage-encoded RuvC nuclease, a protein required for resolution of phage DNA structures (Bidnenko et al., 1995).

Toxin-antitoxin (TA) systems are known to stabilize plasmids. TA systems act as a programmed cell death mechanism by post-segregational killing (Gerdes et al., 1985; Ogura and Hiraga, 1983). Typically, a plasmid encoded toxin and antitoxin are continuously produced. Antitoxins usually have a shorter half-life than their cognate toxin leading to addiction to the TA system. Cells that lose the plasmids during cell division will not survive because of the lack of the antitoxin .Although most TA systems identified have a role in plasmid maintenance, several systems have been shown to abort phage infection. *Escherichia coli* K38 strains contain the chromosomal addictive toxin-antitoxin module *mazEF* (Aizenman et al., 1996). Continuous production of

the unstable MazE protein is required to inhibit the activity of MazF, a stable toxin that cleaves specific sites of translated mRNAs (Christensen et al., 2003) The mazEF module protects bacterial cultures against the spreads of P1 phages (Hazan and Engelberg-Kulka, 2004). Some TA systems to fulfill dual roles, a role in plasmid maintenance and a role in anti-viral defense. The *hok* (host killing) and *sok* (suppressor of killing) genes encoded on plasmid R1 in *E. coli* was first identified as a post-segregational killer system (Gerdes et al., 1990) but was later shown to exhibit inhibition of infection by phage T4 (Pecota and Wood, 1996). Short-lived Shok mRNA is antisense to Hok mRNA and prevents translation of Hok protein. Upon infection of T4, host mRNA production is halted, allowing translation of the more stable Hok mRNA. The gene product of *hok* is a membrane associated polypeptide that results in a loss of cell membrane potential followed by cessation of respiration leading to cell death (Gerdes et al., 1986).

Modified DNA bases

Nucleic acids are not boring long polymers of only four types of nucleotides (Grosjean, 2009). Until halfway the 20th century, it was thought that nucleic acid polymers (DNA or RNA) only contain four canonical nucleosides: guanosine, adenosine, thymidine, and cytosine. The first report of modified nucleobases was published in 1925, when Johnson and Coghill detected small amounts of methylated cytosine in DNA of avian tubercule bacilli (Johnson and Coghill, 1925). A few decades later, N6-methyl adenine was detected in the DNA of *E. coli* although its function remained unknown (Dunn and Smith, 1958). Unlike methylation in prokaryotes, which is a post-replicative enzymatic process, pre-replicative modifications of bases was discovered in bacteriophages. Replacement of thymine by 5-hydroxymethyluracil was detected in the DNA of phages for *Bacillus subtilis* (Kallen et al., 1962). Since the discovery of modified DNA of phages, many other types of modifications have been found. The most common modifications are simple methylations of cytosine or adenine (Grosjean, 2009). Although less abundant, many exotic nucleobase modifications have been identified in bacteriophages and viruses (Gommers-Ampt and Borst, 1995; Warren, 1980). From biosynthesis to biological functions, modified bases have surprising features. It is well established that DNA modifications affect a wide range of biological processes. DNA modifications in prokaryotes are mainly used by R-M systems as a manner to distinguish between foreign or non-foreign DNA. Sinuously, bacteriophages use elaborate DNA modifications to evade prokaryotic defense mechanisms, such as R-M systems.

DNA modifications in eukaryotes

DNA modifications in eukaryotes play important roles in genetic regulations and have functions that vary from early development in embryonic cells to cellular maintenance pathways in adults. During cellular development the DNA methylation patterns change as a result of dynamic processes involving *de novo* methylation and demethylation. As a result, differentiated cells establish a unique and stable DNA methylation pattern that regulates tissue-specific gene transcription (Moore et al., 2013). 5-methylcytosine formation is catalyzed by a

family of DNA methyl transferases (DNMTs) that transfer a methyl group from S-adenyl methionine to the fifth carbon of cytosine in a post-replicative process and is regarded as a stable, highly heritable mark. Methylation of cytosines predominantly occurs in the dinucleotide CpG, often localized in so-called CpG islands, and is used as a transcriptional silencing mechanism (Bonasio et al., 2010; Jones, 2001). Differences in gene expression are highly correlated with alteration in the epigenetic landscape. The dynamic nature of methylation patterns is realized by both active methylation and demethylation however no enzymes are known that can cleave a methyl group from 5-mC. The recent discovery of 5-hydroxymethylcytosine (5-hmC) in eukaryotes has led to the hypothesis that 5-hmC serves as an intermediate in the demethylation of 5mC in a pathway that involves further oxidation from 5-hmC to 5-formylcytosine (5fC) and 5-carboxylcytosine (5-caC) (Branco et al., 2011). Oxidation of 5mC to 5-hmC is catalyzed by ten-eleven translocation (Tet) enzymes (Tahiliani et al., 2009). 5-hmC can be further oxidized to 5fC and then to 5caC by Tet enzymes or alternatively 5-hmC can be deaminated by activation-induced cytidine deaminase (AID)/apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like (APOBEC) enzyme complexes to 5-hydroxymethyluracil (5-hmU) (Guo et al., 2011). Resulting modified residues can be cleaved off by thymine DNA glycosylase (TDG) in the base excision repair (BER) pathway (Cortellino et al., 2011; He et al., 2011). Efforts to quantify the levels of 5-hmC have resulted in several extremely sensitive methods to detect low levels in various tissues (Chen et al., 2017; Shahal et al., 2014; Yuan et al., 2016). It is not fully understood whether 5-hmC merely acts as an intermediate in demethylation pathways or has a distinct regulatory role but its importance in epigenetics made 5-hmC to be considered as the sixth base in higher eukaryotes (Münzel et al., 2011).

DNA modifications in prokaryotes

Modified bases were first identified in the DNA of bacteria and for a few decades the only known functional roles were in host defense against bacteriophages. Later additional roles of DNA methylation were identified when inactivation of the methyl-directed mismatch repair system in *E. coli* was shown to cause a strong mutator phenotype (Modrich and Lahue, 1996). Multiple adenine methylation-dependent regulatory systems were later identified in various bacterial species(Casadesus and Low, 2006; Wion and Casadesús, 2006). Extensive analysis of the genomes of 230 diverse bacterial and archaeal species revealed that DNA methylation is observed in 93% of analyzed genomes but strikingly 48% of organisms harbor active MTases with no apparent cognate restriction enzyme (Blow et al., 2016). Interestingly, these 'orphan' MTase are more strictly conserved across prokaryotic phyla than R-M system-associated MTases (Blow et al., 2016). A well-studied example of an orphan MTase that plays an essential role in regulating epigenetic circuits is the DNA adenine methylase (Dam) in *E. coli*. Dam moves along the genomic DNA by linear diffusion and specifically methylates 5'-GATC-3' sequences by transferring a methyl group to the exocyclic N6 nitrogen of adenine (Urig et al., 2002). 5'-GATC-3' sequences that are methylated on both strands are converted into two hemimethylated DNA duplexes after DNA replication which provides a signal that DNA replication just has occurred and plays a role in diverse cellular processes (Casadesus and Low, 2006; Marinus and Casadesus, 2009). It has remained unknown how common

methylation-associated regulatory pathways are and also it is unknown if R-M-associated methylases can play a regulatory role.

DNA modifications in bacteriophages

Modifications of nucleobases are not only found in eukaryotes but also in prokaryotes and their viruses. While the functional roles of DNA modifications in eukaryotes is mostly attributed to genetic regulation, DNA modifications in prokaryotes and viruses are primarily related to processes the differentiate DNA between self and non-self. One well-studied DNA modification is β-glucosyl-5-hydroxymethylcytosine (5ghmC), a base modification that is found in T4. T-even bacteriophages were instrumental in the first formulations of many fundamental biological concepts, however early efforts to clone T4 genes were largely thwarted by the inability of most restriction endonucleases to cleave 5-ghmC containing DNA (Miller et al., 2003). T4 mutants that contain cytosine instead of 5-ghmC were constructed (Snyder et al., 1976)which allowed the construction of an detailed restriction map of T4 and further advancing of the field of molecular biology (Carlson and Nicolaisen, 1979; Snyder et al., 1976). Already in the 1950s it was discovered that T4 degrades the host genome and shuttles free dCTP though a protein complex that produces 5-hydroxymethyl-deoxycytidine triphosphate (5-hmdCTP). This complex, named dNTP synthesis complex (DSC) consists of at least 8 phage encoded proteins and 2 host encoded proteins. 5-hmdCTP is used by the DNA polymerase and other proteins involved in DNA synthesis and is further glucosylated by T4 glucosyltransferases. Approximately 30% of 5-hmC is glucosylated by α-glucosyltransferases and the remaining 70% is glucosylated by β-glucosyltransferases. Interestingly, glucosylation seems to be not completely sequence independent and α-glucosylation is hindered by neighboring 5-hmC residues (de Waard et al., 1967). Although glucosylation is a post-replicative process, it is shown to be tightly coupled to replication (McNicol and Goldberg, 1973).

Substitution of thymine by 5-hydroxymethyluracil (5hmU) is found in phage φe of *Bacillus subtilis* (Warren, 1980). Like 5-hmC synthesis in T4, the route of biosynthesis of 5hmU in φe ensures the exclusion of thymine from the DNA. Unlike T4, φe does not degrade the host DNA in order to shuttle breakdown products to 5hmU, meaning that deoxyribonucleotides in φe DNA are synthesized purely *de novo*. To limit incorporation of thymine in the DNA, Φe encodes a protein inhibitor of the host thymidylate synthase to channel dUMP towards 5hmdUMP, which is further phosphorylated to 5hmdUTP (Roscoe and Tucker, 1966). Substitution of thymine by 5-hmU confers resistance to the restriction enzymes produced by the bacterial host (Samson et al., 2013).

Some bacteriophages that substitute nucleobases by modified bases further modify them through secondary modification, a feature commonly referred to as hypermodification. One example of a bacteriophage with hypermodified DNA is *Bacillus subtilis* phage SP10 in which 15-20 % of thymine is substituted by α-glutamylthymine (Gommers-Ampt and Borst, 1995; Witmer and Wiatr, 1985). Upon infection by SP10, host thymidylate synthase activity is reduced by 70% and is replaced by deoxyuridylate hydroxymethyltransferase activity followed by incorporation of 5-hmU in the DNA (Witmer and Dosmar, 1978). In a presumably post-

replicative process 5hmU is further modified to α-glutamylthymine. In *Delftia acidovorans* phage φW14, thymine is for 50% substituted by the hypermodified base 5-(4-aminobutylaminomethyl)uracil, also called α-putrescinylthymine (Miller et al., 1983, 1985)
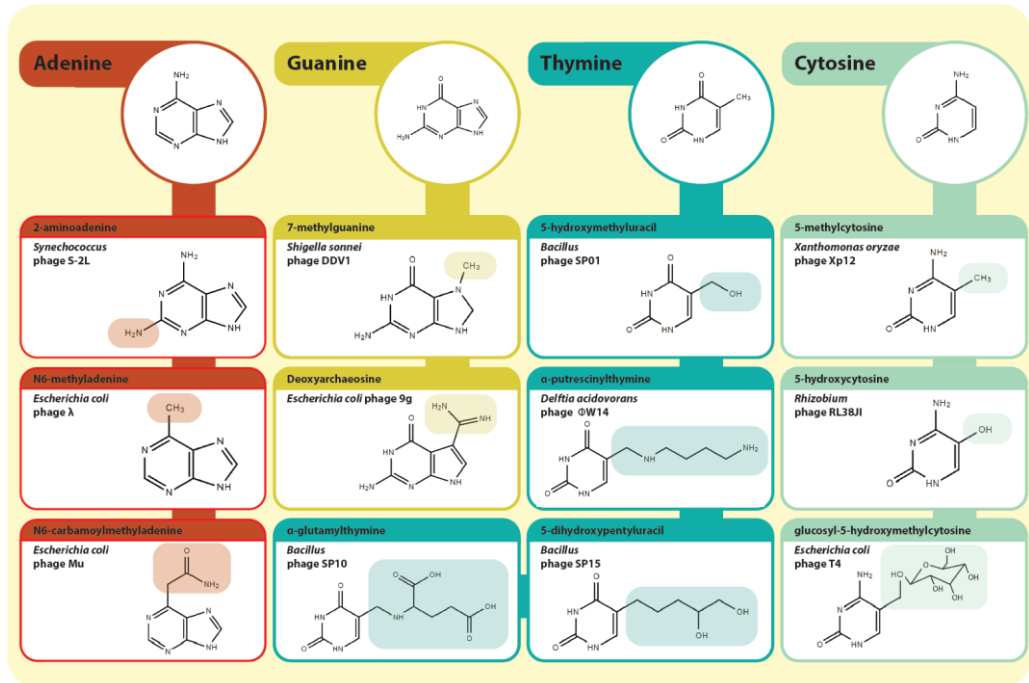

Figure 4. Modified bases of bacteriophages. Modifications are highlighted.

Roles of DNA modifications in bacteriophage DNA

The primary role of DNA modifications of bacteriophage DNA is to escape anti-viral defense systems in their hosts. Base modifications in virulent phages can serve two purposes. Firstly, base modifications prevent site specific recognition of DNA by restriction enzymes, thereby preventing cleavage. Secondly, base modifications serve to protect from phage-encoded nucleases that attack the unmodified host DNA. T4 encodes nucleases EndoII and EndoIV (gene product of *den*A and *den*B) that initiated degradation of cytosine-containing host DNA. EndoII is a sequence specific endonuclease that restricts 5'-CCGC-3' motifs by sequential nicking of both bottom and top-strands (Carlson and Kosturko, 1999). DNA fragments are further degraded to single nucleotides by 3'-5' exonuclease DexA and shuttled to the DSC (Warner et al., 1972). Not all bacteriophages that degrade host DNA protect their own DNA using base modifications: *E. coli* phages T5, T7, and *L. lactis* phage c6A all degrade host DNA but do not contain base modifications (Mitsunobu et al., 2014; Powell et al., 1992; Warner et al., 1975).

Beside the roles of DNA modifications in the arms race between prokaryotes and their viruses, base modifications also play key roles in packaging of DNA into the viral capsid. The hypermodified bases in the DNA of *Delftia acidovorans* phage φW14 facilitate the exceptionally high density of packaged DNA in the capsid (Benbasat, 1984). The hypermodified bases (α-putrescinylthymine) exhibit a decreased linear charge density compared to B-form DNA, resulting in compression of the DNA double helix. By increasing the packable length, φW14 base modifications allow more information to be carried on the genome. Besides influencing packaged DNA densities base modifications can have a role in the initiation of DNA packaging. Packaging of *E. coli* phage P1 is initiated after cleavage of the phage packaging site (*pac*) (Sternberg and Coulby, 1987). Cleavage of unmethylated *pac* is blocked and stalls initiation of DNA packaging and shows that the phage life cycle is regulated by base modifications (Sternberg and Coulby, 1988).

Single Molecule Localization Microscopy

Microscopy has provided the biological sciences with many insights in the inner workings of living organisms. From the time that Antoni van Leeuwenhoek and Robert Hooke started using a single lens to study simple biological samples, to current state of the art microscopy techniques, the field of microscopy has seen large progression. Most notable was the development of fluorescence microscopy. Already in 1852, Stokes discovered that substance specific fluorescence could be used to identify compounds (Stokes, 1852). This molecular feature could be captured using the first fluorescence microscopes, developed by Carl Zeiss in the 20th century. Not all biological components of interest have fluorescent properties. For this reason, biological substances, such as proteins and DNA, are often labeled with specific fluorophores. Intracellular distribution, and the biological processes in which they are involved, of fluorescently labeled molecules can be studied.
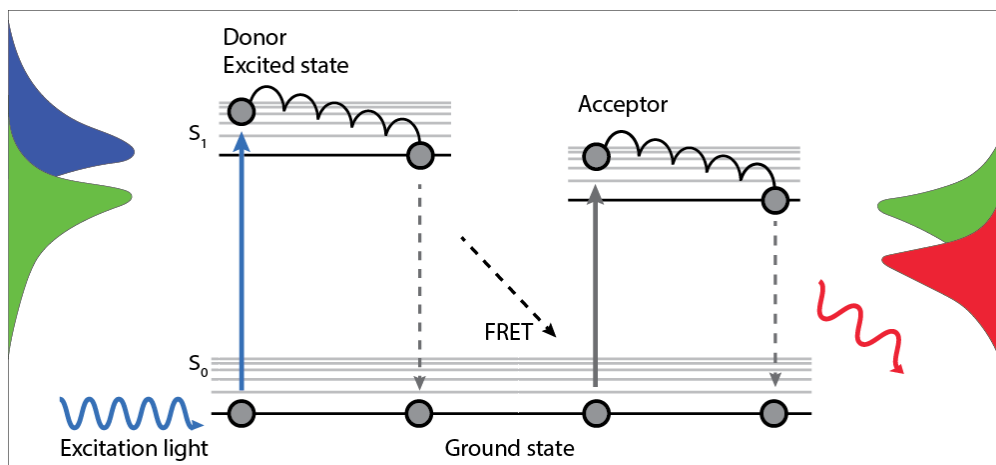
Förster resonance energy transfer

Figure 5. Energy transitions of FRET pairs. Excitation light can cause transition of a donor from the non-fluorescent ground state into an excited state. The excited donor can return to the ground state by either fluorescent emmission or by resonance energy transfer to the acceptor molecule.

Förster resonance energy transfer (FRET) is a technique to measure distances between two fluorescent molecules. A donor fluorophore that is excited by laser light may transfer its excitation energy to a nearby acceptor fluorophore in a non-radiative fashion. Non-radiative energy transfer occurs through long-range dipole-dipole interactions (Förster, 1959). Energy transfer decreases the lifetime of the excited state of the donor fluorophore but increases the fluorescence intensity of the donor fluorophore. The efficiency of this transfer (FRET efficiency) is dependent on the distance between the donor and acceptor fluorophores. FRET is a useful tool in molecular biology and provides an ideal probe of inter- and intramolecular distances in a range of 10-100Å. Conventional FRET measurements involve high concentrations of donor and acceptor pairs,

resulting in average values in which it is difficult to distinguish subpopulations. Recent developments in single molecule FRET (smFRET) have made it possible to single fluorophores under biological relevant conditions (Ha, 2001). Usually, molecules of interest are immobilized on a surface and acquired trajectories are analyzed using high throughput data analysis software (Rahul et al., 2008). smFRET can thus be used to resolve heterogeneous populations providing association and dissociation kinetics on binding constants.

PALM and STORM

Technical advances have greatly increased the resolution of microscopy images. However, achievable resolutions are limited by the diffraction limit (also called Abbe-limit) of the imaging light. Because of this diffraction limit, molecules with an inter-particle distance of less than 250 nm cannot be differentiated and thus are observed as one. The diffraction limit was considered an unbreakable rule for more than a century. In 1994, Hell and Wichmann introduced the stimulated emission depletion (STED) technique (Hell and Wichmann, 1994). By stimulating a small spot on the surface of a sample and depleting everything around that spot, Hell was able to excite individual fluorophores 65 nm apart from each other, thereby overcoming the diffraction limit. Similar techniques providing an even higher resolution, such as photo activated localization microscopy (PALM) and stochastic optical reconstruction microscopy (STORM) were later developed (Hess et al., 2006; Rust et al., 2006). Both PALM and STORM techniques use the particular ability of some fluorescent molecules to switch between a fluorescent ''on-state'' and dark ''off-state''. Fluorophores reside in a non-fluorescent dark state and can be stochastically activated using low intensity laser light. During that excitation, the molecule has a certain probability to switch back into the dark state or bleach, thereby losing its ability to fluoresce. Fluorophores in the dark state can be recovered to its ground state using short wavelength light. Photons emitted during the fluorescent phase are observed with a camera and a super resolution image can be fitted. STORM originally made use of fluorophore pairs but improvements enabled to directly switch only the fluorophore of interest (van de Linde et al., 2011). Later improvements involved alterations of the switching properties of dyes by changing the concentrations of reducing compounds and oxygen in the solution and thereby reducing the number of visible molecules, a technique called direct STORM (dSTORM) (Ehmann et al., 2014) Figure 6. PALM techniques use a similar imaging procedure as STORM, but instead of photo-switchable dyes they use photo-switchable proteins (Betzig et al., 2006; Hess et al., 2006).
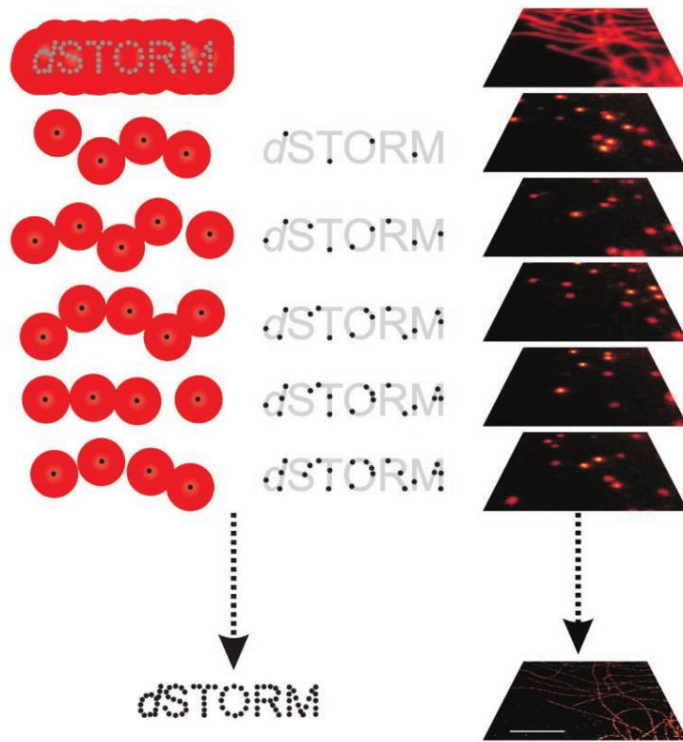
Figure 6. The *d*STORM concept. (van de Linde et al., 2011)

Single molecule localization microscopy (SMLM) provides incredible possibilities to study *in vivo* behavior of proteins and other cellular components. Next to information on diffusion, single-particle tracking reveals cellular localization patterns that cannot be identified using conventional microscopy and can reconcile molecular models with seemingly contradictory observations. For example, single-particle tracking of ribosomal subunits indicated that free subunit precursors are not excluded from the *E. coli* nucleoid, allowing for direct translation after transcription. This finding allowed reconciliation of the requirements of several gene-regulation mechanisms with previous observations that ribosomes are excluded from the nucleoid (Sanamrad et al., 2014). The unique opportunities of SMLM to provide insights in protein stoichiometry and spatial relations at a nanoscopic resolution have also been harnessed to study the molecular mechanisms of CRISPR-Cas systems. In a study on DNA targeting by crRNA-Cas9 complexes, proteins labeled with Q-dots were observed to find their target by a three-dimensional diffusion binding mechanism (Sternberg et al., 2014).

Thesis Outline

This thesis is about the interaction of bacteriophages and their hosts and aims to explore mechanistic details of host defense by CRISPR-Cas systems. The described studies address questions concerning physical phenomena of target search and binding by CRISPR complexes. Besides exploring the surveillance behavior of CRISPR-Cas components we investigated defense evasion mechanisms employed by bacteriophages. We found opportunity to exploit one of the discovered features of defense evasion and developed a widely customizable method for labeling macromolecular DNA.

In **chapter 2** we describe the mechanism of probing for targets by the Cascade complex in *E. coli*. The CRISPR-Cas type I-E system in *E. coli* encodes several Cas proteins of which five different proteins for a surveillance complex that binds target DNA based on complementarity with its crRNA. Using single molecule FRET binding, events of *bona fide* targets and mutated targets were visualized. We characterized the binding modes by analysis of FRET states and binding times. We developed a model explaining that the binding mode of Cascade leads to different functional outcomes.

In **chapter 3** we provide a perspective of target search by crRNA-effector complexes in a cellular environment. We describe how we have used single-particle tracking methods to analyze the surveillance behavior of CRISPR complexes and give insight in how these systems are able to provide protection from invaders in molecular crowded cellular environments. We show that these complexes are obligate ribonucleoprotein complexes that spend their time scanning the host nucleoid for PAM sites. Surprisingly, we found that type I-E Cascade complexes expel the large Cas8e subunit when encountering CRISPR arrays in order to avoid self-targeting, while type I-F Cascade complexes are oblivious to CRISPR arrays.

As described in the introduction, bacteriophages have developed diverse strategies to escape bacterial defense systems of which the production of anti-CRISPR proteins is one example. In an effort to identify novel anti-CRISPR proteins we isolated a number of *E. coli* phages and screened them for anti-CRISPR activity. In **chapter 4** we describe the isolation and genome sequencing of one of the screened phages, phage Ayreon. Even though anti-CRISPR activity could not be confirmed, the characterization of phage Ayreon is a valuable addition to the body of knowledge about bacteriophage biology.

In order to broaden our search for escape strategies employed by bacteriophages we investigated the effect of bacteriophage DNA modifications on CRISPR-Cas interference. In **chapter 5** we show that glucosyl-5-hydroxymethylation of cytosine in the DNA of T4 inhibits interference by both type I-E and type II-A CRISPR systems. CRISPR-Cas interference of a type V system is not affected by glucosyl-5-hydroxymethylation of DNA. Our results suggest that CRISPR-Cas systems have contributed to the selective pressure on phages to develop more generic solutions to escape sequence specific host defense systems.

The finding that glucosyl-5-hydroxymethylcytosine inhibits CRISPR-Cas interference raised the question if this feature could be used for analytical or biotechnological purposes. In **chapter 6** we demonstrate that introduction of several T4 genes into *E. coli* results in synthesis of modified DNA. We show that modified plasmid DNA can be specifically labeled with groups of interest, such as fluorophores. We demonstrate that single particle tracking of plasmid DNA can be used to study plasmid uptake and stability after transformation. We demonstrate that this method can be used to identify bottlenecks in efforts of genetic transformation of bacterial species

References

Ackermann, H.W. (2003). Bacteriophage observations and evolution. Res. Microbiol. *154*, 245–251.

Aizenman, E., Engelberg-Kulka, H., and Glaser, G. (1996). An Escherichia coli chromosomal addiction module regulated by 3′,5′-bispyrophosphate: a model for programmed bacterial cell death. Proc. Natl. Acad. Sci. U. S. A. *93*, 6059–6063.

Alicja Niewiadomska, J.K., Niewiadomska, A., and Klama, J. (2005). Polish Journal of Microbiology. Polish J. Microbiol. *54*, 43–48.

Arisaka, F., Kanamaru, S., Leiman, P., and Rossmann, M.G. (2003). The tail lysozyme complex of bacteriophage T4. Int. J. Biochem. Cell Biol. *35*, 16–21.

Atanasiu, C., Byron, O., McMiken, H., Sturrock, S.S., and Dryden, D.T. (2001). Characterisation of the structure of ocr, the gene 0.3 protein of bacteriophage T7. Nucleic Acids Res. *29*, 3059–3068.

Barrangou, R. CRISPR-Cas systems and RNA-guided interference. Wiley Interdiscip. Rev. RNA *4*, 267–278.

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.

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.

Benbasat, J.A. (1984). Condensation of Bacteriophage W 14 DNA of Varying Charge Densities by Trivalent Counterions. 3609–3619.

Betzig, E., Patterson, G.H., Sougrat, R., Lindwasser, O.W., Olenych, S., Bonifacino, J.S., Davidson, M.W., Lippincott-Schwartz, J., and Hess, H.F. (2006). Imaging Intracellular Fluorescent Proteins at Nanometer Resolution. Science (80-.). *313*, 1642–1645.

Bickle, T.A., and Kruger, D.H. (1993). Biology of DNA restriction. Microbiol Rev *57*, 434–450.

Bidnenko, E., Ehrlich, D., and Chopin, M.C. (1995). Phage operon involved in sensitivity to the Lactococcus lactis abortive infection mechanism AbiD1. J. Bacteriol. *177*, 3824–3829.

Bidnenko, E., Chopin, M.C., Ehrlich, S.D., and Anba, J. (2002). Lactococcus lactis AbiD1 abortive infection efficiency is drastically increased by a phage protein. FEMS Microbiol. Lett. *214*, 283–287.

Blow, M.J., Clark, T.A., Daum, C.G., Deutschbauer, A.M., Fomenkov, A., Fries, R., Froula, J., Kang, D.D., Malmstrom, R.R., Morgan, R.D., et al. (2016). The Epigenomic Landscape of Prokaryotes. PLoS Genet. *12*, 1–28.

Bonasio, R., Tu, S., and Reinberg, D. (2010). Molecular Signals of Epigenetic States. Science (80-.). *330*, 612–616.

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

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

Branco, M.R., Ficz, G., and Reik, W. (2011). Uncovering the role of 5-hydroxymethylcytosine in the epigenome. Nat. Rev. Genet. *13*, 7–13.

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*, 960–964.

Brüssow, H. (2009). The not so universal tree of life or the place of viruses in the living world. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *364*, 2263–2274.

Carlson, K., and Kosturko, L.D. (1999). Sequence-speci ® c cleavage by bacteriophage T4 endonuclease II in vitro. *31*, 1395–1405.

Carlson, K., and Nicolaisen, B. (1979). Cleavage Map of Bacteriophage T4 Cytosine-Containing DNA by Sequence-Specific Endonucleases SalI and KpnI. *31*, 112–123.

Casadesus, J., and Low, D. (2006). Epigenetic Gene Regulation in the Bacterial World. Microbiol. Mol. Biol. Rev. *70*, 830–856.

Chen, H., Wei, J.-R., Pan, J.-X., Zhang, W., Dang, F., Zhang, Z.-Q., and Zhang, J. (2017). Spectroscopic quantification of 5-hydroxymethylcytosine in genomic DNA using boric acid-functionalized nano-microsphere fluorescent probes. Biosens. Bioelectron. *91*, 328–333.

Chopin, M.C., Chopin, A., and Bidnenko, E. (2005). Phage abortive infection in lactococci: Variations on a theme. Curr. Opin. Microbiol. *8*, 473–479.

Christensen, S.K., Pedersen, K., Hansen, F.G., and Gerdes, K. (2003). Toxin-antitoxin loci as stress-response-elements: ChpAK/MazF and ChpBK cleave translated RNAs and are counteracted by tmRNA. J. Mol. Biol. *332*, 809–819.

Clokie, M.R.J., Millard, A.D., Letarov, A. V, and Heaphy, S. (2011). Phages in nature. 31–45.

Cortellino, S., Xu, J., Sannai, M., Moore, R., Caretti, E., Cigliano, A., Le Coz, M., Devarajan, K., Wessels, A., Soprano, D., et al. (2011). Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. Cell *146*, 67–79.

D'Herelle, F. (1917). An invisible microbe that is antagonistic to the dysentery bacillus.

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.

Dawkins, R. (1976). The selfish gene (New edition. Oxford ; New York : Oxford University Press, 1989.).

Demerec, M., and Fano, U. (1945). Bacteriophage-Resistant Mutants in Escherichia coli. Genetics *30*, 119.

Doudna, J.A., and Charpentier, E. (2014). The new frontier of genome engineering with CRISPR-Cas9. Science (80-.). *346*, 1258096–1258096.

Dulbecco, R. (1952). Mutual exclusion between related phages. J. Bacteriol. *63*, 209–217.

Dunn, D.B., and Smith, J.D. (1958). The occurrence of 6-methylaminopurine in deoxyribonucleic acids. Biochem. J. *68*, 627–636.

Ehmann, N., van de Linde, S., Alon, A., Ljaschenko, D., Keung, X.Z., Holm, T., Rings, A., DiAntonio, A., Hallermann, S., Ashery, U., et al. (2014). Quantitative super-resolution imaging of Bruchpilot distinguishes active zone states. Nat. Commun. *5*, 1–12.

Fineran, P.C., Gerritzen, M.J.H., Suarez-Diez, M., Kunne, T., Boekhorst, van Hijum, F.T., Staals, R.H.J., Brouns, S.J.J., Suárez-Diez, M., Künne, T., et al. (2014). Degenerate target sites mediate rapid primed CRISPR adaptation. Proc. Natl. Acad. Sci. U. S. A. *111*, E1629--38.

Forde, A., and Fitzgerald, G.F. (1999). Bacteriophage defence systems in lactic acid bacteria. Antonie van Leeuwenhoek, Int. J. Gen. Mol. Microbiol. *76*, 89–113.

Förster, T. (1959). 10Th Spiers Memorial Lecture. Discuss. Faraday Soc. *27*, 7–17.

Garneau, J.E., and Moineau, S. (2011). Bacteriophages of lactic acid bacteria and their impact on milk fermentations. Microb. Cell Fact. *10*, S20.

Georgiou, T., Yu, Y.N., Ekunwe, S., Buttner, M.J., Zuurmond, a, Kraal, B., Kleanthous, C., and Snyder, L. (1998). Specific peptide-activated proteolytic cleavage of Escherichia coli elongation factor Tu. Proc. Natl. Acad. Sci. U. S. A. *95*, 2891–2895.

Gerdes, K., Larsen, J.E.L., and Molin, S. (1985). Stable inheritance of plasmid R1 requires two different loci. J. Bacteriol. *161*, 292–298.

Gerdes, K., Bech, F.W., Jørgensen, S.T., Løbner-Olesen, A., Rasmussen, P.B., Atlung, T., Boe, L., Karlstrom, O., Molin, S., and von Meyenburg, K. (1986). Mechanism of postsegregational killing by the hok gene product of the parB system of plasmid R1 and its homology with the relF gene product of the E. coli relB operon. EMBO J. *5*, 2023–2029.

Gerdes, K., Poulsen, L.K., Thisted, T., Nielsen, A.K., Martinussen, J., and Andreasen, P.H. (1990). The hok killer gene family in gram-negative bacteria. New Biol. *2*, 946–956.

Glacer, L., Lonesco, H. & Shaefer, P. (1966). Teichoic acids as components of a specific phage receptor in Bacillus subtilis. *124*, 415–417.

Goldberg, E.B., and Dharmalingam, K. (1976). 1976 Nature Publishing Group. Nature *263*, 326–328.

Goldberg, E.B., and Dharmalingam, K. (1979). Restriction in Vivo. *411*, 404–411.

Goldfarb, T., Sberro, H., Weinstock, E., Cohen, O., Doron, S., Charpak-Amikam, Y., Afik, S., Ofir, G., Sorek, R., Bikard, D., et al. (2015). BREX is a novel phage resistance system widespread in microbial genomes. EMBO J. *34*, 169–183.

Gommers-Ampt, J.H., and Borst, P. (1995). Hypermodified bases in DNA. FASEB J. *9*, 1034–1042.

Grosjean, H. (2009). Nucleic acids are not boring long polymers of only four types of nucleotides: a guided tour. DNA RNA Modif. Enzym. Struct. Mech. Funct. Evol. 1–18.

Guo, J.U., Su, Y., Zhong, C., Ming, G.L., and Song, H. (2011). Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. Cell *145*, 423–434.

H.Kruger, D., J.Barcakl, G., Reuter, M., and O.Smith, H. (1979). EcoRII can be activated to cleave refractory DNA recognition sites. *6*, 715–732.

Ha, T. (2001). Single-Molecule Fluorescence Resonance Energy Transfer. Methods *25*, 78–86.

Hantke, K. (1978). Major outer membrane proteins of E. coli K12 serve as receptors for the phages T2 (protein Ia) and 434 (protein Ib). Mol. Gen. Genet. *164*, 131–135.

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*, 227–234.

He, Y.-F., Li, B.-Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., et al. (2011). Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. Science (80-.). *333*, 1303–1307.

Hell, S.W., and Wichmann, J. (1994). Stimulated-Emission-Depletion Fluorescence Microscopy. Opt. Lett. *19*, 780–782.

Hess, S.T., Girirajan, T.P.K., and Mason, M.D. (2006). Ultra-High Resolution Imaging by Fluorescence Photoactivation Localization Microscopy. Biophys. J. *91*, 4258–4272.

Ho, T.D., and Slauch, J.M. (2001). OmpC is the receptor for gifsy-1 and gifsy-2 bacteriophages of salmonella. J. Bacteriol. *183*, 1495–1498.

Ho, C.-H., Wang, H.-C., Ko, T.-P., Chang, Y.-C., and Wang, A.H.-J. (2014). The T4 Phage DNA Mimic Protein Arn Inhibits the DNA-binding Activity of the Bacterial Histone-like Protein H-NS. J. Biol. Chem.

Hoffman, L.M., Haskins, D.J., and Jendrisak, J. (2002). TypeOne™ Inhibitor Improves Transformation Efficiencies by Blocking Type I Restriction and Modification Systems In Vivo. Epic. Forum *9*, 8.

Horvath, P., and Barrangou, R. (2010). CRISPR/Cas, the immune system of bacteria and archaea. Science *327*, 167–170.

Houte, S. van, Ekroth, A.K.E., Broniewski, J.M., Chabas, H., Ashby, B., Gandon, S., Mike Boots, S.P., Buckling, A.J., and Westra, E.R. (2016). The diversity-generating benefits of a prokaryotic adaptive immune system. Nature *532*, 385–388.

Iida, S., Streiff, M.B., Bickle, T.A., and Arber, W. (1987). Two DNA antirestriction systems of bacteriophage P1, darA, and darB: characterization of darA- phages. Virology *157*, 156–166.

Johnson, T.B., and Coghill, R.D. (1925). RESEARCHES ON PYRIMIDINES. C111. THE DISCOVERY OF 5-METHYL-CYTOSINE IN TUBERCULINIC ACID, THE NUCLEIC ACID OF THE TUBERCLE BACILLUS [1]. J. Am. Chem. Soc. *47*, 2838–2844.

Jones, P.A. (2001). The Role of DNA Methylation in Mammalian Epigenetics. Science (80-.). *293*, 1068–1070.

Jore, M.M., Lundgren, M., van Duijn, E., Bultema, J.B., Westra, E.R., Waghmare, S.P., Wiedenheft, B., Pul, U., Wurm, R., Wagner, R., et al. (2011). Structural basis for CRISPR RNA-guided DNA recognition by Cascade. Nat. Struct. Mol. Biol. *18*, 529–536.

Kallen, R.G., Simon, M., and Marmur, J. (1962). The occurrence of a new pyrimidine base replacing thymine in a bacteriophage DNA: 5-hydroxymethyl uracil. J. Mol. Biol. *5*, 248–250.

Koonin, E. V, and Makarova, K.S. (2013). CRISPR-Cas: evolution of an RNA-based adaptive immunity system in prokaryotes. RNA Biol. *10*, 679–686.

Kühnlein, U., and Arber, W. (1972). Host specificity of DNA produced by Escherichia coli. J. Mol. Biol. *63*, 9–19.

Kunne, T., Kieper, S.N., Bannenberg, J.W., Depken, M., Suarez-diez, M., Brouns, S.J.J., Ku, T., Kieper, S.N., Bannenberg, J.W., Vogel, A.I.M., et al. (2016). Cas3-Derived Target DNA Degradation Fragments Fuel Primed CRISPR Adaptation Article Cas3-Derived Target DNA Degradation Fragments Fuel Primed CRISPR Adaptation. 1–13.

van de Linde, S., Löschberger, A., Klein, T., Heidbreder, M., Wolter, S., Heilemann, M., and Sauer, M. (2011). Direct stochastic optical reconstruction microscopy with standard fluorescent probes. Nat. Protoc. *6*, 991–1009.

Loenen, W.A.M., and Raleigh, E.A. (2014). The other face of restriction: Modification-dependent enzymes. Nucleic Acids Res. *42*, 56–69.

Lu, M.J., and Henning, U. (1994). Superinfection exclusion by T-even-type coliphages. Trends Microbiol. *2*, 137–139.

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

Marinus, M.G., and Casadesus, J. (2009). Roles of DNA adenine methylation in host-pathogen interactions: Mismatch repair, transcriptional regulation, and more. FEMS Microbiol. Rev. *33*, 488–503.

Marraffini, L.A., and Sontheimer, E.J. (2010). CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. Nat. Rev. Genet. *11*, 181–190.

Marraffini, L. a, and Sontheimer, E.J. (2008). CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science *322*, 1843–1845.

Marti, R., Zurfluh, K., Hagens, S., Pianezzi, J., Klumpp, J., and Loessner, M.J. (2013). Long tail fibres of the novel broad-host-range T-even bacteriophage S16 specifically recognize Salmonella OmpC. Mol. Microbiol. *87*, 818–834.

McNicol, L.A., and Goldberg, E.B. (1973). Immunochemical Characterization of Glucosylation in Bacteriophage-T4. J. Mol. Biol. *76*, 285–301.

Meisel, a, Bickle, T. a, Krüger, D.H., and Schroeder, C. (1992). Type III restriction enzymes need two inversely oriented recognition sites for DNA cleavage. Nature *355*, 467–469.

Meyer, J.R., Dobias, D.T., Weitz, J.S., Barrick, J.E., Ryan, T., and Lenski, R.E. (2012). Repeatability and Contingency in the Evolution of a Key Innovation in Phage Lambda. *335*, 428–432.

Miller, E.S., Kutter, E., Mosig, G., Kunisawa, T., Rüger, W., Arisaka, F., and Ru, W. (2003). Bacteriophage T4 Genome Bacteriophage T4 Genome †. Microbiol. Mol. Biol. Rev. *67*, 86–156.

Miller, P.B., Scraba, D.G., Leyritz-Wills, M., Maltman, K.L., and Warren, R. a. J. (1983). Formation and Possible Functions of alpha-Putrescinylthymine in Bacteriophage W-14 DNA. J. Virol. *47*, 399–405.

Miller, P.B., Wakarchuk, W.W., and Warren, R.A.J. (1985). IRL Press Limited , Oxford , England . Nucleic Acids Research. Nucleic Acids Res. *11*, 1475–1489.

Mitsunobu, H., Zhu, B., Lee, S.J., Tabor, S., and Richardson, C.C. (2014). Flap endonuclease activity of gene 6 exonuclease of bacteriophage T7. J. Biol. Chem. *289*, 5860–5875.

Modrich, P., and Lahue, R. (1996). Mismatch repair in replication fidelity, genetic recombination, and cancer biology. Annu. Rev. Biochem. *65*, 101–133.

Molineux, P.G. and I.J. (2008). Is phage DNA "injected" into cells - biologists and physicists can agree. *10*, 401–409.

Mondigler, M., Ayoub, A.T., and Heller, K.J. (2006). The DNA region of phage BF23 encoding receptor binding protein and receptor blocking lipoprotein lacks homology to the corresponding region of closely related phage T5. J. Basic Microbiol. *46*, 116–125.

Moore, L.D., Le, T., and Fan, G. (2013). DNA Methylation and Its Basic Function. Neuropsychopharmacology *38*, 23–38.

Münzel, M., Globisch, D., and Carell, T. (2011). 5-Hydroxymethylcytosine, the Sixth Base of the Genome. Angew. Chemie - Int. Ed. *50*, 6460–6468.

Murray, N.E. (2000). Type I Restriction Systems: Sophisticated Molecular Machines (a Legacy of Bertani and

Weigle). Microbiol. Mol. Biol. Rev. *64*, 412–434.

Nanda, A.M., Heyer, A., Krämer, C., Grünberger, A., Kohlheyer, D., and Frunzke, J. (2014). Analysis of SOS-induced spontaneous prophage induction in Corynebacterium glutamicum at the single-cell level. J. Bacteriol. *196*, 180–188.

Ofir, G., Melamed, S., Sberro, H., Mukamel, Z., Silverman, S., Yaakov, G., Doron, S., and Sorek, R. (2017). DISARM is a widespread bacterial defence system with broad anti-phage activities. Nat. Microbiol. 1–9.

Ogura, T., and Hiraga, S. (1983). Mini-F plasmid genes that couple host cell division to plasmid proliferation. Proc. Natl. Acad. Sci. U. S. A. *80*, 4784–4788.

Oliveira, P.H., Touchon, M., and Rocha, E.P.C. (2014). The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. Nucleic Acids Res. 1–14.

Paez-Espino, D., Sharon, I., Morovic, W., Stahl, B., Thomas, B.C., Barrangou, R., Banfield, J.F., and Banfielda, J.F. (2015). CRISPR Immunity Drives Rapid Phage Genome Evolution in Streptococcus thermophilus. MBio *6*, e00262-15-.

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

Pawluk, A., Amrani, N., Zhang, Y., Garcia, B., Hidalgo-Reyes, Y., Lee, J., Edraki, A., Shah, M., Sontheimer, E.J., Maxwell, K.L., et al. (2016b). Naturally Occurring Off-Switches for CRISPR-Cas9. Cell *167*, 1829–1838.e9.

Pecota, D.C., and Wood, T.K. (1996). Exclusion of T4 phage by the hok/sok killer locus from plasmid R1. J. Bacteriol. *178*, 2044–2050.

Pingoud, A., and Jeltsch, A. (2001). Structure and function of type II restriction endonucleases. Nucleic Acids Res. *29*, 3705–3727.

Powell, I.B., Tulloch, D.L., Hillier, A.J., and Davidson, B.E. (1992). Phage DNA synthesis and host DNA degradation in the life cycle of Lactococcus lactis bacteriophage c6A. J. Gen. Microbiol. 945–950.

Pupo, E., Lindner, B., Brade, H., and Schromm, A.B. (2013). Intact rough- and smooth-form lipopolysaccharides from Escherichia coli separated by preparative gel electrophoresis exhibit differential biologic activity in human macrophages. FEBS J. *280*, 1095–1111.

Rahul, R., Hohng, S., and Ha, T. (2008). A Practical Guide to Single Molecule FRET. Nat. Methods *5*, 507–516.

Rakieten, M.L., and Rakieten, T.L. (1937). Relationships between Staphylococci and Bacilli Belonging to the Subtilis Group as Shown by Bacteriophage Absorption. J. Bacteriol. *34*, 285–300.

Rauch, B.J., Silvis, M.R., Hultquist, J.F., Waters, C.S., McGregor, M.J., Krogan, N.J., and Bondy-Denomy, J. (2017). Inhibition of CRISPR-Cas9 with Bacteriophage Proteins. Cell *168*, 150–158.e10.

Ricci, V., and Piddock, L.J V (2010). Exploiting the role of TolC in pathogenicity: Identification of a bacteriophage for eradication of Salmonella serovars from poultry. Appl. Environ. Microbiol. *76*, 1704–1706.

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*, 1805–1812.

Roberts, R.J., Vincze, T., Posfai, J., and Macelis, D. (2015). REBASE-a database for DNA restriction and modification: Enzymes, genes and genomes. Nucleic Acids Res. *43*, D298–D299.

Roscoe, D.H., and Tucker, R.G. (1966). The Biosynthesis of SHydroxymethyldeoxyuridylic Bacillus Acid subtilis in Bacteriophage-Infected. *166*, 157–166.

Rust, M.J., Bates, M., and Zhuang, X.W. (2006). Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). Nat Methods *3*, 793–795.

Sain, B., and Murray, N.E. (1980). The hsd (host specificity) genes of E. coli K 12. Mol. Gen. Genet. *180*, 35–46.

Sanamrad, A., Persson, F., Lundius, E.G., Fange, D., Gynnå, A.H., and Elf, J. (2014). Single-particle tracking reveals that free ribosomal subunits are not excluded from the Escherichia coli nucleoid. Proc. Natl. Acad. Sci. U. S. A. *111*, 11413–11418.

São-José, C., Lhuillier, S., Lurz, R., Melki, R., Lepault, J., Santos, M.A., and Tavares, P. (2006). The ectodomain of the viral receptor YueB forms a fiber that triggers ejection of bacteriophage SPP1 DNA. J. Biol. Chem. *281*,

11464–11470.

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.

Schrödinger, E. (1967). Mind and matter. What Is Life? Phys. Asp. Living Cell 90–165.

Shahal, T., Gilat, N., Michaeli, Y., Redy-Keisar, O., Shabat, D., and Ebenstein, Y. (2014). Spectroscopic quantification of 5-hydroxymethylcytosine in genomic DNA. Anal. Chem. *86*, 8231–8237.

