The type I-E CRISPR-Cas system

Biology and applications of an adaptive immune system in bacteria

LINA AMLINGER
Abstract


CRISPR-Cas systems are adaptive immune systems in bacteria and archaea, consisting of a clustered regularly interspaced short palindromic repeats (CRISPR) array and CRISPR associated (Cas) proteins. In this work, the type I-E CRISPR-Cas system of *Escherichia coli* was studied.

CRISPR-Cas immunity is divided into three stages. In the first stage, adaptation, Cas1 and Cas2 store memory of invaders in the CRISPR array as short intervening sequences, called spacers. During the expression stage, the array is transcribed, and subsequently processed into small CRISPR RNAs (crRNA), each consisting of one spacer and one repeat. The crRNAs are bound by the Cascade multi-protein complex. During the interference step, Cascade searches for DNA molecules complementary to the crRNA spacer. When a match is found, the target DNA is degraded by the recruited Cas3 nuclease.

Host factors required for integration of new spacers into the CRISPR array were first investigated. Deleting *recD*, involved in DNA repair, abolished memory formation by reducing the concentration of the Cas1-Cas2 expression plasmid, leading to decreased amounts of Cas1 to levels likely insufficient for spacer integration. Deletion of RecD has an indirect effect on adaptation. To facilitate detection of adaptation, a sensitive fluorescent reporter was developed where an out-of-frame *yfp* reporter gene is moved into frame when a new spacer is integrated, enabling fluorescent detection of adaptation. Integration can be detected in single cells by a variety of fluorescence-based methods. A second aspect of this thesis aimed at investigating spacer elements affecting target interference. Spacers with predicted secondary structures in the crRNA impaired the ability of the CRISPR-Cas system to prevent transformation of targeted plasmids. Lastly, in absence of Cas3, Cascade was successfully used to inhibit transcription of specific genes by preventing RNA polymerase access to the promoter.

The CRISPR-Cas field has seen rapid development since the first demonstration of immunity almost ten years ago. However, much research remains to fully understand these interesting adaptive immune systems and the research presented here increases our understanding of the type I-E CRISPR-Cas system.

Keywords: CRISPR, CRISPR-Cas, virus defense, bacteria, bacteriophage, adaptation, spacer integration, interference, gene silencing, fluorescent reporter, *Escherichia coli*

Lina Amlinger, Department of Cell and Molecular Biology, Microbiology, Box 596, Uppsala University, SE-75124 Uppsala, Sweden.

© Lina Amlinger 2017

ISSN 1651-6214
urn:nbn:se:uu:diva-312234 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-312234)
Till Fredrik
This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

I  **Amlinger, L.**, Larsson, M., Koskiniemi, S., Lundgren, M. Deletion of *recD* indirectly reduce adaptation in the type I-E CRISPR-Cas system. *Manuscript.*


* Shared first authors.

Reprints were made with permission from the publisher.
Published papers not included in the thesis

## Contents

Svensk sammanfattning ............................................................................................................. 11

Introduction .......................................................................................................................... 15
  The CRISPR-Cas systems ............................................................................................... 15
    Diversity, classification, and evolution ................................................................. 18
  Mechanism of CRISPR-Cas Immunity ...................................................................... 19
  Adaptation ...................................................................................................................... 19
    Spacer integration ...................................................................................................... 19
    Spacer capture ............................................................................................................. 20
    Primed adaptation ...................................................................................................... 23
  Adaptation in other CRISPR-Cas systems ............................................................... 24
  Methods to study adaptation in vivo ............................................................................ 25
  Expression and processing ......................................................................................... 26
    Processing and maturation in other CRISPR-Cas systems ...................................... 28
  Interference .................................................................................................................. 29
    Target recognition .................................................................................................... 30
    Cas3-mediated target degradation ........................................................................... 32
    Interference in other systems ................................................................................... 33
  Applications of the CRISPR-Cas system .................................................................... 34
    Applications in eukaryotes ....................................................................................... 35
    Applications in bacteria ............................................................................................ 36

Current investigations ....................................................................................................... 38
  Aim .................................................................................................................................. 38
  Requirements for adaptation (Paper I) ................................................................. 38
    A fluorescent reporter for detection of adaptation (Paper II) .................................. 40
  Spacer structure affects interference (Paper III) .................................................. 42
  Gene silencing by Cascade (Paper IV) ....................................................................... 44

Conclusion and future perspectives ...................................................................................... 46
  Adaption ......................................................................................................................... 46
  Expression, processing, and interference ................................................................... 47
  Applications .................................................................................................................... 49

Acknowledgements ............................................................................................................ 51

References .......................................................................................................................... 53
Abbreviations

AAM  Acquisition affecting motif
ATP  Adenosine triphosphate
Bfp  Blue fluorescent protein
bp   base pair
Cas  CRISPR associated
Cascade CRISPR-associated complex for antiviral defense
CRISPR Clustered regularly interspaced short palindromic repeats
crRNA CRISPR RNA
dCas9 Catalytically dead Cas9
dsDNA double-stranded DNA
Gfp  Green fluorescent protein
gRNA Guide RNA
HDR  Homology directed repair
H-NS Histone-like nucleoid-structuring
IHF  Integration host factor
NHEJ Non-homologous end joining
nt   nucleotide
ORF  Open reading frame
PAM  Protospacer adjacent motif
Pre-crRNA Precursor CRISPR RNA
ssDNA single-stranded DNA
ssRNA single-stranded RNA
NTE  Normalized transformation efficiency
TIR   Terminal inverted repeats
TracrRNA Trans-activating CRISPR RNA
Yfp  Yellow fluorescent protein

CRISPR-Cas-systemet består av en CRISPR-matris (clustered regularly interspaced short palindromic repeats) och de Cas-proteiner (CRISPR associated) som behövs för immunitet. CRISPR-matrisen innehåller korta repetrade sekvenser som är separerade av korta mellansekvenser. CRISPR-Cas-systemen är vitt utbredda i bakterier och arkéer, och återfinns i ungefär 50 % av alla studerade bakterier och nästan 90 % av alla studerade arkéer. Det finns flera olika varianter av CRISPR-Cas-system, men alla fungerar enligt samma generella mekanism. Arbetet som presenteras här fokuserar på typ I-E systemet från *Escherichia coli*.


Olika mellansekvenser varierar i hur effektivt de skyddar cellerna mot infektion. Vi undersökte om form och struktur i mellansekvensen i crRNA orsakar skillnader i CRISPR-Cas-effektivitet (papper III). Fem olika mellansekvenser testades i hur effektivt crRNA kunde användas av Cascade för att hindra upptag av plasmider målsöpta av de olika crRNA mellansekvenserna. De förändrade mellansekvenserna hindrade inte plasmidupptag lika bra som den ursprungliga mellansekvensen, trots att plasmiden innehöll en sekvens som matchade mellansekvensen i crRNA. Effektiviteten av CRISPR-Cas-systemet påverkades alltså negativt av att crRNA mellansekvenserna var strukturerade. Resultaten visar att mellansekvenserna och deras RNA-struktur påverkar hur väl CRISPR-Cas-interferens fungerar.

CRISPR-Cas-systemen har utvecklats till olika tillämpningar, och vi använde Cascade för att hindra att målsöka gener översätts till RNA (papper

Introduction

All organisms are subjects to their environment, from which they gain nutrients necessary for survival and interact with other organisms. However, environmental agents can also be harmful to the organisms, causing infections, disease, and death. Organisms have therefore developed numerous ways of protecting themselves against potentially dangerous interactions. Eukaryotes, such as humans, have developed sophisticated immune systems, both innate and adaptive, for protection against viruses and bacteria. Plants employ e.g. interfering RNA in the protection against RNA viruses (Westra et al., 2012a).

Bacteria are no exceptions, and have developed several protective systems against their main threat; mobile genetic elements. Mobile genetic elements can contain genes that are beneficial for the bacteria, e.g. plasmid-borne antibiotic resistance genes, but not uncommonly are the invaders parasitic plasmids or bacteriophages (Westra et al., 2012a). Bacteriophages, or phages, are very abundant, and in some environments outnumber bacteria ten to one (Stern & Sorek, 2011). To protect themselves, bacteria have developed several innate immune systems, such as the restriction-modification systems, receptor masking, and abortive infection. Most recently, an adaptive bacterial immune system was discovered; the CRISPR-Cas system (Westra et al., 2012a).

In this thesis, I will describe the prokaryotic, adaptive CRISPR-Cas immune system, its mechanisms, and the advances I have made in increasing our understanding of the type I-E CRISPR-Cas system of Escherichia coli.

The CRISPR-Cas systems

The CRISPR-Cas systems are adaptive immune systems found in prokaryotes. They are composed of an AT-rich leader followed by an array of short repeats; the clustered regularly interspaced short palindromic repeats (CRISPR), and the CRISPR associated (cas) genes, encoding the proteins required for immunity. The repeats are separated by unique, similarly short, sequences, called spacers (Figure 1) (Makarova et al., 2011). The first observation of the genetic structure of the CRISPR loci was made in 1987 (Ishino et al., 1987), but it took almost 20 years before it was recognized as
an immune system upon observing that the spacers in CRISPR arrays largely match extrachromosomal elements such as phages and plasmids (Bolotin et al., 2005, Mojica et al., 2005, Pourcel et al., 2005). Immunity function of a CRISPR-Cas system was first confirmed in vivo in *Streptococcus thermophilus* where the CRISPR-Cas system provided acquired resistance to phage infection (Barrangou et al., 2007). The CRISPR-Cas systems are now known to be adaptive immune systems, capable of storing memory of invaders and passing this memory on to progeny in a Lamarckian fashion. Non-immunity functions of the CRISPR-Cas systems have also been described, including gene regulation, DNA repair, and virulence regulation (reviewed in Westra et al., 2014). Immunity is generally considered the primary function of the system, and is the focus of this work.

The CRISPR-Cas systems are widespread in prokaryotes, found in approximately 50% of all sequenced bacterial genomes and 87% of all sequenced archaeal genomes (Makarova et al., 2015). The CRISPR-Cas systems are currently divided into two major classes which are further divided into six types with several subtypes (Makarova et al., 2015, Shmakov et al., 2015). CRISPR-Cas immunity thus comes in many flavors. All systems share the same overall mechanism, but specific details and participating components differ. CRISPR-Cas immunity can be divided into three main stages: adaptation, expression and maturation, and interference (Figure 1).

