2016

CPISPR-CAS: From a Prokaryotic Immune System to a Gene Editing Tool

Wenyan Jiang

CRISPR-CAS: FROM A PROKARYOTIC IMMUNE SYSTEM TO A GENOME EDITING TOOL

A Thesis Presented to the Faculty of

The Rockefeller University

in Partial Fulfillment of the Requirements for

the degree of Doctor of Philosophy

by

Wenyan Jiang

June 2016

CRISPR-CAS: FROM A PROKARYOTIC IMMUNE SYSTEM TO A GENOME EDITING TOOL

Wenyan Jiang, Ph.D.

The Rockefeller University 2016

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and their associated genes (*cas*) encode an adaptive, small-RNA-based immune system that protects prokaryotes from infectious phages and plasmids. CRISPR-Cas systems can be classified into three types based on their *cas* gene content. My thesis work focused on two parts. First, I investigated the mechanism and function of RNA cleavage in type III CRISPR-Cas immunity. Secondly, I developed a tool to manipulate prokaryotic genomes and gene expression by using an engineered type II CRISPR-Cas system.

To date, all three types of CRISPR-Cas systems target DNA. Type III CRISPR-Cas immunity displays an elaborate targeting mechanism and distinguishes itself from type I and type II systems in at least two ways. My previous work in collaboration with other members of the lab helped to discover that the system cleaves both DNA and RNA molecules, and that active transcription across the target is necessary for targeting. Whereas DNA cleavage is required for phage DNA clearance and essential for immunity against infection, the function of RNA cleavage is unknown. It is well established that gene expression of many phages is temporally regulated. Using a type III-A CRISPR-Cas system of *Staphylococcus epidermidis* as a model, I first identified a new CRISPR-associated RNase, Csm6. The transcriptional requirement for DNA

cleavage created a challenge for host bacteria. When the target was located in a late-expressed phage gene, the phage infection cycle can proceed unchecked until the target was transcribed, resulting in a sharp increase of viral genomes in host cells. In this targeting condition, genetic inactivation of the type III-A RNases Csm3 and Csm6 led to the accumulation of the target phage mRNA and abrogated immunity. Csm6 was also required to provide defense in the presence of mutated phage targets, when DNA cleavage efficiency was reduced. My results showed that the degradation of phage transcripts by CRISPR-associated RNases ensures robust immunity in situations that lead to a slow clearance of the target DNA.

Recent work on the type II CRISPR-Cas adaptive immune systems has led to the discovery of Cas9, a dsDNA nuclease whose sequence specificity is programmed by small CRSPR RNAs (crRNAs). In collaboration with Dr. David Bikard, a former colleague, we found when reprogrammed to target the genomes of host bacteria, CRISPR can target and kill the cells. The lethal consequence upon targeting made CRISPR-Cas a novel tool for sequence-specific counter-selection. When combined with editing templates, we demonstrated fast and efficient genome editing in *Streptococcus pneumoniae*, *Escherichia coli* (when used in combination with λ-Red recombination) and *Staphylococcus aureus*. Later, we inactivated the nuclease domains of Cas9, creating a catalytically dead Cas9 (dCas9) which retained DNA binding activity. We demonstrated that dCas9, when programmed with appropriate crRNAs, acted as a transcription repressor by preventing the binding of the RNA polymerase (RNAP) to promoter

sequences or as a transcription terminator by blocking the elongating RNAP. In addition, a fusion between the ω subunit of the RNA polymerase and dCas9 allowed for programmable transcription activation. The easy programmability and high specificity of crRNA-guided Cas9 and dCas9 greatly facilitates both genome editing and modulation of gene expression, and is likely to substantially advance our capability to decipher gene function in prokaryotes and manipulate them for biotechnological purposes.

ACKNOWLEDGEMENTS

First and foremost I would like to express my deepest gratitude to my Ph.D. advisor, Dr. Luciano Marraffini. I joined Dr. Marraffini's lab in January 2011 as his first graduate student. Dr. Marraffini is intelligent, hardworking, open-minded and inspirational. Throughout the years, his careful and patient mentorship has been instrumental to my Ph.D. training. In his lab, I am very grateful to have the opportunity to both study basic biology and extract the knowledge there for new technological development. In both cases, Dr. Marraffini always asks me to test hypotheses with rigorous scientific methods and actively inspires me to explore new frontiers. Dr. Marraffini supports me in many different levels. In addition to his mentorship in the lab, he generously supports me to go to conferences on a yearly basis; he gives due advice, support and respect to all of my proposals, ideas and issues. I am truly indebted to him for accepting me as a student and mentoring me along the way, as I learned a great deal about science, collaboration, perseverance as well as family, friendship and things that will impact my life for the many years to come.