Shipman, S.L., Shipman, S.L., Nivala, J., Macklis, J.D., and Church, G.M. (2016). Molecular recordings by directed CRISPR spacer acquisition. *1175*, 1–16.

Shmakov, S., Smargon, A., Scott, D., Cox, D., Pyzocha, N., Yan, W., Abudayyeh, O.O., Gootenberg, J.S., Makarova, K.S., Wolf, Y.I., et al. (2017). Diversity and evolution of class 2 CRISPR–Cas systems. Nat. Rev. Microbiol.

Snyder, L., Gold, L., and Kutter, E. (1976). A gene of bacteriophage T4 whose product prevents true late transcription on cytosine-containing T4 DNA. Proc. Natl. Acad. Sci. U. S. A. *73*, 3098–3102.

Sorek, R., Lawrence, C.M., and Wiedenheft, B. (2013). CRISPR-mediated adaptive immune systems in bacteria and archaea. Annu Rev Biochem *82*, 237–266.

Sternberg, N., and Coulby, J. (1987). Recognition and Cleavage of the Bacteriophage Pl Packaging Site Qm) I . Differential Processing of the Cleaved Ends in Vivo. 469–479.

Sternberg, N., and Coulby, J.N. (1988). Processing of the bacteriophage P1 packaging site (pac) is regulated by adenine methylation. Gene *74*, 203.

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*, 62–67.

Stokes, G.G.. (1852). On the Change of Refrangibility of Light. Philos. Trans. R. Soc. London *6*, 195–200.

Stotz, K. (2004). Gordon Graham: Genes: A Philosophical Inquiry,. Isis *95*, 172–173.

Sutherland, I.W. (1995). Polysaccharide lyases. FEMS Microbiol. Rev. *16*, 323–347.

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*.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., et al. (2009). Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. Science (80-.). *324*, 930–935.

Thomas, M., White, R.L., and Davis, R.W. (1976). Hybridization of RNA to double-stranded DNA: formation of R-loops. Proc. Natl. Acad. Sci. U. S. A. *73*, 2294–2298.

Tock, M.R., and Dryden, D.T.F. (2005). The biology of restriction and anti-restriction. Curr. Opin. Microbiol. *8*, 466–472.

Urig, S., Gowher, H., Hermann, A., Beck, C., Fatemi, M., Humeny, A., and Jeltsch, A. (2002). The Escherichia coli dam DNA methyltransferase modifies DNA in a highly processive reaction. J. Mol. Biol. *319*, 1085–1096.

van Valen, L. (1973). A new evolutionary law. Evol. Theory *1*, 1–30.

de Waard, a, Ubbink, T.E., and Beukman, W. (1967). On the specificity of bacteriophage-induced hydroxymethylcytosine glucosyltransferases. II. Specificities of hydroxymethylcytosine alphaand beta-glucosyltransferases induced by bacteriophage T4. Eur. J. Biochem. *2*, 303–308.

Walkinshaw, M.D., Taylor, P., Sturrock, S.S., Atanasiu, C., Berge, T., Henderson, R.M., Edwardson, J.M., and Dryden, D.T.F. (2002). Structure of Ocr from bacteriophage T7, a protein that mimics b-form DNA. Mol. Cell *9*, 187–194.

Warner, H.R., Snustad, D.P., Koerner, J.F., and Childs, J.D. (1972). Identification and genetic characterization of mutants of bacteriophage T4 defective in the ability to induce exonuclease A. J. Virol. *9*, 399–407.

Warner, H.R., Drong, R.F., and Berget, S.M. (1975). Early events after infection of Escherichia coli by bacteriophage T5. Induction of a 5′-nucleotidase activity and excretion of free bases. J. Virol. *15*, 273–280.

Warren, R. a (1980). Modified bases in bacteriophage DNAs. Annu. Rev. Microbiol. *34*, 137–158.

Washizaki, A., Yonesaki, T., and Otsuka, Y. (2016). Characterization of the interactions between Escherichia coli receptors, LPS and OmpC, and bacteriophage T4 long tail fibers. Microbiologyopen.

Westra, E.R., Swarts, D.C., Staals, R.H.J., Jore, M.M., Brouns, S.J.J., and van der Oost, J. (2012). The CRISPRs,

they are a-changin': how prokaryotes generate adaptive immunity. Annu. Rev. Genet. *46*, 311–339.

Whitehead, H.R., and Cox, G.A. (1935). The occurrence of bacteriophage in cultures of lactic streptococci: A preliminary note (Government Printer).

Wiedenheft, B., Sternberg, S.H., and Doudna, J. a. (2012). RNA-guided genetic silencing systems in bacteria and archaea. Nature *482*, 331–338.

Wigington, C.H., Sonderegger, D., Brussaard, C.P.D., Buchan, A., Finke, J.F., Fuhrman, J.A., Lennon, J.T., Middelboe, M., Suttle, C.A., Stock, C., et al. (2016). Re-examination of the relationship between marine virus and microbial cell abundances. Nat. Microbiol. 15024.

Wion, D., and Casadesús, J. (2006). N6-methyl-adenine: an epigenetic signal for DNA–protein interactions. Nat. Rev. Microbiol. *4*, 183–192.

Witmer, H., and Dosmar, M. (1978). Synthesis of 5-hydroxymethyldeoxyuridine triphosphate in extracts of SP10 c phage-infected Bacillus subtilis W23. Curr. Microbiol. *1*, 289–292.

Witmer, H., and Wiatr, C. (1985). Polymer-Level Synthesis of Oxopyrimidine Deoxynucleotides by Bacillus subtilis Phage SPlO : Characterization of Modification-Defective Mutants. *53*, 522–527.

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.

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.

Young, F.E. (1967). Requirement of Glucosylated Teichoic Acid for adsorption of phage in Bacillus subtillis 168. Pnas *58*, 2377–2384.

Yuan, F., Zhang, X.-H., Nie, J., Chen, H.-X., Zhou, Y.-L., and Zhang, X.-X. (2016). Ultrasensitive determination of 5-methylcytosine and 5-hydroxymethylcytosine in genomic DNA by sheathless interfaced capillary electrophoresis-mass spectrometry. Chem. Commun. *52*, 2698–2700.

Chapter 2

Two Distinct DNA Binding Modes Guide Dual Roles of a CRISPR-Cas Protein Complex

Timothy R. Blosser[1,3,5], Luuk Loeff[1,5], Edze R. Westra[2,4], Marnix Vlot[2], Tim Künne[2], Małgorzata Sobota[2], Cees Dekker[1], Stan J.J. Brouns[2], and Chirlmin Joo[1]

[1]Kavli Institute of NanoScience and Department of BioNanoScience, Delft University of Technology, 2628 CJ, Delft, The Netherlands
[2]Laboratory of Microbiology, Department of Agrotechnology and Food Sciences, Wageningen University, 6703 HB, Wageningen,the Netherlands
[3]Present address: Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA
[4]Present address: The School of Biosciences, University of Exeter, Penryn, Cornwall TR10 9EZ, UK

Abstract

Small RNA-guided protein complexes play an essential role in CRISPR-mediated immunity in prokaryotes. While these complexes initiate interference by flagging cognate invader DNA for destruction, recent evidence has implicated their involvement in new CRISPR memory formation, called priming, against mutated invader sequences. The mechanism by which the target recognition complex mediates these disparate responses—interference and priming—remains poorly understood. Using single-molecule FRET, we visualize how bona fide and mutated targets are differentially probed by *E. coli* Cascade. We observe that the recognition of bona fide targets is an ordered process that is tightly controlled for high fidelity. Mutated targets are recognized with low fidelity, which is featured by short-lived and PAM- and seed-independent binding by any segment of the crRNA. These dual roles of Cascade in immunity with distinct fidelities underpin CRISPR- Cas robustness, allowing for efficient degradation of bona fide targets and priming of mutated DNA targets

Introduction

Clustered regularly interspaced short palindromic repeat (CRISPR) loci are widely spread throughout prokaryotic genomes and provide an inheritable RNA-guided adaptive immune system against bacteriophages and mobile genetic elements (Barrangou and Van der Oost, 2013; Charpentier and Marraffini, 2014; Fineran and Charpentier, 2012; Reeks et al., 2013; Samson et al., 2013; Sorek et al., 2013; Westra et al., 2014) In response to invading phages or mobile genetic elements, CRISPR-associated (Cas) proteins integrate small fragments of foreign DNA into the CRISPR array, which are subsequently processed into mature CRISPR RNAs (crRNAs). crRNAs form a complex with one Cas protein (Cas9 from type II) or multiple Cas proteins (types I and III), which utilizes the crRNA as a guide to trigger degradation of cognate invading nucleic acids. While it is DNA that is targeted in types I and II (van der Oost et al., 2014), recent studies suggest that both DNA and RNA are targeted in type III (Goldberg et al., 2014; Hale et al., 2014; Rouillon et al., 2013; Staals et al., 2013; Tamulaitis et al., 2014). Among the target recognition complexes, Cas9 has been widely applied as a versatile tool for genome engineering in a broad spectrum of organisms (Hsu et al., 2014; Terns and Terns, 2014). In the CRISPR-Cas/I-E system of Escherichia coli, mature crRNAs are incorporated into CRISPR-associated complex for antiviral defense (Cascade), an 11-subunit complex comprised of five different Cas proteins (Cse11, Cse22, Cas76, Cas51, and Cas61)(Jore et al., 2011)(Figure 1A). In the CRISPR interference pathway, Cascade generates an R-loop between the crRNA and its double-stranded DNA (dsDNA) target (protospacer), which subsequently leads to target degradation by the nuclease-helicase Cas3 (Mulepati and Bailey, 2013; Sinkunas et al., 2013; Westra et al., 2012a). The first 8 nt (with exception of the sixth nucleotide) of the protospacer, or ''seed'' region, must be a perfect match for efficient R-loop formation (Künne et al., 2014). Additionally, R-loop formation requires an immediately neighboring trinucleotide protospacer adjacent motif (PAM). This conserved PAM sequence at the seed end of the proto- spacer is recognized by the Cse1 subunit and is essential for the discrimination between targets and non-targets (Sashital et al., 2012; Westra et al., 2013). The mechanism by which Cascade finds its target among the vast amount of DNA in the cell remains elusive. It has been hypothesized that Cascade transiently associates with PAM sequences, interrogating neighboring sequences for a complementary seed, followed by directional R-loop formation (Künne et al., 2014). A recent single-molecule study has visualized the transient interactions of Cas9 with PAM-rich sequences in real time (Sternberg et al., 2014). Another study with Cascade and Cas9 has shown directional R-loop formation and how PAM and protospacer complementarity influences its stability (Szczelkun et al., 2014). However, it is yet to be shown how the stepwise interaction between PAM, seed, and protospacer is coordinated and how off-targeting is avoided during target recognition.
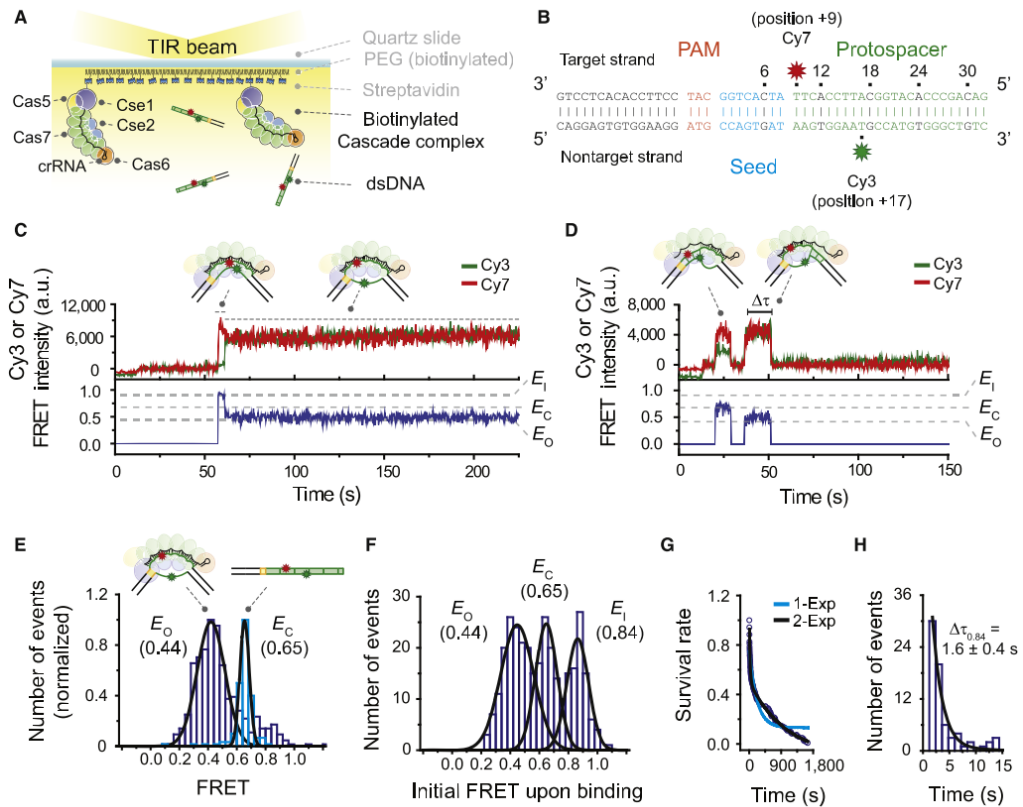
Figure 1. Two Binding Modes of Cascade Revealed by a Single-Molecule FRET Assay. (A) Schematic of a single-molecule FRET experiment used to monitor binding of Cascade to target DNA substrates. (B) The bona fide target construct consists of a 15 bp flank (black), a PAM(orange), and a protospacer (green), with its seed highlighted in blue. Cy7 (red star) was attached to position +9 of the target strand and Cy3 (green star) to position +17 of the nontarget strand. (C) A representative time trace of donor (Cy3, green) and acceptor (Cy7, red) fluorescence and corresponding FRET (blue) exhibiting the long-lived binding of the bona fide target. High FRET (~0.84, named E_I for FRET efficiency of an intermediate state) exhibited upon binding is followed by low FRET (~0.44, named E_O for FRET efficiency of an open state). DNA was added at time = 10 s. (D) A representative time trace exhibiting the short-lived binding of the bona fide target exhibits two FRET states (E_O~0.44 and E_C~0.65; E_C is for FRET efficiency of a closed state). The duration of each state is measured as the dwell time ($\Delta\tau$). DNA was added at time = 10 s. (E) The FRET distribution of the bona fide target DNA alone (light blue) or after equilibration with immobilized Cascade (purple) with peaks at E_C (0.65) and E_O (0.44), respectively (derived from Gaussian fit, black line). Data were obtained from five fields of view each. (F) A histogram of the initial FRET upon binding (average of first 1.5 s of each event) of the bona fide target exhibits three peaks at FRET = E_O (0.44), E_C (0.65), E_I (0.84) (derived from Gaussian fit, black line). (G) The survival rate of events that start at E_I (0.84) was fitted using a single (light blue color) and a double (black color) exponential curve. The double

exponential fit resulted in two characteristic times (25.9 and 1040 s). (H) The dwell time distribution of E_I (0.84) state of bona fide target binding with mean $\Delta\tau_{0.84}$ (derived from single exponential fit, black line). Error represents standard deviation (three individual data sets).

Recent in vivo studies have revealed an additional functionality of CRISPR-Cas immunity. When facing ''escape mutants,'' previously targeted sequences that bear mutations in their PAM and/or protospacer, Cascade initiates a response called priming wherein the CRISPR-Cas system acquires new spacer sequences from the mutant at an elevated rate to restore immunity (Datsenko et al., 2012; Fineran et al., 2014; Li et al., 2014; Richter et al., 2014). High-throughput plasmid loss assays of a randomized PAM and protospacer library have revealed that priming is a robust process, tolerating up to 13 mutations in the PAM and protospacer sequence (Fineran et al., 2014). Even though Cascade is essential for priming, its role in this process is poorly understood. Intriguingly, biochemical studies have shown that a single-point mutation in the PAM or seed sequence leads to a drastic decrease in the binding affinity of Cascade (Semenova et al., 2011). Therefore, it is puzzling how Cascade is able to associate with these mutated substrates despite its low affinity and, further, how it distinguishes these mutated substrates from bona fide targets to initiate priming.

Single-molecule fluorescence is a powerful tool for elucidating the intricate mechanistic details of complex protein-nucleic acid interactions (Ha, 2014; Joo et al., 2013; Juette et al., 2014; Robinson and van Oijen, 2013; Schuler and Hofmann, 2013). To dissect Cascade's two distinct functional roles, we developed a single-molecule FRET assay to monitor the interaction of Cascade with bona fide and mutated substrates. Real-time observation of Cascade-target interactions revealed that an initial recognition complex proceeds to a stable R-loop only if the crRNA makes an extensive match with the target. In addition to this ''canonical binding mode,'' we identified an alternative binding mode of Cascade that is triggered by partial complementarity to a target. Using an in vivo assay, we validated that this binding mode enables Cascade to probe mutated DNA substrates and consequently initiate priming.

Results

Single-Molecule Observation of Cascade Target Binding

For single-molecule measurements, Cascade was labeled with a biotin on the N terminus of its Cse1 subunit and immobilized to the surface of a microscope slide via a biotin-streptavidin linkage (Figure 1A). Dye-labeled dsDNA targets were added to the slide, and individual binding events were imaged in real time with a total internal reflection fluorescence (TIRF) microscope (Figure 1A). DNA constructs consisted of a protospacer, a PAM, and an additional 15 base pair flank (Figure 1B). The target strand (complementary to the crRNA) was labeled with an acceptor dye (Cy7) at protospacer position +9, whereas the nontarget strand was labeled with a donor dye (Cy3) at protospacer position +17. These labeling positions yielded a FRET value of ~0.65 (named

E_C for a FRET state which represents a closed conformation of dsDNA between nt 9 and 17) (Figure 1E) as measured by immobilization of the DNA alone (see Experimental Procedures and Table S1). Control experiments showed that dye labeling of the DNA at protospacer positions +9 and +17 did not appreciably affect the target binding reaction of Cascade (Figure S1F). We first explored Cascade's interaction with a bona fide target DNA, a substrate that triggers interference in vivo. This substrate contains a protospacer with perfect complementarity to the crRNA and an interference-permissive PAM (named ''interfering PAM'') (Figure 1B) (Fineran et al., 2014; Westra et al., 2012a). After equilibration of the DNA with the immobilized Cascade, the measured FRET distribution exhibited one major peak centered at 0.44 (named E_O for a FRET state which represents an open conformation of dsDNA between nucleotides 9 and 17), a decrease from the starting value of E_C (0.65) (Figure 1E). This decrease in FRET is consistent with the expected open DNA conformation resulting from R-loop formation upon Cascade binding. A similar decrease in FRET was observed upon exchanging the position of the donor and acceptor dyes (Figure S1C) or when Cascade was prebound to the DNA prior to immobilization (Figure S1D), indicating that the observed decrease in FRET was not due to a protein- or surface-induced photophysical effect.

Two Distinct Binding Modes of Cascade

Next we characterized the kinetics and structural dynamics of Cascade binding in real time by adding a bona fide target substrate to immobilized Cascade during data acquisition. Interestingly, time trajectories exhibited disparate binding events that varied in their dwell time and FRET value. The dwell time distribution followed a double-exponential decay curve (Figure 1G, a fit in black), suggesting heterogeneity in binding. A histogram of the initial FRET of binding events exhibited three distinct peaks (centered at EO [0.44], EC [0.65], and 0.84) (Figure 1F), which, combined with dwell time analysis, led us to divide the events into two distinct types. The first type of binding event initiated at a FRET of 0.84, persisted over the entire duration of our observation time (30 min) (Figure 1C), and was therefore considered to be irreversible over the timescale of our experiment (Figure 1G). Interestingly, events of this type did not remain at their initial FRET of 0.84, but exhibited a transition after 1.6 ± 0.4 s (Figure 1H) to a final FRET of 0.44 (Figure 1C). This observation is consistent with the single FRET peak centered at 0.44 (EO) observed at equilibrium (Figure 1E). The initial transient state (0.84, named EI for an initial transient state) may represent a target-recognition complex wherein the crRNA interacts with the dsDNA before full displacement of the nontarget strand (schematic, Figure 1C). Notably, the FRET of the initial state is higher than that of the DNA alone (EC, 0.65, Figure 1E), likely arising from a subtle conformational change of the dsDNA upon target recognition (e.g., twisting or bending) (Hochstrasser et al., 2014; Westra et al., 2012a). The observed transition (EI/EO) may represent a previously hypothesized locking process, wherein Cascade slides its Cse2 dimer toward its Cse1 subunit upon target recognition (Szczelkun et al., 2014; Wiedenheft et al., 2011a), ultimately resulting in the displacement of the nontarget strand and stable R-loop formation (schematic, Figure 1C). Taken together, considering Cascade's strong target association and

observed conformational change (Semenova et al., 2011; Szczelkun et al., 2014; Wiedenheft et al., 2011a), we interpret the first type of binding event to correspond to Cascade's canonical mode of target binding that leads to interference in vivo. We therefore refer to this event type as Cascade's interference mode of binding. Unlike the interference mode, the second type of binding event was short-lived (25.9 s, Figure 1G) and exhibited an initial FRET of either EO or EC (Figure 1D). These states were further distinguished from the interference mode, as they did not exhibit any kinetic intermediates, nor did they show transitions to other FRET states. As a substrate containing no complementarity (Mut[S1-6]) to the crRNA showed negligible binding (Figure S1E), we speculate that these short binding events (named ''noncanonical mode'') arise from sequence-specific interactions wherein the probed region of the target DNA is either opened up in a locally formed R-loop (EO) or remains closed (EC). To explore the origin of Cascade's disparate binding interactions, we first focused on the role of the PAM. We repeated our assay with a DNA substrate containing a point mutation in the PAM (Mut[PAM], Table S2) that represents one of the dominant mutant phenotypes of bacteriophages that escape CRISPR interference (Semenova et al., 2011) and subsequently trigger priming in vivo (Datsenko et al., 2012; Fineran et al., 2014). Notably, while Cascade was still able to interact with Mut[PAM], only binding events characteristic of its noncanonical mode were observed (Figure 2A). A histogram of the initial FRET of each event exhibited only two peaks, centered at EC and EO (Figure 2B), identical to the peak positions observed for the noncanonical binding mode (Figure 1F). In addition, the binding events observed for Mut[PAM] were short-lived, exhibiting a dwell time of 24.8 ± 8.9 s (Figure 2C), similar to that of the noncanonical binding mode (Figure 1G). These results indicate that Cascade's interaction with target substrates through its noncanonical binding mode does not require an interfering PAM. Given the results above, we hypothesize that the observed binding states represent two functional modes of Cascade. The first is the interference mode, in which Cascade binds a bona fide DNA target (i.e., interfering PAM and complementary protospacer) and triggers Cas3-mediated target degradation. The second is the priming mode (noncanonical mode), in which Cascade is able to associate with targets harboring a PAM mutation to initiate primed spacer acquisition.
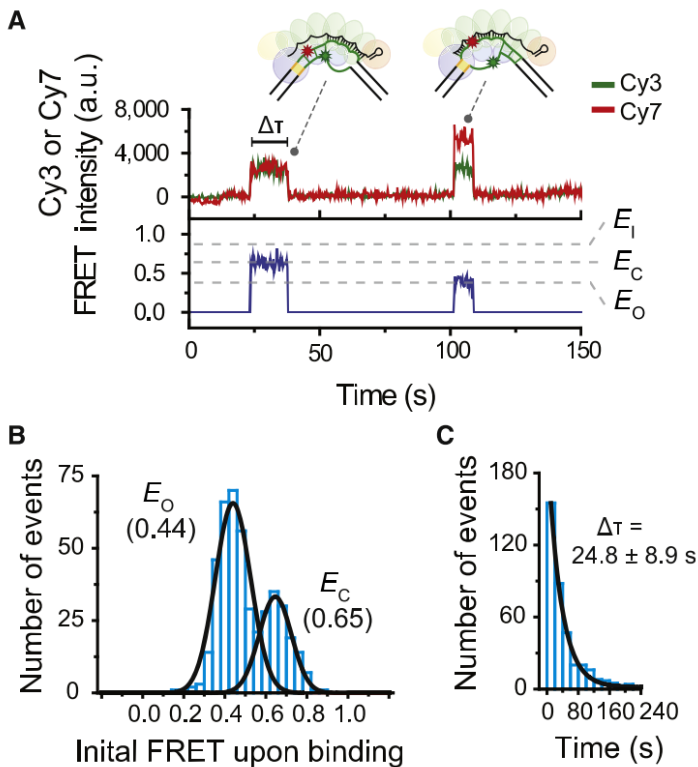
Figure 2. Short Binding of Cascade to PAM-Mutated Targets. (A) A representative time trace exhibiting the short-lived binding of the PAM- mutated target exhibits two FRET states, E_O (0.44) and E_O (0.65). The duration of each state is measured as the dwell time ($\Delta\tau$).DNA was added at time = 10 s. (B) A histogram of the initial FRET upon binding (average of first 1.5 s of each event) of Mut[PAM] exhibits peaks at E_O (0.44) and E_C (0.65) (derived from Gaussian fit, black line). (C) The dwell time distribution of Mut[PAM] binding events with mean $\Delta\tau\tau$ (derived from single exponential fit, black line). Error represents SD (three individual data sets).

Structural Elements of Two Distinct Binding Modes

To investigate the structural elements of Cascade's two different binding modes, we employed a series of target DNA substrates bearing mutations in their PAM and/or protospacer sequence(s). Recent studies have reported that base pairing between Cas- cade's crRNA and the protospacer occurs over five segments of 5 nt (segments 1–5) and one segment of 2 nt (segment 6) (Fineran et al., 2014; Jackson et al., 2014; Mulepati et al., 2014; Zhao et al., 2014) We therefore chose to systematically mutate the protospacer in segments, starting from either the PAM-proximal or PAM-distal end of the protospacer (Figures 3 and S2; Table S2). Upon mutation of the first segment of the protospacer (Mut[S1], Figure 3A), which comprises the majority of the seed region, the noncanonical binding mode persisted as binding events exhibited nearly identical FRET values and dwell times

to the Mut[PAM] targets (Figures 3B and 3C). The same was observed for a DNA substrate containing both the PAM and seed mutations (Mut[PAM+S1], Figures 3A–3C), indicating that the noncanonical binding mode is largely insensitive to the PAM and seed sequence. This observation is in stark contrast to the canonical binding mode, which requires both an intact seed sequence and an interfering PAM. Remarkably, when the first two PAM-proximal segments, including the entire seed, were mutated (Mut[S1-2]), the noncanonical binding mode was still evident, with initial FRET values centered at E_O or E_C and an average dwell time of 19.6 ± 0.4 s (Figures 3B, 3C, and S2A).

Intriguingly, when the first three (Mut[S1-3], Figure 3A) PAM- proximal segments were mutated, the binding events exhibited only one major initial FRET population centered at E_C, with an average dwell time of 10.5 ± 1.9 s (Figures 3B, 3C, and S2A), indicating that these events arise from sequence-specific interactions confined outside of the probed region of the protospacer (segments 4–6, Figure 3A). Removal of complementarity in the first four segments (Mut[S1-4], Figures 3A and 3B) or all segments (Mut[S1-6], Figure S1E) disrupted binding to background levels. Taken together, the series of PAM-proximal mutations indicate that the noncanonical binding mode of Cascade comprises sequence-specific interactions with a minimum requirement of three full segments for target recognition. The PAM-distal mutation series showed complementary behavior, consistent with the structural features of the non- canonical binding mode observed above (Figures 3D–3F). Upon mutation of the last two segments of the protospacer (Mut[S5-6]), the noncanonical binding mode persisted with two peaks centered at E_O and E_C. When three segments (Mut [S4-6]) were mutated, the noncanonical binding mode exhibited only one peak centered at E_O, indicating that these interactions are confined within the probed region (segments 1–3). Further removal of complementarity disrupted binding to background levels, confirming that a minimum of three consecutive segments are required for noncanonical binding. Besides the noncanonical mode, a fraction of binding events in the PAM-distal mutation series exhibited the signature initial FRET of the interference mode (E_I, left column, Figure 3E). Even though this initial FRET was identical to that of the canonical binding mode, binding events were transient and did not exhibit any FRET transitions until dissociation after 24.8 ± 7.3 s (Figures 3F and S2C). This state reports on the formation of an interference-like target-recognition complex that cannot be locked and is in line with a previous observation that the PAM- distal region is required for stable R-loop formation in the interference model (Szczelkun et al., 2014). Finally, to evaluate the role of the PAM in Cascade's noncanonical binding mode, we repeated both series of protospacer mutations in the presence of the escape mutant PAM (named ''priming PAM,'' Figures 3 and S2). Overall, mutation of the PAM substantially reduced the number of binding events for each mutant compared to its interfering PAM counterpart (compare columns, Figures 3B and 3E), indicating that the PAM is not strictly required for, but rather facilitates, noncanonical binding. In addition, the E_I state observed in Mut[S5-6] and Mut[S4-6] was completely abrogated upon PAM mutation, suggesting that this intermediate requires the coordinated ternary interaction of Cascade with the PAM and the seed. In summary, our single-molecule results show that the noncanonical binding mode of Cascade is much more robust than its canonical

mode, capable of binding a wide variety of mutated targets, yet still exhibiting sequence specificity. Such versatility could facilitate primed spacer acquisition, in which invading DNA variants that harbor mutations in their PAM or protospacer can still be detected by the CRISPR-Cas immune system.
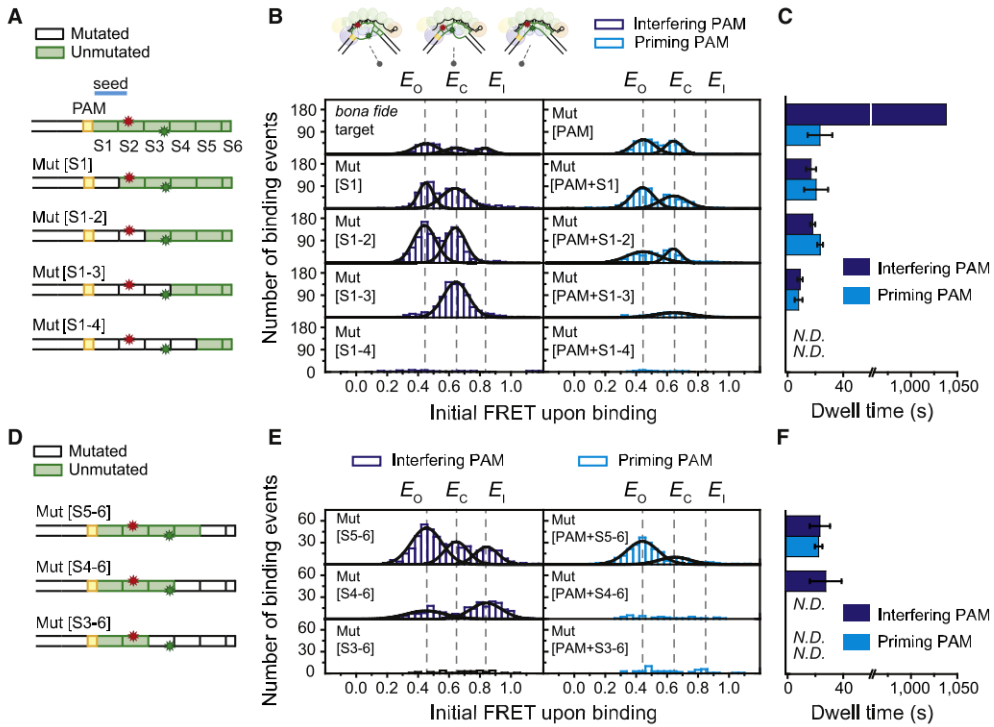

Figure 3. Cascade Exhibits Noncanonical Binding to Protospacers with PAM-Proximal or PAM-Distal Segmented Mutations. (A) Schematics of DNA targets in the PAM-proximal mutation series illustrating mutated (white) or unmutated (green) segments (S1–S6) of the protospacer. Mut [S1], Mut[S1-2], Mut[S1-3], and Mut[S1-4] have segments 1, 1-2, 1-3, and 1-4 mutated, respectively. (B) Histograms of the initial FRET upon binding (average of first 1.5 s of each event) of each PAM-proximal mutant from (A) bearing either an interfering (purple bars, left column) or priming (light blue bars, right column) PAM exhibit peaks (Gaussian fits, black lines) positioned similar to that of the bona fide and Mut[PAM] targets (top row, same as Figures 1F and 2B) at E_O (0.44), E_C (0.65), or E_I (0.84) (dashed black lines). The recorded events are from one field of view of the detector. (C) Mean binding dwell time of each PAM-proximal mutant from (A) bearing either an interfering (purple bars) or a priming (light blue bars)PAM(derived from dwell time distributions, see Figure S2A). Error represents SD (three individual data sets). The dwell time of the bona fide target could not be measured accurately due to the photobleaching and thus we arbitrarily set 1,040 s to represent the longer characteristic timescale in Figure 1G. (D) Schematics of DNA targets in the PAM-distal mutation series illustrated as in (A). Mut[S5-6], Mut[S4-6], and Mut[S3-6] have segments 5-6, 4-6, and 3-6 mutated, respectively. (E) Histograms of the initial FRET upon binding of each PAM-distal mutant from (D) displayed in a similar fashion to that in (B). (F) Mean binding dwell

of each PAM-distal mutant from (D) bearing either an interfering (purple bars) or a priming (light blue bars) PAM (derived from dwell time distributions, see Figure S2A). Error represents SD (three individual data sets). *N.D.*, ''not determined.'' See also Table S2.
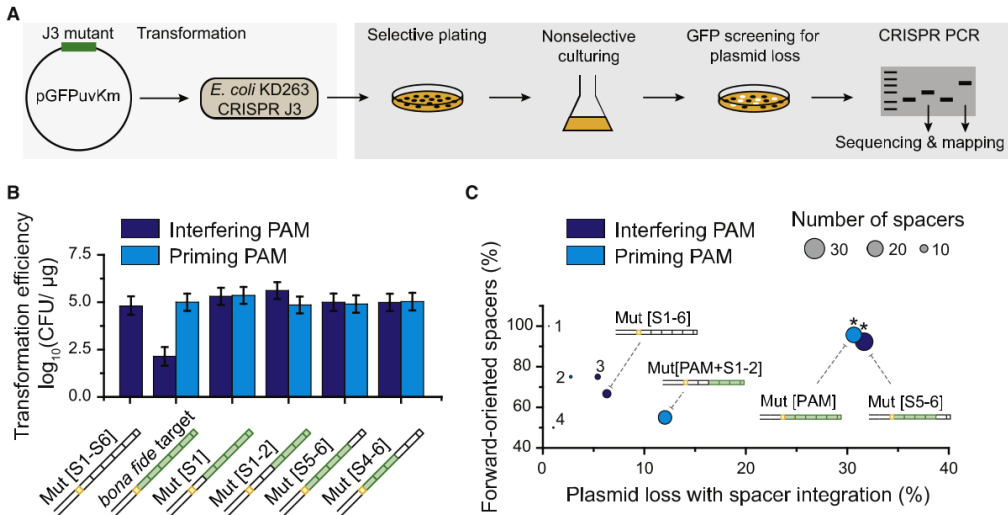

Figure 4. Noncanonical Binding Leads to Primed Spacer Acquisition. (A) Cartoon representation of the in vivo assay used to determine primed plasmid loss and spacer acquisition. (B) Transformation efficiencies of plasmids harboring different target sequences (see schematics) with an interfering (purple bar) or a priming (light blue bar) PAM. CFU, ''colony-forming unit.'' Error is SD of three individual measurements. (C) A two-dimensional bubble plot showing the fraction of forward-oriented spacers acquired versus the percentage of plasmid loss for those targets in (B) that exhibited spacer integration. Circle size represents the total number of spacers that were acquired, and circle color represents an interfering (purple) or a priming (light blue) PAM. Asterisk indicates a forward directional bias (relative to random) with a p value $< 1 \times 10^{-5}$ based on binomial statistics. The numbers of 1, 2, 3, and 4 indicate data points from constructs Mut[PAM+S5-6], Mut[PAM+S1], Mut[S4-6], and Mut[PAM+S4-6], respectively. See also Figure S3, Table S1, and Table S3.

Functional Roles of Two Distinct Binding Modes

To investigate whether the canonical and noncanonical binding modes of Cascade lead to different functional outcomes, we reconstituted CRISPR interference in vitro. We cloned the segmented mutants that showed binding in our single-molecule experiments into plasmids (Table S3) and tested the plasmids for Cascade-directed degradation by Cas3. Our assay revealed that only the plasmid with a perfectly complementary protospacer accompanied by an interfering PAM led to target degradation, whereas target plasmids containing either an escape PAM mutation and/or segmented mutations proximal or distal to the PAM were unaffected by Cas3 (Figure S3). These results suggest that only Cascade's canonical binding mode (E_I / E_O) generates an R-

loop structure that supports target degradation by Cas3. Next, we sought to determine if the noncanonical binding mode of Cascade results in primed spacer acquisition in vivo. To assess primed spacer acquisition, we first transformed the target plasmids with segmented mutations into *E. coli* containing a targeting CRISPR array plasmid (Figure 4A). Notably, only the target with a perfectly complementary protospacer and interfering PAM led to a reduced transformation efficiency (Figure 4B), confirming that the CRISPR-Cas system exclusively targets the R-loops generated through the canonical binding mode of Cascade. Next, transformants were transferred to nonselective media, which allowed the CRISPR-Cas system to mount a primed response. After 2 days of cell growth, three mutant constructs (Mut[PAM], Mut[S5-6], and Mut[PAM+S1-2]) showed a higher degree of plasmid loss than the negative control construct Mut[S1-6] did (Figure 4C). To identify if these plasmids were lost through primed spacer acquisition, the genomic CRISPR- array was amplified by PCR, and amplicons with increased size were sequenced (Figure 4A). In total, 23, 26, and 20 new spacers were obtained that originated from the target plasmids Mut[PAM], Mut[S5-6], and Mut[PAM+S1-2], respectively. Sequencing of the genomic CRISPR array also allowed us to determine whether the acquired spacers showed any strand bias that is typical of the priming process in type I-E systems (Datsenko et al., 2012; Fineran et al., 2014; Swarts et al., 2012). Among the three constructs, Mut[PAM] and Mut[S5-6] exhibited bias in spacer acquisition toward the target strand (p value < 1 x 10^{-5}, Figure 4C), suggesting that these spacers were obtained by primed spacer acquisition. Taken together, the high frequency of plasmid loss and strand bias in the acquired spacers suggests that the noncanonical binding mode acts as a gateway to priming in vivo.
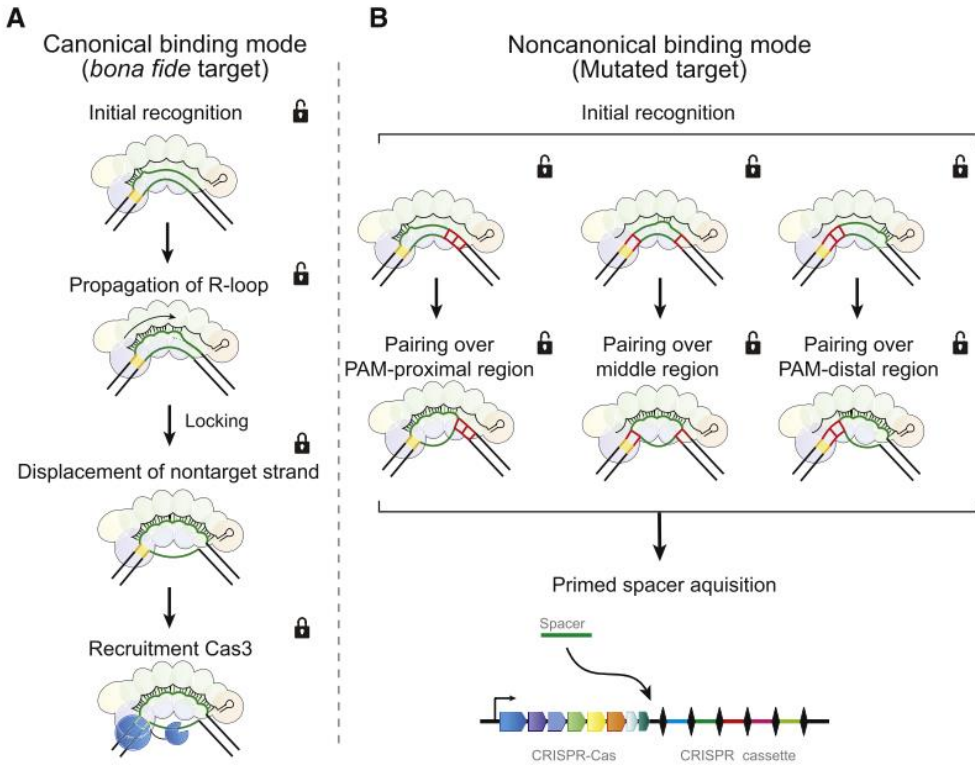
Figure 5. Two Binding Modes of Cascade Lead to Different Functional Outcomes. Cascade employs two distinct target-DNA binding modes that trigger (A) interference or (B) priming. (A) In the interference pathway, target recognition initiates from the PAM and PAM-proximal region. R-loop formation then propagates toward the PAM- distal region. When Cascade senses the fully paired structure, it brings this complex into a lower energy state ("locking") that displaces the nontarget strand out of Cascade. This exposed strand is then cleaved by Cas3. (B) In the priming pathway, DNA is probed through brief interactions. PAM recognition facilitates this priming pathway but is not required. The brief interactions may initiate from the PAM-proximal (left), the PAM-distal region (right), or the middle of the protospacer (middle), which becomes stable when paired over three or more segments. This noncanonical ("unlocked") binding mode leads to a unique conformation of the R-loop and signals for primed spacer acquisition.

Discussion

Adaptive immune systems are found in both vertebrates and prokaryotes and provide specific defense against invading pathogens. The high specificity of this immunity is important for distinguishing self from nonself (Gandon and Vale, 2014), yet it brings a downside that it can be readily overcome by rapidly evolving pathogens

(Koel et al., 2013). However, both vertebrates and prokaryotes have developed sophisticated fail-safe mechanisms to target these pathogens. For example, when vertebrates face invaders bearing mutated antigens, they may still be recognized by a pool of polyclonal antibodies (Purtha et al., 2011). The resulting secondary response proceeds more quickly and efficiently than the primary response, which allows vertebrate hosts to keep pace with their evolving pathogens (Tarlinton and Good-Jacobson, 2013). The prokaryotic adaptive immune system faces similar challenges. Rapidly evolving pathogens readily overcome sequence-specific CRISPR-Cas-mediated host defense (Deveau et al., 2010; Semenova et al., 2011), exposing a major limitation to prokaryotic adaptive immunity (Weinberger et al., 2012). However, analogous to vertebrate adaptive immunity, once pre-exposed to an ancestral invader, CRISPR-Cas responds more rapidly and efficiently to future variants than it can to a novel invader (Datsenko et al., 2012; Fineran et al., 2014; Li et al., 2014; Richter et al., 2014; Samson et al., 2013; Tamulaitis et al., 2014; Westra et al., 2012b). Although Cascade was shown to be essential for this ''primed'' response (Datsenko et al., 2012), the underlying mechanism has remained enigmatic. Here we provide in- sights into this puzzle by showing that Cascade binds mutated targets through a distinct noncanonical mode with low-fidelity compared to the high-fidelity binding mode used for unmutated targets. We show that the canonical, high-fidelity binding mode is a stepwise process that locks, triggering recruitment of nuclease/helicase Cas3 only when all criteria are met, including an interfering PAM, a matching seed, and pairing of all segments of the crRNA guide. In contrast, the noncanonical, low-fidelity binding mode initiates a downstream pathway that results in rapid spacer acquisition through the priming process (Figure 5).

Protein-Mediated High-Fidelity Target Recognition

Our single-molecule data demonstrate in real time that high- fidelity target-DNA binding is a multistep process and occurs in a directional manner from the PAM-proximal to PAM-distal end of the protospacer. Previous studies have shown that the recognition process is initiated when the Cse1 subunit recognizes the PAM (Sashital et al., 2012) and the crRNA hybridizes with the seed sequence. After this initial recognition complex is formed, the R-loop propagates toward the PAM-distal region of the protospacer (Semenova et al., 2011; Szczelkun et al., 2014; Wiedenheft et al., 2011b). When the pairing of the crRNA reaches the PAM distal-end of the protospacer, Cascade senses the fully paired structure and stabilizes this complex into a lower energy state (''locking'') (Szczelkun et al., 2014). This state acts as a flag for the destruction of the target DNA by Cas3 (Mulepati and Bailey, 2013; Sinkunas et al., 2013; Westra et al., 2012a). This stepwise mechanism involves both protein-nucleic acid interactions (Cse1-PAM) and progressive crRNA-DNA base pairing, ensuring efficient and high-fidelity recognition, and degradation of targeted DNA. Our study shows how Cascade maintains a strict regime to prevent nonspecific cleavage by controlling the pathway toward the proposed locking process (Szczelkun et al., 2014). When Cascade encounters a target with mismatches (e.g., Mut [S5-6], Figure 3F), the initial recognition complex forms, but the R-loop does not propagate throughout the full protospacer (absence

of a transition of E_I to E_O)(Figure S2C). As a result, Cascade will not lock the R-loop, and the initiation complex can disassemble using thermal energy. This process cannot be explained by the thermodynamic properties of base pairing alone, since a target with mismatches often forms a far larger number of consecutive base pairs than seven (e.g., Mut[S5- 6]), which has been shown to be the minimal number of base pairs required for stable binding (Cisse et al., 2012). Instead, the last step of stepwise recognition (locking) must involve protein-nucleic acid interactions that verify base pairing over the entire protospacer. This model is analogous to the stepwise conformational change observed with Argonaute proteins during its target search process (Schirle et al., 2014; Wang et al., 2009) and contrasts with the low-fidelity RecA-mediated target search that does not seem to use protein-nucleic acid interaction in promoting specificity (Chen et al., 2008).

Structural View of the Priming Mode

The structure of Cascade supports our finding that low-fidelity target-DNA interactions can initiate from any segment of the crRNA (Figure 5B). Cascade is composed of five different Cas protein subunits assembled into a highly interlocked, crRNA- containing protein complex (Jackson et al., 2014; Mulepati et al., 2014; Zhao et al., 2014). The backbone of the complex consists of six Cas7 subunits with a hand-like architecture. Each hand uses its thumb to hold and position the crRNA at 6 nt intervals. As a consequence, every sixth base is flipped out of the plane and is unable to interact with the target DNA. This unusual configuration permits the crRNA to pair with a target in segments of five nucleotides in an underwound, ribbon-like structure (Mulepati et al., 2014). Interestingly, individual segments of the crRNA in the apo-Cascade structure are already preordered in a pseudo A-form helix with their nucleobases facing the solvent (Jackson et al., 2014; Zhao et al., 2014). Structural preordering is a common strategy to facilitate target binding of nucleic acid guided complexes (e.g., Argonaute and RecA) (Chen et al., 2008; Künne et al., 2014; Wang et al., 2009), and thus the preordered segments of crRNA of Cascade are in line with the idea that low-fidelity interactions can nucleate from any crRNA segment (Figure 5B). Although the low-fidelity binding mode leads to relatively short-lived R-loops, it is their distinct conformation that likely signals for a primed spacer acquisition response in the cell. The DNA recognition mechanism of Cascade contrasts that of Cas9, which has recently been shown to be strictly dependent on the PAM (Sternberg et al., 2014). Furthermore, Cas9 does not base pair its crRNA in segments to the target DNA (Nishimasu et al., 2014) but rather forms a contiguous double helix, making it more difficult to imagine that PAM-distal regions of Cas9's crRNA can initiate an interaction with the target DNA. Off-target cleavage analysis of Cas9 during genome editing clearly indicates that Cas9 also tolerates mutations (Kuscu et al., 2014), but whether this leads to a priming response in bacteria with Type II CRISPR-Cas systems remains to be shown.