The first stage, adaptation, lays the foundation of adaptive immunity by integration of pieces of invader DNA, or RNA, into the CRISPR array on the chromosome, thereby forming the memory of the system. In the second stage, the CRISPR array and the associated *cas* genes are expressed. The transcribed CRISPR array, the precursor CRISPR RNA (pre-crRNA), is processed into unit crRNAs composed of, or parts thereof, one spacer and one repeat. During interference, the crRNA associates with one or several Cas proteins for form a ribonucleoprotein complex, which uses the crRNA spacer to search for invaders. When a sequence complementary to the crRNA spacer is found, the target nucleic acid is cleaved or degraded by the Cas protein(s), e.g. by Cas3 in the type I-E CRISPR-Cas system. The target DNA or RNA sequence is termed protospacer. Several systems also use a short motif next to the protospacer, the protospacer adjacent motif (PAM), for efficient target recognition (Figure 1) (van der Oost et al., 2014).
Figure 1. Overview of CRISPR-Cas immunity. A schematic CRISPR-Cas locus consisting of cas genes and a CRISPR array with a leader followed by short repeated sequences (R) separated by spacers (S). During adaptation, a piece of foreign nucleic acid is integrated into the CRISPR array (S0) from e.g. an invading phage (top). The cas genes are transcribed and translated. The CRISPR array is transcribed to pre-crRNA and processed into mature crRNA. One or several Cas proteins associate with a crRNA to form a crRNA-loaded ribonucleoprotein complex (blue). The complex uses the crRNA as a guide to find a cognate target. Upon infection by a phage (bottom), the DNA or RNA is identified by binding of the crRNA to the target, complementarity leads to degradation of the targeted nucleic acid. The term protospacer refers to the sequence in the DNA, or RNA, targeted by the crRNA spacer. Indicated next to the protospacer is also the PAM, which is important for target recognition in several CRISPR-Cas systems.
Diversity, classification, and evolution

The CRISPR-Cas systems are highly diverse, evolving rapidly through constant interactions with phages and other mobile elements. They are mobile themselves, frequently undergoing horizontal gene transfer (van der Oost et al., 2014), classification of the systems is therefore not trivial. The CRISPR-Cas systems are currently classified based on signature proteins and cas locus architecture. The CRISPR-Cas systems are divided into two classes, class 1 and 2. Systems of class 1 possess multi-subunit crRNA-effector complexes whereas class 2 systems have one main protein responsible for immunity. Class 1 is further divided into three types, I, III, and the putative IV, with seven (I-A-F and I-U), four (III-A-D), and one subtype respectively. Class 2 is also divided into three types; type II, type V, and type VI, with three (II-A-C, V-A-C), and one subtype respectively. The types are frequently referred to with the name of their respective effector complex, i.e. type I systems with effector complexes called CRISPR-associated complex for antiviral defense (Cascade), type III is represented by the Cascade-like Cmr and Csm complexes, type II by Cas9, and type V by Cpf1 (Makarova et al., 2011, Makarova et al., 2015). The most recently characterized system, type VI, is currently represented by the C2c2 effector protein (Shmakov et al., 2015). Classification is continuously updated as systems are further characterized and new systems are discovered.

Class 2 systems are almost exclusively found in bacteria, with only one exception to date. Class 1 systems are present both in bacteria and archaea. The most abundant CRISPR-Cas systems are type I systems; subtype I-B followed by subtypes I-C and I-E (Makarova et al., 2015).

The mobile nature of the CRISPR-Cas systems also provides challenges for evolutionary studies. Bioinformatical analyses have offered some clues to the evolutionary origin of the CRISPR-Cas systems through the discovery of transposons encoding cas1, named casposons. Most casposons are flanked by terminal inverted repeats (TIR) (Krupovic et al., 2014), that have been proposed to be the origin of the CRISPR repeats. It has been suggested that a Cascade-like effector was present as an innate immune system and that the first CRISPR-Cas system was formed through integration of a casposon adjacent to the Cascade genes. The TIR of the casposon could have duplicated to form the CRISPR array and the casposon itself supplied the proteins required for memory formation. Type II systems could have evolved through a second integration event inserting a transposon with Cas9 and replacing Cascade in a CRISPR-Cas locus (Koonin & Krupovic, 2015).

To counteract the CRISPR-Cas systems, some phages encode anti-CRISPR systems. The anti-CRISPR systems of Pseudomonas phages are best studied, and target the type I-E or type I-F system of Pseudomonas aeruginosa by inhibiting the different stages of CRISPR-Cas immunity. For a recent review, see Maxwell (2016).
Mechanism of CRISPR-Cas Immunity

The work presented in this thesis has been performed in the type I-E CRISPR-Cas system from *E. coli*, one of the best studied CRISPR-Cas systems. The following sections therefore describe the mechanistic features of this system unless otherwise stated.

Type I-E CRISPR-Cas systems consists of the Cas3 target-degrading nuclease, the multi-subunit effector complex Cascade, Cas1, Cas2, and the associated CRISPR array. The array contains palindromic repeats of 29 base pairs (bp) and intervening spacers of 32 bp. In the crRNA, the palindromic repeat sequence form a stem-loop structure (Brouns *et al.*, 2008).

Adaptation

Adaptation in the type I-E CRISPR-Cas system is the best understood CRISPR-Cas adaptation process (Figure 2). Despite well-studied, adaptation remains the least understood of the stages in CRISPR-Cas immunity.

Two modes of adaptation have been described; naïve and primed adaptation. Naïve adaption occurs without previous immunization, while an existing memory of the invader leads to an accelerated adaptation process called primed adaptation. During primed adaptation, the interference machinery directs adaptation to the targeted DNA and primed adaptation therefore requires all components of the CRISPR-Cas system (Datsenko *et al.*, 2012). The newly integrated spacer is preferentially integrated at the leader-proximal end of the array, and extends the array by one spacer-repeat unit (61 bp) (Yosef *et al.*, 2012).

Naïve adaptation requires only two Cas proteins; Cas1 and Cas2, the leader, and one repeat (Yosef *et al.*, 2012). Additionally, certain host proteins are also involved in naïve adaptation. DNA polymerase I is thought to provide the polymerase activity required for synthesis of the new repeat (Ivančić-Bačić *et al.*, 2015). Additionally, integration host factor (IHF) and its consensus binding site in the leader are needed for adaptation (Nuñez *et al.*, 2016). There are conflicting data on the strict requirement of RecB for adaptation. One study reported its requirement (Ivančić-Bačić *et al.*, 2015), while a second demonstrated functional, albeit lower, spacer integration in ΔrecB cells (Levy *et al.*, 2015).

Spacer integration

Cas1 and Cas2 form a complex consisting of two Cas1 dimers and one Cas2 dimer. Complex formation as well as the nuclease activity of Cas1 are required for adaptation. Cas2 nuclease activity is dispensable, indicating a
more structural role for Cas2 (Núñez et al., 2014). The spacer-ends are coordinated on each side of the complex by the Cas1 dimers, and the double-stranded spacer spans the central Cas2 dimer (Figure 2) (Núñez et al., 2015a, Wang et al., 2015). The new spacer is integrated into the array through an integrase-like mechanism coordinated by Cas1-Cas2 (Núñez et al., 2015b). IHF binding most likely introduces bending of the array DNA, enabling access of Cas1-Cas2 to the integration site, as well as provides specificity for the leader-proximal end of the array. The PAM aids in orienting the newly integrated spacer so that all spacers are in the same orientation vis-a-vis the leader (Mojica et al., 2009, Shmakov et al., 2014). The first nucleotide of the PAM, C, is integrated with the spacer, but as it is almost invariably a C, it is usually considered as the last base of the repeat once integrated (Figure 2) (Datsenko et al., 2012, Swarts et al., 2012).

Integration proceeds through two nucleophilic attacks, the first occurring in the leader-repeat boundary in a sequence-dependent manner (Núñez et al., 2016). The second attack is determined by a ruled mechanism and targets the repeat-spacer border (Goren et al., 2016). The result is a stable intermediate with a double-stranded spacer flanked by two single-stranded repeats (Arslan et al., 2014). The reaction is presumed to be completed by fill-in of the repeat, likely by DNA polymerase I (Ivančić-Barče et al., 2015), and unknown ligases (Figure 2).

The exact sequence requirements for spacer integration are currently unknown. To date, the requirement of the repeat-proximal 60 bp of the leader has been demonstrated (Yosef et al., 2012), which comprises the IHF binding site from -9 to -35 (Núñez et al., 2016). The requirement of the remaining 20 bp still has to be elucidated. The array must contain at least one repeat, since the leader-proximal repeat is used as template for the new repeat (Yosef et al., 2012). Mutations in the leader-repeat boundary abolish integration (Núñez et al., 2016) and cruciform DNA secondary structures potentially formed in the repeats may also play a role in Cas1-Cas2 recognition (Núñez et al., 2015b).

Spacer capture

Not much is known about the first steps of adaptation; spacer capture and how new spacers are generated. During naïve adaptation, the majority of the integrated spacers originated from the plasmid expressing Cas1 and Cas2, rather than from the chromosome (Levy et al., 2015, Yosef et al., 2012). This indicates a selection mechanism for exogenic DNA, as the bias occurred even without the selective pressure of having an active interference system, which would remove such self-targeting spacers from the population. The bias can neither be easily explained by excess of plasmid DNA nor by differences in PAM-abundance (Yosef et al., 2012), but could be created by primary defense systems that generate spacer precursor molecules that are
scavenged by Cas1-Cas2. Two mechanisms for spacer generation have been described, primed adaptation (see below) and RecBCD-dependent spacer generation (Künne et al., 2016, Levy et al., 2015).

Figure 2. Schematic illustration of the adaptation process. Spacer fragments, generated by Cas1 and Cas2 or other immunity systems, are captured by the Cas1-Cas2 complex. The spacer is inserted in the leader-proximal end of the CRISPR array through two nucleophilic attacks (1 and 2). The PAM (here: CTT) helps orient the new spacer within the array and the last base of the repeat originates from the integrated invader DNA. The new repeat is synthesized and the integration is complete. Dashed arrows are used for steps that are not yet fully elucidated.

The RecBCD-complex is involved in repair of double-stranded DNA (dsDNA) breaks and Cas1 has been shown to interact with RecB and RecC (Babu et al., 2011). These DNA-repair proteins were therefore hypothesized to aid
spacer integration. Upon double-stranded breaks, RecBCD binds the dsDNA and unwinds it, subsequently degrading the single-stranded DNA (ssDNA). When the complex reaches a specific sequence, called a Chi site, RecBCD switches modes, RecA is now loaded onto the ssDNA, and the DNA is repaired through RecA-dependent homologous recombination (Dillingham & Kowalczykowski, 2008). As Chi sites are overrepresented on the chromosome compared to a random sequence, RecBCD will rapidly encounter a Chi site on the chromosome and only a short stretch of DNA will be degraded. As extrachromosomal DNA, such as phages and plasmids, contains less frequent Chi sites, RecBCD-mediated degradation of such DNA will generate more degradation fragments.

With overproduction of Cas1 and Cas2, the fraction of acquired spacers originating from the chromosome is higher compared to acquisition with moderate Cas1-Cas2 levels (Levy et al., 2015). This may indicate that Cas1-Cas2 do not have an inherent selection mechanism for exogenic DNA. However, the fraction of spacers with chromosomal origin increased when genes encoding for the components of the RecBCD complex were deleted. Furthermore, DNA segments between double-stranded breaks and Chi sites were found to be hotspots for protos pacer sampling, and RecBCD-mediated DNA degradation was proposed as a mechanism for pre-spacer generation. Cas1-Cas2 could sample the single-stranded degradation products of RecBCD for spacer material, leading to the observed biased acquisition from extrachromosomal DNA (Levy et al., 2015). This, however, is not easily reconciled with in vivo and in vitro data supporting integration of a double-stranded spacer (Arslan et al., 2014, Nuñez et al., 2015b). It nonetheless remains an attractive model as RecBCD has emerged as part of an innate immune system, by targeting double-stranded ends of e.g. infecting phages and subsequently degrading the invading DNA (Dillingham & Kowalczykowski, 2008). It remains to be elucidated how Cas1-Cas2 captures the spacer, whether the captured spacer is double- or single-stranded, and how the complex trims the pre-spacer prior to integration into the CRISPR array (Künne et al., 2016, Wang et al., 2015).