I would like to give much credit to my thesis committee – Drs. Vincent Fischetti, Cori Bargmann and Hermann Steller for contributing their expertise to my thesis work. They constantly challenged and encouraged me, which helped me to foster a rigorous and agile scientific mind. I also thank them for patiently listening to me and giving me frank career advice.

I would like to thank Dr. Martin Jinek from University of Zurich for serving as my external examiner. Dr. Jinek's previous works include biochemical characterization of Cas9 and Csm6, the two CRISPR-associated proteins that are the core of my thesis project.

Next I would like to thank all current and previous members of the Marraffini lab for being wonderful colleagues and friends. In particular, I want to mention the people who trained and collaborated with me on my thesis work. Dr. David Bikard is a former post-doc who trained me and introduced me to the fascinating world of synthetic biology. He is always sharp, insightful and resourceful and I constantly learn new facts and ideas from him. Importantly, our collaboration led to the development of the type II CRISPR-Cas as a genome editing and gene regulation tool in bacteria. Dr. Poulami Samai, a former post-doc, is a skillful biochemist who trained me on many techniques including protein purification, Southern blot, and various in vitro DNA/RNA cleavage assays. In the lab, we were studying similar projects with regard to the targeting stage of type III CRISPR-Cas immunity and often shared data and gave mutual encouragement. She helped me set up a few important in vitro assays that allowed me to interrogate my thesis question. Gregory Goldberg is a graduate student who collaborated with me on various projects. I thank him for always being critical and rigorous toward my work and gave me many useful suggestions for my thesis project. I want to thank Dr. Joshua Modell, a methodical and inspirational post-doc who is currently collaborating with me on multiple new projects. I want to thank Dr. Asma Hatoum-Aslan, a former post-doc for teaching me the primer

extension assay and Inbal Maniv, our former lab manager for collaborating on a CRISPR “escaper” project.

Kudos to the David Rockefeller graduate program and the Deans’ Office (current members: Drs. Sidney Strickland, Emily Harms and Andrea Morris, Kristen Cullen, Marta Delgado, Cristian Rosario and Stephanie Fernandez) for supporting me both financially and academically. My graduate school experience would not have been so smooth without help and input from them.

I would like to thank my former mentors at Rockefeller, Drs. Seth Darst and Robert Darnell. In 2009 through the Rockefeller Summer Undergraduate Research Fellowship program, I worked in Dr. Darst’s lab where I learned a great deal about protein purification. In 2010, I rotated in Dr. Darnell’s lab and learned about antibody internalization and the HITS-CLIP technique. These learning experiences are important parts of my Ph.D. training, and I thank members from both labs for helping me then and afterwards.

I would like to thank Dr. Connie Zhao, Bin Zhang and Sophie Huang from the genomics, Drs. Henrik Molina, Milica Tešić Mark from the proteomics and Drs. Alison North and Kaye Thomas from the imaging resource centers at Rockefeller for helping me with various of my projects. I would like to thank all my past and current collaborators outside Rockefeller University for contributing their expertise and reagents. They are Dr. Bruce Levin from Emory University, Dr. Feng Zhang from the Broad Institute and MIT, Dr. Donald Court from National Cancer Institute and Dr. Ann Hochschild from Harvard Medical School.

Last but not least, I would like to extend my gratitude to my parents, Chao Jiang and Yinhong Lou for their unconditional love and support; for coming to the U.S. together with me and giving me the freedom to pursue what I love. Equally I want to thank all my friends here at Rockefeller, Weil Cornell and Sloan-Kettering for spending their times with me and making my graduate school career an unforgettable experience.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 INTRODUCTION