Mechanisms of the Priming Mode

Although the interference response of CRISPR immunity is a relatively well-characterized phenomenon, the molecular mechanism of priming remains poorly understood. First, our data show that Cascade distinguishes

mutated targets from bona fide targets using a low-fidelity binding mode that can initiate priming. A recent study showed that priming in *E. coli* is robust, tolerating up to 13 mutations throughout the 32 nt protospacer and 3 nt PAM (Fineran et al., 2014). Even when mutations were clustered in any of the crRNA defined segments, priming was not abolished. The low-fidelity binding mode of Cascade, in which individual segments may initiate pairing with a target, can explain the reported high tolerance for distributed and clustered mutations in a target during priming. In this mode, Cascade can probe DNA for complementarity to any of its crRNA segments, and extend such an interaction in either direction, thereby achieving sequence-specific detection of targets with limited base complementarity. However, the minimal number of base pairs required for priming (Fineran et al., 2014) ensures that detrimental self-priming of the bacterial genome at random sites is unlikely. Second, we observed that the noncanonical binding mode occurs even for substrates containing an interfering PAM and an intact seed, suggesting that direct interference and priming may occur simultaneously. Indeed, we have previously observed that *E. coli* is cured from high copy number plasmids by using existing spacers to expand the CRISPR array with a range of new spacers against the same target (Fineran and Charpentier, 2012; Swarts et al., 2012). For a host this is a highly advantageous strategy; by simultaneously using interference and priming, the CRISPR interference effect is amplified while the chance that invaders evade immunity through point mutations in their protospacers is reduced. Even though it remains to be seen how priming is coordinated in the presence of the remaining Cas protein machinery (Cas1, Cas2, and Cas3), the relatively short time that Cascade spends on a target in the priming mode suggests that other factors might stabilize this relatively weak interaction. Finally, in CRISPR-Cas/I-E systems priming is a DNA strand- dependent process in which approximately 90% of new spacers are integrated from the same strand as the spacer triggering priming (Shmakov et al., 2014). Our results with Mut[PAM] and Mut[S5-6] in Figure 4 are consistent with this strand bias. In contrast, primed spacer acquisition in Type I-B and I-F systems does not exhibit such strand bias ((Li et al., 2014; Richter et al., 2014). Interestingly, for protospacers mutated in the PAM and segments 1 and 2 (Mut[PAM+S1-2]), including the seed, we observed a higher degree of spacer acquisition without the typical strand bias, suggesting that these types of targets lead to a priming behavior in which strand specificity is lost.

Conclusion

Faithful copying and decoding of genetic information is central to the most important processes in the cell, including DNA replication (Kunkel, 2004), RNA transcription (Xu et al., 2014), and protein translation (Zaher and Green, 2009). But high fidelity always comes at the cost of reduced processing speed. Here we show how a crRNA guided complex solves this dilemma by employing both high- and low-fidelity target-DNA recognition modes. While the high-fidelity mode ensures destruction of only perfectly matching targets, the low-fidelity priming mode enables detection of a whole range of mutated invaders to initiate the priming process. The

unique combination of these two properties in a single RNA-guided complex not only makes CRISPR immunity robust, but also reveals versatility of adaptive immunity against rapidly mutating pathogens.

Experimental Procedures

Preparation of Cascade, Biotinylated Cascade, and Cas3 Cascade and Cascade lacking subunit Cse1 were affinity purified from *E. coli* strain BL21(DE3) as described previously (Jore et al., 2011). For biotinylated Cascade, Cascade was site-specifically labeled at the N terminus of subunit Cse1 with an aldehyde moiety using the formylglycine-generating enzyme (FGE) method described previously (Rabuka et al., 2012), reacted with biotin-hydrazide, and purified by size exclusion chromatography. Cas3 was purified by size-exclusion chromatography as described previously (Mulepati and Bailey, 2013) (see ''Protein Preparation'' in Supplemental Experimental Procedures).

Single-Molecule FRET Biotinylated Cascade complexes were anchored to polyethylene glycol- coated quartz microscope slides by biotin-streptavidin linkage Dye-labeled (Cy3 and Cy7) dsDNA targets (see ''DNA Preparation'' in Supplemental Experimental Procedures) were added to the immobilized Cascade complexes and detected by a prism-type TIRF microscope. In a typical field of view, 200–300 molecules were detected. dsDNA targets were excited with a 532 nm laser and fluorescence emissions from Cy3 and Cy7 were separated by dichroic mirrors and imaged onto two halves of a CCD camera after passing through various filters. Imaging buffer consisted of Cascade buffer (50 mM HEPES [pH 7.5], 75 mM NaCl, 2 mM MgCl2), an oxygen-scavenging system (1% glucose [v/v], 0.1 mg/mL glucose oxidase [Sigma], 17 mg/mL Catalase [Roche]) to reduce photobleaching, and 1 mM Trolox (Sigma) to reduce photoblinking of the dyes (Rasnik et al., 2006). Imaging was performed at room temperature (23 C±2 C). Fluorescence time traces of individual binding events were identified in recorded movies and subsequently analyzed using custom software developed in IDL and MATLAB, respectively. The FRET value was defined as IA/(ID+IA), where ID and IA represent the fluorescence signals detected in the Cy3 and Cy7 channels, respectively. See ''Single-Molecule Fluorescence'' in Supplemental Experimental Procedures.

Direct interference and priming Direct interference was assessed by determining the transformation efficiency of target plasmids to E. coli strain KD263 (Shmakov et al., 2014) containing pWUR564 (Table S3). Cas gene expression was induced 30 min prior to making cells chemically competent. Priming was assessed using plasmid loss assays as previously described (Fineran et al., 2014). Briefly, *E. coli* trans- formants containing the target plasmids (pWUR738-747) were grown in LB for 48 hr, plated on LB agar, and imaged under mild UV light. GFP-negative colonies were screened for spacer integration by PCR. Newly acquired spacers were sequenced and were strand-specifically mapped onto the target plasmid sequence to verify priming (see ''Direct Interference and Priming'' in Supplemental Experimental Procedures).

Supplemental Information

Supplemental Information includes three figures, three tables, and Supplemental Experimental Procedures and can be found with this article at http:// dx.doi.org/10.1016/j.molcel.2015.01.028.

References

Barrangou, R., and Van der Oost, J. (2013). CRISPR-Cas Systems. RNA-Mediated Adapt. Immun. Bact. Archaea.

Charpentier, E., and Marraffini, L.A. (2014). Harnessing CRISPR-Cas9 immunity for genetic engineering. Curr. Opin. Microbiol. *19*, 114–119.

Chen, Z., Yang, H., and Pavletich, N.P. (2008). Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures. Nature *453*, 489–4.

Cisse, I.I., Kim, H., and Ha, T. (2012). A rule of seven in Watson-Crick base-pairing of mismatched sequences. Nat. Struct. Mol. Biol. *19*, 623–627.

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.

Deveau, H., Garneau, J.E., and Moineau, S. (2010). CRISPR/Cas system and its role in phage-bacteria interactions. In Annual Review of Microbiology, S. Gottesman, and C.S. Harwood, eds. (Palo Alto: Annual Reviews), pp. 475–493.

Fineran, P.C., and Charpentier, E. (2012). Memory of viral infections by CRISPR-Cas adaptive immune systems: Acquisition of new information. Virology *434*, 202–209.

Fineran, P.C., Gerritzen, M.J.H., Suárez-Diez, M.M.M., Künne, 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. U. S. A. *111*, E1629--38.

Gandon, S., and Vale, P.F. (2014). The evolution of resistance against good and bad infections. J. Evol. Biol. *27*, 303–312.

Goldberg, G.W., Jiang, W., Bikard, D., and Marraffini, L. a (2014). Conditional tolerance of temperate phages via transcription-dependent CRISPR-Cas targeting. Nature *514*, 633–637.

Ha, T. (2014). Single-molecule methods leap ahead. Nat. Methods *11*, 1015–1018.

Hale, C.R., Cocozaki, A., Li, H., Terns, R.M., and Terns, M.P. (2014). Target RNA capture and cleavage by the Cmr type III-B CRISPR-Cas effector complex. Genes Dev. *28*, 2432–2443.

Hochstrasser, M.L., Taylor, D.W., Bhat, P., Guegler, C.K., Sternberg, S.H., Nogales, E., and Doudna, J. a (2014). CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. Proc. Natl. Acad. Sci. U. S. A. *111*, 6618–6623.

Hsu, P.D., Lander, E.S., and Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. Cell *157*, 1262–1278.

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

Joo, C., Fareh, M., and Narry Kim, V. (2013). Bringing single-molecule spectroscopy to macromolecular protein complexes. Trends Biochem. Sci. *38*, 30–37.

Jore, M.M., Lundgren, M., van Duijn, E., Bultema, J.B., Westra, E.R., Waghmare, S.P., Wiedenheft, B., Pul, U., Wurm, R., Wagner, R., et al. (2011). Structural basis for CRISPR RNA-guided DNA recognition by Cascade. Nat. Struct. Mol. Biol. *18*, 529–536.

Juette, M.F., Terry, D.S., Wasserman, M.R., Zhou, Z., Altman, R.B., Zheng, Q., and Blanchard, S.C. (2014). The bright future of single-molecule fluorescence imaging. Curr. Opin. Chem. Biol. *20*, 103–111.

Koel, B.F., Burke, D.F., Bestebroer, T.M., van der Vliet, S., Zondag, G.C.M., Vervaet, G., Skepner, E., Lewis, N.S., Spronken, M.I.J., Russell, C.A., et al. (2013). Substitutions near the receptor binding site determine major antigenic change during influenza virus evolution. Science *342*, 976–979.

Kunkel, T.A. (2004). DNA Replication Fidelity. J. Biol. Chem. *279*, 16895–16898.

Künne, T., Swarts, D.C., and Brouns, S.J.J. (2014). Planting the seed: target recognition of short guide RNAs. Trends Microbiol. *22*, 74–83.

Kuscu, C., Arslan, S., Singh, R., Thorpe, J., and Adli, M. (2014). Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. Nat. Biotechnol. *32*, 677–683.

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.

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

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

Nishimasu, H., Ran, F.A., Hsu, P.D., Konermann, S., Shehata, S.I., Dohmae, N., Ishitani, R., Zhang, F., and Nureki, O. (2014). Crystal structure of Cas9 in complex with guide RNA and target DNA. Cell *156*, 935–949.

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.

Purtha, W.E., Tedder, T.F., Johnson, S., Bhattacharya, D., and Diamond, M.S. (2011). Memory B cells, but not long-lived plasma cells, possess antigen specificities for viral escape mutants. J. Exp. Med. *208*, 2599–2606.

Rabuka, D., Rush, J.S., deHart, G.W., Wu, P., and Bertozzi, C.R. (2012). Site-specific chemical protein conjugation using genetically encoded aldehyde tags. Nat. Protoc. *7*, 1052–1067.

Rasnik, I., McKinney, S. a, and Ha, T. (2006). Nonblinking and long-lasting single-molecule fluorescence imaging. Nat. Methods *3*, 891–893.

Reeks, J., Naismith, J.H., and White, M.F. (2013). CRISPR interference: a structural perspective. Biochem. J *453*, 155–166.

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.

Robinson, A., and van Oijen, A.M. (2013). Bacterial replication, transcription and translation: mechanistic insights from single-molecule biochemical studies. Nat. Rev. Microbiol. *11*, 303–315.

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.

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

Sashital, D.G., Wiedenheft, B., and Doudna, J. a. (2012). Mechanism of Foreign DNA Selection in a Bacterial Adaptive Immune System. Mol. Cell *46*, 606–615.

Schirle, N.T., Sheu-Gruttadauria, J., and MacRae, I.J. (2014). Structural basis for microRNA targeting. Science (80-.). *346*, 608–613.

Schuler, B., and Hofmann, H. (2013). Single-molecule spectroscopy of protein folding dynamics-expanding scope and timescales. Curr. Opin. Struct. Biol. *23*, 36–47.

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. U. S. A. *108*, 10098–10103.

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.

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.

Sorek, R., Lawrence, C.M., and Wiedenheft, B. (2013). CRISPR-mediated adaptive immune systems in bacteria and archaea. Annu Rev Biochem *82*, 237–266.

Staals, R.H.J.H.J., Agari, Y., Maki-Yonekura, S., Zhu, Y., Taylor, D.W.W., van Duijn, E., Barendregt, A., Vlot, M., Koehorst, J.J.J., Sakamoto, K., et al. (2013). Structure and Activity of the RNA-Targeting Type III-B CRISPR-Cas Complex of Thermus thermophilus. Mol. Cell *52*, 135–145.

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*, 62–67.

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*.

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. U. S. A. *111*, 9798–9803.

Tamulaitis, G., Kazlauskiene, 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.

Tarlinton, D., and Good-Jacobson, K. (2013). Diversity among memory B cells: origin, consequences, and utility. Science *341*, 1205–1211.

Terns, R.M., and Terns, M.P. (2014). CRISPR-based technologies: Prokaryotic defense weapons repurposed. Trends Genet. *30*, 111–118.

Wang, Y., Juranek, S., Li, H., Sheng, G., Wardle, G.S., Tuschl, T., and Patel, D.J. (2009). Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. Nature *461*, 754–761.

Weinberger, A.D., Wolf, Y.I., Lobkovsky, A.E., Gilmore, M.S., and Koonin, E. V. (2012). Viral diversity threshold for adaptive immunity in prokaryotes. MBio *3*.

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. (2012a). CRISPR Immunity Relies on the Consecutive Binding and Degradation of Negatively Supercoiled Invader DNA by Cascade and Cas3. Mol Cell *46*, 595–605.

Westra, E.R., Swarts, D.C., Staals, R.H.J., Jore, M.M., Brouns, S.J.J., and van der Oost, J. (2012b). The CRISPRs, they are a-changin': how prokaryotes generate adaptive immunity. Annu. Rev. Genet. *46*, 311–339.

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*, e1003742.

Westra, E.R., Buckling, A., and Fineran, P.C. (2014). CRISPR–Cas systems: beyond adaptive immunity. Nat. Rev. Microbiol. *12*, 317–326.

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

Wiedenheft, B., van Duijn, E., Bultema, J.B., Bultema, J., Waghmare, S.P., Waghmare, S., Zhou, K., Barendregt, A., Westphal, W., Heck, A.J.R., et al. (2011b). RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. Proc. Natl. Acad. Sci. U. S. A. *108*, 10092–10097.

Xu, L., Da, L., Plouffe, S.W., Chong, J., Kool, E., and Wang, D. (2014). Molecular basis of transcriptional fidelity and DNA lesion-induced transcriptional mutagenesis. DNA Repair (Amst). *19*, 71–83.

Zaher, H.S., and Green, R. (2009). Fidelity at the Molecular Level: Lessons from Protein Synthesis. Cell *136*, 746–762.

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

Chapter 3

Direct visualization of CRISPR target search in live bacteria

Jochem N.A. Vink[1]*, Marnix Vlot[2]*, Rebecca E. McKenzie[1], Koen Martens[3], Cristobal Almendros[1], Boris Estrada Bonilla[1], Daan J.W. Brocken[2,4], Johannes Hohlbein[3], Stan J.J. Brouns[1,2]

[1] Kavli Institute of Nanoscience, Department of Bionanoscience, Delft University of Technology, Van der Maasweg 9, 2629 HZ Delft, Netherlands
[2] Laboratory of Microbiology, Wageningen University, Stippeneng 4, 6708 WE Wageningen, Netherlands.
[3] Laboratory of Biophysics, Wageningen University, Stippeneng 4, 6708 WE Wageningen, Netherlands.
[4] Leiden Institute of Chemistry, Gorlaeus Laboratories, Leiden University, 2333 CC Leiden, Netherlands.

* These authors contributed equally

Abstract

CRISPR-Cas systems encode versatile machineries that have evolved to store, recognize and cleave specific DNA sequences in prokaryotic cells. We visualized a CRISPR-Cas system in its native bacterial host to understand how RNA-effector complexes find foreign DNA targets in a crowded, DNA-packed, cellular environment containing ~700,000 native PAMs. By tracking individual RNA-effector complexes in live cells, we show these complexes interrogate potential sites much faster than reported before, enabling 100 RNA-effector complexes to find a single invader DNA sequence within 20 minutes. We discovered that RNA-effector complexes are affected by the presence of CRISPR arrays which lead to the PAM-scanning subunit Cas8e being expelled from the Cascade complex after binding CRISPR arrays. Though this response avoids self-targeting, it affects the ability of Cascade to interfere with foreign targets and therefore exposes an inherent weakness of CRISPR-Cas systems. Our results reveal new links between target search kinetics, host self-avoidance and CRISPR-Cas interference in its natural environment.

Introduction

RNA-guided CRISPR-Cas surveillance complexes have evolved to specifically and rapidly recognize sequences of previously catalogued mobile genetic elements (MGEs) (Marraffini, 2015). Target recognition depends on RNA-DNA complementarity and, in many CRISPR-Cas systems, on the presence of a protospacer adjacent motif (PAM), a short (3-6 bp) nucleotide sequence flanking the target, that together determine whether interference, priming (accelerated acquisition) or target escape occurs (Fineran et al., 2014; Mojica et al., 2009). In natural environments, CRISPR-Cas complexes have to overcome two major challenges common to all immune systems to work effectively. Firstly, the complexes have to avoid self-targeting, which is complicated by the presence of target-like sequences in the host CRISPR array. Secondly, they need to find their targets fast enough to prevent becoming a bacteriophage factory (Hadas et al., 1997).

Many CRISPR-Cas surveillance complexes use the PAM to overcome these challenges (Blosser et al., 2015; Xue et al., 2016). The absence of the PAM in the genomic CRISPR array allows the complex to differentiate self from foreign DNA, efficiently preventing self-targeting (Westra et al., 2013). PAM scanning has been suggested to speed up the search process (Van Erp et al., 2015), as it can circumvent the need of R-loop formation (strand separation) at every position to check for potential targets (Jore et al., 2011; Rutkauskas et al., 2015). Several studies have shown that CRISPR-Cas complexes spend more time on probing PAM rich regions, which is indicative of its function as the first recognition site (Cooper et al., 2017; Redding et al., 2015; Rollins et al., 2015; Sternberg et al., 2014). Targets in the *E. coli* Type I-E system are most often recognized when flanked by one of six PAMs (Leenay et al., 2016) , but the *E. coli* K12 genome itself, residing in the central region of the cell called the nucleoid (Macvanin and Adhya, 2012), contains already 700,000 of these PAM motifs. Both *in vitro* and *in vivo* studies have suggested that interaction times with PAM motifs occurs on the time scale of seconds (Blosser et al., 2015; Knight et al., 2015; Sternberg et al., 2014), which would mean that it would take > 200 hours for a single complex to probe the entire genome.

Single-Particle Tracking set-up

In this study, we investigate the search process of type I-E CRISPR-Cas surveillance complexes (Cascade) of *E. coli* in their cellular environment. The fusion of the Cascade complex to a photoactivatable red fluorescent protein (PATagRFP) enabled us to trace the paths of thousands of complexes during their search through the bacterial cell (Fig 1A). We minimized changes in the regulation of Cas expression by genomic integration of the red fluorescent protein gene as a fusion to the N-terminus of *cas8e* (Suppl. M&M). Transformation assays confirmed that the interference ability in this strain was retained (Fig. S1). We upregulated *cas* gene expression to an average copy number of approximately 300-600 Cascade complexes by overexpression of the positive regulator LeuO to minimize the influence of autofluorescent entities in our analysis (Fig. S2, M&M (Westra et al., 2010)). Cascade diffusion behaviour was extracted from the displacement of each fluorescent particle for

four consecutive steps, spanning a time interval of 80 ms, and fitted to a varying number of underlying states (Fig. 1B). This allowed us to investigate the abundance and behaviour of complexes and subunits in the cell.

Complex formation of Cascade

To distinguish associated motilities of the fluorescent Cascade complexes from monomeric Cas8e subunits, we first measured the diffusion of the PATagRFP-Cas8e fusion in strains lacking the other Cascade subunits (Δ(*cas11-6e*); Fig. 1C). We found that the majority of tracks in this strain were highly mobile (D_3* = 2.55 µm^2/s, 83%) (2). In the WT strain, this fraction only accounted for 27% of all trajectories (Fig. 1D), the rest being slower showing that the majority of our signal comes from fully formed Cascade complexes. Interestingly, when we used the faster maturing, but less photostable PAmCherry (maturation time of 23 min; (Subach et al., 2009)) instead of PATagRFP (75 min; (Subach et al., 2010)) as a fluorescent tag, the percentage of the subunit Cas8e fraction increased (D* = 2.55 µm^2/s, 40%), indicating that complex formation did not fully complete within the PAmCherry maturation time of 23 minutes (Fig. S3).

It has been hypothesized that Cascade complexes are formed around a mature crRNA (24), although it is possible to isolate complexes without crRNA (25). We deleted both genomic CRISPR arrays ($\Delta CRISPR$) to test whether complex formation is possible in the absence of crRNA. The diffusion behaviour of labeled Cas8e in the $\Delta CRISPR$ fusion strain was the same as in strains expressing only labeled Cas8e Δ(*cas11-cas6e*) (Fig. 1E), demonstrating that these complexes are obligatory ribonucleoprotein complexes that depend on crRNA for their formation.
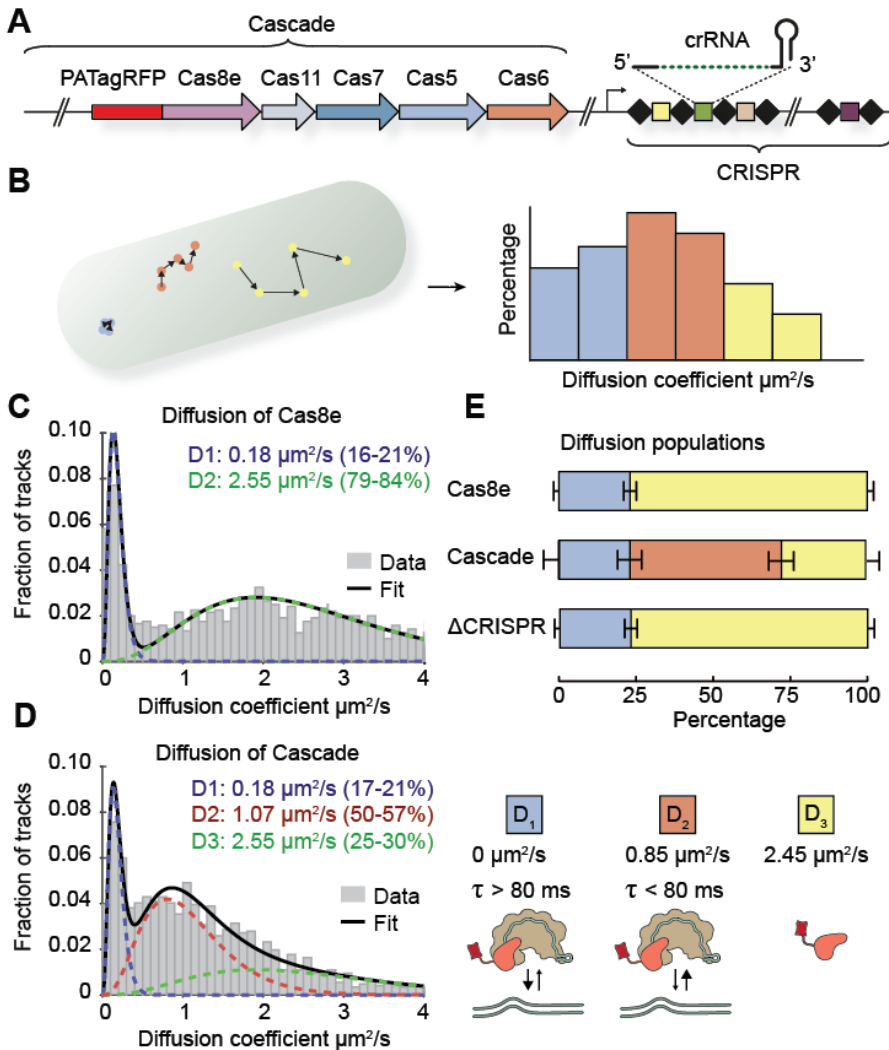
Figure 1. Experimental set-up and complex vs subunit diffusion. (A) Chromosomal locus of the Cascade subunits and integration site of the *patagrfp* gene upstream of *cas8e.* (B) Four subsequent theoretical displacements of particles with different mobility within a bacterial cell are used to infer the different diffusion states. (C) Diffusion states of the Cas8e subunit. (D) Diffusion states of the Cascade complex. (E) Relative fractions of the diffusion states for the Cas8e, Cascade and ΔCRISPR strains and their assignment to different fractions; D_1 represents longer DNA interactions, D_2 represents short DNA interactions and D_3 represents freely diffusing monomeric Cas8e proteins.

Cascade search in absence of targets

The probing behavior of Cascade could be inferred from the remaining trajectories belonging to the fractions slower than monomeric Cas8e found in the Cascade strain. The immobile fraction (D_1*= 0 µm^2/s, 19%) represents molecules tightly bound to DNA for longer than our measured time window (80 ms), whereas the largest fraction (D_2*= 0.85 µm^2/s, 50%) represent mobile Cascade complexes (Fig. 1E). To assess the contribution of PAM interactions of Cascade with DNA, we introduced a G160A mutation in the Cas8e subunit that abolishes PAM interaction. We detected a reduction in short dwell times, but longer dwell times remained relatively abundant, indicating that PAM interactions have very short dwell times (Fig. 2F).

The fraction of very rapid Cascade-DNA interactions is difficult to estimate, due to the limited time resolution of the measurements. To investigate to what extent fast DNA probing events take place, we studied their diffusional behavior in a DNA-free environment created by addition of cephalexin, that causes cells to elongate, forming DNA-free cytoplasmic space between nucleoids (Fig. 2A). We observed a 1.4±0.1 fold enrichment of Cascade in the nucleoid containing regions (Fig. 2B), while the labeled subunit Cas8e was more equally abundant in both regions (1.1±0.06). This suggests that intact Cascade has an inherent affinity for DNA. In order to extract the fraction of DNA binding events of Cascade from its nucleoid enrichment, we simulated Cascade complexes diffusing in cephalexin treated cells (see Materials and Methods; Fig. 2C-E). Based on these simulations, we estimated that Cascade complexes spend 40% of their time bound to DNA (Redding et al., 2015).
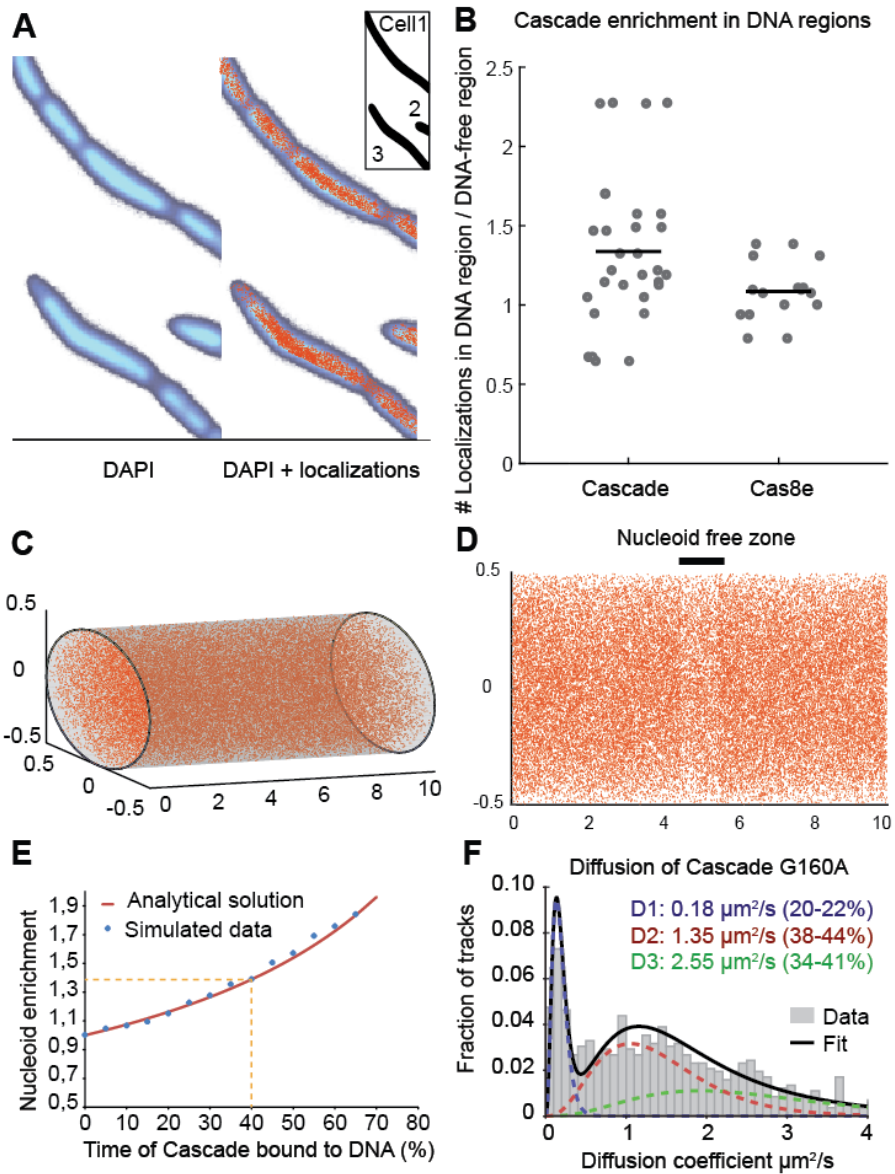
Figure 2. Cascade DNA interactions in the absence of targets. (**A**) Elongated cells in the presence of cephalexin stained with DAPI (left) and the same image overlaid by Cascade localizations (right). (**B**) The average number of localizations in DNA-containing segments divided by the average number of localizations in DNA-free segments (nucleoid enrichment; NE) are plotted per cell (grey dots) and the average is represented by a black bar for the Cascade strain (left) and the Cas8e strain (right). (**C**) Simulation of particles diffusing in an elongated cell without endcaps containing a nucleoid free zone. (**D**) Side view of panel C where the nucleoid free zone (1 µm width) becomes visible (NE:1.4). (**E**) The relationship between the time Cascade is bound to DNA (t_{bound})

and the nucleoid enrichment. Analytical solution: NE = $1/(1-A_{bound})$ where $A_{bound} = t_{bound}A_{Cascade}$ (M&M). (**F**) Diffusion states of the Cas8e^{G160A} strain.

Cascade search in presence of canonical PAM targets

Next we investigated the diffusion behavior of surveillance complexes in the presence of targets. To make sure that all produced Cascades would have a target in the cell, we constructed a plasmid (pTarget) that contained target sites for all 18 spacers found in the genomic arrays of *E. coli* K12 (Fig. 3A). In order to maximize the number of binding events, we used a plasmid with a high copy number (pUC *ori*, ~400), resulting in a total of 7200 potential target sites per cell outnumbering the copy number of Cascade (~ 600) by more than 10 times. We knocked out the nuclease Cas3 that is recruited by the Cascade complex after target recognition (Westra et al., 2012) to prevent target degradation and observed that deletion of *cas3* and addition of a non-targeted control plasmid (pUC19) did not influence Cascade diffusion (Fig. S5 & S6).

Addition of pTarget drastically decreased the fraction of mobile complexes (D_2* = 0.85 µm^2/s; 8%) and increased two very slow fractions (D_1* = 0 µm^2/s; 30%; D_4* = 0.08 µm^2/s; 40%) corresponding to DNA bound states (Fig. 3B). As the immobile fractions could consist of Cascade binding to both chromosome and pTarget, we estimated the fraction of Cascade complexes not bound to pTarget by the relative abundance of mobile Cascade complexes (D_2*) in the presence and absence of pTarget (16%). This number allows a rough estimation of the *in vivo* Kd value (40 nM) and an upper limit for the time required for one Cascade complex to find a target (30 hours) (M&M). This long search time requires a high enough copy number for physiological relevant interference levels. We observed no detectable interference with low Cascade copy numbers (~ 40) in the presence of saturating levels of Cas3, but with induced Cascade copy numbers (~ 600), 1% of the transformed pTarget plasmids escaped targeting (Fig. 3C). Considering our search time estimate, the induced Cascade strain would find its target within three minutes, whereas the WT strain requires 45 minutes, which probably is too slow with plasmid replication highlighting the crucial role for target search kinetics in the efficiency of interference.
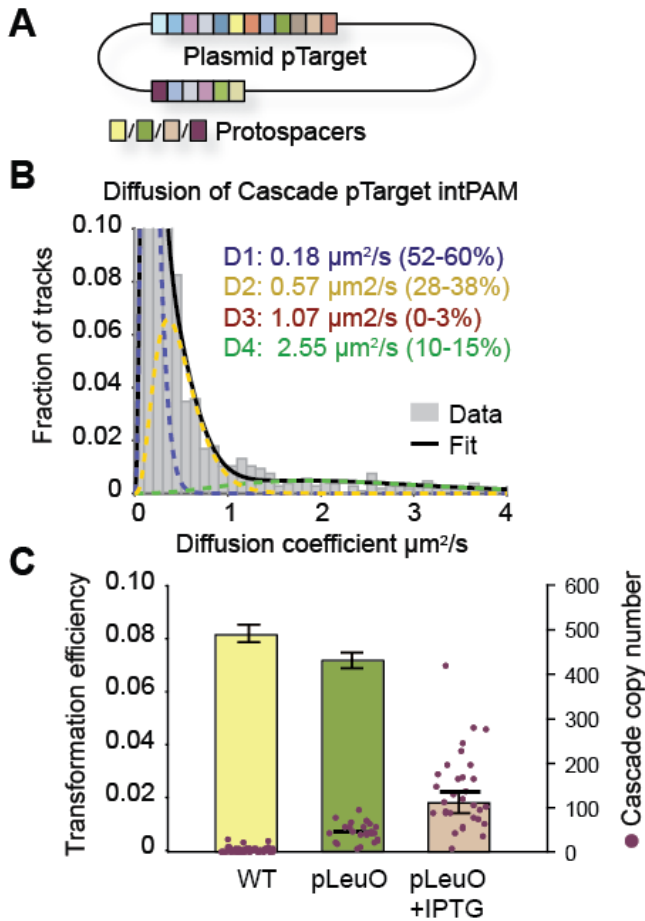
Figure 3. Cascade DNA interactions in the presence of targets. (A) The pTarget plasmid containing 18 targets for the 18 spacers found in the two CRISPR arrays of the *E. coli* K12 genome flanked by a '5-CTT-3' PAM. **(B)** Diffusion states in the presence of pTarget. **(C)** The relationship between the copy number of Cascade and the interference levels as assessed by the relative transformation of pTarget/pGFPuv under different induction conditions (error bars: S.E.M. of three biological replicates).

Influence of CRISPR arrays on target search

We made a variant of pTarget where the target sites are flanked by a 5'-CGG-3', identical to the motif found in the repeats of genomic CRISPR arrays (pDecoy) (Fig. 4A). Strains carrying pDecoy still showed an increase of immobile Cascade complexes albeit to a lesser extent than for pTarget. However, we observed an increase in the fraction of monomeric Cas8e subunits (from 27% to 49%; Fig. 4B) suggesting that Cascade complexes still bind to these targets but that Cas8e subunits dissociate subsequently upon binding to a repeat-like PAM, corroborating recent *in vitro* studies (Jung et al., 2017). Dissociation of Cas8e after binding to targets within a CRISPR array will decrease the number of full complexes and is therefore expected to impair the interference

ability of the whole system. To investigate this hypothesis, we constructed a plasmid containing two CRISPR arrays, but lacking a transcriptional promoter (pArray). Two origin of replication (ori) variants were made that result in different plasmid copy numbers. We found that interference levels of pTarget in the presence of a high copy pArray variant was decreased, but not in a low copy variant or a plasmid lacking the CRISPR arrays (Fig. 4C). This shows that using high copy variants of the CRISPR system can decrease CRISPR interference in the Type I-E system.
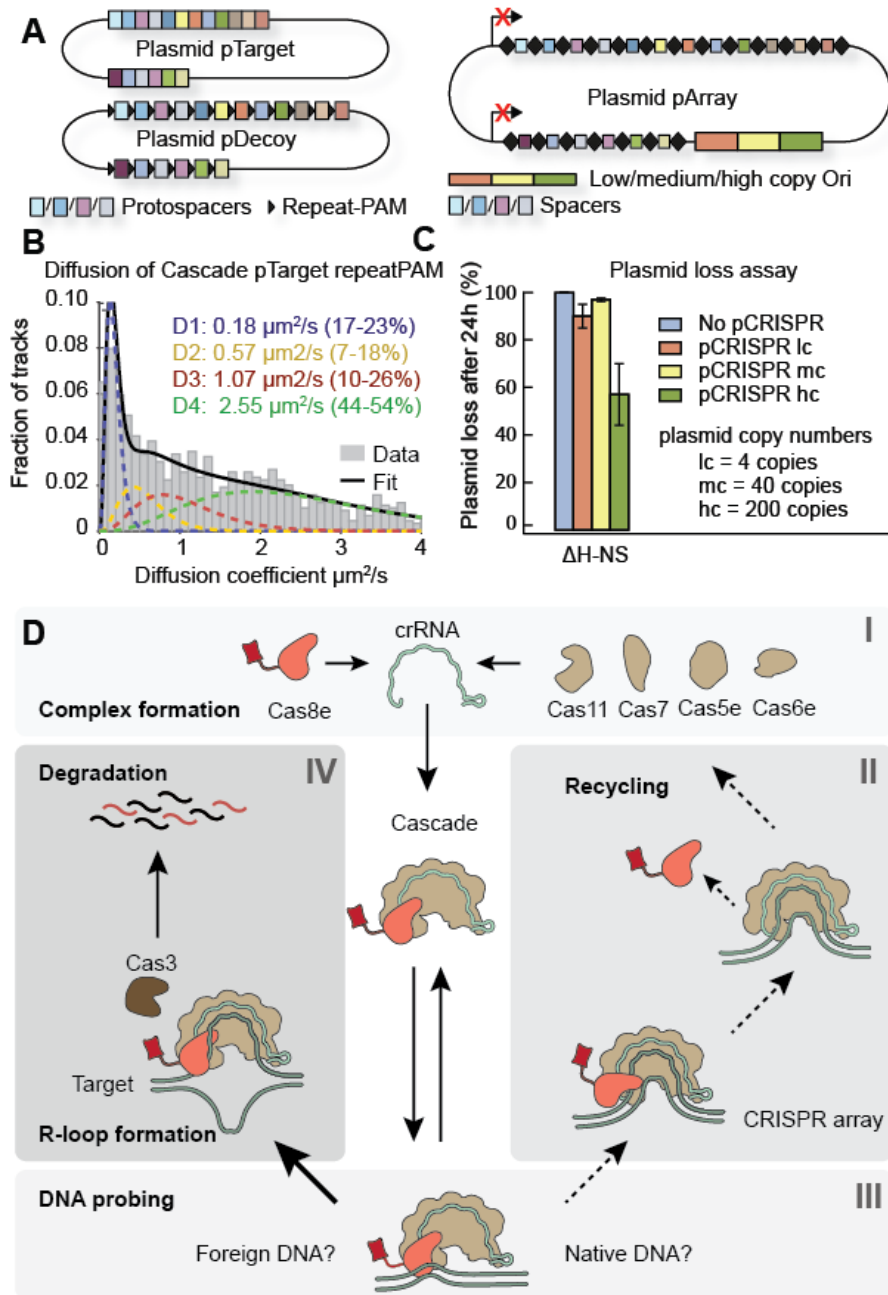
Figure 4. Cascade DNA interactions in the presence of CRISPR arrays. (**A**) The pDecoy plasmid containing 18 targets for the 18 spacers found in the two CRISPR arrays of the *E. coli* K12 genome flanked by the 5'-CGG-3' motif also found in CRISPR repeats. (**B**) Diffusion states in the presence of pDecoy. (**C**) The relationship between the copy number of Cascade and the interference levels as assessed by the relative transformation of pTarget/pGFPuv in the presence of different copy numbers of CRISPR arrays (error bars: S.E.M. of three

biological replicates). (**D**) A model of Cascade target search where I) the Cascade complex is formed in presence of crRNA, II) probes the DNA via rapid DNA binding events III) it is most likely to find the target as it is directed by PAM interactions, which leads to target degradation, IV) but it also has PAM-independent DNA-binding modes that could lead to interaction with the chromosomal CRISPR array, upon which the Cas8e subunit is expelled and target degradation prevented.

Conclusions

Our results provide new insights in the search behavior of RNA-guided CRISPR-Cas complexes in their native cellular environment. Aspecific DNA binding of the Cascade complex occurs rapidly, especially during PAM probing interactions. The timescales observed here indicate that Cascade requires several hours to scan all PAMs, instead of several hundred hours inferred from PAM binding times observed *in vitro*. Our study more closely matches a previously reported *in vivo* study of non-native dCas9 search (Jones et al., 2017). Differences in these observations might result from a limited time resolution of *in vitro* studies or the influence of specific conditions only found in cells. The hour-long search time explains the necessity to express a minimum amount of Cascade to overcome phage/plasmid replication speed. We showed that 300-600 Cascade molecules were required to defend the majority of cells to an incoming plasmid upon transformation. We still observed expression of Cascade molecules in the WT strain, but the level was too low to give a measurable interference level.

Our data shows that Cascade interacts with CRISPR arrays without correct PAM sequence, suggesting the presence of a second, PAM-independent DNA-binding mode of Cascade (Blosser et al., 2015; Xue et al., 2016, 2017). Interestingly, this mode puts the host chromosome in danger, as it does not fully protect the CRISPR array from degradation. Perhaps to this end, the system has evolved to expel the Cas8e subunit upon binding of a CRISPR array. This mechanism is potentially accommodated by conformational changes in Cas8e (Xue et al., 2016). Alternatively, the absence of PAM locking could drive out the Cas8e subunit when the rest of the complex changes conformation. The expelling mechanism, however, comes at the cost of decreasing concentrations of functional Cascade complexes.

References

Blosser, T.R.R., Loeff, L., Westra, E.R.R., Vlot, M., Künne, T., Sobota, M., Dekker, C., Brouns, S.J.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.

Cooper, L.A., Stringer, A.M., and Wade, J.T. (2017). Determining the Specificity of Cascade Binding, Interference, and Priming in vivo 2 3 4 5.

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.

Fineran, P.C., Gerritzen, M.J.H., Suarez-Diez, M., Kunne, T., Boekhorst, van Hijum, F.T., Staals, R.H.J., Brouns, S.J.J., Suárez-Diez, M., Künne, T., et al. (2014). Degenerate target sites mediate rapid primed CRISPR adaptation. Proc. Natl. Acad. Sci. U. S. A. *111*, E1629--38.

Hadas, H., Einav, M., Fishov, I., and Zaritsky, A. (1997). Bacteriophage T4 Development Depends on the Physiology of its Host Escherichia Coli. Microbiology *143*, 179–185.

Jones, D.L., Leroy, P., Unoson, C., Fange, D., Ćurić, V., Lawson, M.J., and Elf, J. (2017). Kinetics of dCas9 target search in Escherichia coli. Science (80-.). *357*, 1420–1424.

Jore, M.M., Lundgren, M., van Duijn, E., Bultema, J.B., Westra, E.R., Waghmare, S.P., Wiedenheft, B., Pul, U., Wurm, R., Wagner, R., et al. (2011). Structural basis for CRISPR RNA-guided DNA recognition by Cascade. Nat. Struct. Mol. Biol. *18*, 529–536.

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.

Knight, S.C.C., Xie, L., Deng, W., Guglielmi, B., Witkowsky, L.B.B., Bosanac, L., Zhang, E.T.T., El Beheiry, M., Masson, J.-B.J.-B., Dahan, M., et al. (2015). Dynamics of CRISPR-Cas9 genome interrogation in living cells. Science (80-.). *350*, 823–826.

Leenay, R.T.T., Maksimchuk, K.R.R., Slotkowski, R.A.A., Agrawal, R.N.N., Gomaa, A.A.A., Briner, A.E.E., Barrangou, R., and Beisel, C.L.L. (2016). Identifying and Visualizing Functional PAM Diversity across CRISPR-Cas Systems. Mol. Cell *62*, 137–147.

Macvanin, M., and Adhya, S. (2012). Architectural organization in E. coli nucleoid. Biochim. Biophys. Acta - Gene Regul. Mech. *1819*, 830–835.

Marraffini, L.A. (2015). CRISPR-Cas immunity in prokaryotes. Nature *526*, 55–61.

Mojica, F.J.M., Dı, C., and Garcı, J. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. 733–740.

Redding, S., Sternberg, S.H., Wiedenheft, B., Jennifer, A., Greene, E.C., Marshall, M., Gibb, B., Bhat, P., and Guegler, C.K. (2015). Surveillance and Processing of Foreign DNA by the Escherichia coli CRISPR-Cas System. Cell *163*, 1–12.

Rollins, M.F., Schuman, J.T., Paulus, K., Bukhari, H.S.T., and Wiedenheft, B. (2015). Mechanism of foreign DNA recognition by a CRISPR RNA-guided surveillance complex from Pseudomonas aeruginosa. Nucleic Acids Res. *43*, 2216–2222.

Rutkauskas, M., Sinkunas, T., Siksnys, V., Seidel, R., Songailiene, I., and Tikhomirova, M.S. (2015). Directional R-Loop Formation by the CRISPR-Cas Surveillance Complex Cascade Provides Efficient Off-Target Site Rejection. CellReports *10*, 1534–1543.

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*, 62–67.

Subach, F. V, Patterson, G.H., Renz, M., Lippincott-Schwartz, J., and Verkhusha, V. V (2010). Bright

monomeric photoactivatable red fluorescent protein for two-color super-resolution sptPALM of live cells. J. Am. Chem. Soc. *132*, 6481–6491.

Subach, F. V., Patterson, G.H., Manley, S., Gillette, J.M., Lippincott-Schwartz, J., and Verkhusha, V. V. (2009). Photoactivatable mCherry for high-resolution two-color fluorescence microscopy. Nat. Methods *6*, 153–159.

Westra, E.R., Pul, Ü., Heidrich, N., Jore, M.M., Lundgren, M., Stratmann, T., Wurm, R., Raine, A., Mescher, M., Van Heereveld, L., et al. (2010). H-NS-mediated repression of CRISPR-based immunity in Escherichia coli K12 can be relieved by the transcription activator LeuO. Mol. Microbiol. *77*, 1380–1393.

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.

Westra, E.R., Semenova, E., Datsenko, K.A., Jackson, R.N.N., Wiedenheft, B., Severinov, K., Brouns, S.J.J., Labrie, S., Samson, J., Moineau, S., et al. (2013). Type I-E CRISPR-Cas Systems Discriminate Target from Non-Target DNA through Base Pairing-Independent PAM Recognition. PLoS Genet. *9*, e1003742.

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.