**Spacer sequence**

Sequence requirements of the integrated spacer are poorly characterized. Structures of Cas1-Cas2 in complex with pre-spacers show that the preferred substrate in vitro is dsDNA with 3’ single-stranded overhangs (Nuñez et al., 2015a, Wang et al., 2015). Cas1-Cas2 functions as a molecular ruler, specifying the length of the spacer and probably trims the pre-spacer before integration (Wang et al., 2015). Interactions between the spacer and the Cas1-Cas2 complex are sequence-independent, allowing for integration of any spacer sequence (Nuñez et al., 2015a). Half of the spacers acquired through naïve adaptation are associated with the consensus PAM; 5’-protospacer-CWT-3’ (W = A or T), with the first base (C) as the most conserved in adap-
tation (Yosef et al., 2012). However, conclusions about the importance of PAM in naïve integration are hampered by the ambiguity in the definition of a PAM. When the data from Yosef et al. (2012) was reanalyzed, the majority of the new spacers were found to be associated with a functional PAM (Xue et al., 2015). A recent structure of Cas1-Cas2 in complex with a pre-spacer provides evidence for sequence-specific PAM recognition in the single-stranded 3’ overhangs by Cas1 (Wang et al., 2015). This is in contrast to the in vivo data where different PAMs are observed (Xue et al., 2015, Yosef et al., 2012). The structure of Cas1-Cas2 (Wang et al., 2015) does not, however, explain why the C in the first position of the PAM would be more conserved during adaptation than the other two base pairs.

Another sequence motif, the acquisition affecting motif (AAM), in the 3’ end of the spacer, was proposed as a motif important for adaptation rather than interference (Yosef et al., 2013). However, the significance of the AAM still remains to be corroborated by other studies.

**Primed adaptation**

Spacers integrated during prior infections are used in primed adaptation to rapidly generate new spacers targeting the same, or a similar, invader. Primed adaptation widens the targeting potential of the CRISPR-Cas system as older spacers can be used to generate new spacers which, in turn, can be used to clear the invader. It limits the opportunity for the phage to avoid interference by escape mutations. It also enables broad-spectrum targeting as a spacer targeting one phage could lead to priming against closely related phages (Swarts et al., 2012), and invader DNA is more easily cleared when targeted by multiple spacers. Involvement of the interference machinery will also further aid in targeting acquisition to exogenic DNA molecules (Barrangou et al., 2007, Swarts et al., 2012).

Primed adaptation in the type I-E system requires all components of the CRISPR-Cas system; Cas3, Cascade, and a targeting crRNA (Datsenko et al., 2012), as well as other host factors (Ivančić-Baćić et al., 2015). As mentioned above, primed adaptation leads to acquisition of multiple spacers targeting the same invader. Integration itself seems to occur independently of the interference machinery (Nuñez et al., 2015b, Rollie et al., 2015), and primed adaptation is likely important for generation of spacer-donor fragments, thereby facilitating adaptation.

Primed adaptation is thought to be driven by “escape mutants” (Datsenko et al., 2012, Fineran et al., 2014). Spacers that do not elicit direct interference, or delayed direct interference, display the most efficient priming (Künne et al., 2016). Recent data, however, shows that perfectly matching spacers also can be used for efficient primed adaptation (Semenova et al., 2016), in agreement with previous in vitro data (Blosser et al., 2015). The observation that partially matching spacers are more efficient in eliciting
primed adaptation probably reflects differences in target clearance rates. Delayed clearance would allow the target to replicate and persist within the cell, giving plenty of opportunities to acquire spacers from the target before it is cleared (Künne et al., 2016, Semenova et al., 2016). Delayed target clearance could be due to slower target degradation by Cas3. The fate of the protospacer-containing DNA would thus be determined by Cas3 activity (Künne et al., 2016), which may be specified by differential target-binding by Cascade (Blosser et al., 2015). Attempts to characterize the sequence requirements for primed adaptation have so far failed to provide clear patterns (Fineran et al., 2014), and may be highly dependent on PAM and spacer sequence (Xue et al., 2015).

In vitro, Cas3 generates 30-100 nucleotide (nt) degradation products that can be integrated into the CRISPR array by Cas1-Cas2 (Künne et al., 2016). Cas3, however, has single-stranded endonuclease activity (Sinkunas et al., 2011), while the pre-spacer most likely is double-stranded (Arslan et al., 2014, Nuñez et al., 2015b). The single-stranded degradation products generated by Cas3 may re-anneal before being scavenged by Cas1-Cas2 for spacer integration (Künne et al., 2016). Further experiments are required to explore the Cas3 mechanism and its in vivo degradation products to determine if these depend on Cascade target-binding and/or target clearance rate.

The interference machinery may provide the PAM-specificity, as most integrated spacers in a primed setting are associated with a consensus PAM (Savitskaya et al., 2013). Cas3 preferentially cleaves T-rich sequences, and therefore generates fragments enriched for NTT in the 3’-end. Such cleavage preference may contribute to the high prevalence of acquired spacers with consensus PAM in primed adaptation (Künne et al., 2016).

Whether the AAM motif affects primed adaptation is so far unclear; one study confirmed the presence of the motif (Fineran et al., 2014), while other studies failed to detect it in primed adaptation (Künne et al., 2016, Shmakov et al., 2014).

Adaptation in other CRISPR-Cas systems

Adaptation is best studied in the type I-E system, but has been observed in several other CRISPR-Cas systems. The first observation of adaptation was in fact made in the type II-A system of S. thermophilus. For adaptation in the type II-A system, Cas1, Cas2 and Csn2 are required (Barrangou et al., 2007). Additionally, Cas9 and a second RNA, the trans-activating CRISPR RNA (tracrRNA) are needed for adaptation (Wei et al., 2015b), but the crRNA seems to be dispensable (Heler et al., 2015). Cas9 targets the system to acquire spacers associated with correct PAM but, interestingly, adaptation itself does not require Cas9 nuclease activity (Heler et al., 2015, Wei et al., 2015b). Similar to type I-E adaptation, type II-A adaptation requires parts of the leader as well as one repeat. The sequence of the leader-repeat junction is
also of importance for integration (Wei et al., 2015a). In the absence of Cas9 nuclease activity, more chromosomal spacers are integrated, indicating that spacers are sampled from both plasmid and chromosome but that cells with chromosomal spacers are removed from the population through self-targeting (Wei et al., 2015b).

Adaptation has been observed for several type I systems; type I-A (Erdmann & Garrett, 2012, Liu et al., 2015), I-B (Li et al., 2014), and I-F (Cady et al., 2012, Richter et al., 2014). Both naïve and primed adaptation have been reported, but not for all systems (Sternberg et al., 2016). A recent study of adaptation in the type I-F system in Pectobacterium atrosepticum studied both types of adaptation simultaneously, showing that priming was dominant over naïve adaptation. Furthermore, newly acquired spacers promoted subsequent integration events, indicating that interference enhances adaptation in the type I-F system. Both primed and naïve adaptation led, to comparable extent, to integration of spacers with canonical PAMs (Staals et al., 2016).

Studies of adaptation in type III systems are complicated by the co-occurrence of type III and type I CRISPR-Cas systems, and their potential cross-talk (Sternberg et al., 2016). Some evidence for spacer integration exists for type III-B, but integration activity is however not exclusive to this system (Erdmann et al., 2014).

Recently, the type III-B system from Marinomonas mediterranea was shown to integrate spacers from an RNA target. Here, Cas1 is fused to a reverse transcriptase, enabling sampling RNA for new spacers. This system acquires spacers both from DNA and RNA, and in absence of RT activity, it continues to acquire spacers of DNA origin. However, overexpression of the components involved was required, and the biological relevance of such adaptation activity therefore remains to be established (Silas et al., 2016). Unexpectedly, no reverse transcriptase has yet been identified in the RNA-targeting type VI system (Abudayyeh et al., 2016).

Methods to study adaptation in vivo

To study spacer integration in vivo, several detection methods have been employed. The most commonly used is PCR-based, exploiting the fact that the array is extended by one repeat-spacer unit upon integration. After amplification of the CRISPR array, integration can therefore be detected as a shift in migration of the PCR product (Yosef et al., 2012). PCR can be performed on single colonies and on cell populations and is often used in combination with other assays. When done on a population, several bands will be observed, one for the parental, unexpanded, array and one – or several – from the expanded array(s) with new spacer-repeat unit(s). Although simple and applicable to most experimental set-ups, PCR is subject to amplification bias and may give different results depending on the template used in the
reaction as well as the amount of analyzed PCR product. Furthermore, the sensitivity is low and PCR can only detect 0.4% expanded arrays or more (Yosef et al., 2012). In a development of this method to detect integration of a particular spacer sequence, spacer-specific primers was used together with a primer outside of the CRISPR array, giving a PCR-product only if the specific spacer had been integrated (Yosef et al., 2013).

A quantitative method for studying spacer integration is to perform deep sequencing on the PCR product or extracted DNA from the population. It has a lower detection limit at less than 0.01% expanded arrays in a population (Sternberg et al., 2016). Spacer sequences are simultaneously obtained, allowing for a more refined analysis of acquired spacers as well as enabling further studies based on spacer sequence (Levy et al., 2015, Savitskaya et al., 2013). Such analyses may, however, be biased due to PCR amplification, and require substantial post-processing before obtaining the results.

A reporter for spacer integration in *E. coli* has been developed, enabling positive selection of bacteria that has undergone adaptation. It utilizes an antibiotic resistance gene out-of-frame of a translational start codon, with a CRISPR array placed between the start codon and the reporter gene. Upon integration, extension of the array by 61 bp moves the gene into frame. Cells with an expanded array will therefore be resistant to the antibiotic while cells with unexpanded arrays will not survive on selective plates. This reporter system allows for detection of rare integration events, but does not easily provide the integration frequency (Díez-Villaseñor et al., 2013).

Other assays rely on the functionality of the integrated spacers, studying survivors after phage infection (Datsenko et al., 2012, Ivančić-Baće et al., 2015) or plasmid removal (Fineran et al., 2014, Swarts et al., 2012). This requires the activity of the complete immune system and only detects functional spacers, thus excluding non-functional, or partially functional, spacers as well as self-targeting spacers from the analysis.

**Expression and processing**

**Expression**

*E. coli* K-12 encodes two CRISPR-Cas loci, one containing the *cas* genes and CRISPR-I with 13 spacers while the second only encodes a CRISPR array, CRISPR-II, with six spacers (Díez-Villaseñor et al., 2010). Cas3 is transcribed from its own promoter, while the genes encoding Cascade, Cas1, and Cas2 are part of the same operon (Figure 3) (Pul et al., 2010).

Functionality of the type I-E CRISPR-Cas system was demonstrated early, but despite all components being functional (Brouns et al., 2008), native gene expression is not enough to provide phage protection in *E. coli* K-12 (Westra et al., 2010). *Cas* gene expression is silenced by the histone-like
nucleoid-structuring (H-NS) protein, that normally functions as an innate immune system. H-NS does not destroy the invader DNA but silences its transcription, thereby enabling integration of e.g. horizontally transferred genes into the host chromosome without harming the host. The origin of the CRISPR-Cas system can therefore be the reason for its silencing by H-NS. H-NS binds the promoter of the cascade-cas1-cas2 operon (Pul et al., 2010) and, by deleting hns, cas gene expression is elevated so that CRISPR-Cas immunity can be studied without heterologous promoters. H-NS repression can be relieved by LeuO, which also binds within in the cascade-cas1-cas2 promoter region. The CRISPR array is constitutively transcribed at low levels (Westra et al., 2010), but CRISPR expression is also upregulated in Δhns (Pul et al., 2010). The cas3 promoter is less characterized than that of cascade-cas1-cas2. When hns is deleted, cas3 transcripts are elevated in stationary phase (Majsec et al., 2016).

Figure 3. The CRISPR-Cas locus with CRISPR-I of E. coli K-12. Cas3 is transcribed by its own promoter, while the genes encoding for the proteins composing Cascade (blue) and for proteins responsible for adaptation (orange) are transcribed as one operon. The promoter of the CRISPR array resides within the leader. Indicated is also the consensus sequence of the E. coli K-12 repeat, nucleotides participating in base paring in the stem-loop are underlined.

It has been speculated that H-NS-mediated repression could be relieved upon phage infection, as H-NS could be titrated off the cas promoter to instead bind incoming phage DNA (Pul et al., 2010). However, this remains to be demonstrated experimentally. Phage infection could also cause envelope stress, shown to induce CRISPR-Cas expression, but the exact mechanism for this induction currently unknown (Perez-Rodriguez et al., 2011). Neither of these events seems to occur upon Lambda phage infection, since introduction of a spacer targeting phage Lambda in an otherwise unperturbed cell was not sufficient to protect the cells from phage infection (Westra et al., 2010).

Processing
The CRISPR array is transcribed into a long pre-crRNA which is processed into mature crRNAs by Cas6e (Brouns et al., 2008). Cas6e remains associat-
ed to the crRNA after cleavage (Jore et al., 2011). The pre-crRNA is short lived, but the processed crRNAs are more stable and accumulate in the cell (Pougach et al., 2010). The palindromic sequence in the repeats base pairs and forms a stem-loop within the CRISPR repeats (Figures 1 and 3) (Brouns et al., 2008).

Cas6e is a metal-independent endoribonuclease (Brouns et al., 2008), processing the pre-crRNA with single-turnover kinetics (Sashital et al., 2011). Two structures of Cas6e from Thermus thermophilus provide mechanistic insights into the binding and subsequent cleavage of the pre-crRNA. Cas6e binds the stem-loop of the crRNA repeat in a positively charged groove in the enzyme (Sashital et al., 2011) and crRNA-recognition is both sequence- and structure-specific, with most interactions taking place in the 3’ strand of the stem (Gesner et al., 2011, Sashital et al., 2011). Cas6e cleaves the pre-crRNA at the bottom of the 3’ strand of the stem-loop. In the resulting crRNA, the spacer is flanked by a defined 8 nt 5’ handle and a 3’ handle comprising the stem-loop structure (Figure 1) (Brouns et al., 2008, Jore et al., 2011). Apart from Cas6e, cleavage is stimulated by Cse2, Cas7, and Cas5e (Brouns et al., 2008).

Processing and maturation in other CRISPR-Cas systems

Most of the class 1 systems utilizes Cas6 for sequence- and structure-specific metal-independent cleavage of pre-crRNAs (Charpentier et al., 2015). Although differing in amino acid sequence, the different Cas6 ribonucleases adopt similar folds, indicating common ancestry. Specific recognition and cleavage of the cognate pre-crRNA are established by different amino acids participating in the interaction (Carte et al., 2008, Haurwitz et al., 2010, Niewoehner et al., 2014). The type I-C system stands out as the only type I system without a Cas6-homologue; instead Cas5d cleaves the pre-crRNA. Despite using a different Cas protein, pre-crRNA cleavage by Cas5d shares many of the features of Cas6-mediated maturation (Nam et al., 2012).

In type III systems, crRNAs are further matured by 3’ trimming after the primary processing by Cas6. The length of the mature crRNA is independent of spacer length (Hale et al., 2012) and is likely determined by a ruler-based mechanism (Hatoum-Aslan et al., 2011). The size of the mature crRNA could be determined by the number of subunits in the Csm complex, protecting the crRNA from further 3’ trimming (Hatoum-Aslan et al., 2011, Rouillon et al., 2013).

Type II systems rely on RNaseIII for crRNA maturation, but also require the tracrRNA and Cas9. The tracrRNA, encoded in proximity to the CRISPR array, is complementary to part of the repeat. The resulting crRNA-tracrRNA duplex is cleaved by host factor RNaseIII. Contrary to the type III systems, the crRNA is trimmed at the 5’ end. The mechanism of the second maturation step remains to be elucidated (Deltcheva et al., 2011). The need
for RNaseIII processing may explain the underrepresentation of type II systems in archaea which lack RNaseIII homologues (Charpentier et al., 2015).

Some type II-C system, such as the one found in Neisseria meningitides, have leader-less CRISPR arrays. The crRNAs are instead transcribed from promoters within the repeats themselves. Processing of these crRNAs is RNaseIII-dependent, but processing is dispensable for interference (Zhang et al., 2013).

Pre-crRNA processing in the recently characterized type V and VI systems appears to be a mixture between the processing in type I and III, and that of the type II CRISPR-Cas systems. Processing requires the effector proteins, i.e. Cpf1, C2c1, or C2c2, but differs in the requirement of the tra-crRNA for pre-crRNA processing (Fonfara et al., 2016, Shmakov et al., 2015, Zetsche et al., 2015). More work is required to further elucidate the mechanisms of pre-crRNA processing in these newly discovered CRISPR-Cas systems.

**Interference**

The effector complex of the type I-E CRISPR-Cas system is a multi-subunit complex called Cascade, composed of one Cse1, two Cse2, six Cas7, one Cas5e, and one Cas6e subunit (Figure 4) (Jore et al., 2011). Nine of the eleven Cas proteins in the complex have direct contacts with the crRNA (Jackson et al., 2014). Several structures of the crRNA-Cascade ribonucleoprotein complex bound to ss- or dsDNA have provided insights into the configuration of Cascade as well as the observed crRNA and target requirements.

Cascade adopts a sea horse-like shape (Jore et al., 2011), where Cas6e retains the 3’ stem-loop of the mature crRNA at the top of the complex (Figure 4). Six Cas7 subunits form the backbone of the complex, without sequence-specific interactions with the crRNA spacer. The Cas7 subunits exhibit a fold that can be likened to a right hand, each subunit with a “thumb” folded over the crRNA in 6 nt regular intervals. Consequently, mismatches between the crRNA and the target DNA at every sixth position does not affect target interference as the nucleotides of the crRNA are not available for base pairing (Jackson et al., 2014). The conformation of the last Cas7 subunit is altered by the interaction with the crRNA, and the Cas7 backbone does not extend further (Wiedenheft et al., 2011). Cas5e, coordinating the 5’ end of the crRNA at the other end of the complex, may also limit the extension of the Cas7 backbone (Jackson et al., 2014). The 5’ handle of the crRNA inserts into a pocket between Cas5e and Cse1 (Wiedenheft et al., 2011), where Cas5e makes several sequence-specific contacts with the crRNA (Jackson et al., 2014, Mulepati et al., 2014). The 5’ handle is of particular importance for Cascade-crRNA binding, and removal of even one nucleotide in the han-
dle impairs the interaction (Beloglazova et al., 2015). Two Cse2 subunits span the belly of the complex, connecting its head to its tail (Jackson et al., 2014). Cse1 is most loosely associated with the complex (Brouns et al., 2008) and is responsible for PAM recognition during target surveillance (Sashital et al., 2012).

It is not yet known how Cascade assembles, and if pre-crRNA cleavage by Cas6e preludes Cascade assembly. Shorter or longer crRNAs are successfully incorporated into Cascade, coupled to extension or shortening of the Cas7 backbone and Cse2 belly, suggesting that Cascade assembles in a crRNA-dependent manner (Kuznedelov et al., 2016, Luo et al., 2016). In contrast, if processed crRNA of correct size were supplied, partial activity was retained by Cascade, indicating that Cas6e may not be strictly required for Cascade assembly (Semenova et al., 2015).

**Figure 4.** Schematic representation of Cascade bound to a crRNA. For simplicity, the crRNA (grey) is drawn on the outside of the complex. Cas6e (light green) binds the 3’ stem-loop of the crRNA at the top of the complex. The six Cas7 subunits (blue) form the backbone, holding the spacer of the crRNA. Two Cse2 subunits (yellow) cover the belly, connecting its head to its tail. Cas5e (dark green) and Cse1 (red) cap the bottom of the complex. The 5’ handle of the crRNA sits in a pocket between Cas5e and Cse1. Adapted from Jackson et al. (2014) and Jore et al. (2011).

**Target recognition**

Two different target binding modes of Cascade, one PAM-dependent and one PAM-independent, have been described. PAM-dependent recognition is used in *bona fide* target binding (see below), while PAM-independent binding appears to initiate primed adaptation (Blosser et al., 2015, Redding et al., 2015). Binding studies and *in vivo* experiments indicate that both binding modes can occur on the same target (Blosser et al., 2015, Semenova et al., 2016).
PAM-independent target recognition requires Cas1 and Cas2 in order to recruit Cas3 to Cascade. The interaction attenuates the nuclease activity of Cas3 and enables Cas3 translocation in both directions from the protospacer. During this translocation, however, Cas3 does not cleave the target (Redding et al., 2015). It therefore remains to be elucidated how primed adaptation is promoted, as Cas3-mediated target degradation is proposed to generate spacer fragments for Cas1-Cas2 during primed adaptation. Two independent studies describe four different fates of invader DNA (Künne et al., 2016, Xue et al., 2015), and it has been proposed that the fate of the invader is determined by the rate of direct interference, dependent on spacer sequence (Xue et al., 2015). The underlying molecular mechanisms are still unclear, especially whether Cascade binding specifies the outcome of targeting and how Cas3 action on different targets is regulated.

**Protospacer adjacent motif**

For efficient target destruction, several CRISPR-Cas systems use a PAM, present next to the target protospacer but not in the CRISPR array on the chromosome. Self-targeting can thus be avoided despite the presence of an exact match to the spacer on the chromosome. In the type I-E system, the PAM is located at the 3’ end of the protospacer (i.e. 5’-protospacer-PAM-3’), and throughout this thesis the PAM sequence is designated as the sequence on the target strand (See also Figure 1). The consensus PAM found by alignment of multiple protospacers is 5’-CWT-3’ (Mojica et al., 2009). Several PAMs are tolerated in vivo, and in a screen of 26 PAM mutants, four were found to lead to phage resistance and efficient prevention of plasmid transformation; 5’-CAT, CTT, CCT, CTC-3’. More PAMs were tolerated for plasmid transformation prevention than for phage infection protection, indicating a greater flexibility depending on the mode of invasion (Westra et al., 2013).

The PAM is recognized by Cse1 in the target strand of the DNA (Sashital et al., 2012), independent of base pairing with the crRNA (Westra et al., 2013). The PAM is read in the minor groove of the DNA duplex (Hayes et al., 2016) and must be in double-stranded form for efficient target cleavage (Hochstrasser et al., 2014). A recent crystal structure of Cascade bound to a dsDNA target provides insights to PAM recognition by Cse1. A combination of specific protein interactions and steric hindrance led to proposal of a more general PAM, 5’-CHH-3’ (H = A, C or T) (Hayes et al., 2016), which includes all the previously described PAMs (Mojica et al., 2009, Westra et al., 2013). Combined, a 5’-CGG-3’ triplet would be the least favorable PAM which, interestingly, would be the PAM if the spacer in the chromosomal CRISPR array was targeted. Upon PAM recognition and binding by Cse1, the first nucleotides of the protospacer is flipped out (Hayes et al., 2016), thus destabilizing the dsDNA duplex and enabling crRNA-protospacer base pairing (Sashital et al., 2012).
**Seed sequence**
In addition to the PAM, the first bases of the protospacer have been reported to be of greater importance during target recognition. This sequence, bases 1-5, and 7-8, is referred to as the seed sequence. Mismatches between the target and the crRNA spacer in these positions are most deleterious for interference (Semenova et al., 2011). The discontinuity in the seed sequence at the sixth base is explained by the structure of Cascade where the sixth nucleotide of the crRNA is kinked out by one of the Cas7 thumbs (Jackson et al., 2014, Mulepati et al., 2014).