Xue, C., Zhu, Y., Zhang, X., Shin, Y.K., and Sashital, D.G. (2017). Real-Time Observation of Target Search by the CRISPR Surveillance Complex Cascade. Cell Rep. *21*, 3717–3727.
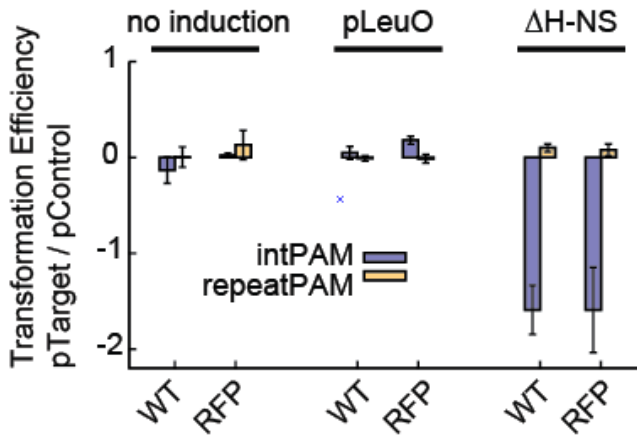
Supplementary Figures

Figure S1. Transformation efficiency assays of pTarget plasmids relative to pControl plasmids. Transformation efficiencies show no difference between wild-type (WT) strains and the strain in which *patagrfp* was fused to *cas8e* (RFP).
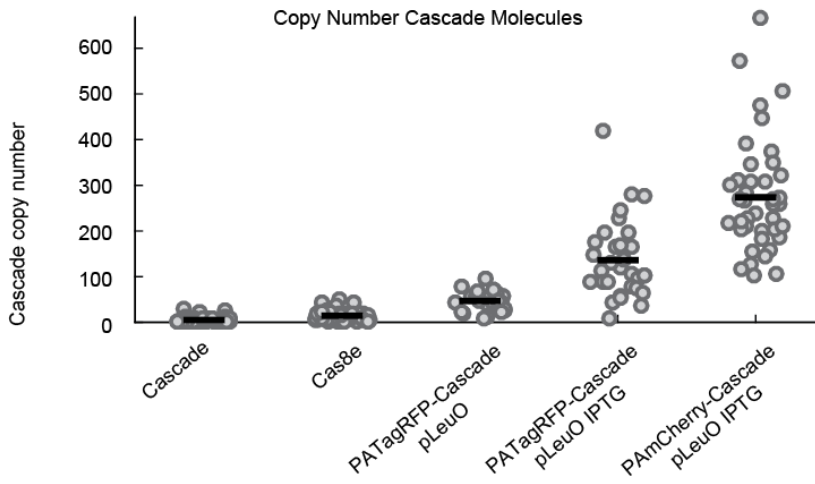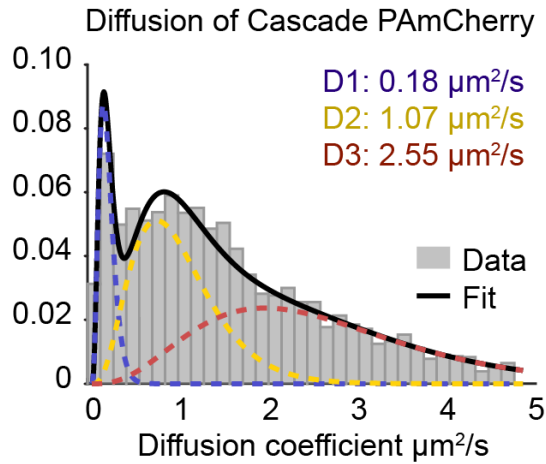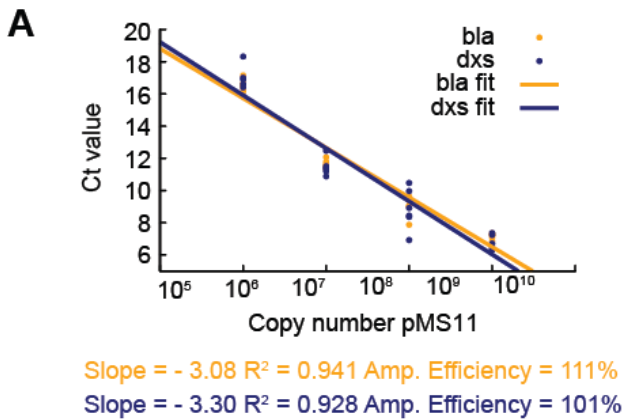

Figure S2. Cascade copy numbers per cell

Figure S3. Diffusion of Cascade PAmCherry. Relative fractions of the diffusion states for Cascade PAmCherry.; D_1 represents longer DNA interactions, D_2 represents short DNA interactions and D_3 represents freely diffusing monomeric Cas8e proteins

Strain	Ct (CN) bla	Ct (CN) dxs	PCN
K12 PATagRFP-Cascade ΔCas3 + pLeuO + pTarget intPAM	8.82 ± 0.09 (-5.75 ± 0.03)	15.15 ± 0.11 (-7.76 ± 0.04)	102 (92-114)
K12 PATagRFP-Cascade ΔCas3 + pLeuO + pTarget repeatPAM	8.85 ± 0.07 (-5.76 ± 0.02)	15.20 ± 0.22 (-7.78 ± 0.07)	104 (89-122)

Figure S4. Plasmid copy number determination. A) Calibration curve of dxs and bla primer amplification with dilution series of pMS11 (plasmid containing both dxs and bla gene). The regression of six technical replicates was used to make (regression parameters of bla and dxs gene in orange and purple respectively) The amplification efficiency was calculated from the formula: Amp.efficiency= 〖10〗 ^(-1/slope)-1. B) The Ct values of bla and dxs gene amplifications were calculated from biological triplicates. These Ct values were converted to absolute copy numbers (CN; log10 scale) by using the regression values from the calibration curve. The

plasmid copy number (PCN) was calculated by dividing the copy number of the bla gene by the copy number of the dxs gene
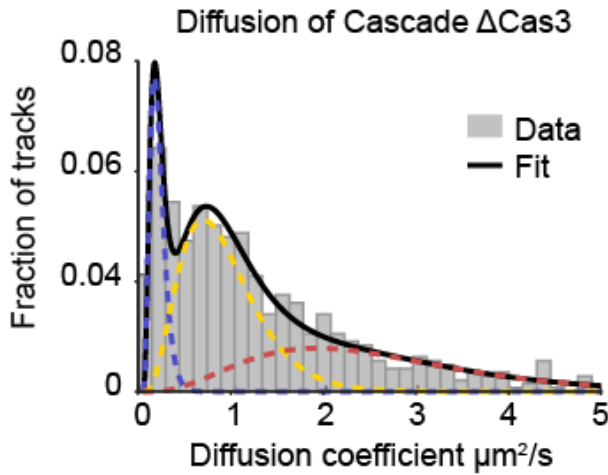

Figure S5. Diffusion of Cascade ΔCas3. Relative fractions of the diffusion states for Cascade ΔCas3.; D_1 represents longer DNA interactions, D_2 represents short DNA interactions and D_3 represents freely diffusing monomeric Cas8e proteins.
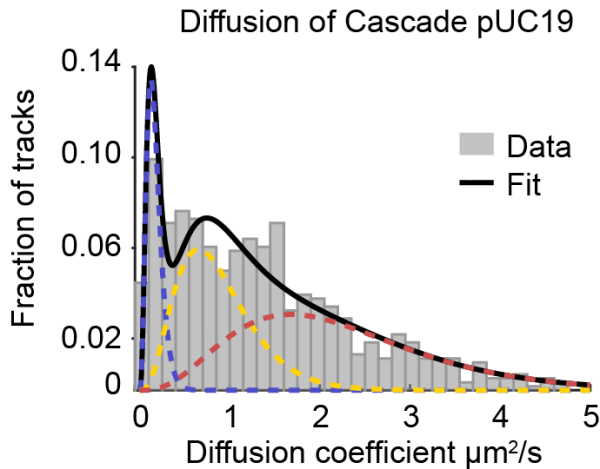

Figure S6. Diffusion of Cascade in the presence of non-target plasmid pUC19. Relative fractions of the diffusion states for Cascade; D_1 represents longer DNA interactions, D_2 represents short DNA interactions and D_3 represents freely diffusing monomeric Cas8e proteins.

Chapter 4

Complete Genome Sequence of the *Escherichia coli* Phage Ayreon

Marnix Vlot[1], Franklin L. Nobrega[1,2], Che F. A. Wong[1], Yue Liu[1],and Stan J. J. Brouns[1,2]

[1]Laboratory of Microbiology, Department of Agrotechnology and Food Sciences, Wageningen University, Wageningen, The Netherlands
[2]Department of Bionanoscience, Kavli Institute of Nanoscience, Delft University of Technology, Delft, The Netherlands

ABSTRACT

We report the whole-genome sequence of a new *Escherichia coli* temperate phage, Ayreon, comprising a linear dsDNA genome of 44,708 bp.

The phage was isolated from pond water samples in Wageningen, The Netherlands (51°58'9.2''N 5°40'43.1''E) using *Escherichia coli* strain KD471, a derivative of *E. coli* K12 derivative strain KD263 (Shmakov et al., 2014) lacking CRISPR-associated genes Cas1 and Cas2, as the host. Infected cultures were given time to allow for lysogeny. Formed lysogens could be induced both using UV exposure and Mitomycin C, demonstrating that phage Ayreon is a temperate phage. Transmission electron microscopy revealed icosahedral capsids (±57 nm) and long flexible tails (±120 nm) which are characteristic of *Siphoviridae* (Ackermann, 2003).

Phage DNA was extracted using the SDS-proteinase K protocol previously described (Sambrook and Russell, 2001). Library preparation and sequencing were performed by BaseClear (Netherlands) using the Illumina HiSeq 2500 platform. About 1,282,954 short reads were generated with a mean 3,478-fold coverage of the genome. The resulting sequences were *de novo* assembled using the CLC Genomics Workbench version 8.5.1. The contig sequences were corrected with Pilon version 1.11 (Walker et al., 2014), and linked using the SSPACE Premium scaffolder version 2.3 (Boetzer et al., 2011). Gapped regions within the scaffolds were partially closed in an automated manner using GapFiller version 1.10 (Boetzer and Pirovano, 2012). Annotation and identification of open reading frame (ORFs) were performed using the RAST annotation server (Aziz et al., 2008) followed by manual curation of all predicted proteins against the NCBI protein database using BLASTp (Altschul et al., 1990) and Pfam domain searches (Finn et al., 2016). tRNAs were predicted with tRNAscan-SE version 1.21 (Schattner et al., 2005), and promoters and terminators were identified using motif search for TTGACAN(15,18)TATAAT with a maximum of two mismatches and ARNold (Naville et al., 2011), respectively. The genome packaging strategy was predicted by phylogenetic analysis of the large terminase subunit (Merrill et al., 2016).

Phage Ayreon has a linear double-stranded DNA with a genome size of 44,708 bp and a G+C content of 50.1%, very similar to the G+C content of its bacterial host (50.8%). The phage has 59 predicted ORFs, of which 25 could not be assigned to a function. Three predicted promoters and 13 predicted Rho-independent terminators were identified; no tRNAs could be identified. Phylogenetics analysis of the large terminase subunit suggests phage Ayreon uses cohesive end site (*cos*) packaging; *cos* sites are expected to be located within ≈1 kbp upstream of the small terminase subunit (Casjens and Gilcrease, 2009). Considering this, the genome of phage Ayreon was opened so that it would begin with the small terminase subunit and end with the expected location of the *cos* site.

The highest degree of similarity was observed with phage Cdt-1 (accession no. AB285204), a cyclomodulin producing prophage (Asakura et al., 2007), with 78% coverage and 96% identity. Phage Cdt-1 contains a gene cluster associated with virulence, encoding the CdtA, CdtB, and CdtC subunits of the Cdt-1

holotoxin. Whole-genome alignment of phage Ayreon and Cdt-I prophage demonstrates that the *cdtI* gene cluster is absent from phage Ayreon. Global alignment with the attP integration site of phage Cdt-1 shows pairwise identity of >92%, indicating that phage Ayreon integrates into the gene coding for peptide chain release factor RF-3 in the host genome (Asakura et al., 2007).

Accession number. The complete genome sequence of Ayreon has been deposited in GenBank under the accession no. MF807953.

4

ACKNOWLEDGMENTS

References

Ackermann, H.W. (2003). Bacteriophage observations and evolution. Res. Microbiol. *154*, 245–251.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

Asakura, M., Hinenoya, A., Alam, M.S., Shima, K., Zahid, S.H., Shi, L., Sugimoto, N., Ghosh, A.N., Ramamurthy, T., Faruque, S.M., et al. (2007). An inducible lambdoid prophage encoding cytolethal distending toxin (Cdt-I) and a type III effector protein in enteropathogenic Escherichia coli. Proc. Natl. Acad. Sci. U. S. A. *104*, 14483–14488.

Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., et al. (2008). The RAST Server: rapid annotations using subsystems technology. BMC Genomics *9*, 75.

Boetzer, M., and Pirovano, W. (2012). Toward almost closed genomes with GapFiller. Genome Biol. *13*, R56.

Boetzer, M., Henkel, C. V., Jansen, H.J., Butler, D., and Pirovano, W. (2011). Scaffolding pre-assembled contigs using SSPACE. Bioinformatics *27*, 578–579.

Casjens, S.R., and Gilcrease, E.B. (2009). Determining DNA Packaging Strategy by Analysis of the Termini of the Chromosomes in Tailed-Bacteriophage Virions BT - Bacteriophages: Methods and Protocols, Volume 2 Molecular and Applied Aspects. M.R.J. Clokie, and A.M. Kropinski, eds. (Totowa, NJ: Humana Press), pp. 91–111.

Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., et al. (2016). The Pfam protein families database: Towards a more sustainable future. Nucleic Acids Res. *44*, D279–D285.

Merrill, B.D., Ward, A.T., Grose, J.H., and Hope, S. (2016). Software-based analysis of bacteriophage genomes, physical ends, and packaging strategies. BMC Genomics *17*, 1–16.

Naville, M., Ghuillot-Gaudeffroy, A., Marchais, A., and Gautheret, D. (2011). ARNold: a web tool for the prediction of Rho-independent transcription terminators. RNA Biol. *8*, 11–13.

Sambrook, J.F., and Russell, D.W. (2001). Molecular Cloning 3rd ed. - 3 Volumes - J. Sambrook WW.pdf. 2100.

Schattner, P., Brooks, A.N., and Lowe, T.M. (2005). The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res. *33*, 686–689.

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.

Vlot, M., Nobrega, F.L., Wong, C.F.A., Liu, Y., and Brouns, J.J. (2018). Complete Genome Sequence of the Escherichia coli Phage Ayreon. 1–2.

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., et al. (2014). Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One *9*.

Chapter 5

Bacteriophage DNA glucosylation impairs target DNA binding by type I and II but not by type V CRISPR-Cas effector complexes

Marnix Vlot[1], Joep Houkes[1], Silke J.A. Lochs[1], Daan C. Swarts[2], Peiyuan Zheng[3], Tim Kunne[1], Prarthana Mohanraju[1], Carolin Anders[2], Martin Jinek[2], John van der Oost[1], Mark J. Dickman[3] and Stan J.J. Brouns[1,4]

[1]Laboratory of Microbiology, Department of Agrotechnology and Food Sciences,Wageningen University, Wageningen, The Netherlands
[2]Department of Biochemistry, University of Zurich, Zurich, Switzerland
[3]ChELSI Institute Department of Chemical and Biological Engineering University of Sheffield, Sheffield, UK
[4]Department of Bionanoscience, Kavli Institute of Nanoscience, Van der Maasweg 9, Delft University of Technology, 2629 HZ Delft, The Netherlands

This Chapter is published as:

ABSTRACT

Prokaryotes encode various host defense systems that provide protection against mobile genetic elements. Restriction-modification (R-M) and CRISPR-Cas systems mediate host defense by sequence specific targeting of invasive DNA. T-even bacteriophages employ covalent modifications of nucleobases to avoid binding and therefore cleavage of their DNA by restriction endonucleases. Here we describe that DNA glucosylation of bacteriophage genomes affects interference of some but not all CRISPR-Cas systems. We show that glucosyl modification of 5-hydroxymethylated cytosines in the DNA of bacteriophage T4 interferes with type I-E and type II-A CRISPR-Cas systems by lowering the affinity of the Cascade and Cas9-crRNA complexes for their target DNA. On the contrary, the type V-A nuclease Cas12a (also known as Cpf1) is not impaired in binding and cleavage of glucosylated target DNA, likely due to a more open structural architecture of the protein. Our results suggest that CRISPR-Cas systems have contributed to the selective pressure on phages to develop more generic solutions to escape sequence specific host defense systems.

INTRODUCTION

In many environments, bacteria are subject to strong selective pressure by bacteriophages (phages). The number of phages exceeds that of their hosts in most ecosystems, outnumbering them up to 150-fold (Wigington et al., 2016). In response to this selective pressure, bacteria have developed a diverse palette of defense mechanisms including prevention of phage adsorption, blocking of DNA entry, restriction of phage DNA, and abortive infection mechanisms (Samson et al., 2013; Stern and Sorek, 2011). Two of these defense mechanisms, restriction-modification (R-M) and Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-CRISPR-associated (Cas) systems, act on the DNA level by selectively degrading invading DNA. Distinction between native and foreign DNA by R-M systems is often based on chemical modification (e.g. methylation) of adenines and cytosines in host genomic DNA, which protects host DNA from cleavage by specific restriction endonucleases. *Escherichia coli* K-12 harbors the type I restriction enzyme EcoK encoded by the *hsdR* gene that cleaves unmethylated DNA at 5'-AAC(N_6)GTGC and 5'-GCAC(N_6)GTT sequences and three type IV R-M systems encoded by *McrA*, *McrBC*, and *mrr* that cleave methylated DNA (Loenen and Raleigh, 2014; Sain and Murray, 1980). Together, these R-M systems limit horizontal gene transfer.

Recognition of invading DNA by type I, II and V CRISPR-Cas systems is based on base pairing of the invader DNA with CRISPR RNAs (crRNAs). These host encoded crRNAs guide effector complexes to complementary target DNA (also called protospacer) which is subsequently cleaved by the effector complex or by recruited nucleases (Makarova et al., 2015a; Mohanraju et al., 2016). Even though all different types of CRISPR-Cas systems have a common role in immunity, they are structurally and mechanistically diverse. Based on this diversity, CRISPR systems are divided into two different classes and six different types (Makarova et al., 2015b; Shmakov et al., 2015). Type I CRISPR systems are the most abundant CRISPR type in nature (Makarova et al., 2015a), and

comprise a multiprotein crRNA-effector complex (named Cascade) and a nuclease (Cas3) to bind target DNA. After target DNA recognition, Cascade recruits Cas3 that mediates target DNA degradation (Jore et al., 2011; Westra et al., 2012). The best characterized example of type II CRISPR-Cas systems is the Cas9 nuclease, a single effector protein which facilitates both crRNA-mediated DNA binding and DNA cleavage (Makarova et al., 2015b). Recently, type V-A CRISPR systems were discovered (Zetsche et al., 2015). Like type II systems, type V-A systems employ a single effector enzyme named Cas12a, which provides an interesting alternative to Cas9 in genome editing (Shmakov et al., 2017; Zetsche et al., 2015).

In response to the selective pressure posed by anti-viral defense systems in bacteria, phages have evolved several mechanisms to escape anti-viral defense systems. Phages can evade sequence-specific host defense systems by mutating target sequences (Deveau et al., 2008; McGrath et al., 1999; Semenova et al., 2011). While this allows for efficient escape from restriction endonucleases, mutations in DNA sequences targeted by CRISPR-Cas systems can trigger a process called priming, which leads to an accelerated update of the CRISPR memory repertoire (Jackson et al., 2017). Furthermore, phages may use recombination of their genomes to get rid of CRISPR target sites (Andersson and Banfield, 2008; Paez-Espino et al., 2015). In addition to mutation and recombination of DNA, phages can escape R-M systems by expressing inhibitory proteins. Such proteins can inhibit R-M immunity by degradation of R-M cofactors, masking of restriction sites, or modification of phage DNA (Samson et al., 2013). Additional strategies include DNA mimicry, as exemplified by the Ocr protein of phage T7 that mimics DNA to sequester EcoKI (Krüger and Bickle, 1983; Walkinshaw et al., 2002). Recently, phage-encoded proteins that inhibit CRISPR-Cas systems have been characterized (Bondy-Denomy et al., 2013; Pawluk et al., 2014, 2016a). Inhibition of CRISPR-Cas systems by anti-CRISPR (ACR) proteins occurs via distinct mechanisms. Two Acr proteins (AcrF1 & AcrF2) were found to bind the type I-F Csy complex, inhibiting the binding of target DNA, while another Acr protein (AcrF3) binds to Cas3, blocking its interaction with Cascade to inhibit target degradation (Bondy-Denomy et al., 2015). Three recently identified families of ACRs are encoded by mobile elements in bacteria and are shown to inhibit *Neisseria meningitidis* Cas9 by direct binding (Pawluk et al., 2016b). ACR proteins AcrIIA1-4 encoded by *Listeria monocytogenes* prophages prevent Cas9 binding and can be used to regulate the genome engineering activity of *Streptococcus pyogenes* Cas9 (Rauch et al., 2017). ACR proteins that inhibit SpyCas9 have also been found in virulent bacteriophages (Hynes et al., 2017).

Another way to escape host defense is found in T-even phages that infect *E. coli* (Weigele and Raleigh, 2016). Phage T4 has evolved a pathway to bypass the R-M systems of *E. coli* by substitution of its genomic cytosines with 5-hydroxymethylcytosines (5-hmC) (Figure 1A and 1B). The dNTP synthesis complex (DSC) which comprises both host- and phage-encoded enzymes performs this substitution before DNA synthesis (Kim, 2005). Degradation of the host genome by T4-derived DenA and DenB nucleases provides an increased pool of dCTP, which is converted to dCMP by the phage-encoded dCTPase gp56. Subsequently, dCMP is converted into 5-hydroxymethyl-dCMP (5-hmdCMP) by phage-encoded Deoxycytidylate 5-hydroxymethyltransferase gp42 (Mathews et al., 1979). Conversion of 5-hmdCMP to 5-hmdCDP is then catalyzed by the phage-encoded dNMP

5

kinase gp1. Host-encoded nucleoside diphosphate kinase Ndk catalyzes the final conversion to 5-hmdCTP, which is incorporated by T4 DNA polymerase into the DNA as 5-hmC. To avoid immunity escape by 5-hmC modification, *E. coli* encodes the McrBC system, which specifically degrades 5-hmC modified DNA (Raleigh and Wilson, 1986). However, T4 is impervious to this modification dependent system because 5-hmC residues are further modified with glucose moieties derived from uridine diphosphate glucose (UPDG). These are covalently attached to the 5-hydroxyl group of 5-hmC by phage-encoded alpha- and beta-glucosyl transferases, yielding glucosyl-5-hydroxymethylcytosine (5-ghmC) (Lehman and Pratt, 1960) (Figure 1C). As DNA glucosyl-5-hydroxymethylation provides an effective way to escape immunity by most R-M systems, the question arises whether it also protects against certain types of CRISPR-Cas immunity, and what the mechanistic basis for protection from CRISPR interference would be.
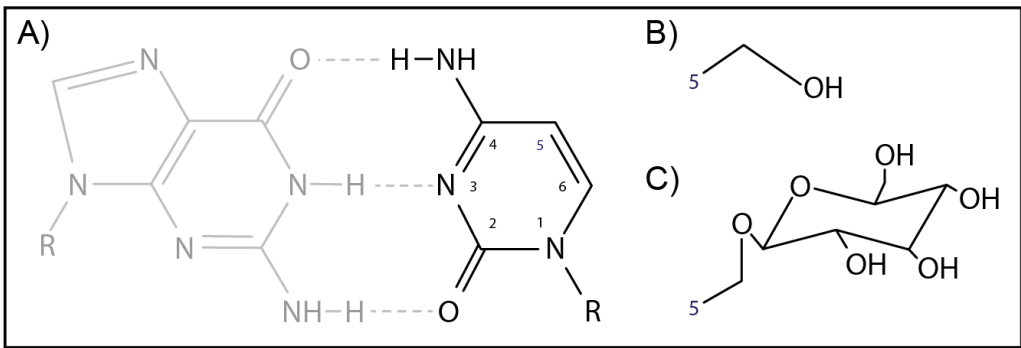

Figure 1. Modifications of nucleobases in phage T4 DNA. (A) Cytosine present in phage T4(C)**. (B)** 5-hydroxymethylation of cytosine present in phage T4(hmC). **(C)** Glucosyl-5-hydroxymethylation present in phage T4(ghmC).

We have tested the effect of DNA glucosylation on the activity of type I-E, type II-A and type V-A CRISPR-Cas systems. We demonstrate that T4 with 5-ghmC DNA can escape type I-E and II-A interference by reducing target binding affinity of Cascade and Cas9, respectively. Interestingly, 5-ghmC modifications do not lower target binding affinity and cleavage efficiency of the type V-A effector nuclease Cas12a. The structural basis for the observed differences as well as potential applications for the described results are discussed.

MATERIAL AND METHODS

Bacterial strains and plasmid construction

Cells were made chemically competent using the RuCl method and transformed by applying a heat-shock as described in the QIA expressionist handbook (QIAGEN). For the experiments with Type I-E CRISPR-Cas, *E. coli* T7 Express (NEB) was transformed with pWUR400, pWUR797 and pWUR800, pWUR801 or pWUR802 (table

S1) and used for plaque assays with phage T4. For the *in vivo* experiments with Type II-A CRISPR-Cas, *E. coli* T7 Express was transformed with pWUR805 and pWUR806, pWUR809 or pWUR810 (table S1).

Bacteriophage strains and propagation

Phage T4(ghmC) (CBS-KNAW Biodiversity Centre, Utrecht, Netherlands) was propagated in *E. coli* B834 (Su⁻). Phage T4(C) (dCTPase⁻, *denA*, *denB*, *alc*⁻) was kindly provided by Prof. Elisabeth Kutter (Evergreen State College) and propagated in *E. coli* B834 or CR63 (Su⁺). T4(hmC) (α-, β-glucosyltransferase) was kindly provided by Peter Weigele (NEB, Ipswich) and propagated in *E. coli* T7 Express (mcrC-mrr). The phage T4(C) mutant we used in our analysis was a quintuple mutant of genes *alc*, *denB*, *denA* and *dCTPase.* Genes *denA* and *denB* encode endonuclease II and IV respectively and *denA* and *denB* mutations allow the phage to contain C DNA. The *alc* mutation allows for the production of late proteins, enabling bursts of phage.

Plaque assays and efficiency of plating calculation

E. coli strains were grown at 37 °C in Luria Broth (LB; 5 g/L yeast extract, 10 g/L tryptone, 5 g/L NaCl) at 180 rpm or on LB-agar plates containing 1.5% (wt/vol) agar. When required, medium was supplemented with the following: ampicillin (Amp; 100 μg/L), chloramphenicol (Cam; 25 μg/mL), kanamycin (Kan; 50 μg/mL) or streptomycin (Str; 50 μg/mL). Bacterial growth was assessed spectrophotometrically at 600 nm (OD_{600}). To induce *cas* gene expression, IPTG (isopropyl β-D-1 thiogalactopyranoside) was added to the final concentration of 1 mM when the bacterial culture reached and OD_{600} of approximately 0.4. After 30 minutes of Cas protein and crRNA production, the cells were used in double agar layer plaque assays with phage T4(ghmC), T4(hmC), and T4(C). Plaque assays were performed in triplicate. The sensitivity of the host to phage infection was calculated as the efficiency of plaquing, which is the plaque count ratio of a strain containing an anti-T4 CRISPR, to that of a strain containing a CRISPR with non-targeting spacers. Anti-T4 spacers were designed by picking a PAM-flanked protospacer in two essential T4 genes (gene 19 and gene 22). CRISPR protospacer sequences are provided in table S2. A two tailed t-test was performed to calculate whether the differences were significant. The presence or absence of modifications of T4 DNA was verified by analysis using restriction enzymes. Isolated DNA was incubated with enzymes that are either sensitive or not sensitive to 5-hmC or 5-ghmC. Our analysis confirmed that our T4 stock contains 5-ghmC DNA and our T4(C) stock contains C DNA. The presence of 5-hmC in phage T4(hmC) was verified by plating on *E. coli* expressing the McrBC restriction enzyme, which is active on 5-hmC DNA but not on 5-ghmC DNA. Inability of T4(hmC) to propagate reveals the presence of 5-hmC DNA.

Design and synthesis of modified dsDNA

All *in vitro* assays were done using a 98 bp dsDNA target containing the spacer 8 (sp8) sequence (table S2). The target was amplified using the primers BG8415 & BG8416 (table S2). PCR products were generated using both Q5 high-fidelity DNA polymerase and Taq DNA polymerase (NEB). hm⁵-dCTP (Bioline) was used in the PCR reactions to obtain hydroxymethylated targets. PCR conditions were optimised for each of the modified dNTPs

used. Typical PCR conditions were as follows: 50 µl PCR reactions with final concentrations dNTPs 200 µM, primers 0.5 µM, DNA template 100 ng, 1.25 U Taq polymerase. The following parameters were used: denaturation at 95 °C for 30 s followed by 30 cycles of 95 °C for 30 s, 46 °C for 60 s, 68 °C for 60 s with a final extension at 68 °C for 3 minutes. For Q5 DNA polymerase (1 U) the following parameters were used; 98 °C 30 s denaturation, followed by 30 cycles at 98 °C for 30 s 61 °C for 30 s, 72 °C for 20 s with a final extension at 72 °C for 2 mins. The PCR products were subsequently purified by QIAprep Spin Miniprep Kit (Qiagen) following the manufacturer's instructions. Purified samples were quantified using a NanoDrop 2000C (Thermo Scientific) and analysed by native PAGE.

Three different spacers were designed for the target sequence for Type I-E, Type II-A and Type V CRISPR systems. Each of these spacers was designed to have a suitable PAM for the respective CRISPR Type (Leenay et al., 2016). The designed spacers can be found in table S2.

Enzymatic glucosylation of the purified 5-hmC dsDNA was used to synthesize DNA with glucosylated 5-hmC. T4 Phage β-glucosyltransferase (NEB) 1 U was incubated with approximately 1 µg of 5-hmC dsDNA, 40 µM UDP-glucose at 37°C for 16 hours. The DNA products were subsequently purified using QIAprep Spin Miniprep Kit (Qiagen) and analysed by native PAGE.

DNA targets containing 5-hmC modifications in the PAM region were produced by annealing oligonucleotides BG6508-6510 to BG6506. Oligonucleotides BG6508-6510, containing 5-hydroxymethyldeoxycytidine, were purchased from Integrated DNA Technologies (IDT).

Target DNA was subsequently glucosylated as described above.

Radiolabeling of target DNA

Target DNA containing unmodified cytosine, 5-hmC and 5-ghmC were 5' radioactively labeled using T4 Polynucleotide Kinase (PNK) (NEB). Conditions are 1x PNK reaction buffer (NEB), 50 pmol [γ-32P] ATP (Perkin Elmer) and 25 pmol target DNA in a 50 μL reaction. The reaction was carried out at 37 °C for 30 minutes and subsequently inactivated at 65 °C for 20 minutes.

Preparation of crRNA and sgRNA

The Cas9 sgRNA was made by *in vitro* transcription (IVT) of partly overlapping primers. DNA oligonucleotides (oligos) used for IVT were PAGE purified (table S3). IVT was performed using the HiScribe™ T7 Quick High Yield RNA Synthesis Kit (NEB) in 30 μL containing 10 μL NTP buffer mix, 1 μg DNA and 2 μL T7 polymerase. The reaction was incubated for 4 hours at 37 °C, thereafter 20 μL Milli-Q and 2 units DNase I (NEB) were added to remove target DNA for 15 minutes at 37 °C. 10 μL 6x loading dye (Thermo Scientific) was added to both samples for separation and isolation from a denaturing PAGE gel. The samples were heated for 5 minutes at 95 °C and loaded on a 7 M urea 8% PAGE gel, together with a Low Range ssRNA Ladder (NEB). The gel was run for 3 hours at 15 mA in 1x TBE. After 3 hours the gel was stained using 1x SYBR gold (Thermo Scientific) in 1x TBE for 10 minutes. For both RNA oligos the band at 103 nt was cut from gel.

The gel fragments were ground using a pipet tip. 1 mL of RNA elution buffer (0.5 M NaAc, 10 mM $MgCl_2$, 1 mM EDTA, 0.1% SDS) was added and incubated for 2 hours at 37 °C. The sample was transferred to a gel extraction column (Zymoclean) and centrifuged at 13,200 x g for 1 minute. The flow-through was purified on a Microcon 30 column (Millipore) by centrifugation at 16,000 x g for 20 minutes at 4 °C. RNA was washed twice using 500 μL Milli-Q and centrifuged at 16,000 x g for 20 minutes at 4 °C. RNA was resuspended on a filter column in 50 μL Milli-Q, and retrieved by centrifuging the column upside down for 5 minutes at 6,000 x g. RNA quantity and quality were analysed using NanoDrop and Denaturing PAGE. The 48 nt crRNA and non-complimentary crRNA for Cas12a assays were synthesized RNA Oligos (Sigma, table S3).

Cascade and Cas3 degradation reactions

Cascade purification was performed as described in (Jore et al., 2011). Cas3 was purified as described in (Mulepati and Bailey, 2013). Target DNA constructs were incubated with 100 μM Cascade and 10 μM Cas3 in reaction buffer (10 mM HEPES pH 7.5, 60 mM KCl, 50 μM $CoCl_2$, 10 mM $MgCl_2$ and 2 mM ATP) at 37 °C for 2 hours. Reactions were stopped by transferring the tubes to ice and addition of 4 μL 6x loading dye (Thermo Scientific). Reaction products were run on a 6% acrylamide gel (with 7M Urea and 1x TBE). The gel was run in 1x TBE for approximately 4 hours at 15 mA and subsequently exposed for 48 hours in a phosphor imaging cassette (Molecular Dynamics) at -20 °C. The phosphor imaging cassette was scanned using a Personal Molecular Imager (Bio-Rad).

Cas9 cleavage reactions

For Cas9 Cleavage reactions 33 nM Cas9 (NEB), 120 nM sgRNA and 1x Cas9 buffer (NEB) were pre-incubated for 20 minutes at 37 °C. Subsequently 5' ^{32}P-radiolabeled target DNA was added to a final concentration of 3 nM and the reaction was incubated for 2 hours at 37 °C. 30 µL 2x formamide loading dye (95% formamide, 0.125% bromophenol blue) was added and the complete samples were heated to 95 °C for 5 minutes and loaded on an 8% acrylamide gel (with 7M Urea and 1x TBE). The gel was run in 1x TBE for approximately 4 hours at 15 mA and subsequently exposed for 16-48 hours in a phosphor imaging cassette (Molecular Dynamics) at -20 °C. The phosphor imaging cassette was scanned using a Personal Molecular Imager (Bio-Rad).

Cas12a protein purification

A codon optimized Cas12a gene was cloned into a bacterial expression vector [6-His-TEV-Cas12a, a pET-based vector that was a gift from Scott Gradia (Addgene plasmid # 29653)]. One liter of LB growth media with 100 µg/mL ampicillin was inoculated with 10 mL overnight culture Rosetta (DE3) (EMD Millipore) cells containing the Cas12a expression construct. Growth media plus inoculant was grown at 37 °C until the cell density reached 0.5 OD600, then the temperature was decreased to 20°C. Growth was continued until OD600 reached 0.6 when a final concentration of 500 µM IPTG was added to induce Cas12a expression. Expression took place for 14-18 hours before harvesting cells and freezing them at -20 °C until purification.

Cell paste was suspended in 20 mL of Lysis Buffer (50 mM NaH2PO4 pH 8, 500 mM NaCl, 1 mM 2-Mercaptoethanol, 10 mM imidazole) supplemented with protease inhibitors (Roche complete, EDTA-free) and lysozyme. Once homogenized, cells were lysed by sonication (Bandelin Sonoplus) and then centrifuged at 16,000g for 1 hour at 4 C to clear the lysate. The lysate was filtered through 0.22 µm filters (Mdi membrane technologies) and applied to a nickel column (HisTrap HP, GE lifesciences), washed and then eluted with 250mM imidazole. Fractions containing protein of the expected size were pooled and dialyzed overnight into the dialysis buffer (250mM KCl, 20mM HEPES/KOH, 1mM DTT). After dialysis, sample was diluted 1:1 in 10 mM HEPES/KOH pH 8, and loaded on a heparin FF column pre-equilibrated in IEX-A buffer (150mM KCl, 20mM HEPES/KOH pH 8). Column was washed with IEX-A and then eluted with a gradient of IEX-C (2 M KCl, 20 mM HEPES/KOH pH 8). The sample was concentrated to 700 µL prior to loading on a gel filtration column (HiLoad 16/600 Superdex 200) via FPLC (AKTA Pure). Fractions from gel filtration were analyzed by SDS-PAGE; fractions containing Cas12a were pooled and concentrated to 200 µL (50 mM Tris-HCl pH 8, 2 mM DTT, 5% glycerol, 500 mM NaCl) and either used directly for biochemical assays or frozen at -80°C for storage.

Cas12a cleavage reactions

Cleavage assays were done by pre-incubating 100 nM Cas12a with 400 nM crRNA in 1x Cas9 buffer (NEB) for 20 minutes at 37 °C. Thereafter 8 nM 5' ^{32}P radiolabeled target DNA was added in a 30 µL reaction and incubation was done at 37 °C for 2 hours. 30 µL 2x formamide loading dye was added and the complete samples were heated to 95 °C for 5 minutes and loaded on an 8% acrylamide gel (with 7 M Urea and 1x TBE). The gel

was run in 1x TBE for approximately 4 hours at 15 mA and subsequently exposed for 16-48 hours in a phosphor imaging cassette (Molecular Dynamics) at -20 °C. The phosphor imaging cassette was scanned using a Personal Molecular Imager (Biorad).

Cascade Electrophoretic mobility shift assays

Cascade electrophoretic mobility shift assays were performed based on (Künne et al., 2015). Incubation reactions were performed with Cascade-crRNA (sp8) complex concentrations of 0, 1, 3, 10, 30, 100, 300 and 1000 nM. Dilutions of the Cascade stock solution (2.78 mg/ml) were made in equilibration buffer (20 mM HEPES pH=7.5, 75 mM NaCl, 1 mM DTT). 3 nM 5' ^{32}P radiolabeled target was incubated with Cascade complex in 1x Equilibration buffer in a 30 μL reaction. The reaction was incubated for 30 minutes at 37°C. 6 μL 6x loading dye (Thermo Scientific) was added and the complete samples were loaded on a native 8% acrylamide gel containing 1x TBE. The gel was run at 15 mA for 2-3 hours, after which the gel was placed in a phosphor imaging cassette and stored for 16-48 hours at -20 °C. The phosphor imaging cassette was scanned using a Personal Molecular Imager (Bio-Rad).

dCas9 and dCas12a electrophoretic mobility shift assays

Electrophoretic mobility shift assays with Cas9 and Cas12a were done using catalytically dead versions of the proteins, dCas9 and dCas12a (Jinek et al., 2012; Zetsche et al., 2015). 1 μM dCas9 and dCas12a were incubated with 4 μM sgRNA and crRNA respectively in binding buffer (20 mM HEPES, pH 7.5, 250 mM KCl, 2 mM MgCl$_2$, 0.01% Triton X-100, 0.1 mg/ml bovine serum albumin, 10% glycerol) for 10 minutes at 25 °C as described in (Anders et al., 2015). A dilution series was made with 0.1, 0.3, 1, 3, 10, 30, 100 and 300 nM final protein concentration using binding buffer. 3 nM 5' ^{32}P-radiolabeled target DNA was added in a 30 μL reaction and incubated for 10 minutes at 37 °C. 10 μL 6x loading dye (Thermo Scientific) was added and the complete samples were loaded on a 12% native acrylamide gel containing 1x TBE. The gel was run in 1x TBE at 15 mA for 2-3 hours. Subsequently, the gel was exposed for 16-48 hours in a phosphor imaging cassette (Molecular Dynamics) at -20 °C. The phosphor imaging cassette was scanned using a Personal Molecular Imager (Biorad)

Structural Modeling of modified target DNA binding to Cascade, Cas9 and Cas12a

Structural coordinates of glucosyl-5-hydroxymethyl modified thymine was isolated from an existing model (PDB 308D; (Gao et al., 1997)). The modified nucleotide was statically modelled onto target strand (TS) and non-target strand (NTS) nucleotides in structures of *E. coli* Cascade (PDB: 5H9F; (Hayes et al., 2016)), *S. pyogenes* Cas9 (PDB: 5F9R;(Jiang et al., 2016)), and *F. novicida* Cas12a (PDB: 5NFV;(Swarts et al., 2017)) in WinCoot 0.8.3 using Least Squares Fit Superpose. Clashes between the DNA modifications and the polypeptide chain were identified using the MolProbity (Chen et al., 2010) output of the Comprehensive Validation Tool within Phenix (Adams et al., 2010). Identified clashes between the glucosyl-5-hydroxymethyl modification and polypeptide residues are schematically indicated in Figure .

RESULTS

Phages escape CRISPR-Cas interference by DNA glucosylation

The majority of phages that contain 5-ghmC modification in their DNA infect gammaproteobacteria (Warren, 1980). Hence, anti-viral defense systems that are effective against 5-ghmC DNA-containing phages may be expected amongst this class of bacteria. To this end we assessed the ability of the native type I-E CRISPR-Cas system of the gammaproteobacterium *E. coli* to provide resistance against phage T4. Two plasmids expressing crRNAs against essential T4 genes were constructed (table S2). The crRNAs target T4 gene 19 (crRNA19), encoding a tail tube protein, and T4 gene 22 (crRNA22), encoding the prohead core protein. *E. coli* cells expressing Cascade and Cas3 were transformed with the plasmid encoding the crRNAs and were subsequently tested for their sensitivity to infection by a panel of T4 phages. Phage T4(C) contains unmodified cytosine, T4(hmC) contains 5-hmC, and T4(ghmC) contains 5-ghmC. Infectivity of the phages was normalized against *E. coli* cells containing a CRISPR array harboring a non-targeting spacer (Figure). Efficiency of plaquing (EOP) by T4(C) was reduced 10^4-fold by cells expressing targeting Cascade as compared to cells expressing non-targeting Cascade. Similarly, EOP by T4(hmC) was reduced 10^5-fold by cells expressing targeting Cascade. In contrast, EOP by T4(ghmC) was lowered only 10-fold by cells expressing targeting Cascade. We therefore conclude that CRISPR-Cas type I-E resistance is severely inhibited by 5-ghmC modifications, but not by 5-hmC modifications of phage DNA.
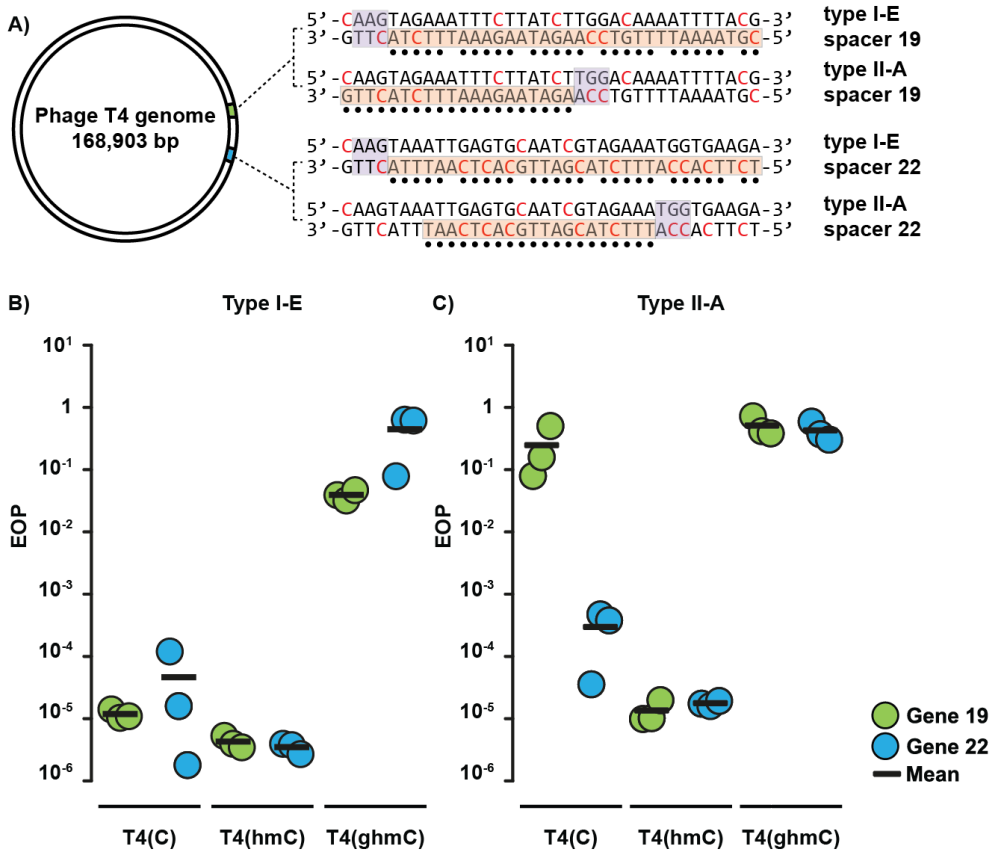
Figure 2. Efficiency of plaquing (EOP) assays of T4 phages on *E. coli* cells expressing CRISPR-Cas systems targeting T4 genes 19 or 22. (A) Representation of the circular genome of bacteriophage T4. Positions of protospacer sequences and PAMs are shown (indicated by orange bars and purple bars respectively). Glucosyl-5-hydroxymethylcytosines are indicated in red and nucleotides that base pair with crRNA are indicated by black dots. **(B)** EOP of T4 phages on *E. coli* cells expressing targeting Cascade complexes normalized to EOP on non-targeting strains. **(C)** EOP of T4 phages on *E. coli* cells expressing targeting Cas9 proteins normalized to EOP on non-targeting strains.

Next, we tested whether the type II-A CRISPR-Cas system of *Streptococcus pyogenes* mediates immunity against T4 phages in *E. coli*. *E. coli* cells expressing Cas9 and tracrRNA were transformed with plasmids encoding crRNAs targeting T4 genes 19 and 22 and subsequently tested for sensitivity to infection by phages T4(C), T4(hmC), and T4(ghmC) (Figure B). Efficiency of plaquing (EOP) by T4(C) and T4(hmC) is decreased 10^4-fold in cells expressing targeting Cas9 compared to non-targeting Cas9. Although both chosen crRNAs mediated efficient targeting of T4(hmC), crRNA19 provided lower resistance than crRNA22, indicating spacer-specific interference efficiencies on normal DNA. In contrast to EOP of T4(C) and T4(hmC), EOP of T4(ghmC) is similar between cells

expressing targeting or non-targeting Cas9. We therefore conclude that, like type I-E systems, type II-A CRISPR-Cas systems are severely impaired by phage DNA glucosylation.

Escaping phages carry seed mutations

We noticed that each of the three T4 phages generated plaques on CRISPR targeting strains, indicating that these phages escaped CRISPR-Cas immunity. To determine the nature of this escape, we sequenced the target site of the different plaque-forming T4 phages (table S4, S5). The majority (23/34) of T4(C) and T4(hmC) phages that escaped from type I-E or type II-A immunity contained a single mutation in the target sequences. Almost all mutations (20/23) were located in the seed region of the target sequences. This is a region in the target site that is highly intolerant for mismatches during CRISPR mediated immunity (Künne et al., 2014). Surprisingly, for both systems none of the mutations were located in the PAM (Protospacer adjacent motif). The PAM is a short motif adjacent to the DNA target sequence (PAM) important for efficient target recognition and avoidance of self-targeting, and is commonly prone to accumulating escape mutations.