The increased tolerance for mismatches in non-seed positions allow broader targeting as well as limits the phage to few positions for potential escape mutations (Semenova et al., 2011). However, the seed is not displayed differently than other bases in the crRNA spacer, and it has been hypothesized that its importance is mainly due to the proximity to the PAM (Jackson et al., 2014), in line with the observation of directional, stepwise, R-loop formation starting in the PAM-proximal end (Blosser et al., 2015).

The PAM and the seed are well-established concepts in type I-E CRISPR-Cas immunity, but the flexibility in target-recognition by the crRNA-Cascade complex has been shown to be highly dependent on spacer sequence. Target requirements may therefore not be possible to generalize (Xue et al., 2015).

**R-loop formation**
Upon target binding by Cascade, an R-loop is formed, a DNA-RNA hybrid where the crRNA spacer is base paired to the protospacer on the target strand, and the non-target strand is displaced. Cascade-binding protects the target strand of the protospacer from cleavage, as well as parts of the non-target strand (Jore et al., 2011). Binding proceeds in a step-wise manner over the entire protospacer, starting in the PAM-proximal end (Blosser et al., 2015). The non-target strand is looped out and stabilized by the two Cse2 subunits (Hayes et al., 2016). Efficient target recognition also leads to structural rearrangements in Cascade: the two Cse2 subunits move toward the tail of the complex, and Cse1 rotates (Mulepati et al., 2014, Wiedenheft et al., 2011). Movement of Cse2 gives access of the distal end of the crRNA spacer to the protospacer, allowing base pairing over the entire protospacer (Mulepati et al., 2014), while the rearrangements in Cse1 leads to recruitment of Cas3 (Hochstrasser et al., 2014).

**Cas3-mediated target degradation**
Cas3 is an adenosine triphosphate (ATP)-dependent helicase and an ATP-independent nuclease (Sinkunas et al., 2011). Both activities are required for
CRISPR-Cas interference. Cas3 interacts with Cascade only when Cascade is target-bound (Westra et al., 2012b), and the ATPase activity of Cas3 is activated upon interaction with Cascade (Mulepati & Bailey, 2013). Without Cascade-interaction, no Cas3 exo- or endonuclease activity is detected (Howard et al., 2011). Target DNA is rapidly degraded and Cas3 is therefore regarded as a highly processive enzyme (Westra et al., 2012b). Cas3 nicks the available single-stranded non-target strand and subsequently also the target strand at several positions (Mulepati & Bailey, 2013). Cas3 then unwinds the target DNA followed by degrading the generated ssDNA unidirectionally 3’ to 5’ (Sinkunas et al., 2011), starting at the protospacer and proceeding upstream (in the direction of the PAM) on the target strand. No cleavage is observed downstream of the protospacer (Mulepati & Bailey, 2013). Following Cas3 cleavage, Cascade remains bound to the cleaved DNA (Sinkunas et al., 2013, Redding et al., 2015), but additional Cas3 are not recruited (Redding et al., 2015).

Much of our knowledge about Cas3 activity comes from in vitro studies (Mulepati & Bailey, 2013, Sinkunas et al., 2013, Westra et al., 2012b), highly influenced by experimental conditions, e.g. salt concentration and temperature. Such experimental variations can lead to differences in observations. A recent in vitro study reported a different Cas3-behavior, where the nuclease generated 240-300 nt ssDNA gap, followed by 3’ to 5’ translocation of Cas3 but without the expected further DNA unwinding and DNA degradation (Redding et al., 2015). This was corroborated by detection of 30-100 bp fragments in an in vitro target degradation assay, the fragments were proposed to be reannealed single-stranded Cas3 degradation products (Künne et al., 2016). Further research therefore needs to elucidate Cas3 and Cascade behavior in vivo.

Interference in other systems

The interference complexes of the class 1 CRISPR-Cas systems are all multi-subunit complexes similar to Cascade. In type I systems, the complex is called Cascade while the type III-A and III-B counterparts are referred to as the Csm and Cmr complex respectively. Cas7, Csm3 in type III-A systems, or Cmr4 in type III-B systems form the backbone of the complexes. For a comprehensive comparison and description of class 1 interference complexes, see van der Oost et al. (2014). Despite structural similarities of class 1 effector complexes, targeting in type III systems functions distinctly differently from that of type I systems. The two best characterized type III systems, type III-A and III-B, were initially shown to degrade RNA and not DNA in vitro (Hale et al., 2009, Staals et al., 2014, Zhang et al., 2012), but DNA in vivo (Marraffini & Sontheimer, 2010). Later, type III-A was shown to target RNA in vivo (Tamulaitis et al., 2014). These two activities were reconciled by the discovery that DNA-degradation is dependent on transcri-
tion of the target (Deng *et al*., 2013, Goldberg *et al*., 2014, Samai *et al*., 2015). The transcription-dependent DNA degradation was proposed as a mechanism to allow phages to integrate on the chromosome. As long as a targeted phage gene is not expressed, the integrated phage is not cleaved by the CRISPR-Cas system, and self-targeting is avoided. If the phage enters the lytic cycle and actively transcribes the target gene, the CRISPR-Cas system will cleave both the phage RNA and DNA, thereby preventing the phage from spreading (Goldberg *et al*., 2014).

No PAM has been observed for the DNA targets of the type III systems. Instead, the Cmr and Csm complexes rely on self-inactivation where binding between the 5’ handle of the crRNA and the target inactivates targeting (Marraffini & Sontheimer, 2010). Self-inactivation is only used for DNA targets (Samai *et al*., 2015, Tamulaitis *et al*., 2014).

Class 2 systems are defined by a single effector protein; Cas9 in the type II systems, Cpf1, C2c1, and C2c3 in type V, and C2c2 in type VI. Cas9 requires both crRNA and tracrRNA for target interference. Upon target binding, the two different endonuclease domains of Cas9 each cleaves one strand of the target DNA (Jinek *et al*., 2012). Cpf1 requires only a crRNA for target interference and has one lobe that cleaves both strands of the target DNA (Zetsche *et al*., 2015). C2c1 requires a crRNA and tracrRNA for target interference (Shmakov *et al*., 2015). A PAM has been identified for all DNA-targeting class 2 effectors (Jinek *et al*., 2012, Shmakov *et al*., 2015, Zetsche *et al*., 2015), except C2c3 that awaits further characterization (Shmakov *et al*., 2015).

The new type VI system is defined by the C2c2 effector protein. Unlike the other class 2 CRISPR-Cas systems, C2c2 targets ssRNA and not DNA for interference. In contrast to RNA-targeting by the type III systems, C2c2 does not require transcription of the target for interference, nor does it degrade the template DNA, other RNAs, however, are non-specifically co-degraded. As the target is RNA, no PAM is required for differentiation between self and non-self to avoid targeting of the crRNA (Abudayyeh *et al*., 2016).

Applications of the CRISPR-Cas system

Like the restriction-modification defense system, the CRISPR-Cas systems have been harnessed as biotechnology tools. This development has revolutionized many fields of research and created an avalanche of publications in an already rapidly moving field. It is therefore not feasible to describe all applications in detail. Here, an overview of applications in eukaryotes is provided, highlighting the most significant features followed by a more detailed description of some of the applications in bacteria.
Applications in eukaryotes

The most famous application of a CRISPR-Cas system is that of CRISPR-Cas9 for gene editing in eukaryotes. Cas9 is well-suited for various applications as it is a single protein that introduces specific double-stranded breaks in the target DNA without degrading the target. The break can subsequently be repaired by one of two DNA repair pathways, non-homologous end joining (NHEJ) or homology directed repair (HDR). NHEJ, leading to gene disruption, is most efficient while HDR, requiring a donor for homologous recombination, is less efficient but enables specific changes in the target. The characterization of Cas9, coupled to the discovery that the crRNA and the tracrRNA could be combined into a single guide RNA (gRNA) paved the way for applications using Cas9 (Jinek et al., 2012). This was rapidly followed by its application in eukaryotes (Cong et al., 2013, Jinek et al., 2013, Mali et al., 2013) and CRISPR-Cas9 is now widely used for genomic engineering. It has been employed in a wide-range of eukaryotes, from yeast, to plants and mammals. It is inexpensive and the ease by which the spacer of the gRNA can be modified is one of the main advantages of CRISPR-based applications (Jiang & Marraffini, 2015, Sternberg & Doudna, 2015).

The CRISPR-Cas9 technology opens up for treatment of diverse genetic diseases. The implications have spurred an ethical debate, especially concerning germline editing. It has also been questioned how many diseases would benefit from the use of CRISPR-Cas9 in treatment (Baltimore et al., 2015, Bosley et al., 2015). Currently, one clinical trial using CRISPR-Cas9 technology is underway where patient cells are edited ex vivo and then reinserted (Cyranoski, 2016).

Other applications using catalytically inactive Cas9 (dCas9) have been described, exploiting the easy programmable sequence-specific DNA-binding abilities of Cas9, but without the associated target cleavage. Examples of such applications are transcriptional activation or repression, DNA localization in live cells, and epigenetic rearrangements.

For recent reviews on CRISPR-Cas9 applications, see Sternberg & Doudna (2015) and Jiang & Marraffini (2015).

Cpf1 has also been employed for editing in eukaryotic cells. As Cpf1 does not require a tracrRNA for immunity, the crRNA can be transcribed into a long pre-crRNA (Zetsche et al., 2015). Multiplexing is therefore easier compared to Cas9 where each crRNA needs to be expressed from individual promoters. Additionally, the PAM of Cpf1 differs from that of Cas9, expanding the targeting potential using a CRISPR-Cas effector (Zetsche et al., 2015).
Applications in bacteria

Applications of the CRISPR-Cas system in bacteria have so far not had the same impact as those developed for eukaryotes. Nevertheless, several interesting applications have been described in bacteria.

To increase efficiency of genome editing in bacteria, the CRISPR-Cas system has been used in combination with existing methods, e.g. Lambda red recombineering in E. coli (Datsenko & Wanner, 2000). The technique is based on introducing a CRISPR-Cas system targeting the non-edited variant, so that unaltered cells are removed from the population. By selecting for the inserted CRISPR-plasmid instead of the introduced change, mutations could be engineered on the chromosome in a single step (Jiang et al., 2013). It also reduces the background of non-edited cells, especially useful in a host where recombination is less efficient.