Thus, several target DNA sequences in T4(C) and T4(hmC) phages escaping type I-E or II-A systems did not contain mutations. To ensure these phages were not revertants in which the T4(ghmC) genotype was restored, the dCTPase gene of the escape mutants was PCR amplified and sequenced. This confirmed that the T4(C) phages still had the dCTPase deficient genotype, which prevents incorporation of 5-ghmC, and suggests the mutants escaped CRISPR immunity in another way (Kutter and Wiberg, 1969). In contrast to T4(C) and T4(hmC) phages that successfully infected *E. coli*, none of the T4(ghmC) phages infecting *E. coli* expressing type I-E and type II-A systems had mutations in the DNA targeted by the CRISPR systems. Sequencing of escape mutant phages shows that T4(C) and T4(hmC) phages mostly escape CRISPR immunity by mutating the target sequence. In contrast, T4(ghmC) infection does not require mutation of DNA for efficient infection. This indicates that the 5-ghmC modification itself allows efficient escape from type I-E and type II-A CRISPR-Cas systems

DNA glucosylation inhibits Cascade target binding

To understand the molecular basis for escape from CRISPR-Cas immunity, we reconstituted CRISPR interference with type I-E components *in vitro* (Figure 3A). After pre-incubation of crRNA-bound Cascade with unmodified, 5-hmC, or 5-ghmC target DNA, the capacity of Cas3 to degrade these targets was assessed. Cas3, in conjunction with Cascade, was able to degrade 5-hmC DNA and the unmethylated control DNA, but no degradation of 5-ghmC target DNA was observed (Figure 3B). To determine whether this was due to inhibition of Cas3 or rather due to inhibition of target binding by Cascade, we performed electrophoretic mobility shift assays (EMSAs) with Cascade (Figure 3C). The EMSAs showed that sequence-specific binding of targets containing 5-hmC is consistently reduced compared to targets containing unmodified cytosines. In addition, we were unable to observe binding of 5-ghmC targets, indicating that phages with glucosylated DNA escape from type I-E CRISPR immunity by preventing binding of Cascade to the target DNA. Even though reduced binding affinity of Cascade

to targets containing 5-ghmC is reduced, thereby allowing escape from CRISPR-Cas interference, inability of Cas3 to cleave glucosylated DNA cannot be excluded even when stable binding of target DNA by Cascade would occur.
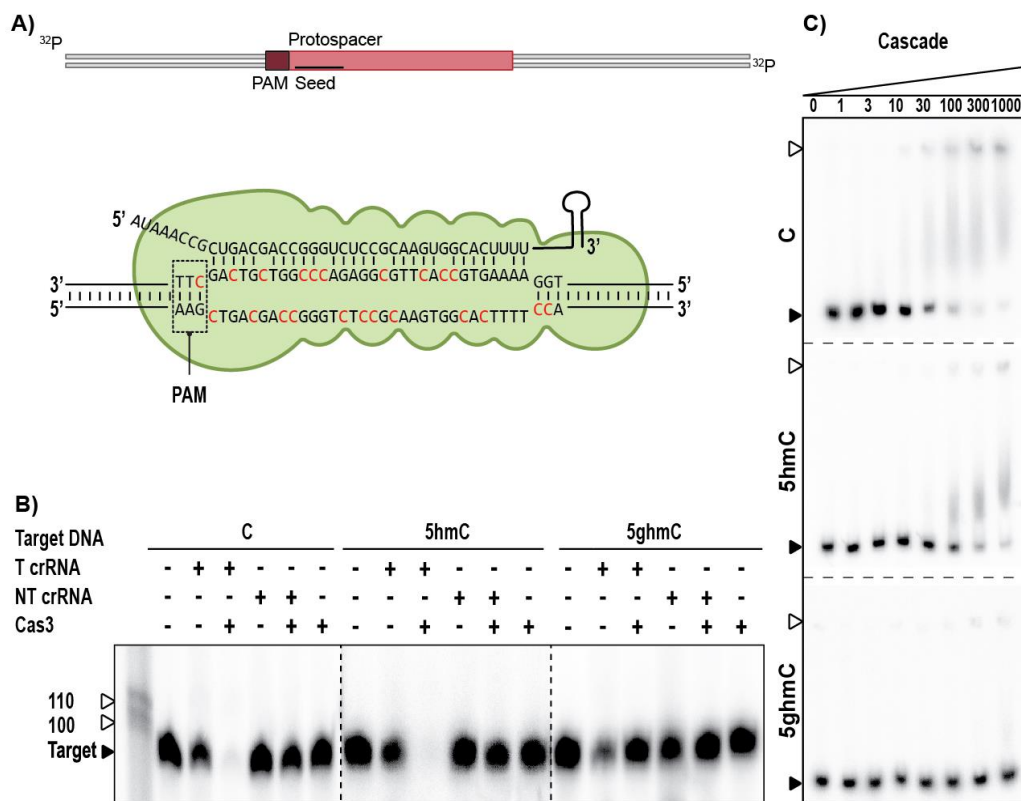

Figure 3. Effect of T4 DNA modifications on type I-E CRISPR-Cas sgRNA mediated DNA targeting. (A) Schematic of DNA targeting and R-loop formation by Cascade. Modified cytosine residues are indicated in red. **(B)** Cleavage assay of Cas3 in conjunction with Cascade on 98 bp modified targets, indicated by black arrow. The marker is indicated by white arrows. Cascade effector complexes are loaded with either targeting crRNA (T crRNA) or non-targeting crRNA (NT crRNA). Restriction products of Cas3 are of undefined length. **(C).** Electrophoretic Mobility Shift Assay (EMSA) of Cascade on target DNA containing C, 5-hmC or 5-ghmC (indicated by black arrow) at increasing protein concentrations [nM]. Fraction of bound target is indicated by white arrows, dotted lines represent separate gels.

Glucosylated DNA affects Cas9 binding and prevents target cleavage

Next, we investigated if DNA modifications can prevent DNA targeting by type II-A Cas9 *in vitro* (Figure 4A). The ability of Cas9 to degrade unmodified, 5-hmC and 5-ghmC targets was assessed by incubation of a reconstituted Cas9-sgRNA-loaded complex with target DNAs. As for Cascade/Cas3, Cas9-sgRNA complexes were able to

cleave the 5-hmC DNA and the unmodified control DNA, but not the 5-ghmC DNA (Figure 4B). To determine the molecular basis for inhibition of 5-ghmC DNA cleavage, EMSAs were conducted using a catalytically inactive version of Cas9 (dCas9) (Figure 4C). Surprisingly, the assays showed that the sequence-specific binding of 5-hmC target DNA is consistently increased compared to binding of unmodified DNA targets. In contrast, the binding affinity of 5-ghmC target DNA is consistently decreased compared to targets with unmodified DNA. Combined, these results indicate that 5-ghmC partially affects type II-A interference by lowering the binding affinity of Cas9-sgRNA complexes to the target DNA. Additionally, glucosylation may affect the ability of Cas9 to restrict target DNA by protecting the scissile bonds from the catalytic residues, rendering Cas9-sgRNA complex-mediated interference even less effective.
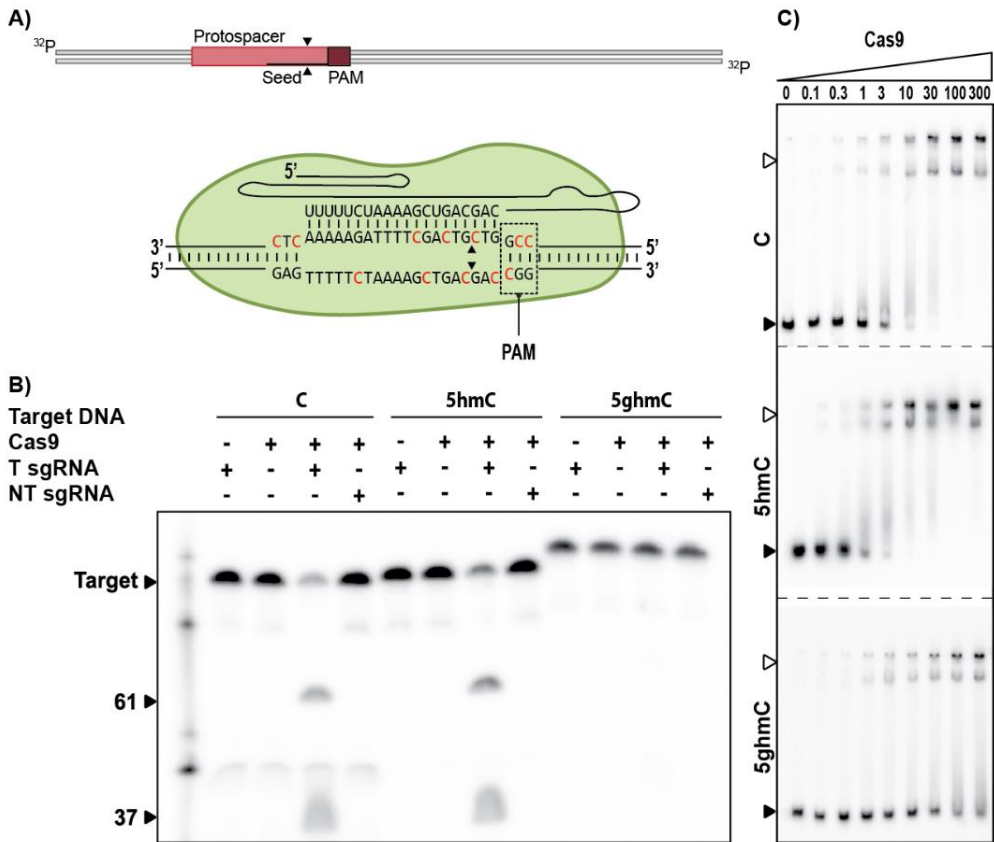

Figure 4. Effect of T4 DNA modifications on type II-A CRISPR-Cas sgRNA mediated DNA targeting. (A) Schematic of DNA targeting by Cas9. Modified cytosine residues are indicated in red. Cleavage sites are indicated by black arrows**. (B)** Cleavage assay of Cas9 on 98 bp modified targets (indicated by black arrow). Cas9 is loaded with either targeting sgRNA (T sgRNA) or non-targeting sgRNA (NT sgRNA). Restriction products of Cas9 are 61 and 37 bp. **(C)** EMSA of dCas9 on target DNA containing C, 5-hmC or 5-ghmC (indicated by black arrow) at increasing protein concentrations [nM]. Fraction of bound target is indicated by white arrows.

Glucosyl-5-hydroxymethylation of the PAM does not abolish cleavage by Cas9

We next investigated whether glucosyl-5-hydroxymethylation of cytosines in the PAM region is sufficient to inhibit DNA cleavage by Cas9. Interactions between the target DNA and Cas9 are initiated at the PAM by two arginines probing the major groove for the two guanines in the 5'-NGG-3' PAM (Sternberg et al., 2014). Upon PAM recognition, Cas9 sequentially unwinds the protospacer dsDNA and forms an R-loop structure between the target DNA strand and the guide RNA. Based on their location in the major groove of the DNA, we hypothesized that glucosyl-5-hydroxymethylation of the two cytosines (3'-NCC-5') complementary to the guanines of the PAM could inhibit the PAM recognition process, and could abolish target cleavage. We utilized target DNA containing 5-hmC or 5-ghmC modifications either proximal to the protospacer (3'-N^{hm}CC-5'), distal to the protospacer (3'-NC^{hm}C) or at both cytosine residues in the PAM-complementary region (3'-N^{hm}C^{hm}C-5'). We found that Cas9-mediated cleavage of target DNA containing 5-hmC or 5-ghmC opposite the guanines in the 5'-NGG-3' PAM was not inhibited by such modifications. Thus, cytosine glucosylating modifications of the PAM-complementary sequence are not sufficient and do not contribute to escape from Cas9-mediated interference (Figure 5).
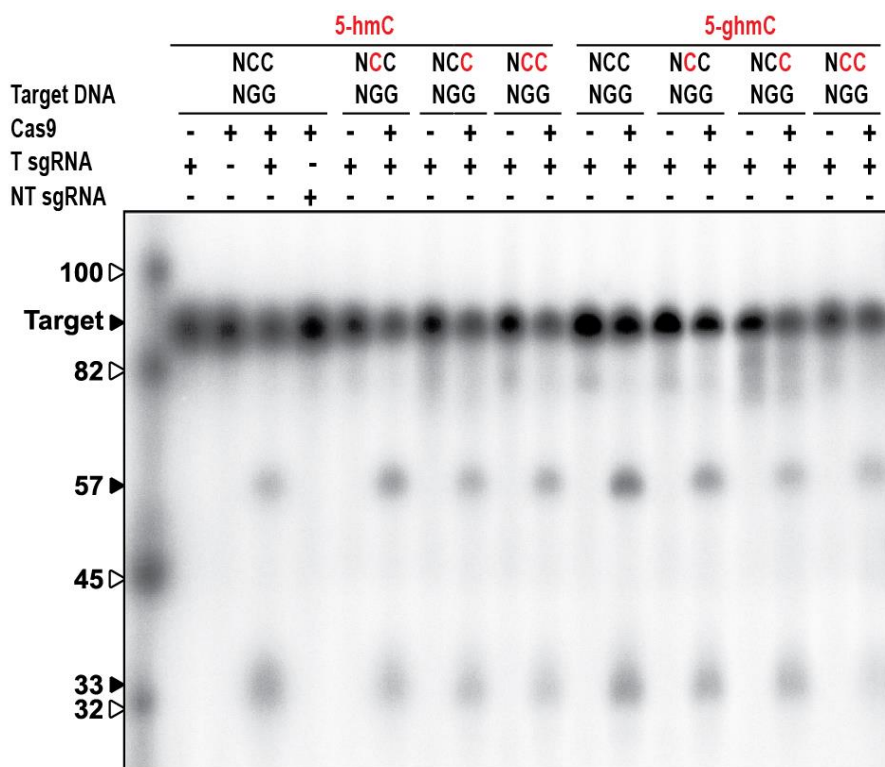

Figure 5. Effect of T4 DNA modifications of PAM-complementary cytosines on type II-A CRISPR-Cas sgRNA mediated DNA targeting. Cleavage assay of Cas9 on target DNA containing 5-hmC (indicated in red). Cas9 is

loaded with either targeting sgRNA (T sgRNA) or non-targeting sgRNA (NT sgRNA). Restriction products of Cas9 are 57 and 33 bp. The marker is indicated by white arrows.

DNA cleavage by Cas12a is not inhibited by glucosyl-5-hydroxymethylation

To investigate the activity of type V-A Cas12a systems to cleave glucosylated DNA, we performed DNA cleavage assays using Cas12a from *Francisella novicida* (FnCas12a; Figure 6A). Strikingly, DNA cleavage assays with Cas12a-crRNA complex and various DNA targets reveal Cas12a activity is not affected by 5-hmC or 5-ghmC modifications of the target (Figure 6B). Furthermore, EMSAs with catalytically inactive Cas12a containing mutations E1006A and R1218A (dCas12a) (Swarts et al., 2017) shows similar binding affinities for unmodified, 5-hmC, and 5-ghmC target DNA (Figure 6C). These results indicate that DNA targeting by Cas12a-crRNA complexes, in contrast to Cascade and Cas9-crRNA effector complexes, is not inhibited by 5-hmC or 5-ghmC modifications in the DNA.
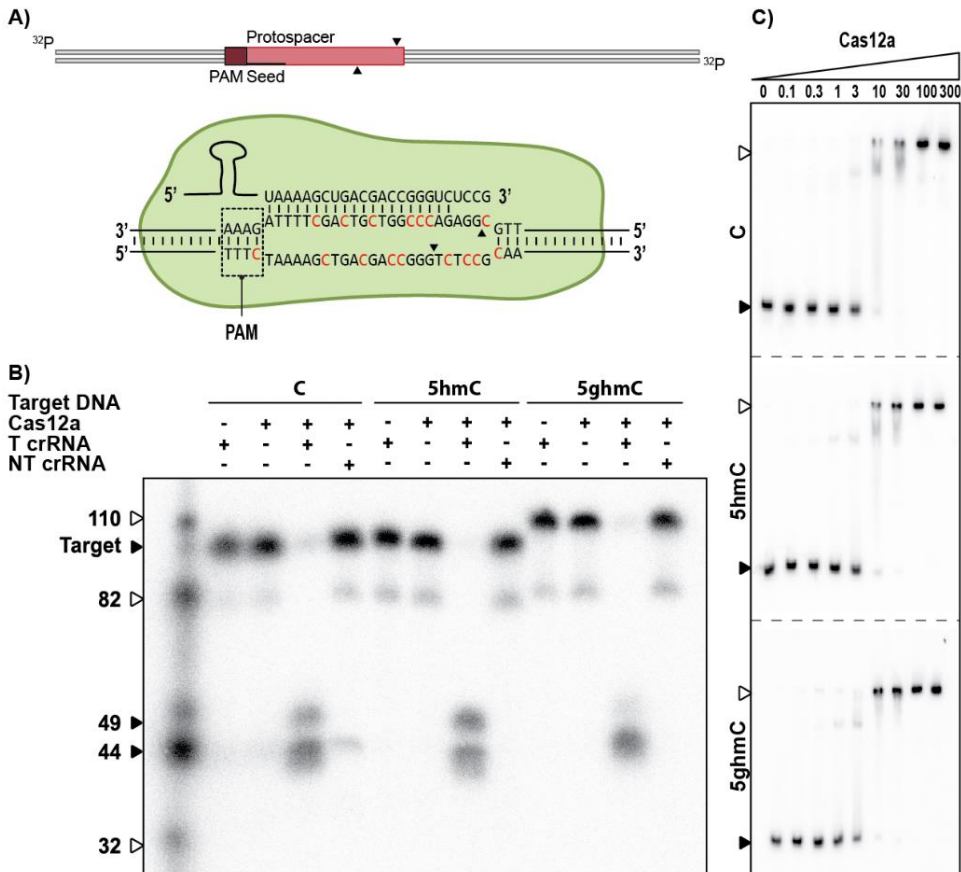

Figure 6. Effect of T4 DNA modifications on type V-A CRISPR-Cas sgRNA mediated DNA targeting. (A) Schematic of DNA targeting by Cas12a. Modified cytosine residues are indicated in red. Cleavage sites are indicated by

black arrows. **(B)** Cleavage assay of Cas12a on 98 bp modified targets (indicated by black arrow). Cas12a is loaded with either targeting crRNA (C crRNA) or non-targeting crRNA (NC crRNA). Cleavage products of Cas12a are 49 and 44 bp. The marker is indicated by white arrows. **(C).** Electrophoretic Mobility Shift Assay (EMSA) of Cas12a on target DNA containing C, 5-hmC or 5-ghmC (indicated by black arrow) at increasing protein concentrations [nM]. Fraction of bound target is indicated by white arrows.

Structural modelling of modified target DNA into crRNA-effector complexes

In order to further understand the contrasting effects of DNA glucosylation on CRISPR-Cas effector complexes we analysed potential steric clashes of DNA modifications in the target and non-target strand of the DNA at each nucleotide position by statically modelling glucosyl-5-hydroxymethyl modified DNA in existing structural models of ternary effector complexes (*E. coli* Cascade (PDB: 5H9F (Hayes et al., 2016)), SpCas9 (PDB: 5F9R (Jiang et al., 2016)), and FnCas12a (PDB: 5NFV (Swarts et al., 2017))). As structures of 5-ghmC are not available, we used the structure of a 5-ghmT (PDB: 3O8D), in which the glucosyl-hydroxymethyl modification is also positioned on carbon 5 of the base. Static modelling does not take into account potential flexibility of polypeptide and nucleic acid chains, but allows for the identification of potential clashes and thereby could provide an idea about how effector complexes might accommodate modified DNA substrates.
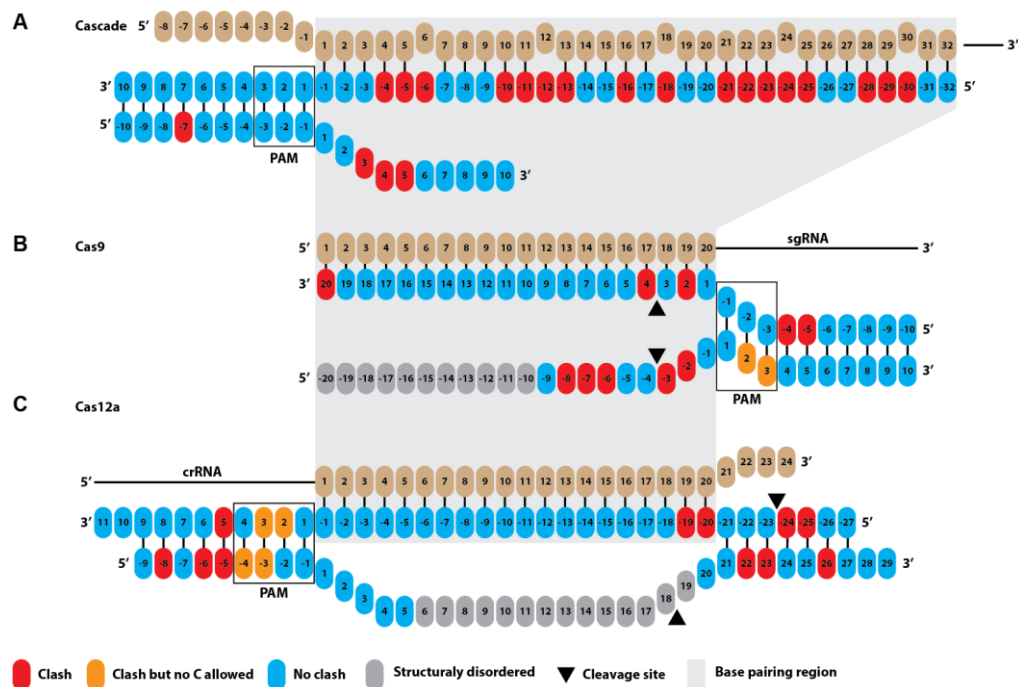

Figure 7. Potential steric clashes between target nucleotide 5-ghmC modifications and CRISPR effector proteins.
(A) Multiple clashes (indicated in red) are observed between the polypeptide chains of Cascade and 5-ghmC

modifications of nucleotides in the target strand (TS, complementary to the crRNA) and non-target strand (NTS). **(B)** Clashes are mostly observed between the polypeptide chains of Cas9 and 5-ghmC modifications of nucleotides in the seed region. (C) No clashes are observed in the seed region for Cas12a.

In both the Cascade and Cas9 models, multiple clashes are observed between the polypeptide chain and 5-ghmC modifications of nucleotides in both target DNA strands (Figure A; 7B, table S6). Clashes with Cascade are observed for 17 out of 32 target strand nucleotides (TS, complementary to the crRNA) and for 3 out of 7 of the non-target strand (NTS) nucleotides of which the position is identified. Clashes with Cas9 are observed for 3 out of 20 TS nucleotides and for 5 out of 9 of the NTS nucleotides that are structurally defined in the crystal structure of Cas9. Several of the modifications in both Cascade and Cas9 that cause steric clashes are located in the segment of the DNA that base-pairs with the crRNA seed sequence. Because interactions with the seed sequence are essential for R-loop formation and efficient targeting by both Cascade and Cas9 (Künne et al., 2014), such clashes can explain why DNA targeting by Cascade and Cas9 is inhibited by 5-ghmC modifications of the target DNA. In contrast, such clashes are not observed in the Cas12a model, in which 5-ghmC modifications can be accommodated at almost all positions within the crRNA-target DNA heteroduplex (Figure C).

It should be noted that for both Cas9 and Cas12a, clashes between the 5-ghmC modifications and the polypeptide chain are observed when 5-ghmC modifications are modelled onto the PAM nucleobases (Figure B; 7C). However, specifically these PAMs cannot contain cytosines at these positions (Fig. 7) since Cas9 and Cas12a recognize 5'-NGG-3' and 5'-(T)TTN-3' PAMs, respectively. In line with our biochemical experiments with Cas9 (Fig. 5), modifications of the cytosine nucleotides opposite to the 5'-NGG-3' PAM do not appear to introduce clashes with the polypeptide chain of Cas9 in our models. In both models, additional clashes can be found upstream the protospacer-PAM segment, and within the PAM-distal protospacer region. Because these segments of the DNA are generally not essential for efficient DNA targeting (Cas9: (Fu et al., 2013)/Cas12a:(Fonfara et al., 2016; Swarts et al., 2017; Zetsche et al., 2016)), we hypothesize that the potential clashes between 5-ghmC modifications and the polypeptide chain in our models will have a limited effect on DNA targeting. In summary, we conclude that 5-ghmC modifications in the target DNA are likely to result in clashes with Cascade and Cas9 and impair their DNA binding activity. However, Cas12a appears to be able to accommodate modified nucleobases due to a much more open structure. These models corroborate our experiments that demonstrate Cascade and Cas9 activity, but not Cas12a activity, is inhibited by 5-ghmC modification of target DNA.

DISCUSSION

Phages have evolved diverse systems that allow them to counteract bacterial defense mechanisms. Evasion from CRISPR-Cas defense can occur by mutation of the protospacer or PAM sequence (Deveau et al., 2008; McGrath et al., 1999; Semenova et al., 2011), target site deletion (Pyenson et al., 2017), genome recombination(Paez-Espino et al., 2015), or by production of anti-CRISPR proteins (Bondy-Denomy et al., 2015; Pawluk et al., 2014, 2016b). Here we show how chemical modifications of DNA nucleotides can render phages resistant to interference by specific CRISPR-Cas types. Our results demonstrate that CRISPR-Cas type I-E and type II-A interference is severely impaired by glucosyl modification of the cytosine bases (5-ghmC) in T4 phage DNA. However, the type V-A effector protein Cas12a remains able to bind and cleave T4 DNA even if it is glucosylated, suggesting the possibility that type V systems have evolved to combat covalently modified phages.

Phage encoded DNA modifications are well known antagonists of microbial defense systems, most notably restriction-modification systems (Labrie et al., 2010). Apart from the known inhibitory effect of DNA modifications on R-M systems, the effect on CRISPR-Cas systems has been studied for some DNA modifications. Small modifications such as N6-methylation of adenine in 5'-GATC-3' sequences do not prevent interference by the type II-A CRISPR-Cas system in *Streptococcus thermophilus* (Dupuis et al., 2013). Later, Yaung and colleagues showed that DNA modifications of T4 phages do not protect the phage from Cas9 interference (Yaung et al., 2014). By contrast and in line with our results, Bryson and colleagues showed that Cas9 interference can be inhibited by glucosylation of DNA when using the naturally occurring combination of crRNA and trans-activating crRNA (tracRNA), and to a lower extent using the engineered single guide RNA (sgRNA) (Bryson et al., 2015). Also, the degree of inhibition is strongly dependent on the selected target DNA sequence, and therefore the positions of base modifications. While this manuscript was in preparation, Strotskaya and colleagues showed that no interference nor CRISPR array expansion was observed when T4 was targeted by the type I-E system in *E. coli* (Strotskaya et al., 2017).The insensitivity of T4 to CRISPR interference was attributed to either DNA modifications or to yet unknown phage mechanisms. Here we demonstrate that the insensitivity is a result of DNA modification.

The inability of Cascade and Cas9 to bind and cleave 5-ghmC DNA is likely due to steric clashes between the glucose modifications of the protospacer segment of the target DNA and the polypeptide chain of the effector complexes. Our results show that Cas9 can still read the PAM nucleotides if only the PAM-complementary cytosines are glucosylated, which is in agreement with structural modeling. The inability of Cas9 to cleave fully glucosylated targets suggests that after successful PAM recognition, subsequent recognition of the protospacer is hampered.

5

The effect of DNA modifications on CRISPR-Cas interference is most likely dependent on the target DNA sequence and thereby the degree of DNA modifications. Although Cas9 and Cas12a target roughly the same region of our tested target DNA, the chosen target sites are different and result in different degrees of DNA modifications in the seed regions of the protospacer (Cong et al., 2013; Jiang et al., 2013; Zetsche et al., 2015). However, our structural modeling suggests that Cas12a has an intrinsically more open architecture and we predict that it can accommodate and better deal with bulky DNA modifications regardless of their position in the protospacer.

In addition to steric hindrance effects, changes in chemical presentation of nucleobases can result in decreased binding affinities. Changes in the chemical signature of the major and minor groove by replacement of guanines in the PAM with purine analogs is shown to attenuate binding affinities of the type I-F Csy complex in *Pseudomonas aeruginosa* (Rollins et al., 2015). Similarly, glucosylation of DNA is known to change the chemical signature and stability of the DNA duplex. Hydrogen bonds can be formed between the side groups of the glucosyl moieties and neighboring bases (Hunter, 1996). These interactions alter the roll, slide, and twist angles of base pairs (el Hassan and Calladine, 1996) which could impair R-loop formation and thereby prevent cleavage of the modified targets.

While phage T4 contains 70% α-glucosylated and 30% β-glucosylated DNA, this study only examined the effect of β-glucosyl linkages *in vitro* (Lehman and Pratt, 1960). The effect of α-glucosyl-5-hydroxymethylation on CRISPR-Cas interference remains to be determined. Nevertheless, based on a comparable occupation of the major groove by α-5-ghmC, we expect there to be little or no differences between the effects of α- or β-coupled glucosyls on CRISPR interference.

Our finding that DNA containing glucosyl-5-hydroxymethylated cytosine bases can be cleaved by a CRISPR-Cas type V-A system, but inhibits binding by crRNA-effector complexes from type I-E and II-A systems, has important implications that could be exploited for genome engineering applications (Vlot and Brouns, 2017). Firstly, Cas9 is not inhibited by 5-hmC modifications which allows editing of cells with 5-hmC modified regions such as some neuronal cells (Kriaucionis and Heintz, 2009; Münzel et al., 2011). Secondly, the inhibitory effect of 5-ghmC on Cas9 targeting could be used to introduce non-cleavable DNA templates for homology directed repair (HDR) applications. Lastly, the seemingly robust activity of Cas12a could be harnessed for genome engineering of organisms and bacteriophages that contain hypermodified bases such as the pathogen *Trypanosoma brucei* (Gommers-Ampt and Borst, 1995) and phage T4.

FUNDING

5

REFERENCES

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

Anders, C., Niewoehner, O., and Jinek, M. (2015). In Vitro Reconstitution and Crystallization of Cas9 Endonuclease Bound to a Guide RNA and a DNA Target. In Methods in Enzymology, (Elsevier Inc.), pp. 515–537.

Andersson, A.F., and Banfield, J.F. (2008). Virus population dynamics and acquired virus resistance in natural microbial communities. Science *320*, 1047–1050.

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

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

Bryson, A.L., Hwang, Y., Sherrill-Mix, S., Wu, G.D., Lewis, J.D., Black, L., Tyson A. Clark, F., and Bushman, rederic D. (2015). Covalent Modification of Bacteriophage T4 DNA Inhibits CRISPR- Cas9. MBio *6*, 1–9.

Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: All-atom structure validation for macromolecular crystallography. Acta Crystallogr. Sect. D Biol. Crystallogr. *66*, 12–21.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. Science (80-.). *339*, 819–823.

Deveau, H., Barrangou, R., Garneau, J.E., Labonté, 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.

Dupuis, M.-È., Villion, M., Magadán, A.H., and Moineau, S. (2013). CRISPR-Cas and restriction-modification systems are compatible and increase phage resistance. Nat. Commun. *4*, 2087.

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 1–19.

Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K., and Sander, J.D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat. Biotechnol. *31*, 822+.

Gao, Y.-G., Robinson, H., Wijsman ER, Van Der Marel, G., Van Boom, J., and Wang, A.-J. (1997). Binding of Daunomycin to beta-D-Glucosylated DNA Found in Protozoa Trypanosoma brucei Studied by X-Ray Crystallography. J. Am. Chem. Soc. *119*, 1496–1497.

Gommers-Ampt, J.H., and Borst, P. (1995). Hypermodified bases in DNA. FASEB J. *9*, 1034–1042.

el Hassan, M.A., and Calladine, C.R. (1996). Propeller-twisting of base-pairs and the conformational mobility of

dinucleotide steps in DNA. J. Mol. Biol. *259*, 95–103.

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.

Hunter, C.A. (1996). Sequence-dependent DNA structure. Bioessays *18*, 157–162.

Hynes, A.P., Rousseau, G.M., Lemay, M.-L., Horvath, P., Romero, D.A., Fremaux, C., and Moineau, S. (2017). An anti-CRISPR from a virulent streptococcal phage inhibits Streptococcus pyogenes Cas9. Nat. Microbiol.

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 (80-.). *356*.

Jiang, F., Taylor, D.W., Chen, J.S., Kornfeld, J.E., Zhou, K., Thompson, A.J., Nogales, E., and Doudna, J.A. (2016). Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. Science (80-.). *351*, 867–871.

Jiang, W., Bikard, D., Cox, D., Zhang, F., and Marraffini, L. a (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat Biotechnol *31*, 233–239.

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 (80-.). *337*, 816–821.

Jore, M.M., Lundgren, M., van Duijn, E., Bultema, J.B., Westra, E.R., Waghmare, S.P., Wiedenheft, B., Pul, U., Wurm, R., Wagner, R., et al. (2011). Structural basis for CRISPR RNA-guided DNA recognition by Cascade. Nat. Struct. Mol. Biol. *18*, 529–536.

Kim, J. (2005). Organization of the T4 dNTP Synthetase Complex at DNA Replication Sites. Dctoral Thesis 1–69.

Kriaucionis, S., and Heintz, N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science *324*, 929–930.

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*, 345–360.

Künne, T., Swarts, D.C., and Brouns, S.J.J. (2014). Planting the seed: target recognition of short guide RNAs. Trends Microbiol. *22*, 74–83.

Künne, T., Westra, E.R., and Brouns, S.J.J. (2015). Electrophoretic Mobility Shift Assay of DNA and CRISPR-Cas Ribonucleoprotein Complexes. In CRISPR: Methods and Protocols, M. Lundgren, E. Charpentier, and P.C. Fineran, eds. (New York, NY: Springer New York), pp. 171–184.

Kutter, E.M., and Wiberg, J.S. (1969). Biological Effects of Substituting Cytosine for 5- Hydroxymethylcytosine in the Deoxyribonucleic Acid of Bacteriophage T4. J. Virol. *4*, 439–453.

Labrie, S.J., Samson, J.E., and Moineau, S. (2010). Bacteriophage resistance mechanisms. Nat. Rev. Microbiol. *8*, 317–327.

Leenay, R.T., Maksimchuk, K.R., Slotkowski, R.A., Agrawal, R.N., Gomaa, A.A., Briner, A.E., Barrangou, R., and Beisel, C.L. (2016). Identifying and Visualizing Functional PAM Diversity across CRISPR-Cas Systems. Mol. Cell *62*, 137–147.

Lehman, I.R., and Pratt, E. a. (1960). On the structure of the glucosylated hydroxymethylcytosine nucleotides of coliphages T2, T4, and T6. J. Biol. Chem. *235*, 3254–3259.

5

Loenen, W.A.M., and Raleigh, E.A. (2014). The other face of restriction: Modification-dependent enzymes. Nucleic Acids Res. *42*, 56–69.

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

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

Mathews, C.K., North, T.W., and Prem veer Reddy, G. (1979). Multienzyme complexes in DNA precursor biosynthesis. Adv. Enzyme Regul. *17*, 133–156.

McGrath, S., Seegers, J.F.M.L., Fitzgerald, G.F., and Van Sinderen, D. (1999). Molecular characterization of a phage-encoded resistance system in Lactococcus lactis. Appl. Environ. Microbiol. *65*, 1891–1899.

Mohanraju, P., Makarova, K.S., Zetsche, B., Zhang, F., Koonin, E. V, and Van der Oost, J. (2016). Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. Science (80-.). *353*, aad5147.

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

Münzel, M., Globisch, D., and Carell, T. (2011). 5-Hydroxymethylcytosine, the Sixth Base of the Genome. Angew. Chemie - Int. Ed. *50*, 6460–6468.

Paez-Espino, D., Sharon, I., Morovic, W., Stahl, B., Thomas, B.C., Barrangou, R., Banfield, J.F., and Banfielda, J.F. (2015). CRISPR Immunity Drives Rapid Phage Genome Evolution in Streptococcus thermophilus. MBio *6*, e00262-15-.

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

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

Pawluk, A., Amrani, N., Zhang, Y., Garcia, B., Hidalgo-Reyes, Y., Lee, J., Edraki, A., Shah, M., Sontheimer, E.J., Maxwell, K.L., et al. (2016b). Naturally Occurring Off-Switches for CRISPR-Cas9. Cell *167*, 1829–1838.e9.

Pyenson, N.C., Gayvert, K., Varble, A., Elemento, O., and Marraffini, L.A. (2017). Broad Targeting Specificity during Bacterial Type III CRISPR-Cas Immunity Constrains Viral Escape. Cell Host Microbe 1–11.

Raleigh, E.A., and Wilson, G. (1986). Escherichia coli K-12 restricts DNA containing 5-methylcytosine. Proc. Natl. Acad. Sci. U. S. A. *83*, 9070–9074.

Rauch, B.J., Silvis, M.R., Hultquist, J.F., Waters, C.S., McGregor, M.J., Krogan, N.J., and Bondy-Denomy, J. (2017). Inhibition of CRISPR-Cas9 with Bacteriophage Proteins. Cell *168*, 150–158.e10.

Rollins, M.F., Schuman, J.T., Paulus, K., Bukhari, H.S.T., and Wiedenheft, B. (2015). Mechanism of foreign DNA recognition by a CRISPR RNA-guided surveillance complex from Pseudomonas aeruginosa. Nucleic Acids

Res. *43*, 2216–2222.

Sain, B., and Murray, N.E. (1980). The hsd (host specificity) genes of E. coli K 12. Mol. Gen. Genet. *180*, 35–46.

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

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. U. S. A. *108*, 10098–10103.

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.

Shmakov, S., Smargon, A., Scott, D., Cox, D., Pyzocha, N., Yan, W., Abudayyeh, O.O., Gootenberg, J.S., Makarova,
K.S., Wolf, Y.I., et al. (2017). Diversity and evolution of class 2 CRISPR–Cas systems. Nat. Rev. Microbiol.

Stern, A., and Sorek, R. (2011). The phage-host arms race: Shaping the evolution of microbes. BioEssays *33*, 43–
51.

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*, 62–67.

Strotskaya, A., Savitskaya, E., Metlitskaya, A., Morozova, N., Datsenko, K.A., Semenova, E., and Severinov, K.
(2017). The action of *Escherichia coli* CRISPR–Cas system on lytic bacteriophages with different lifestyles
and development strategies. Nucleic Acids Res. *45*, gkx042.

Swarts, D.C., van der Oost, J., and Jinek, M. (2017). Structural Basis for Guide RNA Processing and Seed-
Dependent DNA Targeting by CRISPR-Cas12a. Mol. Cell *66*, 221–233.e4.

Vlot, M., and Brouns, S.J.J. (2017). Dna modification. Patent WO2017055514 A1.

Walkinshaw, M.D., Taylor, P., Sturrock, S.S., Atanasiu, C., Berge, T., Henderson, R.M., Edwardson, J.M., and
Dryden, D.T.F. (2002). Structure of Ocr from bacteriophage T7, a protein that mimics b-form DNA. Mol.
Cell *9*, 187–194.

Warren, R. a (1980). Modified bases in bacteriophage DNAs. Annu. Rev. Microbiol. *34*, 137–158.

Weigele, P., and Raleigh, E.A. (2016). Biosynthesis and Function of Modified Bases in Bacteria and Their Viruses.

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.

Wigington, C.H., Sonderegger, D., Brussaard, C.P.D., Buchan, A., Finke, J.F., Fuhrman, J.A., Lennon, J.T.,
Middelboe, M., Suttle, C.A., Stock, C., et al. (2016). Re-examination of the relationship between marine
virus and microbial cell abundances. Nat. Microbiol. 15024.

Yaung, S.J., Esvelt, K.M., and Church, G.M. (2014). CRISPR/Cas9-mediated phage resistance is not impeded by
the DNA modifications of phage T4. PLoS One *9*, 3–9.

Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E.,

5

Joung, J., van der Oost, J., Regev, A., et al. (2015). Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. Cell *163*, 759–771.

Zetsche, B., Heidenreich, M., Mohanraju, P., Fedorova, I., Kneppers, J., DeGennaro, E.M., Winblad, N., Choudhury, S.R., Abudayyeh, O.O., Gootenberg, J.S., et al. (2016). Multiplex gene editing by CRISPR–Cpf1 using a single crRNA array. Nat. Biotechnol. *35*.

A DNA labeling approach based on incorporation of modifiable nucleobases using bacteriophage T4 genes

Marnix Vlot[1], Koen Martens[2], Peiyuan Zheng[3], Jan Groen[2], John van der Oost[1], Johannes Hohlbein[2], Mark J. Dickman[3], Stan J.J. Brouns[1,4*]

[1]Laboratory of Microbiology, Department of Agrotechnology and Food Sciences, Wageningen University, Wageningen, The Netherlands
[2]Laboratory of Biophysics, Department of Agrotechnology and Food Sciences, Wageningen University, Wageningen, The Netherlands
[3]ChELSI Institute Department of Chemical and Biological Engineering University of Sheffield, Sheffield, United Kingdom
[4]Department of Bionanoscience, Kavli Institute of Nanoscience, Van der Maasweg 9, 2629 HZ, Delft University of Technology, Delft, The Netherlands

Abstract

Modified nucleobases can be found in all domains of life and in viruses, and play an important role in gene regulation and host defense. A well-studied example of modified nucleobases is the substitution of cytosine by glucosyl-5-hydroxymethylcytosine in bacteriophage T4, which provides the phage protection from host restriction enzymes. Here we introduced four genes from the DNA modification pathway of bacteriophage T4 (gp42, gp1, gp56, and β-gt) into *Escherichia coli* and achieved a modification of approximately 15% of all cytosines. By leaving out the β-glucosyltransferase gene, a 5-hydroxymethylcytosine (5-hmC) intermediate is produced, which can be exploited for internal plasmid DNA labeling approaches *in vitro* using click chemistry-compatible glucosyl groups. Covalent labelling of plasmids with fluorophores enabled single molecule particle analysis of plasmid DNA. We suggest that this method can be used for various DNA modification applications including plasmid labeling, immobilization and tracking.

Introduction

Modified bases in DNA have been found in organisms and their viruses across all domains of life (Gommers-Ampt and Borst, 1995). The covalent chemical modifications of nucleobases range from simple methyl groups to amino acids and oligosaccharides. In eukaryotes, methylation of cytosines in CpG islands is used as a silencing mechanism (Bonasio et al., 2010; Jones, 2001). Recently, the discovery of low levels of 5-hydroxymethylcytosine (5-hmC) in eukaryotes has led to new insights into epigenetic and 5-hmC is now considered to be the sixth base in higher eukaryotes (Münzel et al., 2011). Although base modifications have regained interest due to their role in epigenetic regulation in eukaryotes, they have long been known to be highly present in the genomes of bacteria and their viruses (i.e. bacteriophages). Methylation of host genomic DNA enables differentiation between foreign and non-foreign DNA by restriction and modification (R-M) systems and directs cleavage of non-modified phage DNA, protecting the cell from infection (Wilson and Murray, 1991). The role of DNA modifications in host defense was identified when phage λ propagated in *E. coli* strain C was shown to be unable to infect *E. coli* strain K12 (Arber et al., 1963). Sinuously, phages have harnessed DNA modification as a way to evade R-M systems (Labrie et al., 2010). Phage T4 encodes endonucleases DenA and DenB, which degrade the host DNA upon infection. *De novo* generated 5-hmC is incorporated in phage DNA and subsequently glucosylated by phage-encoded α- and β-glucosyltransferases (Figure) (Lehman and Pratt, 1960). This glucosylation provides protection from cleavage by modification dependent systems (MDS), such as McrBC from *E. coli* that specifically degrades 5-hmC containing DNA (Raleigh and Wilson, 1986). Glucosyl-5-hydroxymethylcytosine (5-ghmC) is also found in the DNA of *Trypanosoma brucei,* the causative agent of sleeping sickness (van Leeuwen et al., 1998).
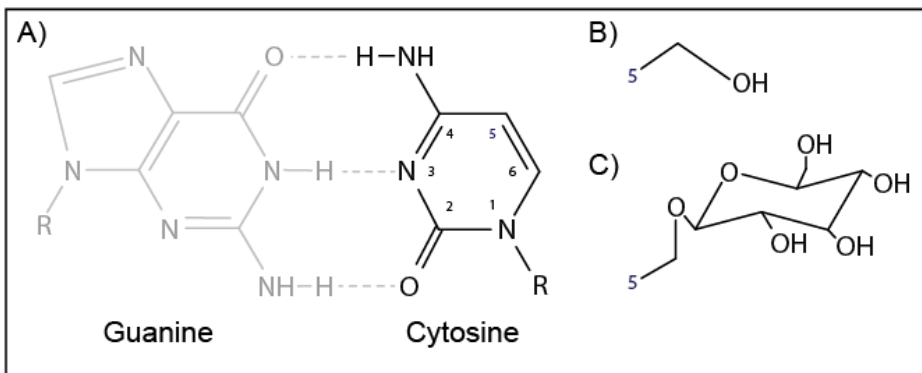
6

Figure 1. Modifications of nucleobases in phage T4 DNA. (**A**) Cytosine base pairing with guanine. (**B**) 5-hydroxymethylation of cytosine (5-hmC). (**C**) β-Glucosyl-5-hydroxymethylcytosine (5-ghmC) present in phage T4.

Biosynthesis of nucleotide precursors in *E. coli* is achieved by two metabolic pathways: the pyrimidine and purine pathways. The pyrimidine pathway leads to the production of dTTP and dCTP, the purine metabolism pathway leads to the production of dGTP and dATP (Fig 2B). T4 encodes a number of enzymes that form the dNTP synthase complex (DSC), a protein complex responsible for the efficient synthesis of 5-hmC (Moen et al., 1988; Reddy et al., 1977; Shen et al., 2004). The DSC consists of at least eight phage-encoded proteins and two host encoded proteins (Supplementary table 1). Phage-encoded dCTPase gp56 converts cellular dCDP and dCTP to dCMP. Hydroxymethylase gp42 converts dCMP to 5-hydroxymethyl-deoxycytidine-monophosphate (5-hmCMP), which is further phosphorylated to 5-hmdCTP by gp1 and incorporated in *de novo* synthesized DNA (Fig 2C). Post-replicative glucosylation by β-glucosyltransferase (β-gt) results in 5-ghmC.

T4 β-glucosyltransferase has recently been applied in the quantification of 5-hmC in eukaryotic DNA (Li et al., 2012; Nifker et al., 2015; Shahal et al., 2016). In these methods, 5-hmC containing DNA is converted to 6-azide-β-glucosyl-5-hydroxymethylcytosine (N_3-5-ghmC) using uridine-diphosphate-6-azide-glucose (UDP-6-N_3-Gluc) after which a fluorescent label is linked to the azide group using click chemistry forming fluorescently labeled DNA. This possibility to couple functional groups to 5-hmC containing DNA inspired us to transfer the 5-hmC synthesis pathway of T4 to *E. coli*.

To exploit the ability to couple click-chemistry compatible functional groups to large DNA molecules such as plasmids, we engineered the DNA modification pathway of bacteriophage T4 into *E. coli* by introducing four T4 genes.

Results

Expression of 5-ghmC synthesis operon in *E. coli*

To engineer the 5-ghmC synthesis pathway in *E. coli*, we constructed a plasmid (pGhmC) carrying the following genes from bacteriophage T4: gp42, gp1, gp56, and β-gt, encoding a dCMP hydroxymethylase, dNMP kinase, dCTPase, and β-glucosyltransferase, respectively. Expression of genes gp42, gp1, and β-gt was controlled by a rhamnose-inducible promoter. To control potential cytotoxicity of expression of dCTPase and uncouple it from the expression of gp42, gp1 and β-gt, we reversed the orientation of gp56, and transcriptionally isolated the gene between two rho-independent transcriptional terminators (Fig 2A). Expression of the four T4 genes encoded on pGhmC substitutes the production of cytosine by the production of 5-ghmC (Fig 2B & 2C). *E. coli* cells were transformed with pGhmC and plasmid DNA was isolated and digested with enzymes that are either sensitive to or selective for 5-hmC or 5-ghmC DNA (Figure). AbaSI is a DNA modification-dependent endonuclease that recognizes 5-ghmC in double strand DNA. Additionally, AbaSI also recognized 5-hmC DNA, but at a lower efficiency than 5-ghmC. When the pGhmC plasmid was incubated with AbaSI, we observed almost complete DNA degradation, demonstrating the presence of 5-ghmC. McrBC is an endonuclease that

cleaves 5-hmC containing DNA. When the pHmC plasmid was incubated with McrBC, we observed DNA degradation, demonstrating the presence of 5-hmC.
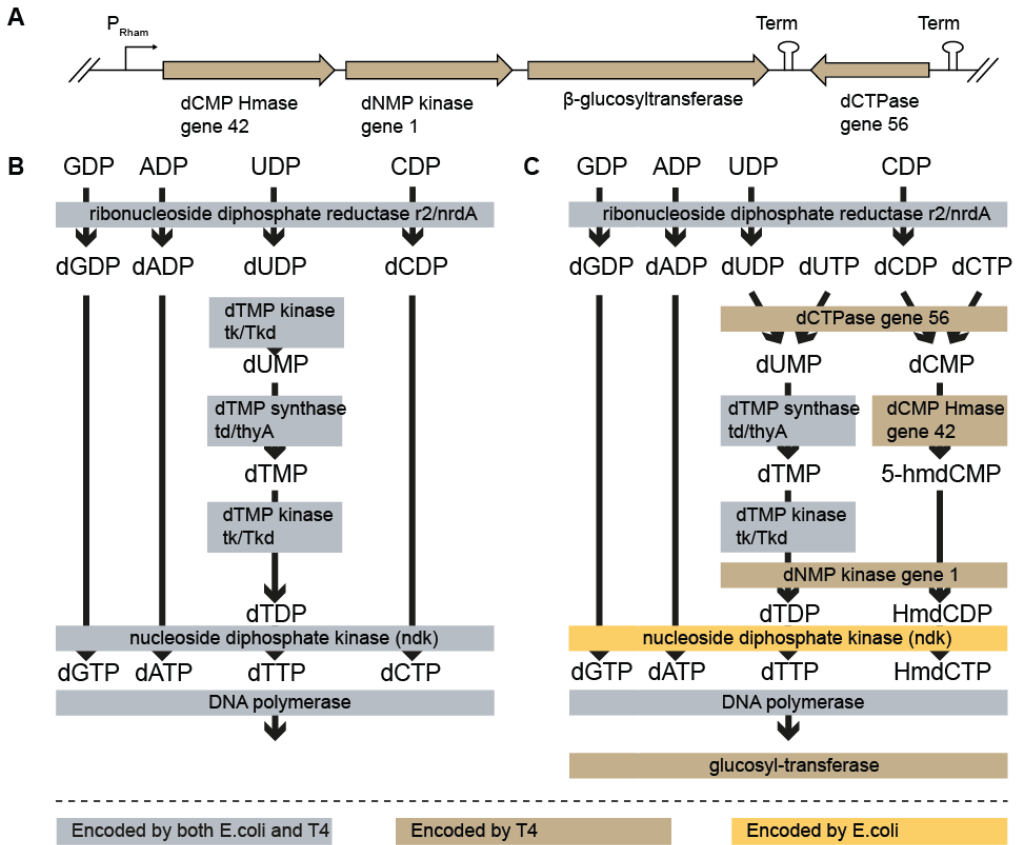

6

Figure 2. Metabolic pathway for DNA synthesis in *E. coli.* (**A**) Four T4 genes encoded on plasmid pGhmC are responsible for the synthesis of 5-ghmC DNA. (**B**) DNA synthesis pathway in *E. coli*. (**C**). DNA synthesis in *E. coli* after introduction of the 5-ghmC synthesis pathway. Three T4 genes are required for the conversion of dCDP and dCTP to 5-hydroxymethyl-2'-cytidine. T4 β-glucosyltransferase glucosylates 5-hmC in a post-replicative process.
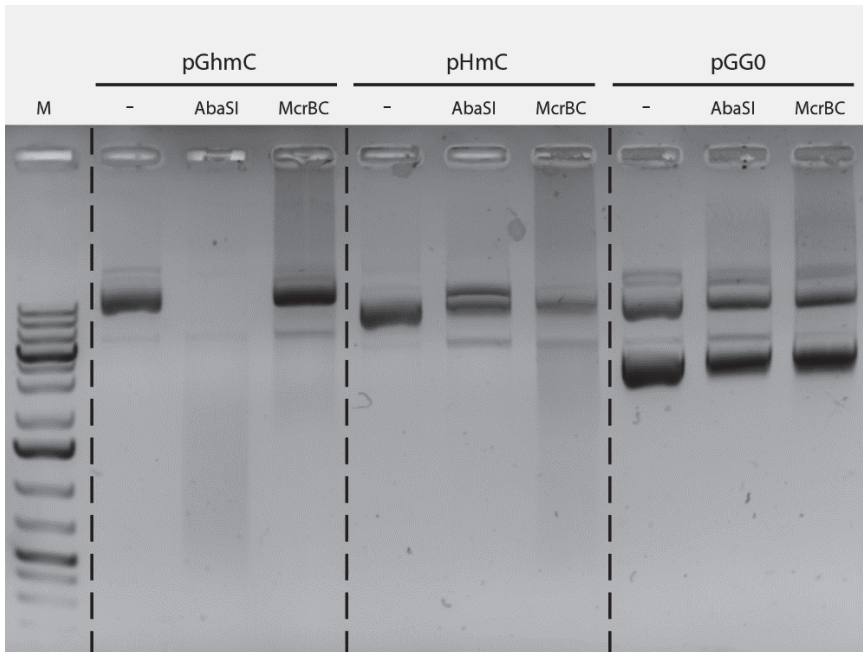
Figure 3. Modification-dependent degradation of plasmid DNA. Degradation of pGhmC DNA by AbaSI demonstrates the presence of 5-ghmC. Degradation of pHmC DNA by McrBC demonstrates the presence of 5-hmC. Unmodified DNA of pGG0 is degraded by neither AbaSI nor McrBC.

To validate that all four genes were essential for the production of 5-ghmC, we constructed plasmids lacking either gene gp42, gp1, gp56, or β-gt. Restriction digest analysis indeed showed that all four genes are required for efficient production of 5-ghmC (Fig S1). As expected, the deletion of gene β-gt results in 5-hmC DNA. The deletion of genes gp1, gp42, and gp56 results in cytosine containing DNA, showing that genes are essential.. We observed that when cells containing pGhmC were cultured for a period longer than one day, isolated plasmid DNA could not be cleaved by AbaSI, suggesting instability of the expression of the construct. To further investigate the toxicity of the expression of the 5-ghmC synthesis operon we performed growth experiments in which we monitored the optical density of the cultures. Cells containing pGhmC or pHmC showed a slower growth rate in the exponential phase and lower cell density in the stationary phase compared with cells containing the control plasmid pMK0 (Fig S2).

To attempt to further increase the substitution levels of cytosine by 5-hmC or 5-ghmC by increasing the dCMP pool we removed the transcriptional terminator upstream the gp56 dCTPase gene. Surprisingly, isolated plasmid DNA showed no degradation by AbaSI, indicating the absence of 5-ghmC in the DNA. This suggests that the expression levels of dCTPase are too high after removal of the transcriptional terminator, leading to cytotoxicity.

Determination of substitution levels

To determine the substitution levels of cytosine by 5-hmC or 5-ghmC, we analyzed the plasmid DNA using HPLC-UV detection. Plasmids pMK0, pHmC, and pGhmC were digested and analyzed using high-performance liquid chromatography ultraviolet (HPLC-UV). 5-hmdC levels were calculated by comparing the peak areas of plasmid digests with a 5-hmdC standard. pHmC DNA contains a cytosine substitution level of approximately 13-15% 5-hmdC under the tested circumstances (Figure 47). Enzymatically digested DNA was further analyzed by liquid chromatography mass spectrometry (LC-MS) (Fig S4). Using this method we were able to confirm the presence of 5-ghmdC by identification of the [M +H]$^+$ ions of 5-ghmdC (m/z 420.2) (Figure). Identification of [M +H]+ ions of 5-ghmC (m/z 304.2) is attributed to the elimination of a 2-deoxyribose moiety from 5-ghmdC caused by cleavage of the N-glycosidic linkage.

Figure 47. Quantification of 5-hmC levels of nucleosides from digested plasmid DNA. (**A**) HLPC analysis of nucleoside standards. (**B**) HPLC analysis of digested pHmC DNA.

Figure 5. LC-MS analysis of 5-ghmC levels of nucleosides from digested 5-ghmC DNA. Presence of 5-ghmC in plasmid pGhmC was confirmed by identification of the [M +H]$^+$ ions of 5-ghmdC (m/z 420.2). Identification of [M +H]+ ions of 5-ghmC (m/z 304.2) is attributed to the elimination of a 2-deoxyribose moiety from 5-ghmdC caused by cleavage of the N-glycosidic linkage.
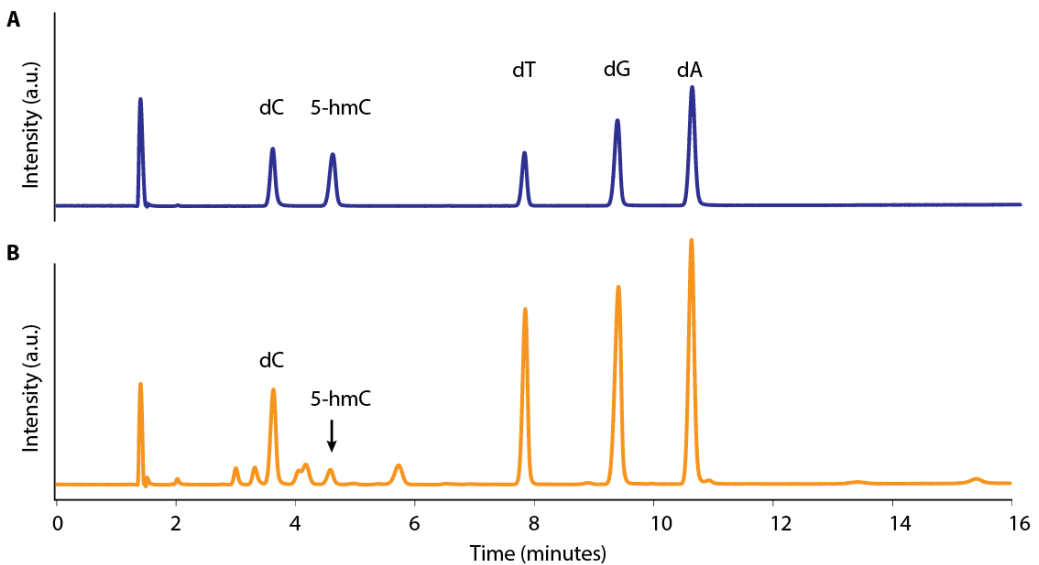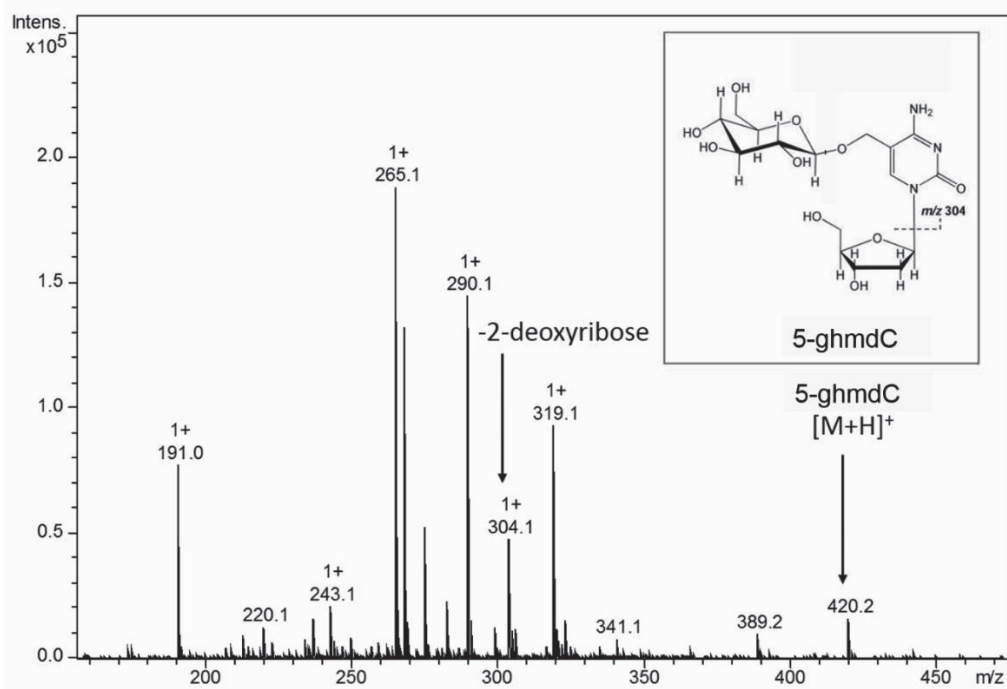
Coupling of functional groups to 5-hmC containing plasmid DNA

To label plasmid DNA with a fluorophore, selective tagging by copper-free click chemistry was used (Baskin et al., 2007). T4 β-glucosyltransferase was used to catalyze the transfer of an azido-sugar from chemically synthesized UDP-6-deoxy-6-azido-glucose (6-N$_3$-UDPG) to the allylic hydroxyl group of 5-hmC. The plasmids were then fluorescently labeled with dibenzocyclooctyne-Cy5 (DCBO-Cy5) and Cy5 fluorescence emission at 670 nm was used to assess the degree of labelling. Based on the relative fluorescence units (RFU) of the spectrum between 590 and 720 nm, we concluded that on average there are 4-5 fluorophores per plasmid molecule (Fig S3). To assess the functionality of the labeling, we immobilized the plasmid DNA on glass slides and imaged the samples using TIRF microscopy. By depositing samples of purified and washed pHmC-Cy5 on glass slides, single fluorescent spots could be visualized. By analysing the point-spread function, information about the number of fluorophores and their intensities could be compared to pMK0-Cy5 and pGhmC-Cy5 negative control plasmids (Fig 6). The pHmC-Cy5 sample contained approximately 2.1*10^2 fluorescent molecules per mm^2, which is approximately 20 times higher than either the pMK0-Cy5and approximately 9 times higher than pGhmC-Cy5 samples. This demonstrates successful labelling of the 5-hmC-containing plasmid

DNA (Fig 6). The low number of localizations when imaging pGhmC-Cy5 plasmids suggests that almost all 5-hmC nucleobases are glucosylated *in vivo* by β-glucosyltransferase.
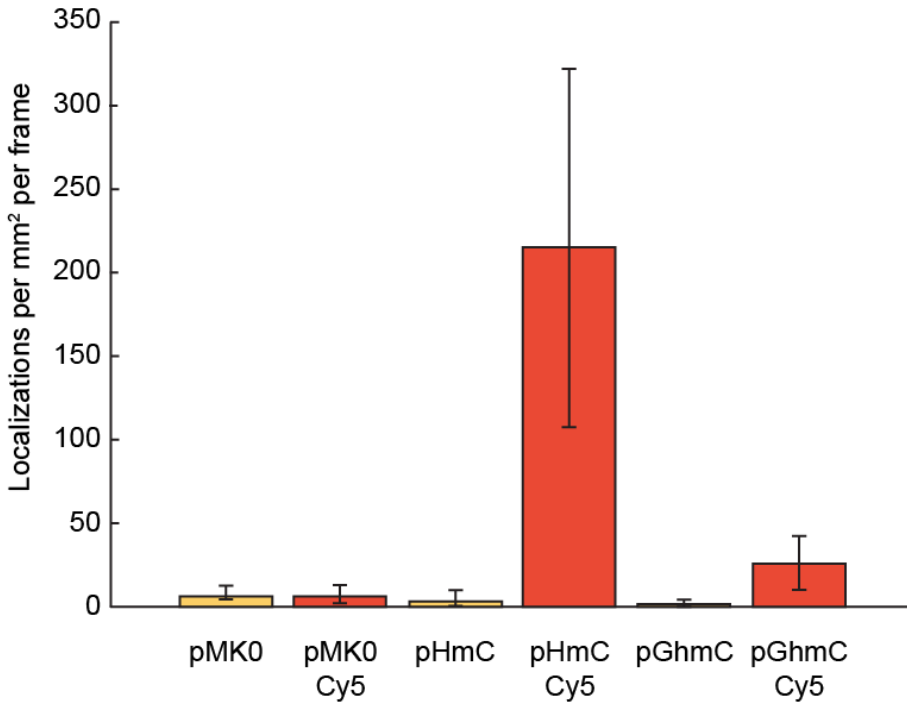
Figure 6. Fluorescence microscopy imaging of plasmids labeled with Cy5. Labeling of plasmid DNA with Cy5 is specific to pHmC.

Single plasmid photobleaching analysis

Illumination of Cy5 with the excitation laser of 642 nm causes photobleaching and switching to the non-fluorescent OFF-state of the fluorophores. Time traces of individual fluorescent spots show the step-wise decrease or increase of fluorescence intensity (Fig 7). The number of steps corresponds to the number of fluorophores labeled in the plasmid. To confirm that the image signal originates from Cy5, we used recovery to the fluorescent ON-state using a 405 nm laser. Fluorescence of the spots was recovered indicating the presence of Cy5.

Figure 7. Step-wise photobleaching and recovery of Cy5 fluorophores labeled to pHmC DNA of two individual time traces. Recovery and bleaching of individual Cy5 fluorophores is indicated by the red lines.
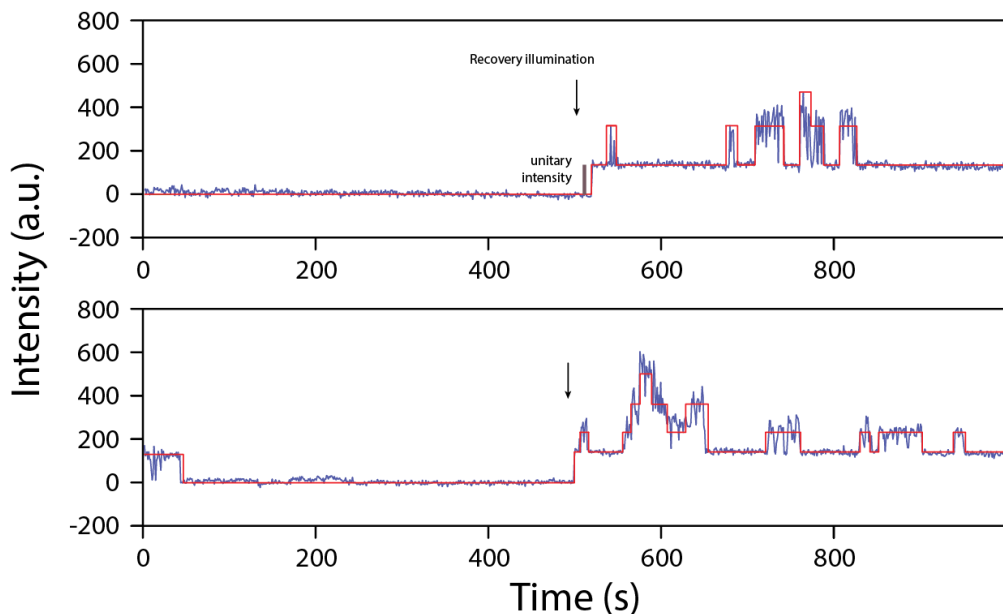
Discussion

In this work we present the introduction the of 5-hmC and 5-ghmC synthesis pathway from T4 into *E. coli* and we demonstrate its utility proof of concept by covalent coupling of functional groups to plasmid DNA. The introduction of 5-hmC and 5-ghmC synthesis pathway from T4 may provide insights into the effect of 5-hmC and 5-ghmC on regulatory processes in *E. coli*. The relatively modest levels of incorporation are beneficial when only low levels of labels are desirable, for example when immobilizing a plasmid on a surface. The levels of substitution of cytosine by 5-hmC can potentially be increased by increasing the expression levels of the introduced T4 genes. While this work was in progress, Mehta and colleagues published a similar approach of introducing the 5-ghmC synthesis pathway in *E. coli* (Mehta et al., 2016). Achieved substitution levels in plasmid DNA were approximately 71% and 45% for 5-hmC and 5-ghmC, respectively. The higher substitution levels, compared to our results, can likely be attributed to the use of codon-optimized T4 genes and higher plasmid copy numbers.

Labeling of DNA with fluorophores is commonly used to visualize DNA molecules, however available methods are often restricted to short oligonucleotides. In contrast our method can be used for labeling DNA molecules of any size, including genomic DNA. Further steps are also easy to implement, making our design a generally applicable method for DNA labeling. In addition, covalent coupling of azide groups to DNA molecules allows

labeling of any click-chemistry compatible functional group. Labeling of 5-hmC-azide with biotin could be used as a method to immobilize DNA. Alternatively, DNA may be labeled with peptides, non-fluorescent dyes, magnetic beads, or gold particles. This work creates a modular method for functionalization of DNA and opens up new DNA-labeling approaches.

6

Materials and Methods

Strains and constructs

E. coli strain cloni 10G (Lucigen) (relevant genotype: endA1 recA1 mcrA Δ(mrr-hsdRMS-mcrBC) galU galK) was used for all analytical experiments. *E. coli* 10G was made chemically competent using the RuCl method and transformed with plasmid DNA by applying a heat-shock as described in the QIAexpressionist handbook (QIAGEN). Cells were grown in Luria Broth (LB; 10 g·L−1 NaCl, 5 g·L−1yeast extract, and 10 g·L−1 tryptone) at 180 rpm or on LB-agar plates containing 1.5% (wt·vol−1) agar. To repress expression of the construct, 0.2% glucose was supplemented to the medium. 0.2% Rhamnose was used for induction of protein expression. When required, medium was supplemented with kanamycin (Km; 50 µg·mL−1). Bacterial growth was measured at 600 nm (OD600).

Sequences of genes 42, 1, B-gt, and 56 were selected from phage T4 (GenBank: AF158101.6) and unwanted restriction sites were deleted. The operon was assembled from synthetic oligonucleotides and/or PCR products and cloned into pMK-RQ (kanR) by GeneArt (Life Technologies). *E. coli* K12 was transformed with the synthetic construct and cells were grown followed by plasmid isolation. All subsequent experiments were performed in E. cloni 10G.

To create the constructs that lack one of the 5-ghmC operon genes, pGhmC was restricted using the enzymes listed in (Supplementary table 2). Following restriction, plasmids were re-ligated using T4 DNA ligase and used to transform E. coli 10G.

Molecular Biology and DNA Sequencing.

All strains and plasmids were confirmed by PCR and sequencing (GATC-Biotech). Plasmids were prepared using GeneJET Plasmid Miniprep Kits (Thermo Scientific) and DNA from PCR and agarose gels was purified using the Thermo Scientific GeneJET PCR Purification and Gel Extraction Kits.

Restriction analysis of DNA

DNA was restricted using 10U of enzyme in the recommended buffers and conditions. The products were visualized on a 1% agarose gel that was stained after running the gel.

HPLC analysis of nucleosides

In order to unambiguously confirm that modified dNTPs are incorporated into the PCR products of dsDNA, we digested the DNA products and identified the presence of the modified nucleosides based on their retention times using HPLC with a UV analysis. Nucleoside standards were generated from dNTPs (10 mM) with 1U shrimp alkaline phosphatase (NEB) at 37°C for 60 minutes and purified using 3K Nanosep Centrifugal Devices (PALL Life

Sciences), centrifuged at 5000 g for 6-15 minutes. The dsDNA PCR products were digested using 5U DNA Degradase (Zymo Research) at 37°C for 120 minutes and 3 U shrimp alkaline phosphatase (NEB), at 37°C for 60 minutes and purified by Nanosep Centrifugal Devices (3K). Analysis of the nucleosides was performed using Dionex UltiMate 3000 (Thermo Scientific) using a 50 X 2.1 mm Accucore C30 column (Thermo Scientific). Buffer A 5% ammonium acetate (HPLC grade, Sigma-Aldrich), buffer B was 40% acetonitrile (Thermo Scientific). The temperature for column compartment was 30°C, flow rate was 0.2 ml/min. Absorbance was recorded at 260 nm.

LC-MS analysis

The online LC part coupled to the MS analysis used was an UltiMate 3000 capillary LC system (Dionex), with different columns used for nucleosides separation. For Pursuit 3u C18 column (150*4.6mm, VARIAN) and Accucore™ C30 (50 × 2.1 mm, 2.6 µm particle size, Thermo Scientific) column, the chromatograph was performed using buffer A (5 mM ammonium acetate, pH 4.5, Sigma) and buffer B (40% acetonitrile (v/v), Fisher Scientific), with linear gradients of buffer B increase form 3% to 20% in 17 minutes, then to 80% in 1 minute, at a constant flow rate of 5 µl/min. For Hypercarb™ porous graphitic carbon column (100 × 2.1 mm, 3 µm particle size, Thermo Scientific) A gradient elution was performed using buffer A (10 mM ammonium acetate, pH 4.5, Sigma) and buffer B (40% acetonitrile (v/v), Fisher Scientific, and 10 mM ammonium acetate, pH 4.5, Sigma), with linear gradients of buffer B increase form 20% to 80% in 15 minutes, then switched to 100% buffer C (95% methanol (v/v), Fisher Scientific) for 3 minutes, then back to 20% B and 80% A in 1.1 minutes, at a constant flow rate of 200 µl/min. MS analysis was conducted on MaXis UHR TOF mass spectrometer (Bruker) equipped with an ESI nano sprayer source, Electrospray needle voltage was set to 4500V. Scan range of MS1 profile was set to m/z 100-800 in positive ion mode. Spectra data were analyzed using a Compass DataAnalysis software (version 4.1.359.0).

Microscopy setup

Imaging was performed on a home-built microscope similar to one described elsewhere [Farooq and Hohlbein, PCCP 17, 27862, 2015]. In short, laser light of 405 nm at and/or 642 nm from a fibre-coupled laser engine (Omicron, Germany) is directed towards an 100x 1.49 NA HP SR TIRF objective (Nikon, Japan) via a set of collimation and focusing lenses (Thorlabs, Germany), in total internal reflection fluorescence (TIRF) or highly inclined illumination (HiLo) mode. Emission light is separated from the excitation light via a dichroic mirror (Chroma, USA), spectrally filtered between 663-738 nm and falls on an Andor Zyla 4.2 plus sCMOS camera (Andor, Northern-Ireland). Only the centre of the chip was used, where a homogeneous excitation laser power was applied.

6

Preparation of Cy5 labeled DNA

Plasmids were labeled using a copper-free click reaction, described by Song and colleagues (2011) with a few adjustments. A stock solution of 250 ng/μL DNA was prepared for the reaction. 20 μL of DNA (5 μg total) was mixed with 2 μL of Uridine 5′-diphospho-6-azide-glucose (UDPG-N3) (150 μM final concentration), 1 μL of T4-β-glycosyl transferase (β-GT) (10 units), 3 μL cut-smart buffer (CS) and 4 μL of MQ, giving a total reaction volume of 30 μL. The samples were incubated at 37 °C overnight. The following day the excess UDPG-N3 was removed using the DNA cleaner & concentrator (5) kit from Zymo. 60 μL binding buffer was added to the reaction mixture (2:1) and loaded on a silica column. The DNA was washed two times with 250 μL washing buffer and eluted in 30 μL of MQ. 5 μL of PBS buffer (10 mM final concentration) and 5 μL of DBCO-Cy5 (0.5 mM final concentration) was added to the DNA. Incubation was performed at 37 °C overnight. The excess DBCO-Cy5 was removed using the DNA clean & concentrator (5) kit from Zymo and eluted in 30 μL MQ. The final DNA concentration was determined using the Qubit DNA quantification kit and the presence of Cy5 in the sample was confirmed and the quantity approximated using the basic fluorometer-app on a DeNovix DS 11 spectrophotometer.

Imaging of fluorescently labeled plasmids using TIRF microscopy

Glass cover slides were first cleaned by burning at 500 °C for 20 minutes, and were coated with Vectabond (Vector Labs, Burlingame, Ca, USA) by immersing them for 5 minutes in a 2% Vectabond solution in acetone. Afterwards, 10 μl of a purified and washed 10 pg/μl DNA solution was placed on the treated glass slides, incubated for 1 minute, and washed with phosphate-buffered saline solution. During measurements, fresh gloxy solution (Nahidiazar et al., 2016) was deposited on the glass slides to minimize bleaching effects.

For the acquisition HiLo illumination with 3.7 W/cm^2 642 nm laser intensity and 2.1 mW/cm^2 405 nm laser intensity were used. The density of fluorophores detected after bleaching the sample for 50 seconds with the 642 nm laser and re-activation with the 405 nm laser is reported for the samples. The acquisition rate of the camera was set to 10 ms. The mean counts of at least three different positions of the sample is presented as data.

Data analysis

The position, intensity, and number of fluorescent particles detected in different conditions was determined by using ThunderSTORM (Ovesny et al., 2014) with the pSMLM plug-in (Martens et al., 2018). The default ThunderSTORM settings were used, but using a local maximum detection equal to 1.5 times the standard deviation of F1 of the wavelet filter and a pSMLM sub-pixel analysis with a radius of 4 pixels. Only fluorescent spots with an intensity count larger than three times the average intensity of non-pHmC-Cy5 were considered in the analysis. Analysis was confined to the last 500 frames of every measurement. Further analysis was performed using custom-written MATLAB (The MathWorks) scripts.

References

Arber, W., Hattman, S., and Dussoix, D. (1963). On the Host-Controlled of Bacteriophage Lambda. Virology *35*, 30–35.

Baskin, J.M., Prescher, J.A., Laughlin, S.T., Agard, N.J., Chang, P. V., Miller, I.A., Lo, A., Codelli, J.A., and Bertozzi, C.R. (2007). Copper-free click chemistry for dynamic in vivo imaging. Proc. Natl. Acad. Sci. *104*, 16793–16797.

Bonasio, R., Tu, S., and Reinberg, D. (2010). Molecular Signals of Epigenetic States. Science (80-.). *330*, 612–616.

Farooq, S., and Hohlbein, J. (2015). Camera-based single-molecule FRET detection with improved time resolution. Phys. Chem. Chem. Phys. *17*, 27862–27872.

Gommers-Ampt, J.H., and Borst, P. (1995). Hypermodified bases in DNA. FASEB J. *9*, 1034–1042.

Jones, P.A. (2001). The Role of DNA Methylation in Mammalian Epigenetics. Science (80-.). *293*, 1068–1070.

Labrie, S.J., Samson, J.E., and Moineau, S. (2010). Bacteriophage resistance mechanisms. Nat. Rev. Microbiol. *8*, 317–327.

van Leeuwen, F., Taylor, M.C., Mondragon, A., Moreau, H., Gibson, W., Kieft, R., and Borst, P. (1998). beta-D-glucosyl-hydroxymethyluracil is a conserved DNA modification in kinetoplastid protozoans and is abundant in their telomeres. Proc. Natl. Acad. Sci. U. S. A. *95*, 2366–2371.

Lehman, I.R., and Pratt, E. a. (1960). On the structure of the glucosylated hydroxymethylcytosine nucleotides of coliphages T2, T4, and T6. J. Biol. Chem. *235*, 3254–3259.

Li, Y., Song, C.-X., He, C., and Jin, P. (2012). Selective Capture of 5-hydroxymethylcytosine from Genomic DNA. J. Vis. Exp. 3–6.

Martens, K.J.A., Bader, A.N., Baas, S., Rieger, B., and Hohlbein, J. (2018). Phasor based single-molecule localization microscopy in 3D (pSMLM-3D): An algorithm for MHz localization rates using standard CPUs. J. Chem. Phys. *148*.

Mehta, A.P., Li, H., Reed, S.A., Supekova, L., Javahishvili, T., and Schultz, P.G. (2016). Replacement of 2′-Deoxycytidine by 2′-Deoxycytidine Analogues in the *E. coli* Genome. J. Am. Chem. Soc. jacs.6b09661.

Moen, L.K., Howell, M.L., Lasser, G.W., and Mathews, C.K. (1988). T4 phage deoxyribonucleoside triphosphate synthetase: purification of an enzyme complex and identification of gene products required for integrity. J. Mol. Recognit. *1*, 48–57.

Münzel, M., Globisch, D., and Carell, T. (2011). 5-Hydroxymethylcytosine, the Sixth Base of the Genome. Angew. Chemie - Int. Ed. *50*, 6460–6468.

Nahidiazar, L., Agronskaia, A. V., Broertjes, J., Van Broek, B. Den, and Jalink, K. (2016). Optimizing imaging conditions for demanding multi-color super resolution localization microscopy. PLoS One *11*, 1–18.

Nifker, G., Levy-Sakin, M., Berkov-Zrihen, Y., Shahal, T., Gabrieli, T., Fridman, M., and Ebenstein, Y. (2015). One-Pot Chemoenzymatic Cascade for Labeling of the Epigenetic Marker 5-Hydroxymethylcytosine. ChemBioChem *16*, 1857–1860.

Ovesny, M., Krizek, P., Borkovec, J., Svindrych, Z., and Hagen, G.M. (2014). ThunderSTORM: a comprehensive ImageJ plug-in for PALM and STORM data analysis and super-resolution imaging. Bioinformatics *30*, 2389–2390.

Raleigh, E.A., and Wilson, G. (1986). Escherichia coli K-12 restricts DNA containing 5-methylcytosine. Proc. Natl. Acad. Sci. U. S. A. *83*, 9070–9074.

Reddy, G.P., Singh, A., Stafford, M.E., and Mathews, C.K. (1977). Enzyme associations in T4 phage DNA precursor synthesis. Proc. Natl. Acad. Sci. U. S. A. *74*, 3152–3156.

Shahal, T., Green, O., Hananel, U., Michaeli, Y., Shabat, D., and Ebenstein, Y. (2016). Simple and cost-effective fluorescent labeling of 5-hydroxymethylcytosine. Methods Appl. Fluoresc. *4*, 44003.

Shen, R., Olcott, M.C., Kim, J., Rajagopal, I., and Mathews, C.K. (2004). Escherichia coli nucleoside diphosphate kinase interactions with T4 phage proteins of deoxyribonucleotide synthesis and possible regulatory functions. J. Biol. Chem. *279*, 32225–32232.

Wilson, G.G., and Murray, N.E. (1991). Restriction and modification systems. Annu. Rev. Genet. *25*, 585–627.

6

Supplementary data
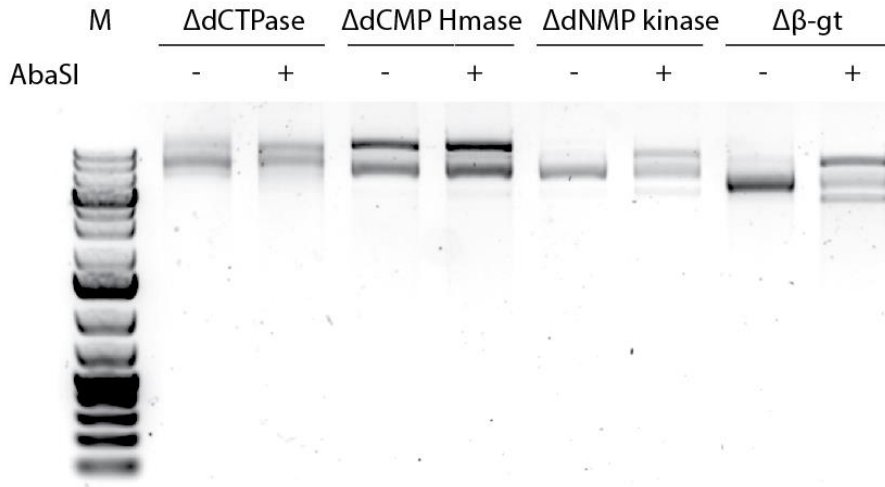
Supplementary figure 1. Restriction analysis of plasmids lacking individual T4 genes. Inability of AbaSI to degrade plasmid DNA demonstrates the absence of 5-ghmC. The minor degradation of the plasmid lacking β-glucosyltransferase (Δβ-gt) indicates the presence of 5-hmC.
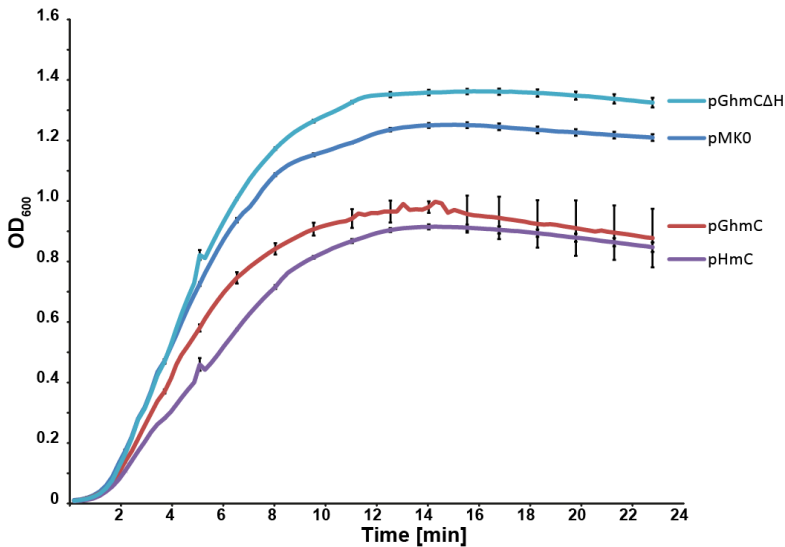
Supplementary figure 2. Growth curves of *E. coli* cells harboring plasmid pGhmC, pHmC or negative control plasmid pMK0 or pGhmC lacking gene 42 (pGhmCΔH).
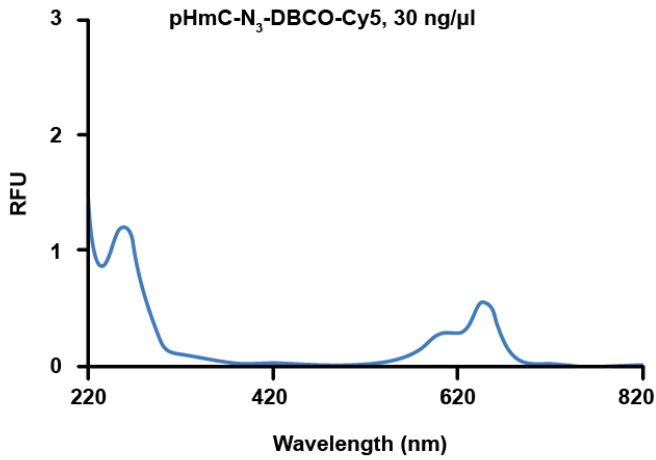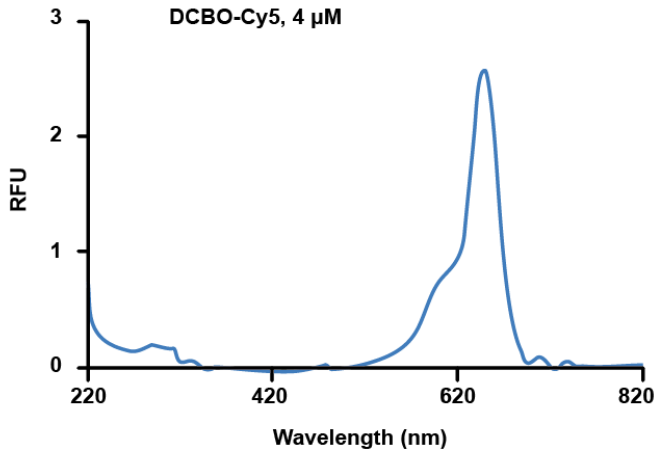
Supplementary figure 3. Absorption spectra of DCBO-Cy5 (top) and plasmid pHmC labeled with Cy5 (bottom). Peaks at 666 nm show emission by Cy5 molecules.
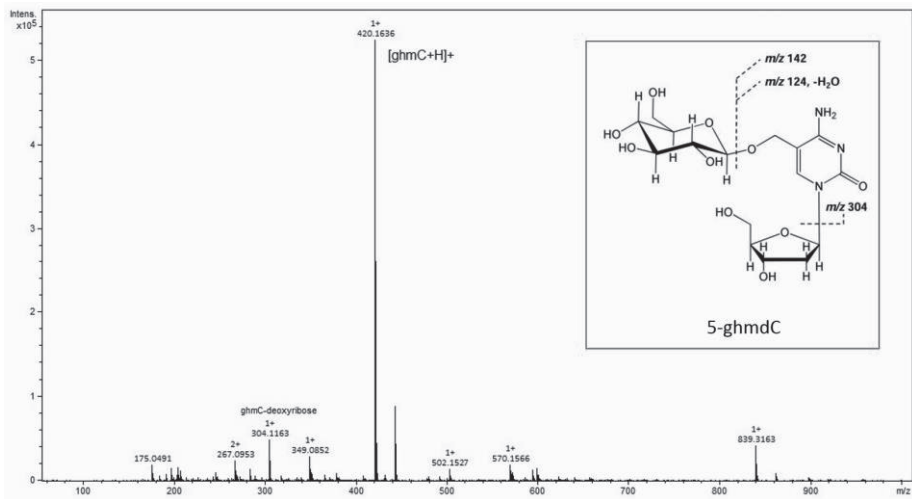
Supplementary figure 4. LC-MS analysis of 5-ghmC levels of nucleosides from digested 5-ghmC DNA. Presence of 5-ghmC in PCR products was confirmed by identification of the [M +H]⁺ ions of 5-ghmdC (m/z 420.2). Identification of [M +H]+ ions of 5-ghmC (m/z 304.2) is attributed to the elimination of a 2-deoxyribose moiety from 5-ghmdC caused by cleavage of the N-glycosidic linkage.

Supplementary table 1. Enzymes of the DSC.

Name	Encoded by	Gene	EC number
dCMP Hmase	T4	42	2.1.2.8
dTMP synthase	T4	td	2.1.1.45
Ribonucleotide reductase	T4	nrdA & nrdB	1.17.4.1
dCTPase	T4	56	3.6.1.12
Dihydrofolate reductase	T4	frd	1.5.1.3
dNMP kinase	T4	1	2.7.4.13
dCMP deaminase	T4	cd	3.5.4.12
Thymidine kinase	T4	tk	2.7.1.21
NDP kinase	E. coli	ndk	2.7.4.6
dAMP kinase	E. coli	adk	2.7.4.3

Supplementary table 2. Enzymes used for construction of plasmids lacking one gene of the ghmC operon.

Construct name	Deleted gene	Enzymes used	Phenotype
pGhmC	-	-	ghmC
pΔ56	dCTPase	MfeI & EcoRI	ghmC
pHmC	B-glucosyltransferase	PvuI & PacI	hmC
p Δ1	dNMP kinase	AscI & MluI	C
p Δ42	dCMP hydroxymethylase	XmaI & BspEI	C

6

Chapter 7

Summary

Summary

This thesis is about molecular studies on CRISPR-Cas systems, the isolation and characterization of a bacteriophage, and the engineering of a pathway in *E. coli* that results in the synthesis of modified nucleotides. Admitted, these topics do initially not seem to have much in common, besides all being interesting subjects of research in the field of molecular microbiology. However, the careful reader will realize that these topics are actually interconnected by a common denominator: bacteriophage-host interactions. Below I will summarize our findings in our efforts to elucidate the mechanisms involved in anti-viral defense systems and explain how this led to the introduction of a nucleotide modification pathway in *E. coli*.

In **chapter 1**, we describe that bacteria and archaea are under constant threat by their viruses, bacteriophages (phages). In response to this selective pressure, prokaryotes have developed multiple strategies to protect from phage infection. Mechanisms that protect prokaryotes from phage infection include inhibition of phage adsorption, superinfection exclusion, restriction-modification, BREX, DISARM, toxin-antitoxin systems, and CRISPR-Cas systems. Clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) systems were the first adaptive and heritable immunity systems found in prokaryotes and provide prokaryotes with resistance against mobile genetic elements. CRISPR-Cas systems function in three stages; adaptation, expression, and interference. In the adaptation process, invader DNA sequences are incorporated into a CRISPR array. In the expression stage CRISPR arrays are transcribed into pre-CRISPR RNA (pre-crRNA) and further processed into CRISPR RNAs (crRNAs). In the interference stage, CRISPR-Cas effector complexes loaded with crRNA bind their cognate foreign DNA targets (termed "protospacers") when flanked by a protospacer-adjacent motif (PAM) followed by degradation of target DNA.

We also discuss several mechanisms by which bacteriophages can evade antiviral defense systems. One common way to evade these systems by phages is by modifying their nucleic acids. Phage DNA can be modified either in the same way as the host DNA, making differentiation between host DNA and phage DNA impossible, or in such a way that the phage DNA is not recognized at all. Besides their role in host defense systems, DNA modification can play a role in regulation of gene expression and other regulatory processes.

Chapter 2, 3, and 6, describe research that makes use of single-molecule microscopy. Considering that these techniques might not be familiar to all readers, we provide a short introduction to these techniques in chapter 1.

In **chapter 2**, we investigate the interaction of the CRISPR-Cas complex Cascade from *E. coli* with bona fide and mutated target DNA. Target binding by Cascade is determined by complementarity of the crRNA and the protospacer, with the additional requirement of an immediately neighboring trinucleotide protospacer adjacent motif (PAM). Upon successful recognition of the PAM and the first 8 nucleotides of the target DNA, a so-called R-loop is formed. We developed a single-molecule FRET assay to monitor the recognition of target

DNA and the formation of the R-loop in real-time. We identified two different modes of binding exhibited by Cascade that are dependent on the target DNA. A long-lived binding event can be detected when a target has a correct PAM and is fully complementary to the crRNA. However, when the target DNA is mutated or does not have a correct PAM, target binding by Cascade takes place as a short-lived event. We propose a model in which the R-loop formation will be "locked" upon successful recognition of a bona fide target. This locked state recruits the CRISPR-associated nuclease-helicase Cas3 which leads to subsequent target degradation.

In case of partial recognition of a mutated target, the R-loop cannot be propagated over the full length of the protospacer and subsequent locking of the R-loop will not occur. This unlocked state is a signal for primed spacer acquisition, in which new spacers are rapidly integrated in the CRISPR array. Using *in vivo* experiments, we validated that these two different binding modes guide the functional outcome of target binding by Cascade.

While experiments such as those described in chapter 2 can be extremely helpful in identifying interactions between Cascade and DNA targets in vitro, there are still many unanswered questions regarding the target search inside the cell. In **chapter 3**, we describe how we visualized Cascade in its native environment by tracking single crRNA-surveillance complexes in live *E. coli* cells. Tracking of the Cas8e subunit of the Cascade complex allowed us to demonstrate that crRNAs are required for assembly of Cascade complexes. We also demonstrate that Cascade spends approximately 40% of its time bound to DNA. By tracking Cascade complexes in which important residues involved in PAM recognition are mutated, we show that the transient DNA-binding behavior of Cascade is largely due to PAM interactions. Furthermore, we demonstrate the underlying mechanism of protection from cell death by cleavage of the genomic CRISPR array using our single-molecule setup. We show that Cascade can still bind to the CRISPR array based on full complementarity of the array to the crRNA, but that recognition of the repeat-PAM leads to dissociation of the Cas8e subunit from the Cascade complex. Dissociation of the Cas8e subunit prevents the recruitment of Cas3, thereby preventing cleavage of the CRISPR array. These results provide new insights in the dynamics of the search behavior of CRISPR-Cas complexes in the molecular crowded environment in the cell.

Bacteriophages have developed diverse strategies to escape bacterial defense systems. One of these strategies is the production of phage-encoded proteins that inhibit CRISPR-interference, called anti-CRISPRs (ACRs). In an effort to identify novel anti-CRISPR proteins we isolated a large number of *E. coli* phages and screened them for ACR activity. Even though ACR activity could not be confirmed, we characterized some of the isolated phages in order to provide an addition to the body of knowledge about bacteriophage biology. In **chapter 4** we describe the isolation and genome sequencing of one of the screened phages, phage Ayreon. Phage Ayreon is a lysogenic phage that is highly similar to phage cdtl, but lacks the gene cluster responsible for cdtl holotoxin production in phage cdtl.

7

The production of ACRs is not the only strategy employed by phages to evade host-defense systems. One of the alternative escape mechanisms is the modification of nucleobases of the phage DNA by methylation. In **chapter 5**, we studied the effect of extensive bacteriophage DNA modifications on CRISPR-Cas interference. We show that glucosyl-5-hydroxymethylation of cytosine in the DNA of bacteriophage T4 inhibits interference by both type I-E and type II-A CRISPR systems. However, CRISPR-Cas interference of a type V system is not affected by glucosyl-5-hydroxymethylation of DNA. We demonstrate that this inhibition of CRISPR-Cas interference is caused by the decreased binding affinity of Cascade and Cas9 to modified DNA targets. Using structural modelling of DNA modifications, we predict which positions of DNA modifications will cause molecular clashes between the protein and the target, thereby identifying key positions of DNA modifications that inhibit interference. Our results also suggest that CRISPR-Cas systems have contributed to the selective pressure on phages to develop more generic solutions to escape sequence specific host defense systems.