The ability of the CRISPR-Cas system to remove specific cells from a population has also been exploited as targeted sequence-specific antimicrobials. Specific crRNAs selectively removed one bacterial species from a population, but not a second, closely related, species (Bikard et al., 2014, Citorik et al., 2014, Gomaa et al., 2014). CRISPR-Cas can also eradicate plasmids carrying antibiotic resistance, as well as prevent spread of such plasmid within a population, without killing the host. Delivery of the CRISPR-Cas system can be achieved by packaging the necessary components in a phagemid (Bikard et al., 2014, Citorik et al., 2014). Additionally, chromosomal targeting by an endogenous type I-F CRISPR-Cas system led to substantial chromosomal rearrangements and growth inhibition (Vercoe et al., 2013), further hinting at the ability of CRISPR-Cas systems to shape bacterial genomes and populations. Although easily designed, and the problems with antibiotic resistance ever increasing, sequence-specific antimicrobials require substantial knowledge about the bacterial population and the genomes present. If to be used e.g. to remove pathogenic bacteria from the gut, the targeting CRISPR-Cas system also have to be efficiently delivered.

Another application of the CRISPR-Cas systems is gene silencing, used both in eukaryotes and bacteria. Gene silencing has been achieved using Cascade (Luo et al., 2015) and dCas9 (Bikard et al., 2013, Qi et al., 2013), in Gram positive (Bikard et al., 2013) and Gram negative bacteria (Luo et al., 2015, Qi et al., 2013). By fusing dCas9 to a transcriptional activator, it is also possible to partly activate gene transcription (Bikard et al., 2013). Ideally, the endogenous CRISPR-Cas system could be used for silencing applications (Luo et al., 2015). Targeting the promoter region of the gene is most efficient (Bikard et al., 2013, Luo et al., 2015), except in one study (Qi et al., 2013). When targeting the open reading frame (ORF), proximity to the promoter determines the efficiency of gene silencing (Qi et al., 2013). Targeting of either strand of the promoter region is equally efficient (Bikard et al., 2013, Luo et al., 2015), but within the ORF, targeting of the non-template
(coding) strand leads to more efficient gene silencing (Bikard et al., 2013, Luo et al., 2015, Qi et al., 2013). CRISPR-Cas gene silencing can also be used to interrogate regulatory networks (Qi et al., 2013). Gene silencing by CRISPR-Cas effector complexes most likely prevents access of the RNA polymerase to the promoter or by blocking its progression during transcription.

Blockage of binding was also used to prevent initiation of replication by targeting dCas9 to the oriC of the bacterial chromosome. Replication becomes stalled and cells synchronized. Interestingly, repression is easily relieved by a temperature shift, hinting at temperature-dependent dCas9-binding (Wiktor et al., 2016).

Finally, dCas9 has been used for gene expression knockdown in Bacillus subtilis. Silencing is titratable, enabling partial knockdown while maintaining viability. By modulating the degree of CRISPR-Cas silencing, it is thus possible to study essential genes, which are otherwise difficult to study in vivo (Peters et al., 2016).

The newly characterized RNA-targeting type VI system, with the C2c2 effector, opens up for several RNA-based applications. It has been shown to efficiently silence expression of a targeted mRNA in E. coli and could be developed for several different applications, such as RNA localization and RNA regulation (Abudayyeh et al., 2016).
Aim

CRISPR-Cas immunity is a comparably recent field of research with many exciting things yet to be discovered. The overall aim of this thesis was therefore to increase the general understanding of CRISPR-Cas immunity in the type I-E system from *Escherichia coli*.

The stages of CRISPR-Cas immunity are complex and dependent on multiple components, all equally important for the function of the system. This is reflected in this work, where all stages of CRISPR-Cas immunity have been investigated. The work done on requirements for adaptation (paper I) was aimed to provide insights into adaptation and during the work on paper I, a tool for studying spacer integration was developed (paper II), aiming to facilitate further studies of adaptation. Next, the effect of crRNA spacer structures on interference and pre-crRNA processing were explored (paper III). Insights gained in paper III may be of use in the design of artificial crRNA for various applications, such as gene silencing by Cascade (paper IV).

Requirements for adaptation (Paper I)

During integration of a new spacer the CRISPR array, commonly located on the chromosome, is first cleaved and subsequently repaired after spacer insertion. We therefore hypothesized that the cellular DNA repair pathways would be involved in adaptation. To this aim, single gene deletions of a selection of DNA repair genes; *recA*, *recB*, *recC*, and *recD* were constructed in isogenic strains of *E. coli*. The RecBCD complex is required for repair of double-stranded DNA breaks. RecA mediates strand invasion during homologous repair and is involved in all major DNA repair pathways in *E. coli*. The engineered strains were derivatives of BW25113, with endogenous *cas* genes and two CRISPR arrays. Cas1 and Cas2 were expressed from a plasmid, pCas1+2, with CloDF13 replication origin (Yosef *et al.*, 2012). Spacer acquisition assays were done, with or without induction of *cas1* and *cas2* expression, for 16 h, and spacer integration was subsequently detected by PCR.
No spacer integration is detected in ΔrecD cells
Deletion of recA, recB, and recC impairs adaptation. Reduced integration was detected in ΔrecB and ΔrecC cells without induction thereby relying on leaky expression of Cas1 and Cas2, and in this condition no integration was detected in ΔrecA. The effect of deleting recA was thus more severe than deleting recB or recC. Most importantly, no spacer integration was detected in the ΔrecD background, even when Cas1 and Cas2 expression was induced. The fact that integration was observed in the recB and recC deletion strains, but not when recD was deleted, was difficult to reconcile with a RecBCD-model. The recD deletion was therefore further investigated.

Cas1 and Cas2 expression plasmid is unstable without RecD
After transformation with pCas1+2, colonies from the ΔrecD strain were smaller than those of wildtype. This difference was most pronounced after transformation but not observed upon re-streaks or in the absence of selection pressure for the plasmid. The small colonies were most likely caused by a longer lag period, indicating that the plasmid had difficulties establishing in the ΔrecD background. Some plasmids are not stably maintained in a recD deletion strain (Biek & Cohen, 1986), and, in agreement, very low plasmid concentration was purified from ΔrecD cells, suggesting that pCas1+2 was not stable without RecD. Additionally, when analyzed by gel electrophoresis, several slower-migrating plasmid species were observed in plasmid preparations from all strains, except in ΔrecA but including wildtype. The formation of the slow-migrating plasmid species is thus RecA-dependent.

To compare plasmid yields in the different strains, pCas1+2 was purified, digested, and analyzed by gel electrophoresis. Plasmids were purified in similar concentrations from wildtype, ΔrecB, and ΔrecC cells. Similar, or slightly lower, plasmid amounts were purified from ΔrecA cells compared to wildtype. Strikingly, very small plasmid amounts were purified from the ΔrecD strain.

Lower plasmid concentration leads to loss of integration
If the Cas1 and Cas2 expression plasmid is stably maintained in ΔrecD cells, either by creating a recA recD double deletion or by using a different, stably maintained, plasmid vector for Cas1 and Cas2 expression, integration was observed in a ΔrecD background. Thus, when using pCas1+2 for Cas protein expression, there is a two-layered effect of deleting recD. One effect is plasmid instability, which may result in less available plasmid material for new spacers as well as less Cas1 and Cas2 production. The second, a RecBCD-dependent effect, previously shown to impair generation of new spacer fragments in a double-stranded break-dependent manner (Levy et al., 2015).
In an effort to determine if Cas1 and Cas2 amounts are limiting for spacer integration in $\Delta$recA and $\Delta$recD cells, western blots using FLAG-tagged Cas1 were performed. In exponential phase, and without induction of Cas protein expression, Cas1$^{\text{FLAG}}$ amounts correlated well with plasmid amounts; Cas1$^{\text{FLAG}}$ amounts were decreased in $\Delta$recA and very low in $\Delta$recD cells. After 16 h of induction of Cas1 and Cas2 expression, comparable Cas1$^{\text{FLAG}}$ amounts were detected in all tested strains. The data therefore indicates that Cas1 and Cas2 amounts are limiting for adaptation. However, it is not yet known when adaptation occurs and further experiments are needed for more conclusive results. In this study, primed adaptation cannot be excluded as the strain encodes endogenous cas genes. Cas gene expression has previously been reported to be too low to support interference and primed adaptation in a similar background (Swarts et al., 2012, Westra et al., 2010). Some results, however, indicate that priming could occur when overexpressing only Cas1 and Cas2, but this was not investigated further. Our results could largely be reproduced in BL21AI, a strain naturally lacking chromosomal cas genes, where primed adaptation cannot occur.

This study demonstrates that the mode of Cas1 and Cas2 expression affects adaptation and that deletion of recD indirectly affects adaptation by decreasing the number of plasmids in the strain. Plasmid-based protein expression is convenient, fast, and easy, but can lead to artefacts. Care should also be taken when comparing different studies using different modes of Cas protein expression as this can also affect adaptation frequency.

A fluorescent reporter for detection of adaptation (Paper II)

The PCR-based assay commonly used for detection of spacer integration is cheap and easy, albeit time consuming and with poor detection limit, and the high detection limit can lead to false negative results (Staals et al., 2016). By PCR, spacer integration is detected on a population level, but provides no insights to adaptation in individual cells. Interpretation of results can be difficult, and nuances are easily overlooked. Therefore, PCR-based detection of spacer integration is best suited for yes-or-no answers and not for quantification. To quantify integration, the PCR products can be subjected to deep sequencing. However, this is more expensive and time consuming and does not alleviate the potential amplification biases.

Here, we used the same principle as the previously described reporter for spacer integration by antibiotic resistance (Díez-Villaseñor et al., 2013) to develop a fluorescent reporter for detection of spacer integration in single cells and easy quantification of adaptation by flow cytometry.
Construction of a fluorescent reporter for spacer integration

A partial CRISPR array was inserted between the start codon (ATG) and the ORF of yellow fluorescent protein (yfp), placing the yfp in the +2 frame of the ATG. The construct was inserted downstream of a constitutive promoter on the MG1655 chromosome. The reporter is thus constitutively transcribed, but Yfp is not translated until a new spacer is integrated into the CRISPR array, expanding it by one spacer-repeat unit (61 bp), thereby moving the yfp into frame with the translational start codon. In the same strain, all endogenous cas genes were deleted. The two endogenous CRISPR arrays were also deleted to target all integration events to the CRISPR array coupled to the fluorescent reporter. Two variants for the fluorescent reporter were constructed, one corresponding to CRISPR-I and the other to CRISPR-II. Integration was achieved in the same way as for paper I.

Spacer acquisition assay was performed using the constructed strains, and cells were subsequently plated or diluted and regrown. After regrowth, spacer integration was detected by PCR on the liquid culture. Integration could be detected in the CRISPR arrays coupled to yfp, demonstrating that the engineered arrays are active for integration. The band representing expanded arrays was very faint for CRISPR-I but stronger for CRISPR-II.

When plated, fluorescent colonies could be detected for both CRISPR-I and CRISPR-II arrays after overnight incubation. Fluorescent and non-fluorescent colonies were investigated by colony PCR. All tested fluorescent colonies had expanded CRISPR arrays while no integration was detected in any of the non-fluorescent colonies. Four PCR products were sequenced, all four integrations corresponded to unique spacers, indicating that the spacers were the result of individual integration events that had taken place in the culture prior to cells being plated.