In **chapter 6**, we further explore the potential of glucosyl-5-hydroxymethylcytosine (5-ghmC), as found in bacteriophage T4. We engineered four genes from the DNA modification pathway of bacteriophage T4 into *Escherichia coli.* Heterologous expression of these genes resulted in the production of 5-ghmC DNA in *E. coli*. Removal of the gene encoding the T4 β-glucosyltransferase resulted in the production of 5-hydroxymethylcytosine (5-hmC). This 5-hmC containing DNA can further be labeled with click-chemistry compatible functional groups, such as fluorophores. As a proof of concept, we labeled 5-hmC containing plasmid DNA with Cy5 fluorophores and studied electro-transformation of *E. coli* cells with these plasmids. Furthermore, we used single-molecule particle tracking to study plasmid uptake and stability after transformation. We propose that this method can be used to identify bottlenecks in efforts to establish genetic transformation protocols for new microbial species

Chapter 8

General Discussion

How bacteriophages shape the ecology in the biosphere

Viruses are the most abundant biological entities in the biosphere and are ubiquitous components in every environment. Our understanding of the biology of viruses continues to expand, making it ever more clear that they play a major role in almost every ecosystem on our planet. The areas influenced by viruses include biogeochemical and ecological processes such as nutrient cycling, bacterial and algal biodiversity distribution, algal bloom control, and genetic transfer, and microbial evolution (Fuhrman, 1999). Interestingly, until around a decade ago, viruses were ignored in studies on microbial loops in marine food webs. Later, studies reported on the significance of viruses on the control of bacteria and phytoplankton (Bergh et al., 1989; Proctor and Fuhrman, 1990; Thingstad, 1990).

Quantitative enumeration of viruses is commonly performed using transmission electron microscopy (TEM), revealing an enormous diversity of morphotypes. Most common morphologies include viruses with contractile tails (such as myoviruses), long non-contractile tails (such as siphoviruses), and short non-contractile tails (such as podoviruses). Lesser known morphologies include bottle-shaped, spindle-shaped, two-tailed, coil-shaped, and turreted icosahedral morphologies, mostly found for archaeal viruses (Prangishvili, 2013). Besides the large diversity in morphologies, viruses are also a reservoir of the greatest genetic diversity on Earth (Suttle, 2005). The remarkable genetic diversity of viruses combined with the absence of conserved evolutionary markers, such as ribosomal genes, makes it difficult to classify viruses (Breitbart et al., 2002; Rohwer and Edwards, 2002).

How anti-viral defense systems shape Life Sciences

Research on bacteriophage biology and anti-viral defense systems have arguably had the largest impact on the field of biology as a whole. The study of several model bacteriophages has led to the elucidation of fundamentals of biology that we take for granted today. Breakthroughs in molecular biology involving bacteriophages include sequencing of the first gene, sequencing of the first RNA genome, and sequencing of the first DNA genome. Some of the most important early discoveries are summarized in Table 1. Another application of phages that has hugely impacted life sciences is phage display, in which peptides are fused to phage particles and enable screening of biological relevant peptides in a high-throughput manner (Smith and Petrenko, 1997). The discovery of restriction-modification systems enabled molecular cloning and manipulation of DNA. Manipulation of specific locations of cellular DNA became available by harnessing the Cre-Lox recombination system from coliphage P1. The Cre-Lox system uses the P1-encoded Cre protein that efficiently promotes recombination of DNA occurring at specific sites, called *lox*, and can be used to create deletions, insertions, translocations and inversions (Sauer, 1987; Sauer and Henderson, 1988). Multiple other phage encoded proteins, including T4 ligase and T7 RNA polymerase, were harnessed to advance molecular cloning and heterologous expression of proteins (Sgaramella et al., 1970; Studier and Moffatt, 1986). Efficiency and simplicity of genetic modification of bacterial genomes increased even more when the λ RED system was

harnessed to use PCR products or synthetic oligonucleotides for homologous recombination (Casjens and Hendrix, 2015; Court et al., 2002). Most recently, the discovery of CRISPR-Cas systems has revolutionized the field of genetic engineering (Wang and Qi, 2016). The type II-A CRISPR-Cas effector protein Cas9 from *S. pyogenes* is nowadays extensively used as a tool to accomplish genome engineering. The CRISPR toolkit for RNA-guided genome engineering has been expanded with Cas9 proteins from various hosts, such as *Streptococcus thermophilus*, *Neisseria meningitidis*, and *Staphylococcus aureus* (Gasiunas et al., 2012; Ran et al., 2015; Zhang et al., 2013). Besides genome engineering, CRISPR-Cas proteins have been developed into tools to accomplish transcriptional regulation and epigenetic regulation (Gilbert et al., 2015, 2013; Konermann et al., 2015). Recent developments even allow the engineering of RNA in the cell (Cox et al., 2017). In conclusion, research on bacteriophages and defense systems that protect prokaryotes from infection shaped to a great extent the way we do science today and continues to do so. Undoubtedly, many more phenomena, pathways, and mechanisms lie in store, waiting to be discovered. Fortunately, with the increasing advance in our knowledge about biology, so do our research technologies and methodologies also advance. In the following paragraphs, I will discuss how advanced microscopy technologies can unveil the mechanisms of anti-phage defense systems.

Table 1 Major early scientific discoveries involving bacteriophages, table adapted from (Keen and Gables, 2016)

Reference	Finding
(Luria and Delbrück, 1943)	Mutation is a spontaneous and random process
(Hershey and Chase, 1952)	DNA is unambiguously the hereditary material of life
(Luria and Human, 1952)	Viruses' ability to infect other hosts is greatly diminished due to DNA cleavage by restriction enzymes
(Hall and Spiegelman, 1961)	Complementary DNA and RNA can hybridize
(Crick et al., 1961)	Nucleotide triplets are read as codons to form proteins
(Brenner et al., 1961)	mRNA is the intermediate between DNA and protein; mRNA is translated into protein by ribosomes
(Goulian et al., 1967)	DNA can be synthesized from its precursors in vitro

8

Single-molecule approaches shed light on CRISPR mechanisms

Adaptive immune systems, such as CRISPR-Cas systems, provide protection from pathogens based on highly specific mechanisms to distinguish self from non-self. This high specificity brings a downside that it can be readily overcome by mutations of rapidly evolving pathogens. However, CRISPR-Cas responds more rapidly and efficiently to a mutated invader by a so-called primed response (Datsenko et al., 2012). This primed response has been studied quite extensively with regards to the genetic processes involved (Fineran et al., 2014; Jackson et al., 2017; Richter et al., 2014; Savitskaya et al., 2013). However, the underlying molecular mechanism of primed spacer acquisition remained enigmatic. Below, I will discuss how single-molecule approaches, including those described in chapter 2 and 3, have provided some of the answers to questions regarding the molecular mechanisms of CRISPR-Cas systems.

Single-molecule methods are well-suited for the investigation of the molecular details of DNA binding and DNA cleavage by CRISPR-Cas systems since they allow detection of transient interactions and intermediate states (Joo et al., 2008; Singh and Ha, 2018). One of the first single-molecule approaches to study DNA interrogation and target binding by Cas9 was performed by Sternberg et al (2014). The authors visualized the target search of Cas9, labeled with quantum dots, on λ DNA that was immobilized on a flow cell by total internal reflection fluorescence (TIRF) imaging. This setup is commonly referred to as ''DNA curtains''. The authors demonstrated that DNA binding is initiated only after detection of the PAM and that DNA strand separation and RNA-DNA heteroduplex formation initiate at the PAM and proceed directionally towards the PAM distal end of the target. Furthermore, they showed that Cas9 samples DNA by 3D diffusion and not by 1D diffusion along the DNA. Next to Cas9, also the DNA interrogation patterns and target search behavior of Cascade was studied using the DNA curtains setup (Redding et al., 2015). In contrast to Cas9, Cascade does bind to protospacers in the absence of an interference PAM, albeit that this recognition is much less efficient than recognition of protospacers with an interference PAM.

Conformational control of CRISPR-Cas-mediated immune responses

The approach taken by Sternberg *et al* and Redding *et al* resulted in valuable data that allowed for refinement of our understanding of DNA interrogation by CRISPR proteins. However, the resolution limit (± 250 bp) of the used technique does not allow identification of the interactions between proteins and nucleic acids that occur on a smaller scale. Dynamic interactions can be observed at a much higher resolution using techniques such as single-molecule Förster resonance energy transfer (smFRET). Typically, two molecules are labeled with a different fluorophore and intermolecular interactions are visualized through FRET between the fluorophores (Joo et al., 2008; Rahul et al., 2008). Differences in FRET can be interpreted as conformational changes of the labeled molecules. In chapter 2, we describe how we studied the target binding by Cascade using smFRET. In short, biotinylated Cascade complexes were immobilized on a surface and dsDNA targets, labeled with a FRET

pair on opposite strands, was added. Changes in FRET could be attributed to binding of Cascade to the target and unwinding of DNA strands. Our single-molecule data demonstrate that high-fidelity target-DNA binding is a multistep process that is initiated at the PAM and continues by DNA unwinding in a PAM-distal direction. When the pairing of the DNA with the crRNA is completed over the full length of the protospacer the complex is stabilized into a "locked" state. This locked state allows destruction of the target DNA by Cas3. However, when Cascade encounters a target with mismatches, the initial recognition complex will be formed but Cascade will not switch into the locked state. By remaining in the "unlocked" state, the complex can disassemble using thermal energy and therefore is usually a short-lived interaction.

Previous research already established that Cse1 subunit adopts an open conformation upon binding of ssDNA (Jackson et al., 2014; Mulepati et al., 2014). This is in contrast with the locked conformation upon binding of dsDNA, in which the C-terminal domain is closer to the N-terminal domain, thereby causing the relocation of the C-terminal domain into the locked position. Our proposed model, describing how different binding modes lead to different functional outcomes, was further explored using smFRET studies (Xue et al., 2016). The authors developed a FRET system that enables measurement of conformational changes of the Cse1 subunit upon binding of the protospacer. Their results demonstrated that targets containing mutations cause conformational changes of Cse1 into the open conformation, thereby exposing the L1 domain. The L1 domain is formed by residues 125-131 of Cse1 and forms a loop that participates in PAM recognition (Van Erp et al., 2015). The L1 domain is critical for controlling Cse1 conformation and L1 mutants can still bind bona fide targets but block recruitment of Cas3 (Xue et al., 2016). Cas1-Cas2 increases the recruitment of Cas3 to bona fide targets and is correlated with primed spacer acquisition (Redding et al., 2015). This suggests that the open Cse1 conformation promotes interference-independent priming (Kunne et al., 2016; Redding et al., 2015; Xue et al., 2016). Together, these results demonstrate a mechanism that guides the direction of the CRISPR-Cas-mediated immune response, resulting in interference and/or priming, depending on target characteristics.

Vires in numeris, strength in numbers

One crucial requirement of anti-viral defense systems is that they must inactivate the invading phage DNA shortly after entry into the cell. The life cycle of many bacteriophages is less than 30 minutes, but even long before completion of the life cycle the phage has overtaken most, if not all, metabolic activities in the cell (Hadas et al., 1994). This provides only a window of mere minutes for the CRISPR-Cas complexes to find and inactivate invading DNA. Our studies, described in chapter 3, show that the concept of "*vires in numeris*" (strength in numbers) clearly applies to CRISPR-mediated immunity. It would take several hours for a single Cascade complex to find its target in the cell. However, we show that 300-600 Cascade complexes are enough to provide protection against plasmid invasion. Obviously, continuous expression of CRISPR-Cas systems comes at a fitness cost. Due to this cost, the availability of resources is a determining factor in the evolution of constitutive defense systems, including CRISPR-Cas, or inducible defences, including cellular surface

8

modification (Westra et al., 2015). While CRISPR-Cas expression in *E. coli* is naturally repressed by Histone-Like Nucleoid Structuring Protein (H-NS), we show that Cascade is functionally expressed, albeit at levels that are too low to provide interference (Westra et al., 2010). Because of the repression of CRISPR-Cas systems by H-NS, and additionally the slow evolution of the CRISPR arrays, it is suspected that the CRISPR-Cas system in *E. coli* fulfils an unknown alternative role (Westra et al., 2014). However, injection and subsequent replication and multiplication of phage DNA will cause redistribution and dilution of H-NS that is bound to the host DNA (Ali et al., 2012; Navarre et al., 2007; Rosen et al., 2006). This redistribution of H-NS might result in decreased repression of the CRISPR locus, thereby allowing functional expression and establishing CRISPR-mediated immunity. Such de-repressing conditions might become even more pronounced by the presence of phage-encoded DNA mimicking proteins such as bacteriophage T4 protein Arn or bacteriophage T7 protein Ocr (Ho et al., 2014).

What can be expected for the next decade of molecular biology?

CRISPR-Cas technology has proven itself as a versatile tool that has revolutionized the field of molecular biology and promises to continue to do so. Over the next decade, CRISPR-Cas derived technologies will expand the toolbox of the molecular biologists and allow many biological questions to be answered. Below, I will discuss the development of several CRISPR-based technologies

Genome Imaging using CRISPR-Cas-based technologies

The spatiotemporal organization of the genome plays an important role in transcriptional regulation and visualizing the dynamic processes that the genome undergoes has the potential to improve our understanding of regulatory processes of living cells. The specificity at which CRISPR proteins are able to target certain sequences makes them excellent tools to study the spatial and temporal behaviour of the genome in live cells. Previously, localization of genomic loci was performed on static organization of the genome using imaging technologies such as fluorescence in situ hybridization (FISH) (Lichter et al., 1990). The first study using CRISPR proteins to aid visualization of genome dynamics made use of a fusion of catalytically inactive Cas9 (dCas9) from *S. pyogenes* with EGFP (Chen et al., 2013). Robust imaging of repetitive elements in living cells was demonstrated and even visualization non-repetitive genomic sequences was achieved by using an array of spacers targeting different regions of a target locus. This approach was also used to design a multicolour visualization technique where three Cas9 orthologs from different species were fused to GFP, RFP, or BFP (Ma et al., 2015). Efficient targeting by these fusion proteins of different genomic loci allow to determine the intranuclear distance between loci, thereby revealing the state of DNA compaction in living cells.

CRISPR-Cas-mediated *in situ* localization of nucleic acids is not restricted to DNA. When sgRNA-guided Cas9 protein is stabilized with a PAMmer oligonucleotide, it can be directed to bind cellular mRNA (O'Connell et al., 2014). This feature has been harnessed in a recent study where a dCas9-GFP fusion protein was used to target mRNA in live cells (Nelles et al., 2016). This novel technology enables the tracking of RNA in livings cells in a programmable manner.

The advances in methods of tracking cellular components in live cells and in real time opens up a wide range of potential applications. Considering the recent developments in technologies of tracking either proteins or mRNA transcripts, it can soon be expected that both techniques will be employed simultaneously in order to study regulatory processes by imaging of both RNA and proteins associated with the translation machinery. The initial applications of RNA-guided CRISPR proteins were primarily focussed on site-directed DNA cleavage. However, CRISPR-based targeting was quickly adapted for purposes like imaging, targeted methylation, and transcription modulation. It can be expected that similar developments will occur for the RNA-guided RNA targeting described above and potential applications include modulation of RNA-processing.

8

## CRISPR-Cas-mediated genetic engineering and gene therapy

CRISPR-Cas-mediated genetic engineering has had enormous impact on the field of molecular biology, including gene therapy research (Dai et al., 2016; Savić and Schwank, 2016). Notable breakthroughs in gene therapy technology include correction of a gene mutation in mouse models of Duchenne muscular dystrophy, thereby rescuing the disease phenotype (Nelson et al., 2017; Tabebordbar et al., 2016). Despite recent advances in CRISPR-mediated gene therapy there are a number of challenges to be addressed before the technology can be used for application in humans. One of the most prevalent concerns raised regarding gene therapy is the fact that the sequence specificity of the RNA-guided CRISPR proteins might not be high enough to prevent off-targeting. Even though there does not seem to be a general consent on the actual prevalence and functional consequences of off-targeting among scientists much effort has been focused on reducing such effects (Chen et al., 2016; Kleinstiver et al., 2016; Shin et al., 2017; Slaymaker et al., 2016).
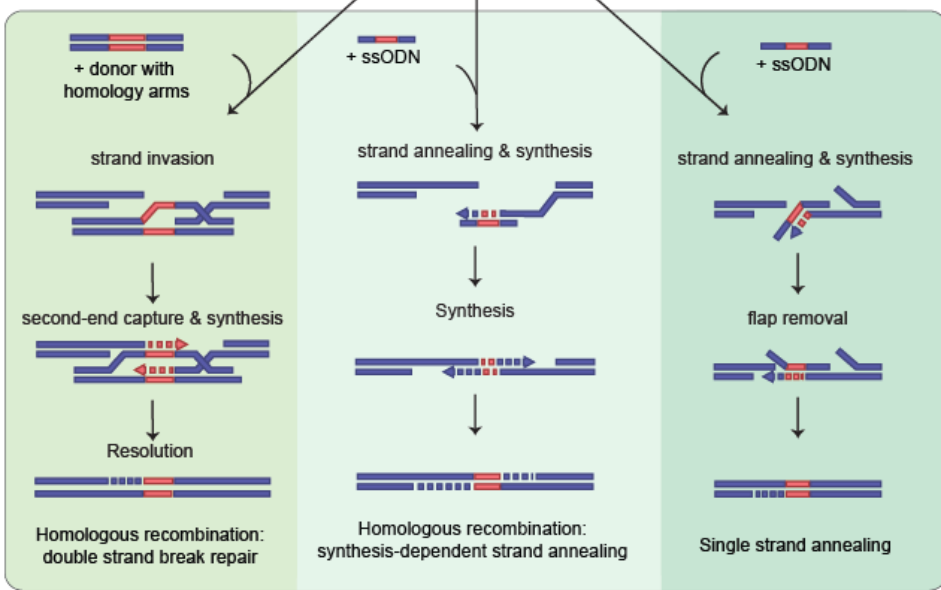
Double stranded DNA repair mechanisms

The application of CRISPR-Cas proteins in genome engineering relies on highly specific and efficient cleavage of DNA followed by repair of the DNA break, introducing genomic alterations. CRISPR-Cas-mediated DNA cleavage will result in a stop of cell-cycle momentum, triggering the DNA damage repair signaling cascade and the activation of endogenous DNA repair mechanisms (Heyer et al., 2010). The repair of a double stranded break can be accomplished by either non-homologous end joining (NHEJ) or homology-directed repair (HDR) (Fig. 1). NHEJ is a mechanism in which the two DNA strands are recoupled by ligation, which will frequently result in the introduction of non-specific insertions and deletions (indels). Indels usually disrupt the targeted gene by a shift in open reading frame (ORF). Due to these error-prone properties, the NHEJ pathway is desirable for creating knockouts, but a less desirable tool for precise genome engineering and gene therapy (Chrenek et al., 2016; Cong et al., 2013; Heyer et al., 2010; Renkawitz et al., 2014; Yang et al., 2013).

Homology-directed repair (HDR), in contrary to NHEJ, uses a homologous sequence as a template for the repair of damaged DNA (Chrenek et al., 2016; Renkawitz et al., 2014). The homologous sequence can be either a sister chromatid or an introduced DNA fragment, but must be in relative close proximity of the double-stranded break (DSB) to promote HDR by establishing DNA base pairing between the desired homologous sequence and the damaged DNA strand. HDR is the most desirable repair mechanism for gene therapy due to its ability to promote highly specific genetic exchange. A drawback of HDR is its relatively low efficiency, especially when compared to NHEJ. This is partly due to the fact that HDR can only occur in the S and the G2 phase of the cell cycle and therefore not active in the post mitotic tissue, whereas NHEJ can occur throughout the complete cellular cycle (Oude Blenke et al., 2016; Richardson et al., 2016; Wright et al., 2016).
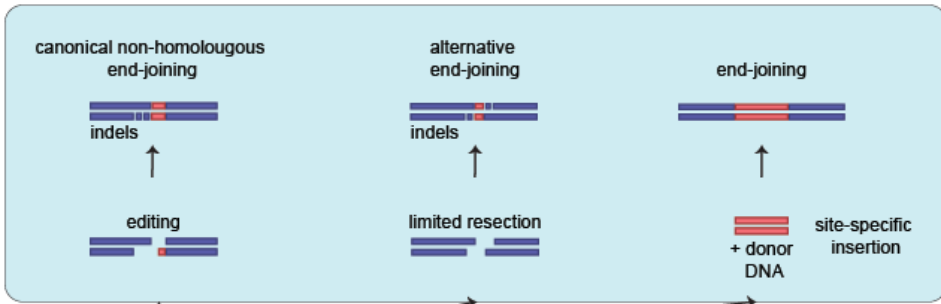
Figure 1. Double-stranded DNA breaks induce endogenous DNA repair mechanisms. Non-homologous end joining (NHEJ) often results in deleterious insertions or deletions. Homologous-directed repair (HDR) makes use of a exogenously introduced homologous template and results in high-fidelity DNA repair. Figure adapted from (Wu et al., 2018)

The mechanism of HDR

The HDR pathway can be divided into three stages: the presynapsis, synapsis and postsynapsis (Heyer et al., 2010; Renkawitz et al., 2014). In the presynapsis stage, species-specific protein complexes will bind to the blunt-end DSB, forming a DNA-protein complex and initiate 5` to 3` resection. DNA-end-resection is a process in which the 5`-end of the dsDNA is cleaved away by exonucleases and forms an extended section of ssDNA. To prevent the formation of secondary structures and to facilitate the binding of recombinase proteins, the ssDNA is coated with a ssDNA-binding protein, replication protein A (RPA). A stable structure is formed by RPA together with the recombinase protein, which is able to perform homology searches by DNA-strand invasion and base pairing with homologous template DNA. In the synapsis stage, homology searching proteins are actively searching for a complementary DNA strand that is homologous to the damaged DNA strand. When the homologous templates has been found, the ssDNA-protein complex invades the template, creating an opening between the two DNA strands of the homologous template and forms the displacement loop (D-loop). In the last stage of HDR, the postsynapsis, previously bound recombinase proteins are disassociated from the now stably paired DNA-DNA complex and the DNA is repaired. The postsynapsis stage of HDR is divided into multiple sub-pathways, which either create crossover or non-crossover products. Consequently, when a non-crossover sub-pathway proceeds, the DNA will be transcribed from the template, in contrast to the crossover sub-pathway in which the homologous DNA template is (partially) incorporated into the desired target DNA strand.

Possible solutions for increased efficiency of HDR by Cas9 and potential obstacles

The CRISPR-Cas9 system and the endogenous HDR system are desirable genome engineering mechanisms, which have been extensively used in research settings and gene therapy, but there are still obstacles to be overcome in order to increase the desired HDR efficiency over NHEJ efficiency (Wright et al., 2016).

One potential solution for increasing HDR efficiencies is to chemically induce cells to inhibit the NHEJ pathway by inhibiting key proteins of the NHEJ pathway (Maruyama et al., 2015). The NHEJ inhibitor promotes the frequency of HDR by inhibiting the NHEJ pathway and therefore favoring incorporation of the desired DNA repair template. However, interfering with NHEJ can be undesirable in gene therapy due to the possible alterations made to the cellular mechanisms and the potential inhibition of other unintentional genomic damages. Alternative strategies to increase HDR efficiencies include engineering the Cas9 proteins, the sgRNA, or the DNA repair template.

The type of DNA template used in genome editing has a drastic impact on HDR efficiency. DNA templates can be either complementary to the target strand or to the non-target strand, have short or long overhangs, and be either double stranded or single stranded. The use of DNA templates that are complementary to the non-target strand demonstrate a higher HDR efficiency, up to 2.6 times the efficiencies when DNA templates that are complementary to the target strand are used (Richardson et al., 2016). Even more so, the use of ssDNA templates demonstrate a 4 times higher HDR efficiency then when dsDNA templates are used.

Glucosyl-hydroxymethylcytosine as a way to improve HDR efficiencies

Even when a sequence specific DSB is repaired via the HDR pathway, potential re-cleavage of that site could nullify the successful gene editing. Re-cleavage of a repaired gene could occur because CRISPR proteins allow a certain degree of imperfect matching with their target. Generally, no mismatches are tolerated in the PAM region and only very few are allowed in the seed region. However, some sgRNA-target DNA mismatches are allowed at the PAM-distal end of the target. When the result of successful gene editing is a single point mutation, the DNA could potentially be cleaved again, creating a *cleave-repair* loop.

Yet another factor limiting the effectiveness of CRISPR-mediated gene editing is the short lifetime of the recombination template DNA. ssDNA templates that are introduced in the cell are anticipated to be subject to various cellular exonucleases. Even though the impact of this degradation is largely unknown, protection of template DNA by phosphorothioate linkages is shown to increase their half-life and to increase the probability of incorporation (Mosberg et al., 2010, 2012).

As described in chapter 5 of this thesis, glucosyl-hydroxymethylation of cytosine (5-ghmC) renders DNA protected against cleavage by various nucleases, including Cas9. This property of 5-ghmC DNA could potentially be harnessed in order to improve gene editing efficiencies (Vlot and Brouns, 2017). In short, modified template DNA can be created by PCR using 5-hydroxymethyldeoxycytidine, resulting in 5-hydroxymethylcytosine (5-hmC) containing DNA. 5-hmC DNA can be further glucosylated to 5-ghmC by T4 β-glucosyltransferase and used as repair template in HDR-mediated gene editing (Figure 2). Current efforts in our laboratory are directed towards improving HDR efficiencies in a RPE65[rd12] mouse model by using 5-ghmC template DNA.
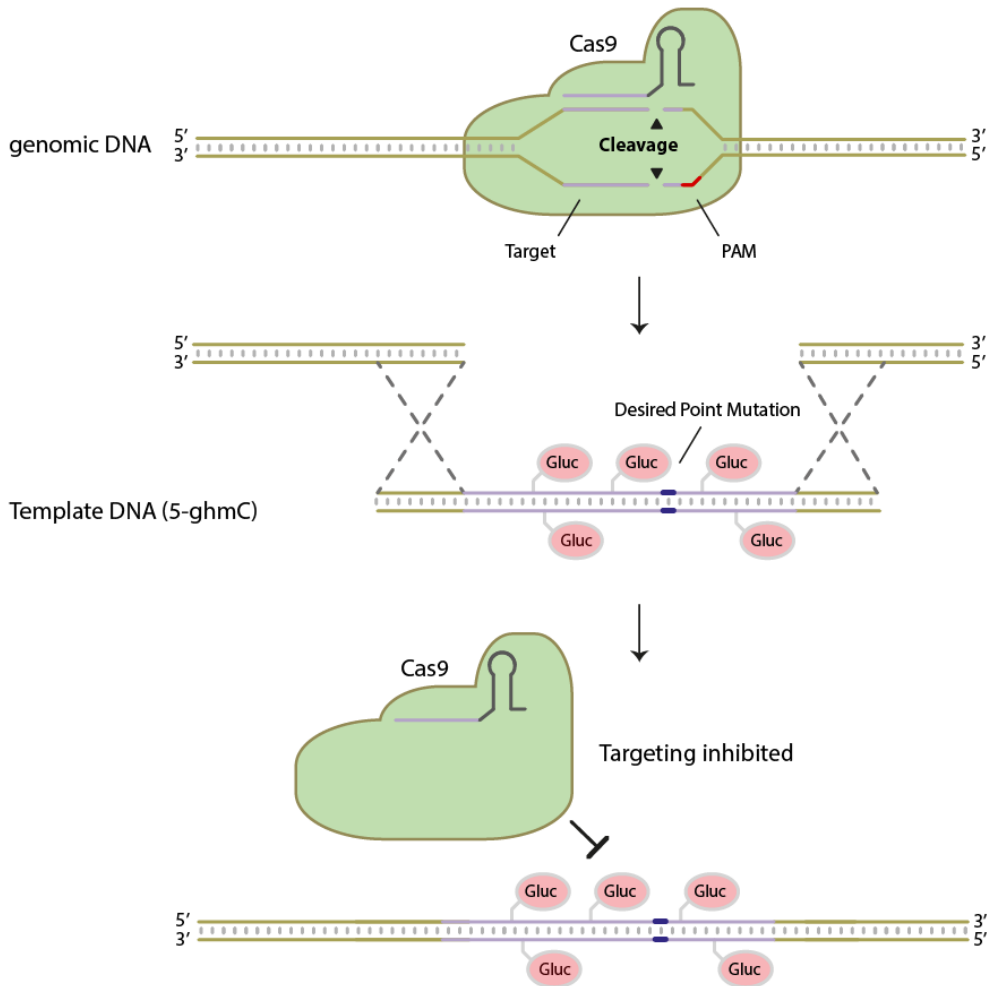
8

Figure 2. Generation of a double-strand break by Cas9, followed by repair via the Homology Directed Repair pathway. Glucosyl-hydroxymethylation of repair template DNA increases its life span and prevents re-cleavage of the DNA after succesfull repair.

Concerns regarding CRISPR-mediated genome engineering

A somewhat discouraging study demonstrated that antibodies against *Staphylococcus aureus* Cas9 (SaCas9) and *Streptococcus pyogenes* Cas9 (SpCas9) were found in human serum (Charlesworth et al., 2018). Upon further investigation, even anti-SaCas9 T-cells were identified in peripheral blood of almost half of the donors. These findings were explained by the fact that both *S. aureus* and *S pyogenes* are common human commensals that cause infections in humans at a high frequency, resulting in pre-existing adaptive immune responses. Pre-existing adaptive immunity against Cas9 likely will limit the effectiveness of Cas9-based gene therapy. As SaCas9

T-cells are able to induce CTL-mediated clearance of cells presenting Cas9 peptides, thereby eliminating edited cells and rendering the therapy ineffective (Kuwano et al., 1993; Mingozzi et al., 2007). Furthermore, this matter raises serious safety concerns because anti-Cas9 CTLs could potentially induce a systemic inflammation. It is clear that findings like these described above raise serious concerns about the efficacy and safety of CRISPR-based gene therapies. However, research into genetic engineering using CRISPR-Cas technologies opens up a wide range of possibilities and should not be discouraged by such concerns, but rather exploit these as stimuli to overcome current limitations.

Bacteriophages with DNA modifications and CRISPR-Cas

Bacteriophages have evolved diverse systems that allow them to counteract bacterial defense mechanisms. Evasion from CRISPR-Cas defense can occur by mutation of the protospacer or PAM sequence (Deveau et al., 2008; McGrath et al., 1999; Semenova et al., 2011), target site deletion (Pyenson et al., 2017), genome recombination (Paez-Espino et al., 2015), or by production of anti-CRISPR proteins (Bondy-Denomy et al., 2015; Pawluk et al., 2014, 2016). Besides the aforementioned escape strategies, phage-encoded DNA modifications, such as methylations, are well-known antagonists of microbial defense systems (Labrie et al., 2010). In chapter 5, we describe that phage-encoded DNA modifications, in this case glucosyl-hydroxymethylation of cytosines, provides a way to escape CRISPR-mediated immunity. We demonstrate that escape from CRISPR-mediated immunity is facilitated by the decreased binding affinity of CRISPR-effector complexes to glucosyl-hydroxymethylated DNA. Notably, this inhibitory effect of DNA modifications is not effective on all CRISPR systems, exemplified by the absence of inhibition on binding of the type V-A Cas12a effector protein. The fact that Cas12a DNA binding is not inhibited by glucosyl-hydroxymethylation of cytosines might raise the question whether there are bacteriophages that infect bacteria that harbor type V-A CRISPR-Cas systems. Phages that potentially contain modifications similar to those studied in chapter 5 (5'-hmC and 5'-ghmC) would at least require a dCMP-hydroxymethylase homolog for the synthesis of dhmCMP. Such homologs are mainly found in phages that infect enterobacteria and a few other families (Weigele and Raleigh, 2016). We have compared a list of phages that potentially contain cytosine modifications with a list of organisms that are currently known to harbor type V-A CRISPR-Cas systems (Zetsche et al., 2015). Unfortunately no match was found. Moreover we could not match any of the spacers to known phages. This might suggest that the prevalence of cytosine modifications of bacteriophage genomes is not high, and that phages that do have these modifications do not infect hosts that have type V-A CRISPR-Cas systems. Therefore the effect of these modifications on the evolution of CRISPR-Cas systems might be very limited.

Even though it is not yet clear if, and if so; how, the evolution of CRISPR-Cas systems and glucosyl-hydroxymethylation pathways are connected, it is worth noting that CRISPR-Cas systems might accelerate phage evolution (Tao et al., 2018). While phage T4 DNA is poorly cleaved by Cas9 and infection rates of T4 on cells expressing Cas9 are virtually unaffected, a remarkable difference in mutation frequencies can be detected

8

(Tao et al., 2018). Mutation frequencies of phage genomes that escape the inefficient cleavage by Cas9 are about six orders of magnitude higher than the spontaneous mutation frequency. While high mutation rates allow fast adaptation, and can thus be considered beneficial, they are not necessarily advantageous once adaptation is achieved (Elena et al., 2005; Giraud et al., 2001; Regoes et al., 2012). Considering that CRISPR-Cas systems exhibit strong selective pressure on adaptation speed, these systems are a driving force for phage evolution.

Bacteriophages with DNA modifications: practical applications

Bacteriophages with DNA modifications in fermentation processes

One of the main industries that has a key interest in studying anti-phage defense systems is the dairy industry. Bacteriophage infection of fermentation cultures, especially phage infection of starter cultures followed by slow fermentation or culture collapse, poses a large problem in this industry. Researchers in the dairy industry played a key role in the discovery of CRISPR-Cas systems and a better understanding of phage-host interactions is expected to contribute to the development of more effective anti-phage measures (Barrangou et al., 2007; Horvath and Barrangou, 2010). Our findings that bacteriophages that have extensively modified nucleobases can inhibit CRISPR-Cas interference should also be taken into account regarding the engineering of phage resistant cultures in the fermentation industry. First of all, it would be highly interesting to study the prevalence of bacteriophage DNA modifications among phages that infect fermentation cultures. If this would be the case, DNA modification insensitive CRISPR-Cas systems, such as Cas12a, could be introduced into the bacteria that are used for these industrial applications, in order to combat these infections.

Bacteriophages with DNA modifications in phage therapy

While bacteriophages that have DNA modifications can pose difficulties in fermentation processes, they might be favorable in applications where phage resistance is undesirable. One of such applications is bacteriophage therapy. Bacteriophage therapy is a promising solution to treat patients that are infected with antibiotic-resistant bacteria. The emergence of antibiotic-resistant bacteria is becoming an increasingly threatening problem. A recent report estimates that by 2050, every year ten million people will die due to antibiotic-resistant bacteria if no global effort to fight this problem is undertaken (O'Neill, 2014). Even though this estimate may be inaccurate due to the difficulty in determining reliable assumptions, it should be evident that there is a large clinical and public health burden associated with antibiotic resistance (de Kraker et al., 2016). The problem of antibiotic resistance becomes even more worrisome when considering that the number of newly discovered antibiotics is in decline.

Already proposed by the discoverer of bacteriophages, Félix d'Hérelle, bacteriophage therapy involves the infection and killing of bacteria by bacteriophages. The premise of such therapy has been explored in East

European countries for the last hundred years and is well accepted and received by physicians and patients. Until now, phage therapy has been used in only a small number of cases in Western countries. Even though a growing number of patients would potentially benefit from phage therapy, it is still rarely used, mostly because of a missing regulatory framework for the approval of therapeutic use of phages (Pirnay et al., 2015; Verbeken et al., 2014). While Western countries lag behind Eastern European countries regarding phage therapy, several phage-based products have been developed by Western countries. For example: the biotechnology company Intralytix has developed ListShield™, a phage preparation for controlling the foodborne bacterial pathogen *Listeria monocytogenes*, designed for treating foods that have a high risk for *L. monocytogenes* contamination.

One of the disadvantages of bacteriophage therapy over the use of conventional antibiotics is that phage resistance can emerge extremely rapidly (Örmälä and Jalasvuori, 2013). Phage resistance can be evolved by mutation of phage receptors, or by evolving more elaborate defense systems, such as CRISPR-Cas systems. The type of phage resistance that is likely to occur is higly dependent on the availability of resources (Westra et al., 2015). Phage cocktails, which contain several types of phages that infect the same species of bacteria, can be used to prevent the emergence of phage resistance (Salman et al., 2014; Nale et al., 2016). Another strategy would be to use phages against which the development of phage resistance would be more difficult. This is where phages that have DNA modifications come into play. As described in this thesis, DNA modifications can protect against two major defense systems employed by bacteria, namely restriction-modification systems and CRISPR-Cas systems. The genomes of phages that have DNA modifications are poorly cleaved by crRNA-effector proteins and acquisition of new spacers from such phages has never been reported (Strotskaya et al., 2017; Tao et al., 2018; Vlot et al., 2017). This suggests that such phages would be a good choice regarding the use of natural phages in phage therapy. Alternatively, DNA modification pathways could be introduced in bacteriophages using genetic engineering. We have demonstrated that such a pathway can be engineered in *E. coli*, and it would be likely that introduction of such a pathway into other phages would be possible.

The lytic nature of phages used in phage therapy, ironically, can promote the spreading of antibiotic resistance genes by horizontal gene transfer (Keen et al., 2017). Upon phage-induced cell lysis substantial amounts of DNA are released, including intact plasmids that encode antibiotic resistance genes. These plasmids can be taken up by naturally competent species, thereby increasing the spread of antibiotic resistance. However, many phages actively degrade the host DNA, including plasmid DNA, using hydrolytic endonucleases. Degradation of phage DNA is often prevented by nucleobase modifications. In the case of phage T4 the host DNA is degraded by endonucleases denA and denB while T4 DNA is protected from this degradation by glucosyl-hydroxymethylation of the phage DNA. Based on these features it becomes clear that phages that degrade the host genome are more suitable for use in medicine than their non-degrading counterparts.

8

What needs to happen to make phage therapy a success?

The interest in phage therapy in Western countries is mainly fuelled by several success stories, where patients, whose bacterial infection did not respond to antibiotics, were saved by experimental phage therapy. In some cases the notion "experimental'' has to be taken literal. This is exemplified in the successful case of Tom Patterson, who was cured from a *Acinetobacter baumannii* infection using a combination therapy of antibiotics and phages. When asked how the team that prepared the phage cocktail knew what dose to prepare, the answer was: "We didn't know, we are making this up as we go along" (Strathdee, 2017). This directly addresses one of the most relevant issues that requires ongoing thought on how to improve phage therapy. Renowned phage researcher Stephen T. Abedon wrote a report in which he stresses the importance of a more rational approach towards phage preclinical development (Abedon, 2018). Primarily the establishment of phage concentrations that are high enough to achieve adequate killing of bacteria needs attention. Therefore, a proper understanding of Poisson distributions, adsorption kinetics, and actual *in situ* multiplicity of infection (MOI) is required. For one, it is crucial to realize that high ratios of phages to bacteria do not necessarily result in eradication of the bacterial population. Even more so, initial *in situ* phage densities after applications of a particular dose of phages will hardly ever result in the killing of 100% of bacteria. This emphasizes that successful bacterial clearing strictly depends on the ability of the phages to replicate in the infection-causing host.

Despite a century of practical application of phages in Eastern Europe and aforementioned success stories of recent cases in the Western world, bacteriophage therapy is often met with scepticism by physicians, in some regards rightfully so. The lack of randomized double-blind placebo controlled clinical studies puts question marks at some of the claimed successes of phage therapy. However, that should be no reason to not pursue further development of experimental and clinical practices of this promising therapy. Fortunately several initiatives have been started, including a Phage Directory. The Phage Directory is a network of phage researchers that facilitate exchange of bacteriophages for emergence treatments (2017). Also a number of clinical test have been started or are about to start.

Perhaps more important than the practical side of phage therapy is the regulatory framework, or rather the lack thereof, for the approval of clinical use of phage therapy (Huys et al., 2013; Verbeken et al., 2014). Currently, bacteriophages do not have legal approval for therapeutic use and therefore cannot be used except as a last resort according to Article 37 of the Declaration of Helsinki (2013). One of the fundamental challenges that needs to be overcome is the classification of bacteriophages used for therapeutic purposes. Currently, bacteriophages are categorized as a human medicinal product (Art. 1 of Directive 2011/83/EC) due to the extreme broad definition of medicinal products. Despite the poor fit with this classification this results in the need for large clinical studies to demonstrate safety and efficacy, which would require approximately ten years and millions of euros for each phage to be approved. While such a trajectory to achieve legal approval would

be possible, it would be necessary for every individual bacteriophage product. One of the most notable advantages of bacteriophages over antibiotics is their specificity for bacterial species or even strains. This feature can only be fully exploited through a patient-specific approach. To enable the therapeutic use of phages to their full potential, a panel of phage researchers is advocating the adaptation of the regulatory framework. The proposed adaptation includes that phage therapeutics would be considered as advanced-therapy medicinal products (ATMPs). This would enable the use of patient-tailored phage preparations in hospital settings (Huys et al., 2013; Pirnay et al., 2018).

Phage therapy clearly has advantages over antibiotics, making them logical partners for our human immune defenses. While there is an evident need for alternatives to antibiotics, it is important to improve our scientific understanding of phage treatments and put in place a regulatory framework that is necessary to make it a successful endeavor. Further development will also require significantly larger investments than have been made so far, an investment that surely will be worth it.

8

Ongoing Research

Prokaryotic defense systems that protect against mobile genetic elements are mostly clustered in defense islands (Doron et al., 2018; Makarova et al., 2011). Recently, a new restriction and modification (R-M)-like defense systems was discovered and named Defense Island Systems Associated with Restriction and Modification (DISARM) (Ofir et al., 2017). This defense system is widespread in bacteria and archaea and is composed of five genes, including a DNA methyltransferase. DISARM systems can be divided in two types (type I and II). The DISARM type II system found in *Bacillus paralicheniformis* protects against infection by several families of DNA phages. Functionality of a type I DISARM system has not been demonstrated and is the main focus of our ongoing research. Some results of these studies are presented below. As the experiments described here are still in progress, they should be treated as preliminary results.

Discovery of the novel DISARM bacterial defense systems

The newly discovered DISARM gene cassette was initially discovered by analyzing the protein domain DUF1998 (pfam09369). DUF1998 is a protein of unknown function and is thought to be participating in anti-phage defense as it was previously found enriched in defense islands (Makarova et al., 2011). In an attempt to characterize the genetic context in which DUF1998 genes are found, 35,893 genomes in the Integrated Microbial Genomes (IMG) database were scanned. DUF1998 was found in 1,274 of 35,893 scanned microbial genomes. Next, neighboring genes of DUF1998 were analyzed. Results showed that in 1,095 of the 1,370 DUF1998-gene containing cases, DUF1998 was preceded by a large gene that contained a pfam00271 domain. This pfam00271 domain was previously found to be part of the catalytic core of DExx-box helicases, involved in a broad spectrum of activities such as rearranging of ribonucleoprotein complexes and pre-mRNA processing (Staley and Guthrie, 1998; Wang et al., 2006). Furthermore, upstream of DUF1998, a gene with a phospholipase D (PLD) domain (pfam13091) was found abundantly in 370 of the 1,370 cases. This PLD domain was previously found to be associated with enzymes manipulating phosphoester bonds like phospholipases, kinases and endonucleases (Selvy et al., 2011). This triplet of genes, pfam00271, DUF1998 and pfam13091, was defined as the core genes of the defense system, present in 351 of the 35,893 scanned genomes. These core genes were associated with a gene containing a methyltransferase domain in 324 of the 351 cases, suggesting that the systems resemble a novel type of R-M-like bacterial defense systems. This new defense system was therefore named Defense Island System Associated with Restriction Modification (DISARM). The DISARM defense system was divided in two types: type I, the common type, and type II, the rare type. Type I contains next to the triplet core genes a DNA adenine N6 methyltransferase gene (pfam13569), which is annotated as Type IIG RM gene in the restriction enzyme database REBASE, and a COG0553 SNF2 family helicase, which contains domains of SNF2-like ATPase (pfam00176) and of helicase C-terminal (pfam00271). Type II DISARM contains a DNA 5-cytosine methyltransferase (pfam00145), and in most cases a protein of unknown function (Figure 3).
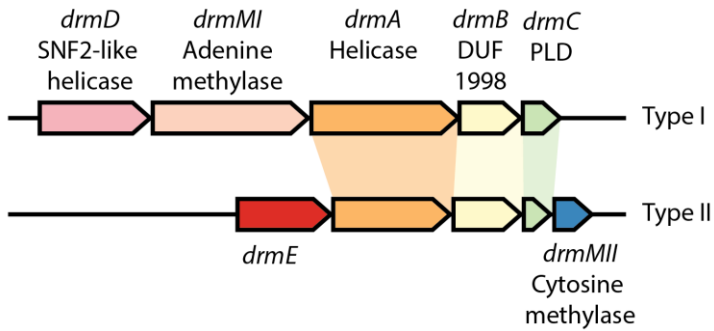
Figure 3. Two types of DISARM systems occur in bacteria and archaea.

The anti-viral activity of a type II DISARM system was demonstrated after Ofir et al. cloned the DISARM locus of *B. paralicheniformis* 9945a into *B. subtilis* BEST7300, an organism that naturally lacks a DISARM system. The engineered *Bacillus* strain was then challenged with phages from various morphological families; Siphophages, Myophages, and Podophages. Their results clearly demonstrate the anti-phage protection, manifested by a delay or absence of culture collapse upon phage infection (Ofir et al., 2017). Further experiments demonstrate that the DISARM-containing strain does allow phage adsorption and DNA injection but phage DNA replication is inhibited. The obvious similarities of the DISARM system with classical R-M systems suggested that methylation of self-DNA plays a role to allow discrimination between self- and non-self DNA. Cloning and expression of *drmMII* revealed that the DISARM system methylates CCWGG motifs in genomic DNA. However, replication inhibition of phages containing CCWGG methylation demonstrated that methylation alone is not sufficient to protect from the DISARM and the system probably uses an additional mechanism to discriminate self and non-self DNA. Sequencing of both bacterial and phage DNA after infection indicated that phage DNA was rapidly depleted in cells containing a functional DISARM system. Furthermore, it was found that phi3T could not form lysogens, suggesting that the DISARM system effectively forms immunity against phage-infections at an early stage.

In short, Type II DISARM showed phage resistance against several phages. After creating DISARM gene knockouts it was found that drmA, drmB and drmE were essential for the defense system. The drmMII gene was important for the survivability of the host, as it methylates the self-genome. These results suggest that DISARM to some extend resembles common R-M systems, however, its restriction modules consisting of multiple components appears to work differently.

8

Investigation of anti-phage activity by a type I DISARM system

To further study and unravel the mechanism of DISARM anti-phage systems we focused on a type I DISARM system. We hypothesize that the functionality of type I systems will be similar to that of type II systems. However, we anticipate that potential differences in activity of both systems might help in unraveling of the molecular mechanism of DISARM systems. To exploit the availability of genetic engineering systems for *E. coli*, we searched for type I DISARM systems present in organisms closely related to *E. coli*. Type I DISARM systems are found in 278 different organisms, of which 8 are classified as Enterobacteriaceae. We selected *Serratia* sp. SCBI (South African Caenorhabditis Briggsae Isolate) and analyzed the gene organization of the DISARM more closely.

This DISARM gene cassette comprises a cytosine methyltransferase (MT) gene, a *drmD* gene, a hypothetical gene, a *drmMI* gene, a *drmA* gene, a *drmB* gene and a *drmC* gene. Interestingly, the gene annotated as *drmC* (IMG Gene_OID: 2518347322) encodes a peptide of 89 amino acids, significantly shorter than the approximately 330 amino acid long homologs in 14 selected organisms containing a type I system. Careful analysis showed that the ORF of *drmC* was incorrectly annotated, and that the gene from bp 549,360-550,121 corresponds to the *drmC* homolog.
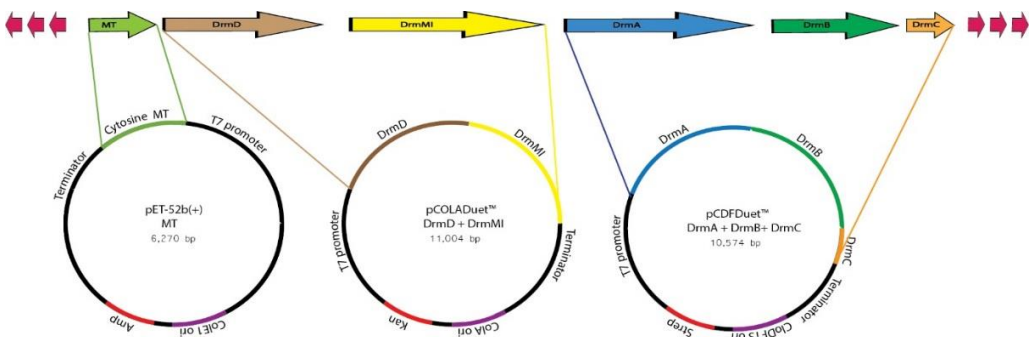

Figure 4. Schematic representation of the three compatible plasmids that contain the DISARM genes.