Fluorescent reporter enables sensitive detection of expanded arrays

The Yfp-positive cells were well-separated from the unexpanded negative control when analyzed by flow cytometry, demonstrating that the reporter can be used for detection of spacer integration by flow cytometry. The detection limit of the fluorescent reporter was tested by mixing Yfp-positive control cells with cells with unexpanded arrays in known ratios. The mixed samples were subjected to PCR and flow cytometry. By PCR, 0.5 % expanded arrays were detected, but it was not possible to distinguish between samples with 0.5, 0.75, and 1 % expanded arrays. Using the fluorescent reporter, as little as 0.05 % expanded arrays could be detected and reproducibly quantified. Detection using the fluorescent reporter is at last ten-fold more sensitive than detection by PCR.
Quantification of spacer integration

Finally, spacer integration into CRISPR-I and CRISPR-II was quantified using the fluorescent reporter and detection by flow cytometry. After the acquisition assay, cells were diluted and regrown to allow development of Yfp fluorescence, especially from late integration events. When using CRISPR-I, 0.6 % of the cells was fluorescent, whereas this number was 3 % for CRISPR-II, correlating well with the difference in intensity of detected PCR-products from expanded arrays.

It is important to note that not all integration events will lead to fluorescence. In some cases, a spacer of aberrant length is integrated, and yfp is not moved into the correct frame. Furthermore, spacers containing in-frame stop codon(s) may be integrated, causing translational termination before yfp. Combined, the probability of detection of spacer integration is 0.59 (Diez-Villaseñor et al., 2013), and the percentage of Yfp-positive cells is therefore not the same as the integration frequency. This does not affect the ability to accurately detect relative differences in adaptation as demonstrated by the dilution tests.

The fluorescent reporter for adaptation can be used for easy detection and quantification of spacer integration in single cells and opens up for a wide range of fluorescence-based techniques. Fluorescence can be detected in single colonies, providing a second method for quantification of integration. Individual spacer sequences can also be analyzed by colony PCR and subsequent sequencing. It can be adapted to detect spacer integration for any CRISPR-Cas system where integration leads to expansion of the array by a number not dividable by three.

Spacer structure affects interference (Paper III)

We investigated whether secondary structures in the crRNA spacer sequence could affect CRISPR-Cas interference. Structures within the crRNA spacer could e.g. affect pre-crRNA processing or Cascade assembly, leading to impaired target interference.

Four different variations of the C1 spacer sequence were designed; C2 predicted to interact with the repeat, C3 with a hairpin in the spacer, C4 predicted to be completely unstructured, and C5 predicted to be moderately structured. The crRNA stem-loop is very stable and was not predicted to be perturbed by any of the designed spacer sequences. Calculated binding energies between the target and its cognate protospacer were comparable.

Minimal CRISPR arrays with 54 bp of the leader, two repeats and one spacer were constructed and low-copy target plasmids were constructed for
each of the crRNA spacers. The CRISPR plasmid, a plasmid coding for Cascade, and a third plasmid coding for Cas3 were inserted in BL21AI, a strain lacking endogenous *cas* genes. CRISPR-Cas interference was investigated by testing how efficiently the system prevented transformation of the target plasmid.

After transformation, the normalized transformation efficiency (NTE) was calculated by dividing the number of transformants with the number of viable cells for each transformation. CRISPR-Cas efficiency was calculated as 1 divided by the normalized transformation efficiency of the target divided by the normalized transformation efficiency of the non-target (1/(NTE_{target}/NTE_{non-target})). These calculations were used as transformation efficiency depends not only on the amount of DNA added but also on the number of cells present in the transformation reaction. *Cas* gene expression was induced at a specific OD for a set time, the OD at harvest varied and normalization with the number of viable cells was therefore necessary. Differences in competence are compensated by division with the normalized transformation efficiency of the non-target. Comparable normalized transformation efficiency was observed for all strains with different spacer variants when transformed with a plasmid not targeted by any of the crRNA spacers, indicating only small differences in competence.

**The CRISPR-Cas efficiency is affected by spacer structure**

The C1 spacer had a CRISPR-Cas efficiency of 10^4-10^5, and efficiently prevented transformation of the target plasmid. For C3 spacer, with an internal hairpin, the CRISPR-Cas efficiency was almost 1, suggesting that the target and the non-target plasmid were transformed equally. C2 and C5 both showed a 100-fold drop in CRISPR-Cas efficiency, whilst C4 showed a 1000-fold drop. For the C5 spacer, more biological replicates are needed to accurately determine the CRISPR-Cas efficiency.

It is surprising that the C4 unstructured spacer had such a considerable effect on CRISPR-Cas efficiency and that the spacer-repeat interaction of C2 is not more detrimental. C5 is moderately structured and is, in this respect, comparable to C1. Interestingly, all spacers except C2 are predicted to engage the single-stranded stretch in the 3’ end of the repeat in binding. This may impair Cas6e-binding or Cascade assembly. However, this single-stranded stretch is engaged in similar interactions in spacer C1 as well, which successfully prevents transformation of the target plasmid. In the event of directional Cascade assembly nucleating from the crRNA 3’-end bound by Cas6e, interactions in the crRNA between the spacer and the 5’ handle could be disrupted by binding of the Cas7 subunits to the spacer. The 5’ handle would subsequently be available to bind Cas5e and Cse1. The 5’ handle of the crRNA is predicted to be partly included in the secondary structures for all the spacer sequences, including C1. As C1 crRNA efficiently prevented transformation, interactions with the 5’ handle do not seem to
be detrimental for Cascade assembly. Further studies are needed to conclude on the effects of spacers binding to the 5’ handle of the crRNA on CRISPR-Cas efficiency.

It should be noted that prediction of secondary structures was done with the already processed crRNA, and not the full pre-crRNA. However, as a minimal CRISPR array was used, the pre-crRNA will not include additional spacers or repeats, and the main difference between the pre-crRNA and the processed crRNA will be the 54 nt leader sequence, identical for all the designed spacers.

A recent study showed that different spacer sequences have different activity and requirements, and differs in target clearance rates (Xue et al., 2015), and it cannot be excluded that the reduced CRISPR-Cas efficiencies observed here are caused by sequence variations. Ideally, more spacer sequences with the same predicted secondary structure would be evaluated to confirm the findings.

The designed spacers all had a detrimental effect on CRISPR-Cas activity. Additional experiments are required to elucidate how structures in the crRNA spacer affect CRISPR-Cas activity. In future studies, pre-crRNA processing could be investigated by northern blot and Cascade assembly could be assessed by in vitro assembly to determined which step of CRISPR-Cas immunity is impaired by the different spacer variants.

Gene silencing by Cascade (Paper IV)

In the last paper, we explored Cascade as a tool to silence gene expression in bacteria. Cascade is a large multi-subunit complex that covers a large region of the target DNA (Jore et al., 2011), and should be able to block access or elongation by RNA polymerase. Like Cas9, Cascade is easily programmed to bind different targets by changing the sequence of the crRNA spacer. In addition, Cascade does not require additional factors for crRNA processing, which facilitates simultaneous targeting of several genes.

Cascade can be used to silence gene expression

We constructed an E. coli strain deleted for cas3 and expressed Cascade from a plasmid, enabling target binding without target destruction. The crRNA was supplied from a second plasmid. We designed eight different spacers, targeting either the constitutive promoter or the ORF of green fluorescent protein (gfp) in a plasmid. Four spacers targeted the template strand, and four the non-template strand. GFP fluorescence was measured in a microplate reader, and, in agreement with other studies (Bikard et al., 2013, Luo et al., 2015), targeting the promoter silenced gene expression more effi-
ciently than targeting the ORF. When targeting the promoter region, at least a 10-fold reduction of Gfp fluorescence was observed, 50-fold at the most. When targeting the template strand of the ORF, Gfp fluorescence was reduced around 5-fold. Targeting of the non-template strand did not reduce Gfp fluorescence to the same extent. The observed reduction in Gfp fluorescence was stable over time.

Silencing was equally efficient when the gfp target was inserted on the chromosome. When a second fluorescent reporter gene, blue fluorescent protein (bfp), was also present, expression of both reporters could be silenced simultaneously using an array with spacers targeting each reporter. The crRNA targeting gfp only reduced Gfp fluorescence while crRNA targeting bfp specifically reduced Bfp fluorescence.

Flow cytometry showed that Gfp expression was similarly reduced in all cells, demonstrating that the reduction in fluorescence observed in bulk measurements was not the result of a mixed population cells.

**Cascade hinders transcription**

Northern blots revealed a decrease in gfp mRNA, demonstrating that silencing was exerted on transcriptional level. When using a probe binding in the beginning of gfp, shorter transcripts were detected in samples with ORF targeting crRNAs. This indicates that, for the ORF-targeting crRNAs, Cascade interferes with transcription elongation.

Compared to Luo et al., who also employed Cascade for gene silencing (Luo et al., 2015), lower levels of silencing were observed here (100-fold vs 50-fold). This may be due to different techniques used for fluorescence measurements. Here, a bulk measurement was employed, whereas Luo et al. used single cell measurements by flow cytometry. The level of observed fluorescence and Cascade-mediated silencing may also depend on the expression level of the reporter, e.g. promoter strength, and the Cascade- and crRNA-levels. If the target gene is located on plasmid, the copy number of the plasmid should also affect silencing, as a higher copy number would require more Cascade-crRNA complexes to bind all available promoters to achieve complete silencing. Other factors could also affect silencing efficiency, such as Cascade binding depending on spacer sequence (see paper III). More efficient silencing was observed when targeting dCas9 or Cascade to the non-template strand than to the template strand (Bikard et al., 2013, Luo et al., 2015, Qi et al., 2013). Here, the opposite tendency was observed. However, the number of tested spacers in this study was too limited to draw conclusions regarding strand bias.

Cascade and Cas9 are comparably efficient in silencing gene expression, and the choice of method for gene silencing would be determined by the studied organisms as well as by any PAM-limitations.
Conclusion and future perspectives

Adaption

Adaption is still the least understood stage of CRISPR-Cas immunity, but our knowledge has rapidly expanded during the last years. Despite the leaps in our understanding of adaption, many mechanistic details remain to be investigated. To discover additional factors needed for adaption, the effect on spacer integration of deleting different host genes has been studied. However, gene deletions often cause secondary effects that inadvertently affect the read-out. In paper I, deletion of recD caused decreased plasmid concentration, which in turn led to decreased integration, providing mechanistic insights into spacer integration and raising several interesting questions concerning limiting factors for adaption.

The data indicates that intracellular Cas1 and Cas2 levels are limiting for spacer integration. However, it is still possible that plasmid concentrations limit the amount of available spacer material, thereby limiting spacer integration. Previous studies have showed that most spacer are sampled from plasmids rather from the chromosome (Yosef et al., 2012, Levy et al., 2015), indicating that the amount of available plasmid is an important parameter. However, the study by Yosef et al. based their conclusions on the copy number listed in the literature. It would be interesting to experimentally determine the plasmid copy number of pCas1+2 or to vary the amount of available plasmid to confirm or rebut such biases inherent to Cas1-Cas2.

A second interesting question concerns the timing of spacer integration. Is any growth phase more favoring for spacer integration? It is possible that e.g. chromosomal structure or replication could aid or counteract integration. RecBCD- and double-stranded break-dependent bias to acquire spacers from replicating DNA was reported, and replication indirectly affect adaption as most double-stranded breaks occur at stalled replication forks (Levy et al., 2015). Most experimental setups to study adaption include extended growth times, 10-24 h, before detection of integration, and therefore fail to reveal temporal information. The fluorescent reporter for spacer integration developed here may be of use to study integration in real-time and thus provide insights to the temporal aspects of adaption.

The fluorescent reporter will be useful in further studies of adaption and it should be applicable to any CRISPR-Cas system where the CRISPR array is expanded by a number of base pairs not dividable by three. It opens up for
several exciting experiments, such as observation of spacer integration in real-time. It could also be used in combination with other assays, for example in studies relying on phage-survival to study adaptation. In such cases, the fluorescent reporter can be used to select for cells that have become phage-resistant through adaptation and not through other mechanisms.