To investigate the functionality of the type I DISARM system from *Serratia* SCBI we cloned the 17 kbp gene cassette unto a set of three compatible plasmids. We transformed *E. coli* BL21-AI with the three plasmids encoding the full type I DISARM system (Figure 4). Cultures of DISARM-containing cells and cells containing empty plasmids were challenged with phage λ. It is important to note that phage λ was propagated in *E. coli* strain SCS110, which lacks both the *dcm* and *dam* methylase genes, resulting in a unmethylated phage stock. Infections were performed at a range of multiple orders of magnitude of multiplicity of infection (MOI). The optical density of the cultures was monitored for 16 hours and measured every ten minutes. The DISARM-containing cultures consistently showed a reduced rate of culture collapse (Figure 5). To quantify the level of protection provided by the DISARM system we calculated the rate of culture collapse after the maximum optical

density was reached as the difference in OD600 per minute [ΔOD600 min-1]. Although the MOI influences the time at which the maximum OD600 is reached after which the optical density decreases due to cell lysis, the rate of culture collapse is not influenced by the MOI (Figure).
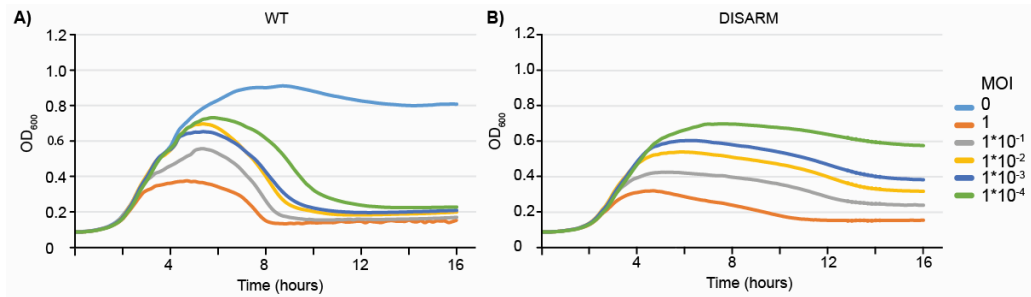

Figure 5. DISARM provides protection agains infection by phage λ. (A) Infection of WT E. coli by phage λ results in culture collapse indicated by a decrease in optical density. **(B)** Infection of E. coli containing DISARM by phage λ results in a delayed culture collapse indicated by a decrease in optical density.
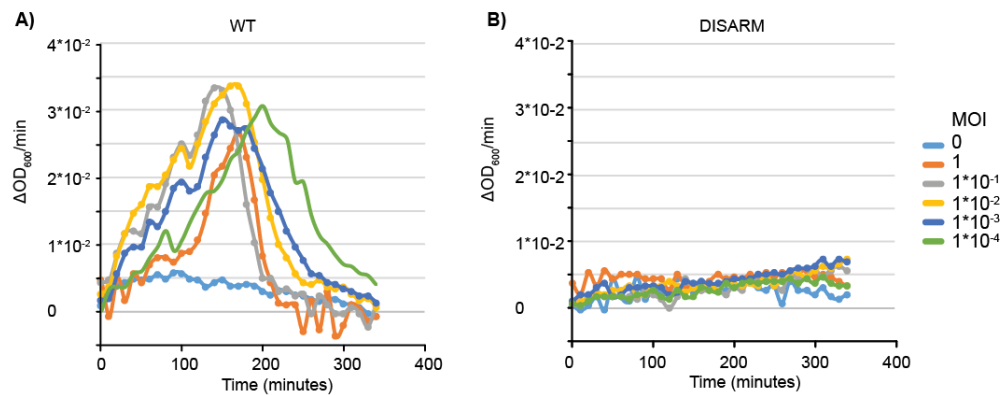

Figure 6. Rate of culture collapse by infection by phage λ. (A) Rate of culture collapse in WT *E. coli*. **(B)** Rate of culture collapse of *E. coli* containing DISARM.

DISARM systems are associated with restriction enzymes and methyltransferases. Until now it is not clear if these associated methyltransferases are essential for functionality of these DISARM systems. The type I DISARM system from Serratia sp. SCBI contains a methyltransferase upstream of *drmD* which we considered to be required for functionality of the system. When we transformed *E. coli* with the plasmid containing *drmABC* followed by the plasmid containing *drmD* and *drmMI*, the plasmids were heavily mutated. This suggests toxicity of the *drmABC* in the absence of *drmD* and/or *drmMI*. To prevent toxicity effects we cloned the methyltransferase (MT) and used this plasmid to transform *E. coli* prior to transformation with *drmABC* and *drmD-MI*. This resulted in a stable plasmid-based DISARM system, able to protect cells from phage infection.

The question still remains if the methyltransferase (MT) is essential, a question that can be answered by analysing DISARM functionality of a strain containing only *drmA-D* and *drmMI*. Furthermore, knockouts of all of the five DISARM individual genes, *drmA, drmB, drmC, drmMI* and *drmD*, need to be created in order to identify which genes are essential.

Recent studies have shown that a fusion between DUF1998 and DEAD/H-helicase, named *dpdJ*, are involved in the Dpd system, a new R-M system consisting of at least 12 genes that modifies DNA with 7-deazaguanin derivatives (Thiaville et al., 2016). Just like in the DISARM systems, *dpdJ* is followed by a PLD-containing gene named *dpdK*. Based on the similarity of the protein domains and the genetic organization thereof, we hypothesize that these proteins fulfill similar roles in the DNA restriction module of the Dpd system and the DISARM system (Figure 7).
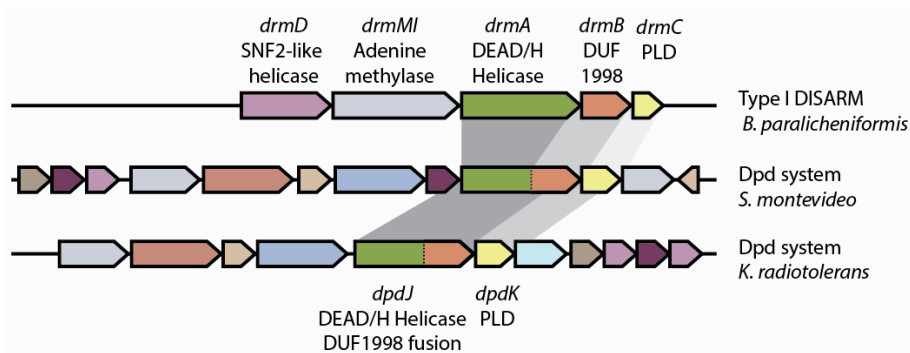

Figure 7. Gene clusters of the type I DISARM system in B. paralicheniformis and the Dpd system in S. montevideo and K. radiotolerans. Similar colors represent homologs.

Coevolution between bacteria and bacteriophages remains a highly interesting topic. There is a great variety in bacterial defense mechanisms, and anti-defense systems developed by bacteriophages. In recent years, more and more defense systems have been discovered suggesting that we only know the tip of the iceberg (Doron et al., 2018; Makarova et al., 2011). DISARM represents a new kind of R-M system with a different recognition and restriction module. The discovery of DISARM represents one of the many more defense systems that might be discovered in the future. Studying the arms race between bacteria and bacteriophages will give us a better understanding of the major evolutionary driving force shaping prokaryotic genomes.

References

Abedon, S.T. (2018). Phage Therapy: Various Perspectives on How to Improve the Art BT - Host-Pathogen Interactions: Methods and Protocols. C. Medina, and F.J. López-Baena, eds. (New York, NY: Springer New York), pp. 113–127.

Ali, S.S., Xia, B., Liu, J., and Navarre, W.W. (2012). Silencing of foreign DNA in bacteria. Curr. Opin. Microbiol. *15*, 175–181.

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.

Bergh, Ø., BØrsheim, K.Y., Bratbak, G., and Heldal, M. (1989). High abundance of viruses found in aquatic environments. Nature *340*, 467–468.

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

Breitbart, M., Salamon, P., Andresen, B., Mahaffy, J.M., Segall, A.M., Mead, D., Azam, F., and Rohwer, F. (2002). Genomic analysis of uncultured marine viral communities. Proc. Natl. Acad. Sci. *99*, 14250–14255.

Brenner, S., Jacob, F., and Meselson, M. (1961). An unstable intermediate carrying information from genes to ribosomes for protein synthesis. Nature *190*, 576–581.

Casjens, S.R., and Hendrix, R.W. (2015). Bacteriophage lambda: Early pioneer and still relevant. Virology *479–480*, 310–330.

Charlesworth, C.T., Deshpande, P.S., Dever, D.P., Dejene, B., Gomez-Ospina, N., Mantri, S., Pavel-Dinu, M., Camarena, J., Weinberg, K.I., and Porteus, M.H. (2018). Identification of Pre-Existing Adaptive Immunity to Cas9 Proteins in Humans. bioRxiv.

Chen, B., Gilbert, L.A., Cimini, B.A., Schnitzbauer, J., Zhang, W., Li, G.W., Park, J., Blackburn, E.H., Weissman, J.S., Qi, L.S., et al. (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. Cell *155*, 1479–1491.

Chen, Y., Liu, X., Zhang, Y., Wang, H., Ying, H., Liu, M., Li, D., Lui, K.O., and Ding, Q. (2016). A Self-restricted CRISPR System to Reduce Off-target Effects. Mol. Ther. *24*, 1508–1510.

Chrenek, M.A., Nickerson, J.M., and Boatright, J.H. (2016). Clustered Regularly Interspaced Short Palindromic Repeats: Challenges in Treating Retinal Disease. Asia-Pacific J. Ophthalmol. *5*, 304–308.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. Science (80-.). *339*, 819–823.

Court, D.L., Sawitzke, J.A., and Thomason, L.C. (2002). Genetic Engineering Using Homologous Recombination. Annu. Rev. Genet. *36*, 361–388.

Cox, D.B.T., Gootenberg, J.S., Abudayyeh, O.O., Franklin, B., Kellner, M.J., Joung, J., and Zhang, F. (2017). RNA editing with CRISPR-Cas13 David. Science (80-.). *1027*, 1019–1027.

Crick, F.H.C., Barnett, L., Brenner, S., and Watts-Tobin, R.J. (1961). General nature of the genetic code for proteins. Nature *192*, 1227–1232.

Dai, W.J., Zhu, L.Y., Yan, Z.Y., Xu, Y., Wang, Q.L., and Lu, X.J. (2016). CRISPR-Cas9 for in vivo Gene Therapy: Promise and Hurdles. Mol. Ther. - Nucleic Acids *5*, e349.

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.

Deveau, H., Barrangou, R., Garneau, J.E., Labonté, 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.

Doron, S., Melamed, S., Ofir, G., Leavitt, A., Lopatina, A., Keren, M., Amitai, G., and Sorek, R. (2018). Systematic discovery of antiphage defense systems in the microbial pangenome. Science (80-.).

Elena, S.F., Sanjuán, R., and Sanjua, R. (2005). Adaptive Value of High Mutation Rates of RNA Viruses : Separating Causes from Consequences J. Virol. *79*, 11555–11558.

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.

Fineran, P.C., Gerritzen, M.J.H., Suarez-Diez, M., Kunne, T., Boekhorst, van Hijum, F.T., Staals, R.H.J., Brouns, S.J.J., Suárez-Diez, M., Künne, T., et al. (2014). Degenerate target sites mediate rapid primed CRISPR adaptation. Proc. Natl. Acad. Sci. U. S. A. *111*, E1629--38.

Fuhrman, J.A. (1999). Marine viruses and their biogeochemical and ecological effects. Nature *399*, 541–548.

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.

Gilbert, L.A., Horlbeck, M.A., Adamson, B., Jacqueline, E., Chen, Y., Whitehead, E.H., Guimaraes, C., Ploegh, H.L., Bassik, M.C., and Qi, L.S. (2015). Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. Cell *159*, 647–661.

Gilbert, L. a, Larson, M.H., Morsut, L., Liu, Z., Gloria, A., Torres, S.E., Stern-ginossar, N., Brandman, O., Whitehead, H., Doudna, J. a, et al. (2013). CRISPR-Mediated Modular RNA-Guided Regualtion of Transcription in Eukaryotes. Cell *154*, 442–451.

Giraud, A., Matic, I., Tenaillon, O., Clare, A., Radman, M., Fons, M., and Taddei, F. (2001). Costs and benefits of high mutation rates: adaptive evolution of bacteria in the\rmouse gut. Science (80-.). *291*, 2606–2608.

Goulian, M., Kornberg, a, and Sinsheimer, R.L. (1967). Enzymatic synthesis of DNA, XXIV. Synthesis of infectious phage phi-X174 DNA. Proc. Natl. Acad. Sci. U. S. A. *58*, 2321–2328.

8

Hadas, H., Einav, M., and Zaritsky, A. (1994). Bacteriophage T4 development depends on the physiology of its host Escherichia coli. 179–185.

Hall, B.D., and Spiegelman, S. (1961). Sequence complementarity of T2-DNA and T2-specific RNA. Proc. Natl. Acad. Sci. *47*, 137–146.

Hershey, A.D., and Chase, M. (1952). Independent functions of viral protein and nucleic acid in growth of bacteriophage. J Gen Physiol. *36*, 39–56.

Heyer, W.-D., Ehmsen, K.T., and Liu, J. (2010). Regulation of homologous recombination in eukaryotes. Annu. Rev. Genet. *44*, 113–139.

Ho, C.-H., Wang, H.-C., Ko, T.-P., Chang, Y.-C., and Wang, A.H.-J. (2014). The T4 Phage DNA Mimic Protein Arn Inhibits the DNA-binding Activity of the Bacterial Histone-like Protein H-NS. J. Biol. Chem.

Horvath, P., and Barrangou, R. (2010). CRISPR/Cas, the immune system of bacteria and archaea. Science *327*, 167–170.

Huys, I., Pirnay, J.-P., Lavigne, R., Jennes, S., De Vos, D., Casteels, M., and Verbeken, G. (2013). Paving a regulatory pathway for phage therapy. EMBO Rep. *14*, 951–954.

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

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 (80-.). *356*.

Joo, C., Balci, H., Ishitsuka, Y., Buranachai, C., and Ha, T. (2008). Advances in Single-Molecule Fluorescence Methods for Molecular Biology. Annu. Rev. Biochem. *77*, 51–76.

Keen, E.C., and Gables, C. (2016). A century of phage research: Bacteriophages and the shaping of modern biology. *37*, 6–9.

Keen, E.C., Bliskovsky, V. V., Malagon, F., Baker, J.D., Prince, J.S., Klaus, J.S., and Adhya, S.L. (2017). Novel "superspreader" bacteriophages promote horizontal gene transfer by transformation. MBio *8*.

Kleinstiver, B.P., Pattanayak, V., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Zheng, Z., and Joung, J.K. (2016). High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. Nature *529*, 490–495.

Konermann, S., Brigham, M.D., Trevino, A.E., Abudayyeh, O.O., Barcena, C., Hsu, P.D., Gootenberg, J.S., Nishimasu, H., and Nureki, O. (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature *517*, 61422–61427.

de Kraker, M.E.A., Stewardson, A.J., and Harbarth, S. (2016). Will 10 Million People Die a Year due to Antimicrobial Resistance by 2050? PLoS Med. *13*, 1–6.

Kunne, T., Kieper, S.N., Bannenberg, J.W., Depken, M., Suarez-diez, M., Brouns, S.J.J., Ku, T., Kieper, S.N., Bannenberg, J.W., Vogel, A.I.M., et al. (2016). Cas3-Derived Target DNA Degradation Fragments Fuel Primed CRISPR Adaptation Article Cas3-Derived Target DNA Degradation Fragments Fuel Primed CRISPR Adaptation. 1–13.

Kuwano, K., Kawashima, T., and Arai, S. (1993). Antiviral Effect of TNF-a and IFN-7 Secreted from a CD8+ Influenza Virus-Specific CTL Clone. VIRAL Immunol. *6*, 1–11.

Labrie, S.J., Samson, J.E., and Moineau, S. (2010). Bacteriophage resistance mechanisms. Nat. Rev. Microbiol. *8*, 317–327.

Lichter, P., Tang, C.J., Call, K., Hermanson, G., Evans, G.A., Housman, D., and Ward, D.C. (1990). High-resolution mapping of human chromosome 11 by in situ hybridization with cosmid clones. Science (80-.). *247*, 64–69.

Luria, S., and Delbrück, M. (1943). Mutations of Bacteria from Virus Sensitivity to Virus Resistance. Genetics *28*, 491–511.

Luria, S.E., and Human, M.L. (1952). A nonhereditary, host-induced variation of bacterial viruses. J. Bacteriol. *64*, 557–569.

Ma, H., Naseri, A., Reyes-Gutierrez, P., Wolfe, S.A., Zhang, S., and Pederson, T. (2015). Multicolor CRISPR labeling of chromosomal loci in human cells. Proc. Natl. Acad. Sci. *112*, 3002–3007.

Makarova, K.S., Wolf, Y.I., Snir, S., and Koonin, E. V. (2011). Defense Islands in Bacterial and Archaeal Genomes and Prediction of Novel Defense Systems. J. Bacteriol. *193*, 6039–6056.

Maruyama, T., Dougan, S.K., Truttmann, M.C., Bilate, A.M., Ingram, J.R., and Ploegh, H.L. (2015). Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. Nat. Biotechnol. *33*, 538–542.

McGrath, S., Seegers, J.F.M.L., Fitzgerald, G.F., and Van Sinderen, D. (1999). Molecular characterization of a phage-encoded resistance system in Lactococcus lactis. Appl. Environ. Microbiol. *65*, 1891–1899.

Mingozzi, F., Maus, M. V Hui, D.J., Sabatino, D.E., Murphy, S.L., Rasko, J.E.J., Ragni, M. V, Manno, C.S., Sommer, J., Jiang, H., et al. (2007). CD8+ T-cell responses to adeno-associated virus capsid in humans. Nat. Med. *13*, 419–422.

Mosberg, J.A., Lajoie, M.J., and Church, G.M. (2010). Lambda red recombineering in Escherichia coli occurs through a fully single-stranded intermediate. Genetics *186*, 791–799.

Mosberg, J.A., Gregg, C.J., Lajoie, M.J., Wang, H.H., and Church, G.M. (2012). Improving Lambda Red Genome Engineering in Escherichia coli via Rational Removal of Endogenous Nucleases. *7*, 1–12.

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

Nale, J.Y., Spencer, J., Hargreaves, K.R., Buckley, A.M., and Trzepin, P. (2016). Bacteriophage Combinations Significantly Reduce Clostridium difficile Growth In Vitro and Proliferation In Vivo. Antimicrob. Agents Chemother. 60, 968–981.

Navarre, W.W., McClelland, M., Libby, S.J., and Fang, F.C. (2007). Silencing of xenogenic DNA by H-NS: facilitation of lateral gene transfer in bacteria by a defence system that recognizes foreign DNA. Genes Dev. *21*, 1456–1471.

Nelles, D.A., Fang, M.Y., O'Connell, M.R., Xu, J.L., Markmiller, S.J., Doudna, J.A., and Yeo, G.W. (2016). Programmable RNA Tracking in Live Cells with CRISPR/Cas9. Cell *165*, 488–496.

Nelson, C.E., Hakim, C.H., Ousterout, D.G., Thakore, P.I., Moreb, E.A., Rivera, R.M.C., Madhavan, S., Pan, X., Ann, F., et al. (2017). In vivo genome editing improves muscle function in a mouse model of Duchenne muscular dystrophy. Science *351*, 403–407.

O'Connell, M.R., Oakes, B.L., Sternberg, S.H., East-Seletsky, A., Kaplan, M., and Doudna, J.A. (2014). Programmable RNA recognition and cleavage by CRISPR/Cas9. Nature *516*, 263–266.

O'Neill, J. (2014). Review on Antimicrobial Resistance. Antimicrobial Resistance: Tackling a Crisis for the Health and Wealth of Nations. 4–16.

Ofir, G., Melamed, S., Sberro, H., Mukamel, Z., Silverman, S., Yaakov, G., Doron, S., and Sorek, R. (2017). DISARM is a widespread bacterial defence system with broad anti-phage activities. Nat. Microbiol. 1–9.

Örmälä, A.-M., and Jalasvuori, M. (2013). Phage therapy. Bacteriophage 3, e24219.

Oude Blenke, E., Evers, M.J.W., Mastrobattista, E., and van der Oost, J. (2016). CRISPR-Cas9 gene editing: Delivery aspects and therapeutic potential. J. Control. Release *244*, 139–148.

Paez-Espino, D., Sharon, I., Morovic, W., Stahl, B., Thomas, B.C., Barrangou, R., Banfield, J.F., and Banfielda, J.F. (2015). CRISPR Immunity Drives Rapid Phage Genome Evolution in Streptococcus thermophilus. MBio *6*, e00262-15-.

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

Pawluk, A., Amrani, N., Zhang, Y., Garcia, B., Hidalgo-Reyes, Y., Lee, J., Edraki, A., Shah, M., Sontheimer, E.J., Maxwell, K.L., et al. (2016). Naturally Occurring Off-Switches for CRISPR-Cas9. Cell *167*, 1829–1838.e9.

Pirnay, J.P., Blasdel, B.G., Bretaudeau, L., Buckling, A., Chanishvili, N., Clark, J.R., Corte-Real, S., Debarbieux, L., Dublanchet, A., De Vos, D., et al. (2015). Quality and safety requirements for sustainable phage therapy products. Pharm. Res. *32*, 2173–2179.

Pirnay, J.-P., Verbeken, G., Ceyssens, P.J., Huys, I., de Vos, D., Ameloot, C., and Fauconnier, A., (2018). The magistral phage. Viruses *10,* 1-7.

Prangishvili, D. (2013). The Wonderful World of Archaeal Viruses. Annu. Rev. Microbiol. *67*, 565–585.

Proctor, L.M., and Fuhrman, J. a. (1990). Viral mortality of marine bacteria and cyanobacteria. Nature *343*, 60–62.

Pyenson, N.C., Gayvert, K., Varble, A., Elemento, O., and Marraffini, L.A. (2017). Broad Targeting Specificity during Bacterial Type III CRISPR-Cas Immunity Constrains Viral Escape. Cell Host Microbe 1–11.

Rahul, R., Hohng, S., and Ha, T. (2008). A Practical Guide to Single Molecule FRET. Nat. Methods *5*, 507–516.

Ran, F.A., Cong, L., Yan, W.X., Scott, D.A., Kriz, J.S.G., J., A., Zetsche, B., Shalem, O., Wu, X., Makarova, K.S., et al. (2015). In vivo genome editing using Staphylococcus aureus Cas9. Nature *520*, 186–191.

Redding, S., Sternberg, S.H., Wiedenheft, B., Jennifer, A., Greene, E.C., Marshall, M., Gibb, B., Bhat, P., and Guegler, C.K. (2015). Surveillance and Processing of Foreign DNA by the Escherichia coli CRISPR-Cas System. Cell *163*, 1–12.

Regoes, R.R., Hamblin, S., and Tanaka, M.M. (2012). Viral mutation rates: modelling the roles of within-host viral dynamics and the trade-off between replication fidelity and speed. Proc. R. Soc. B Biol. Sci. *280*, 20122047–20122047.

Renkawitz, J., Lademann, C. a, and Jentsch, S. (2014). Mechanisms and principles of homology search during recombination. Nat. Rev. Mol. Cell Biol. *15*, 369–383.

Richardson, C.D., Ray, G.J., DeWitt, M.A., Curie, G.L., and Corn, J.E. (2016). Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. Nat. Biotechnol. *34*, 339–344.

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.

Rohwer, F., and Edwards, R. (2002). The phage proteomic tree: A genome-based taxonomy for phage. J. Bacteriol. *184*, 4529–4535.

Rosen, H., Libby, S.J., and Fang, F.C. (2006). Selective Silencing of Foreign DNA Protein in Salmonella. Science (80-.). *682*, 236–238.

Salman, A.E., and Abdulamir, A.S. (2014). Assessment of bacteriophage cocktails used in treating multiple-drug resistant Pseudomonas aeruginosa. 3, 711–722.

Sauer, B. (1987). Functional expression of the cre-lox site-specific recombination system in the yeast Saccharomyces cerevisiae. Mol. Cell. Biol. *7*, 2087–2096.

Sauer, B., and Henderson, N. (1988). Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1. Proc. Natl. Acad. Sci. *85*, 5166–5170.

Savić, N., and Schwank, G. (2016). Advances in therapeutic CRISPR/Cas9 genome editing. Transl. Res. *168*, 15–21.

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.

Selvy, P.E., Lavieri, R.R., Lindsley, C.W., and Brown, H.A. (2011). Phospholipase D: Enzymology, functionality, and chemical modulation. Chem. Rev. *111*, 6064–6119.

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

8

by a seed sequence. Proc. Natl. Acad. Sci. U. S. A. *108*, 10098–10103.

Sgaramella, V., Van de Sande, J.H., and Khorana, H.G. (1970). Studies on polynucleotides. A novel joining reaction catalyzed by the T4-polynucleotide ligase. Proc. Natl. Acad. Sci. U. S. A. *67*, 1468–1475.

Shin, J., Jiang, F., Liu, J.-J., Bray, N.L., Rauch, B.J., Baik, S.H., Nogales, E., Bondy-Denomy, J., Corn, J.E., and Doudna, J.A. (2017). Disabling Cas9 by an anti-CRISPR DNA mimic. Sci. Adv. *3*.

Singh, D., and Ha, T. (2018). Understanding the molecular mechanisms of CRISPR toolbox using single molecule approaches.

Slaymaker, I.M., Gao, L., Zetsche, B., Scott, D.A., Yan, W.X., and Zhang, F. (2016). Rationally engineered Cas9 nucleases with improved specificity. Science *351*, 84–88.

Smith, G.P., and Petrenko, V.A. (1997). Phage display. Chem. Rev. *97*, 391–410.

Staley, J.P., and Guthrie, C. (1998). Mechanical devices of the spliceosome: Motors, clocks, springs, and things. Cell *92*, 315–326.

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*, 62–67.

Strathdee, S. (2017). How Sewage Saved My Husband's Life from a Superbug.

Strotskaya, A., Savitskaya, E., Metlitskaya, A., Morozova, N., Datsenko, K.A., Semenova, E., and Severinov, K. (2017). The action of *Escherichia coli* CRISPR–Cas system on lytic bacteriophages with different lifestyles and development strategies. Nucleic Acids Res. *45*, gkx042.

Studier, F.W., and Moffatt, B.A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J. Mol. Biol. *189*, 113–130.

Suttle, C.A. (2005). Viruses in the sea. Nature *437*, 356–361.

Tabebordbar, M., Zhu, K., Cheng, J.K.W., Chew, W.L., Widrick, J.J., Yan, W.X., Maesner, C., Wu, E.Y., Xiao, R., Ran, F.A., et al. (2016). In vivo gene editing in dystrophic mouse muscle and muscle stem cells. *351*, 407–411.

Tao, P., Wu, X., and Rao, V. (2018). Unexpected evolutionary benefit to phages imparted by bacterial CRISPR-Cas9. 1–10.

Thiaville, J.J., Kellner, S.M., Yuan, Y., Hutinet, G., Thiaville, P.C., Jumpathong, W., Mohapatra, S., Brochier-Armanet, C., Letarov, A. V., Hillebrand, R., et al. (2016). Novel genomic island modifies DNA with 7-deazaguanine derivatives. Proc. Natl. Acad. Sci. 201518570.

Thingstad, T.F. (1990). Partners in Spring Bloom Microbial Trophodynamics. *56*, 1400–1405.

Verbeken, G., Pirnay, J.P., Lavigne, R., Jennes, S., De Vos, D., Casteels, M., and Huys, I. (2014). Call for a dedicated European legal framework for bacteriophage therapy. Arch. Immunol. Ther. Exp. (Warsz). *62*, 117–129.

Vlot, M., and Brouns, S.J.J. (2017). Dna modification. Patent WO2017055514 A1.

Vlot, M., Houkes, J., Lochs, S.J.A., Swarts, D.C., Zheng, P., Kunne, T., Mohanraju, P., Anders, C., Jinek, M., van der Oost, J., et al. (2017). Bacteriophage DNA glucosylation impairs target DNA binding by type I and II but not by type V CRISPR–Cas effector complexes. Nucleic Acids Res. 1–13.

Wang, F., and Qi, L.S. (2016). Applications of CRISPR Genome Engineering in Cell Biology. Trends Cell Biol. *26*, 875–888.

Wang, S., Hu, Y., Overgaard, M.T., Karginov, F. V, Uhlenbeck, O.C., and McKay, D.B. (2006). The domain of the Bacillus subtilis DEAD-box helicase YxiN that is responsible for specific binding of 23S rRNA has an RNA recognition motif fold. RNA *12*, 959–967.

Weigele, P., and Raleigh, E.A. (2016). Biosynthesis and Function of Modified Bases in Bacteria and Their Viruses.

Westra, E.R., Pul, Ü., Heidrich, N., Jore, M.M., Lundgren, M., Stratmann, T., Wurm, R., Raine, A., Mescher, M., Van Heereveld, L., et al. (2010). H-NS-mediated repression of CRISPR-based immunity in Escherichia coli K12 can be relieved by the transcription activator LeuO. Mol. Microbiol. *77*, 1380–1393.

Westra, E.R., Buckling, A., and Fineran, P.C. (2014). CRISPR–Cas systems: beyond adaptive immunity. Nat. Rev. Microbiol. *12*, 317–326.

Westra, E.R., Van houte, S., Oyesiku-Blakemore, S., Makin, B., Broniewski, J.M., Best, A., Bondy-Denomy, J., Davidson, A., Boots, M., and Buckling, A. (2015). Parasite exposure drives selective evolution of constitutive versus inducible defense. Curr. Biol. *25*, 1043–1049.

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.

Wu, W.Y., Lebbink, J.H.G., Kanaar, R., Geijsen, N., Van der Oost, J. (2018) Genome editing by natural and engineered CRISPR-associated nucleases. Nat. Chem. Biol. in press

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.

Yang, L., Guell, M., Byrne, S., Yang, J.L., De Los Angeles, A., Mali, P., Aach, J., Kim-Kiselak, C., Briggs, A.W., Rios, X., et al. (2013). Optimization of scarless human stem cell genome editing. Nucleic Acids Res. *41*, 9049–9061.

Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., van der Oost, J., Regev, A., et al. (2015). Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. Cell *163*, 759–771.

Zhang, Y., Heidrich, N., Ampattu, B.J., Gunderson, C.W., Seifert, S.H., Schoen, C., Vogel, J., and Sontheimer, E.J. (2013). Processing-Independent CRISPR RNAs Limit Natural Transformation in Neisseria meningitidis Yan. Mol Cell *50*, 488–503.

(2013). Wma Declaration of Helsinki – Ethical Principles for Medical. 1–5.

(2017). Phage Directory. Https://phage.directory.

8

Appendices

About the author

List of publications

Overview of completed training activities

Acknowlegdements

About the author

Marnix Vlot was born on the 16th of August, 1990, in Rotterdam, The Netherlands. In 2008, he finished his pre-university education (VWO) at the Gomarus Scholengemeenschap in Gorinchem, and started his Bachelor study Biotechnology at the Wageningen University. During his Bachelor thesis, Marnix studied the root nodule formation of the non-legume *Parasponia andersonii* in the laboratory of Ton Bisseling. During his Master study he specialized in cellular and molecular biotechnology. In 2012, Marnix joined the Wageningen International Genetically Engineered Machine (iGEM) team, which was rewarded with the prize for the best presentation during the European competition and advanced to the World Finals at the Massachusetts Institute of Technology in Boston (USA).

In 2013, he studied the CRISPR-Cas type III-A and III-B systems in *Thermus thermophilus* for his major thesis in the laboratory of John van der Oost. Fascinated by the molecular mechanisms of CRISPR-Cas systems he continued with his minor thesis trying to identify bacteriophage-encoded proteins that could inhibit the CRISPR-Cas type I-E system in *Escherichia coli*. In 2014, Marnix continued to study the CRISPR-Cas type I-E system in *E. coli* under the supervision of Dr. Stan Brouns and Prof. John van der Oost. The findings that resulted from the work in the past 4 years are described in this thesis.

Publication list

1- Vink, J.*, **Vlot, M*.,** McKenzie, R.E., Martens, K.J.A., Almendros, C., Estrada Bonilla, B., Brocken, D.J.W., Hohlbein, J., Brouns, S.J.J. Direct visualization of CRISPR target search and self-avoidance in live bacteria. *Manuscript in preparation*

2- **Vlot, M,** Martens, K.J.A., Groen, J., Hohlbein, J., Zheng, P., Dickman, M.J., Van der Oost, J., Brouns, S.J.J. A DNA labeling approach based on incorporation of modifiable nucleobases using bacteriophage T4 genes. *Manuscript in preparation*

3- Nobrega, F.L., **Vlot, M.,** De Jonge, P.A., Dreesens, L.L., Beaumont H.J.E., Lavigne, R., Dutilh, B.E., Brouns, S.J.J. (2018). Targeting mechanisms of tailed bacteriophages. *Nature Reviews Microbiology, in press*

4- **Vlot, M.,** Nobrega, F.L., Wong, C.F.A., Liu, Y., and Brouns, S.J.J. (2018). Complete Genome Sequence of the Escherichia coli Phage Ayreon. Genome Announc., 10.1128/genomeA.01354-17.

5- **Vlot, M.,** Houkes, J., Lochs, S.J.A., Swarts, D.C., Zheng, P., Kunne, T., Mohanraju, P., Anders, C., Jinek, M., van der Oost, J., et al. (2017). Bacteriophage DNA glucosylation impairs target DNA binding by type I and II but not by type V CRISPR–Cas effector complexes. Nucleic Acids Res. 1–13.

6- **Vlot, M.,** and Brouns, S.J.J. (2015). Dna modification. Patent WO2017055514 A1.

7- Blosser, T.R.R., Loeff, L., Westra, E.R.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.

8- Staals, R.H.J., Zhu, Y., Taylor, D.W., Kornfeld, J.E., Sharma, K., Barendregt, A., Koehorst, J.J., **Vlot, M.,** Neupane, N., Varossieau, K., Sakamoto, K., Suzuki, T., Dohmae, N., Yokoyama, S., Schaap, P.J., Urlaub, H., Heck, A.J.R., Nogales, E., Doudna, J.A., Shinkai, A., Van der Oost, J. (2014). RNA targeting by the type III-A CRISPR-Cas Csm complex of Thermus thermophilus. Mol. Cell 56, 518–530.

9- Staals, R.H.J.H.J., Agari, Y., Maki-Yonekura, S., Zhu, Y., Taylor, D.W.W., van Duijn, E., Barendregt, A., **Vlot, M.,** Koehorst, J.J.J., Sakamoto, K., Masuda, A., Dohmae, N., Schaap, P.J., Doudna, J.A., Heck, A.J.R., Yonekura, K., Van der Oost, J., Shinkai, A. (2013). Structure and Activity of the RNA-Targeting Type III-B CRISPR-Cas Complex of Thermus thermophilus. Mol. Cell 52, 135–145.

10- Musharova, O.*, Sitnik, V., **Vlot, M.,** Savitskaya, E., Datsenko, K.A., Krivoy, A., Fedorov, I., Semenova, E., Brouns, S.J.J., Severinov, K. Systematic analysis of Type I-E *Escherichia coli* CRISPR-Cas PAM sequences ability to promote interference and primed adaptation. *Manuscript in preparation*

11- **Vlot, M*.,** Aparicio-Maldonado, C*,. Ofir, G., Middendorf, P., Nobrega, F.L., Sorek, R., Brouns, S.J.J. Characterization of a novel type I DISARM system from Serratia SCBI. *Manuscript in preparation*

12- Zhu, Y., Klompe, S., **Vlot, M.,** Van der Oost, J., Staals, R.H.J. Shooting the messenger: RNA-targeting CRISPR-Cas systems. *Bioscience Reports, in press*

13- Aparicio-Maldonado, C*., **Vlot, M*.,** Ofir, G., Oktarianti Ainun Jariah, R., Nobrega, F.L., Sorek, R., Brouns, S.J.J. Cleavage of phage DNA by the DISARM-complex of *Bacillus paralicheniformis* 9945a. *Manuscript in preparation*

*Equal contributions

Overview of completed training activities

Discipline-specific activities

Meetings & conferences

- EMBO meeting, New Approaches and Concepts in Microbiology. Heidelberg (DE), 2017

- Microbiology Centennial meeting. Wageningen (NL), 2017

- NWO-ALW molecular genetics meeting. Wageningen (NL), 2017

- NWO-ALW molecular genetics meeting. Wageningen (NL), 2016**

- NBC-16, annual NVB meeting. Wageningen (NL), 2016*

- NWO-ALW molecular genetics meeting. Lunteren (NL), 2015*

- Viruses of Microbes meeting. EMBO, Zurich (CH), 2015*

- KNVM Papendal meeting. Arnhem (NL), 2015

- NWO-ALW molecular genetics meeting. Lunteren (NL), 2014

- CRISPR meeting 2014. Berlin (DE), 2014

- Annual Meeting NWO Protein Research, Nucleic Acids and Lipids & Membranes. Veldhoven (NL), 2014

- CRISPR Meeting, Biochemical Society. St. Andrews (UK), 2014

Courses

- PacBio Seminar, CAT-AGRO. Wageningen (NL), 2016

- PHAGE course, MIT. Braga (PT), 2015

- Spring School Host Microbe Interactions, VLAG. Wageningen (NL), 2014*

- Radiation Course level 5B, WUR. Wageningen (NL), 2014

General courses

- VLAG PhD week. Baarlo (NL), 2014

- Techniques for Writing and Presenting Scientific Papers, WGS/language services. Wageningen (NL), 2015

- Advanced Scientific Artwork, WUR library. Wageningen (NL), 2014

- Essential Skills for Data Intensive Research, SURF eScience. Utrecht (NL), 2016

- Speed Reading and Brain Training, VLAG. Wageningen (NL), 2016

- Presenting with Impact, VLAG. Wageningen (NL), 2016

Optionals

- PhD study trip MIB-SSB. California (USA), 2015*

- Organizing committee, Antiviral defense mechanisms symposium. Wageningen (NL), 2014

- Bacterial Genetics group meetings

- PhD meetings Laboratory of Microbiology

- Microbiology Seminars

*poster presentation, **oral presentation

Acknowledgements

"All good things must come to an end" (Geoffrey Chaucer, 1374). That also applies to doing a PhD. I'm really grateful that I could finish this endeavour and even more grateful for everyone who has helped me accomplish it. I enjoyed the freedom that I had to pursue a diverse set of projects, ranging from phage biology and CRISPR-Cas systems to DNA modifications and single-molecule microscopy.

First of all, I want to thank **Stan**. I enjoyed working with you a lot. Especially the first years, when we shared the same office, were really cool. I admire your intelligence and your ability to remember things about my projects that I already forgot myself. Also, like no other, you can be really enthusiastic about something, while being super calm at the same time. It is great to how see well you do in science. Four years ago you only had to keep an eye on Tim and me, while now you have around a dozen people to keep in line. Besides work, you are also just a really nice guy. Btw. I hope that your ETNs will be worth something someday.

John, thank you for your support during my PhD. When I declined to work on the Argonaut project, I learned soon enough that you cannot hold a grudge for more than 1.3 seconds. Or maybe, deep down, you agree that CRISPR-Cas systems are a lot cooler than these primitive Argonauts. Your way of leading the BacGen group does hardly leave anything for us to complain about. Looking back, we often take the amazing group for granted, but that's maybe because your efforts are 'granted'. Thank you.

A great thank you to all students whom I have supervised. **Alex, Małgorzata, Nicolo, Jochem, Silke, Renée, Kylie, Alicia, Henri, Joep, Herman, Pjotr, Rizka, Jan, and Sebastiaan.** I truly find that working with you was one of the best parts of doing a PhD. The countless unsuccessful experiments were countered by the shear amount of work that you all have done, very impressive. Some of you even find your name back in PubMed as a result. The fact that, as far as I know, nine of you are now pursuing your own PhD, allows me to think that I didn't do such a bad job, as to scare you away from science forever. I wish all of you the very best.

Thanks to all colleagues who shared an office with me. **Daan**, your posture is a good proxy for your savviness. Notwithstanding, your belief that you could resurrect a dead plant by giving it coffee still amazes me. **Tim**, you left some big shoes for me to fill. Luckily, you left 239 pages to draw from. You're a man of few words, but when you talk, it often cannot be more spot-on. **Yifan**, I have known you for some time now, but still I am amazed by how much you care about whiskey. When it comes to Cmr or Csm, I doubt there is a soul who can recall the details of a specific paper more rapidly than you. **Jorrit**, you managed to draw blood from a stone, an eight-armed stone. Well done! Furthermore, I suppose that no one can be more excited about cleavage than you. **Wen**, you have an incredible energy density. Good luck making that decision between becoming rich or remain being a scientist. **Sjoerd**, our walking encyclopaedia, you know something about everything. Thanks for your help in the design of experiments and sharing your countless tips and tricks. **Becca and Prarthana**, my 'normal' colleagueas, it was a pleasure to share an office with you. I wish you all the very best. **Patrick and Sebastian,**

colleagues who turned semi-colleagues, it was good to have you around. **Franklin,** always asking for one more change to the figures, it was a pleasure to work with you and I keep good memories from our stay in Portugal. These tailed inanimate things indeed seem very much alive, and I share your interest in them. **Jochem,** it was really cool to be involved in the single-molecule project with you. The work you did and things we discovered are in one word incredible (but credible at the same time).

Willem, we never talked for more than 20 seconds, but still I appreciated these talks a lot. Thank you for leading the department in the way you did, it has made it a successful and esteemed group. **Romy**, it was fun to organize the MIB seminars together, especially because you did most of the work. Thank you. A big thanks to all the people involved in the iGEM competition. **Rob, Ruben, Stamatios, Christian, Kees, Maarten, Nico,** it was really inspiring to work with you in supervising the iGEM teams.

Tijn, Elleke, Nico, Tom, Servé, Edze, Richard, Mark, Ioannis, Joyshree, Jeroen, Lione, Bas, Jasper, Teunke, Melvin, Kees, Alex, Thijs, Ismael, Raymond, Tessa, Caroline, it was fun to have lunch, drinks, and whatnot together. You all played a good part in making it an enjoyable time. Also **Anja, Philippe, Sjon, Wim, Ton, Tom, Steven, Merlijn, Gosse, Carolien, Mirella,** many thanks for all the support that you provide for the group, it was much appreciated.

Marc, Peiyuan, An-Wen, Gal, Sarah, Rotem, Chirlmin, Luuk, Cees, Tim, Cristian, Koen, Johannes, Martin, Carolin, Olga, Vasily, and Konstantin, I thank you all very much for the fruitful collaborations that we had. You are all experts in your field, and it has been inspiring to work with you.

Pa, Ma, Ruben, Maria, Rhode, Marco, Wulfert, Melvin, Leen, Lydia, Kyra, Jeroen, thanks for your interest in my work. The fact that it was sometimes hard to grasp what I was working on never stopped you from showing your interest. I really appreciate that.

Kristel, you are the most amazing person that I know. And eventhough you started about two years later than I did, I think you know better how to do a PhD than I do. Tricking you into falling in love with me was the biggest achievement in my life, much bigger than finishing a PhD. I feel incredibly lucky that you are in my life, and I am looking forward to all the time that we will share together.
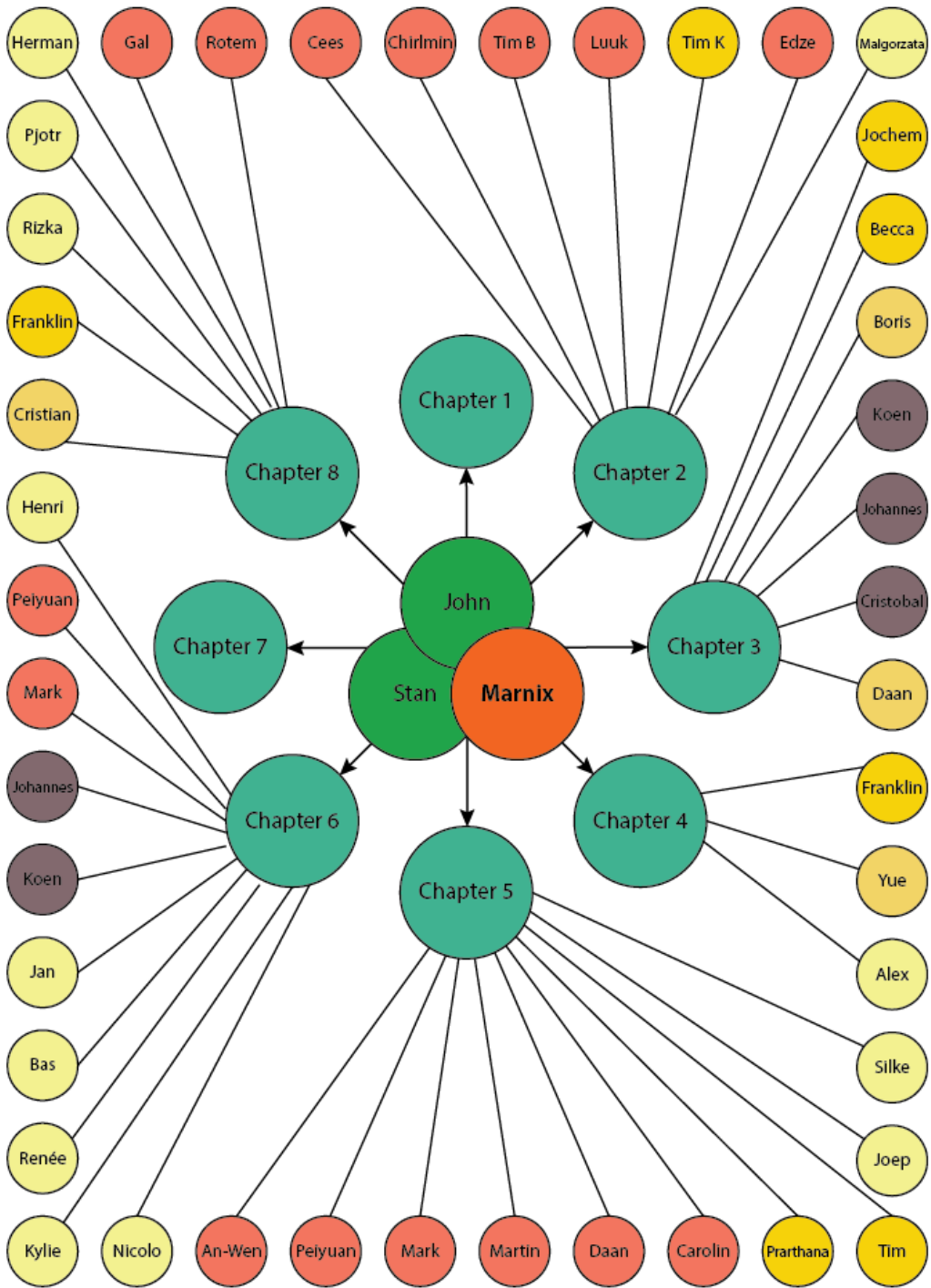
Vizual representation of people who have contributed to the completion of this PhD thesis. Circle sizes do not represent importance of contribution.