The greatest remaining question regarding the adaptation mechanism concerns spacer capture. How are new spacers selected and how are new spacers generated and trimmed? Studies have shown two mechanisms for generating new spacers, one depending on RecBCD and one on CRISPR-Cas interference (primed adaptation). It has not been shown, however, that these two mechanisms are independent of each other. It is possible that Cas3 generates the breaks needed for RecBCD-action which then generates the pre-spacers for Cas1-Cas2. It would therefore be interesting to determine whether primed adaptation is RecBCD-dependent.

Integration is not completely lost even when the described pathways of spacer generation are absent, and the question remains if Cas1-Cas2 themselves are capable of generating new spacers or if other host factors are involved. As more chromosomal spacers were integrated when Cas1 and Cas2 were overexpressed than with more moderate Cas1 and Cas2 levels (Levy et al., 2015), it is possible that Cas1 and Cas2 themselves generate spacers, but without preference for exogenic DNA. With large Cas1-Cas2 amounts, the complex may non-specifically cleave DNA, thereby generating ends susceptible to RecBCD-action.

It also remains to be clarified how the single-stranded sequences generated by RecBCD or Cas3 become double-stranded before being integrated into the CRISPR array. Cas3 cleavage also requires further investigation, not only how Cas3 cleaves target DNA in vivo, but also how the target DNA is degraded, and how Cascade and/or Cas1-Cas2 regulate Cas3 activity.

**Expression, processing, and interference**

We show that spacer sequence, and its predicted secondary structure within the crRNA, affect CRISPR-Cas interference. This may be caused by impairing pre-crRNA processing, Cascade assembly, target binding, or target degradation. Previous studies demonstrated differences in spacer activity depending on spacer sequence, leading to different modes of activity of Cascade and Cas3. The PAMs and number of spacer-crRNA mismatches that supports target interference varies with the spacer sequence (Xue et al., 2015). The basis of these differences remains to be determined, as well as how possible RNA secondary structures fit into the picture. The spacers used in this study all have the same predicted target-binding energies, target affin-
ity should therefore not affect spacer efficiency, but effects of individual spacer sequences on CRISPR-Cas interference cannot be excluded.

Analysis of pre-crRNA processing could also offer insights into how Cascade assembles. If assembly is sequential, nucleating from Cas6e, a strong secondary structure next to the repeat stem-loop (such as spacer C3) could inhibit Cascade formation. Likewise, if Cas5e and Cse1 bind the 5’ handle before assembly of the Cascade backbone, a spacer engaging the 5’ handle of the crRNA in base pairing (such as spacer C2) could inhibit Cascade formation. As neither pre-crRNA processing nor Cascade complex formation have yet been analyzed, it is difficult to speculate on the effects of specific spacers on Cascade assembly. Pre-crRNA processing can be assessed by northern blot and Cascade assembly by e.g. protein pull-out experiments.

A larger tolerance for mismatches between the target and the crRNA within the seed sequence than previously observed was recently reported. The tolerance for mismatches was highly dependent on the spacer sequence (Xue et al., 2015). Some flexibility in the PAM-sequences eliciting efficient interference has already been reported (Westra et al., 2013), and the PAM-sequences supporting interference was further reported to be dependent on spacer sequence (Xue et al., 2015). The PAM- and seed-sequences may therefore not be so easily generalized and may have to be revised. Spacer sequence-dependent interference also hinders comparisons of previous data as different spacers have been used in different studies. It may also hamper analysis of our results in paper III. Much work remains to elucidate the underlying reasons for this spacer sequence dependency and how it affects Cascade action.

Cas3 has been considered a highly efficient exonuclease, rapidly degrading its target, as shown in early in vitro studies (Westra et al., 2012b). However, recent studies by Redding et al. and Künne et al. reported different Cas3 activity and degradation patterns (Künne et al., 2016, Redding et al., 2015). Future research needs to elucidate Cas3 activity in vivo and how Cas3 activity can be regulated by Cascade and Cas1-Cas2. More knowledge about the in vivo target degradation products generated by Cas3 is also needed to further explain the connection between interference and adaptation. It will also be important to investigate the mechanism that leads to different Cascade-targeting depending on the spacer sequence. Is this affected by how Cascade binds the crRNA or is it determined on the level of target-protopspacer binding?

Throughout this thesis, plasmid-based expression of the required CRISPR-Cas components has been used. This is necessary as expression of the CRISPR-Cas system in E. coli K-12 is silenced by H-NS. The absence of CRISPR-Cas immunity, despite functionality of all components, in unper-turbed cells is one of the mysteries of the E. coli CRISPR-Cas system. Without selection for functionality, the cas genes should have accumulated inactivating mutations or been deleted. Perhaps there are conditions that induce
sufficient expression for immunity, but these are yet to be found. Since expression of the CRISPR-Cas components is not enough for immunity, it has been suggested that they are instead important for other processes in *E. coli*, such as DNA repair (Westra *et al.*, 2014).

**Applications**

Although the application of CRISPR-Cas for gene editing in eukaryotes, where such modifications have been difficult or impossible, has attracted most interest in CRISPR-Cas applications, there are nonetheless several interesting applications of the CRISPR-Cas system in bacteria. The DNA-binding abilities of Cascade and dCas9 have enabled several applications in bacteria, such as replication stalling or the use of dCas9 for studies of essential genes in *Bacillus subtilis* (Peters *et al.*, 2016, Wiktor *et al.*, 2016).

We demonstrated strong transcriptional silencing obtained when targeting Cascade to a promoter. Inducible expression could therefore be used for temporary gene silencing to study the effect of certain gene products in *e.g.* different growth phases. It should also be possible to utilize the endogenous Cascade for gene regulation, which is not possible with Cas9 as its nuclease-activity has to be inactivated in order to only silence and not cleave the target. Use of endogenous Cascade for gene regulation requires absent or inactive Cas3. In *P. aeruginosa*, Cas3 inactivation was achieved by prophage-expression of an anti-CRISPR protein inhibiting the nuclease (Bondy-Denomy *et al.*, 2015).

Cascade increases the number of targets that can be silenced as it utilizes a different PAM compared to Cas9. Cascade, unlike Cas9, independently processes the pre-crRNA, and the crRNA can therefore be supplied as one long transcript when using Cascade for gene silencing, facilitating simultaneous targeting of multiple genes.

To further increase the applicability of Cascade and dCas9 in gene silencing, further characterization of off-target effects in bacteria would be needed, as well as more insights about how different spacers could affect Cascade, or dCas9, binding. Insights gained in paper III could thus be of use when designing crRNAs for different applications using Cascade.

Since the first demonstration of immunity of the type I-E CRISPR-Cas system almost ten years ago (Barrangou *et al.*, 2007), considerable knowledge has been gained about the mechanisms of these immune systems. However much remains to be investigated, and some old truths need to be re-examined. In future research, it will be important to continue the search for host factors needed for adaptation, such as the ligases required to finalize spacer integration. Furthermore, generation of spacer materials and spacer capture deserve more attention. In interference, outstanding questions in-
clude how Cascade assembles as well as how different spacer sequences affect Cascade binding and activity and how that – in turn – affects Cas3 activity.

During the last year, several new CRISPR-Cas systems have been discovered and the work of characterizing these new systems has just begun. In the coming years, even more CRISPR-Cas systems will likely be discovered and many exciting discoveries surely await. For other systems, several mechanistic features remain to be characterized. Amongst these are the relevance of RNA-targeting, as well as the mechanism for adaptation, in type III systems. For both type II and type III systems, e.g. the secondary trimming of the processed crRNA remains elusive.

Lastly, with the newly discovered systems, the possibilities of CRISPR-Cas applications are expanded even further. The versatility and simplicity of sequence-specific targeting of the CRISPR-Cas systems will continue to make them an attractive tool for applications. Only the imagination limits the CRISPR-Cas applications, and the continued development of the CRISPR-Cas toolbox will be truly exciting to follow.

To conclude, the work presented here aims to further our understanding of the type I-E CRISPR-Cas system in *E. coli* and describes a new tool to study adaptation. Many mechanistic features yet await characterization and the CRISPR-Cas systems will continue to be an exciting field of research.
Acknowledgements

First and foremost, my supervisor Magnus - thank you for the opportunity to do my PhD with you. Thank you for all the support, freedom, and the great trust you have showed me over the years. I am impressed by your ability to explain our research to the public; you’re a real role model for how to perform the third task.

My co-supervisor Gerhart - thank you for including me in your group, in meetings big and small. Thanks for keeping me grounded and making sure I stay connected to the old-school stuff. I have appreciated all the scientific, and non-scientific, discussions with you.

My second co-supervisor; Sanna - I am so happy that you joined Micro and became my second co-supervisor. Thank you for all the hands-on advice, all the nice collaborations, and intense discussions. I always leave your office feeling enthusiastic about my project(s) and about science.

Mirthe – thank you for teaching me everything I know in the lab. I have loved sharing lab bench with you, all of our discussions, and our joint projects. Without you, half of my projects wouldn’t have been nearly as good and none of my images as pretty.

I also want to thank my students, without whom several projects would not have seen the light of day. I especially want to thank Martin; I really enjoyed having you as my colleague, and discussing all the projects with you. I was proud to see you grow as a scientist and I hope you’ll find your way back to research (if that’s what you want!).

The entire Micro-program: Thank you all for creating such a nice work environment, and for all the help. And last, but not least, thank you for all the fika!

The Wagner group: Thank you for including me as one of your own. Erik, you made a strong impression on me as a young PhD student. Thank you for teaching me things around the lab and for all the feedback. It’s nice to have you back at Micro! Cédric, all your in vitro gels in the group meetings have really taught me a lot - thank you for explaining them to me over and over again. Bork, I really miss having you as an office-mate, I’m sorry for all the interruptions, but I really enjoyed, and miss, our scientific discussions. Sofia -the best office mate I’ve had (sorry Bork…). Thank you for all the chocolates, nice words, and restaurant recommendations. Maaike, thank
you for teaching me all about western blots. The CDI group: Thank you for lending me Sanna from time to time. Marcus, thank you for all your expertise when I’m too lazy to look things up myself. Magnus, keep suiting up for Fridays! Disa, thanks for all the friendly advice and all the help. Anirban and Katrin, good luck with your research. The Giardia and Spiro group; Elin, thanks for all the support during this last year. Thank you for coming out of retirement to be my toastmaster – you will do amazingly! Showgy.

lycka till med din fortsatta karriär! Ásgeir, Thanks for all the tech support! Dimitra and Jingyi, Fellow Väktargatan-people, good luck with your research! The Micro-newcomers: Mikael, good luck with your new group! Maria Letizia, Marie and Stefan, Welcome (back) to Micro, hope you’ll stick around! The Dicty group: Fredrik, thank you for always being ready to help and for all the discussions about teaching. Jonas and Zhen, always nice chatting with you both. Good luck with everything! The worm people: Andrea, it’s been nice having you at micro! Benjamin and Yani, good luck on both of your upcoming dissertations.


Fredrik - min klippa. Tack för allt.
References


Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology 1466

Editor: The Dean of the Faculty of Science and Technology

A doctoral dissertation from the Faculty of Science and Technology, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology. (Prior to January, 2005, the series was published under the title “Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology”.)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-312234

ACTA UNIVERSITATIS UPSALIENSIS
UPPSALA
2017