1.1 Bacteriophages and their life cycles

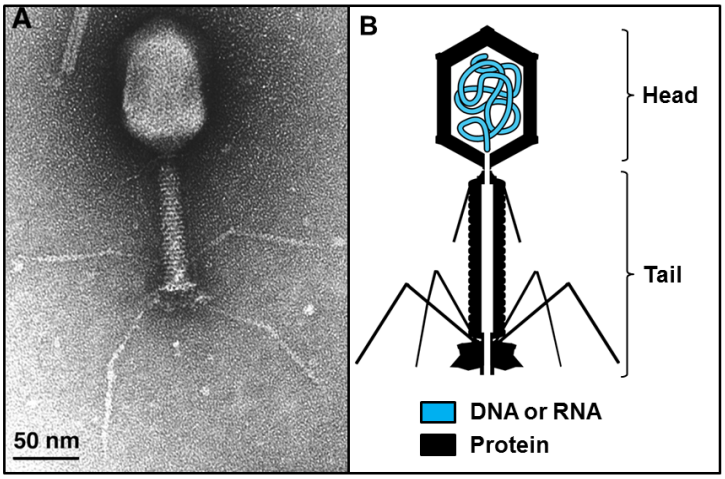


Figure 1-1 Bacteriophage T4

(A) An electron micrograph of bacteriophage T4; adapted from (Miller et al. 2003). **(B)** A phage (in this case T4) is composed of a proteinous head and tail structure. The head encapsulates its DNA or RNA genome. The image is adapted from "https://en.wikipedia.org/wiki/Bacteriophage" with modifications.

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and their associated genes (*cas*) encode an adaptive, small-RNA-based immune system that protects prokaryotes from infectious phages and plasmids. The immunity pathway is divided into three stages: adaptation, CRISPR RNA (crRNA) biogenesis and targeting. On the basis of *cas* gene conservation and operon organization, CRISPR-Cas can be divided into three major types. In addition to being an immune system, CRISPR-Cas has been repurposed as a versatile tool for genome editing. A recent review (Jiang and Marraffini 2015) written by my

advisor, Dr. Luciano Marraffini and me summarizes current understanding of CRISPR-Cas biology in prokaryotes and its many applications in biotechnology.

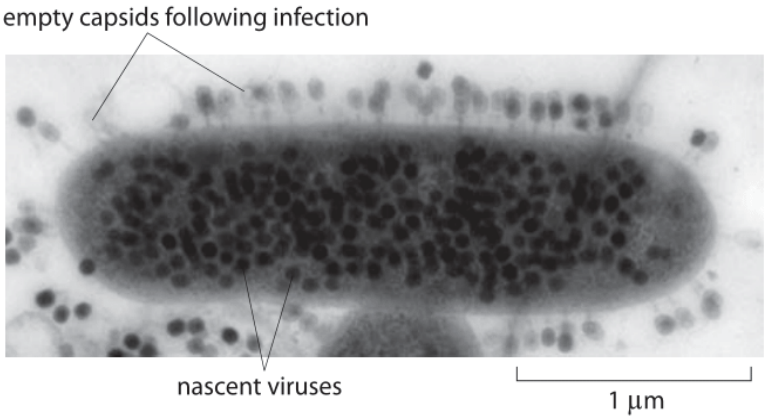


Figure 1-2 Phages predate on bacteria

A transmission electron micrograph of a thin section of *Escherichia coli* K-12 infected with phage T4. Dark materials are phage DNA. Image courtesy of John Wertz.

Bacteriophages (**Figure 1-1**), or phages for short, are viruses that infect and replicate within bacteria (**Figure 1-2**). First discovered independently by Fredrick Twort and Felix d'Herelle in 1915 (Twort 1915) and 1917 (d'Herelle 1917), phages are probably the most abundant and genetically diverse entities known to exist in biology. It is estimated there are 10^{31} phages on Earth (Hatfull 2008), and they infect an estimated 10^{23} bacteria per second (Suttle 2007). Because their massive predation on bacteria that live in the ocean, soil as well as the human gut, phages play a tremendous role on global ecology (Wommack and Colwell 2000) and human health (Reyes et al. 2012). Historically, studies conducted on phages between the years 1940 and 1970 laid much of the most fundamental groundwork for modern molecular genetics. For instance, experiments with

phage T4 uncovered the existence of mRNA, the nature of the genetic code and the function of ribosome (Karam 1994). In addition, basic research on phages has led to the development of many useful tools such as phagemids, phage-plasmid hybrids that can be used as an efficient delivery method of nucleic acids into bacteria (Westwater et al. 2002). Phage display is another powerful technique for the study of protein-protein and protein-DNA interactions (Smith 1985; Kehoe and Kay 2005). With this technique, large libraries of proteins can be screened and amplified through in vitro selection, a process widely used in drug discovery (Smith and Petrenko 1997). The infectious nature of phages is also harnessed as antibacterial agents (Nobrega et al. 2015), which offer a plausible solution to the rise of multi-drug resistant infectious bacteria caused by conventional antibiotic treatment.

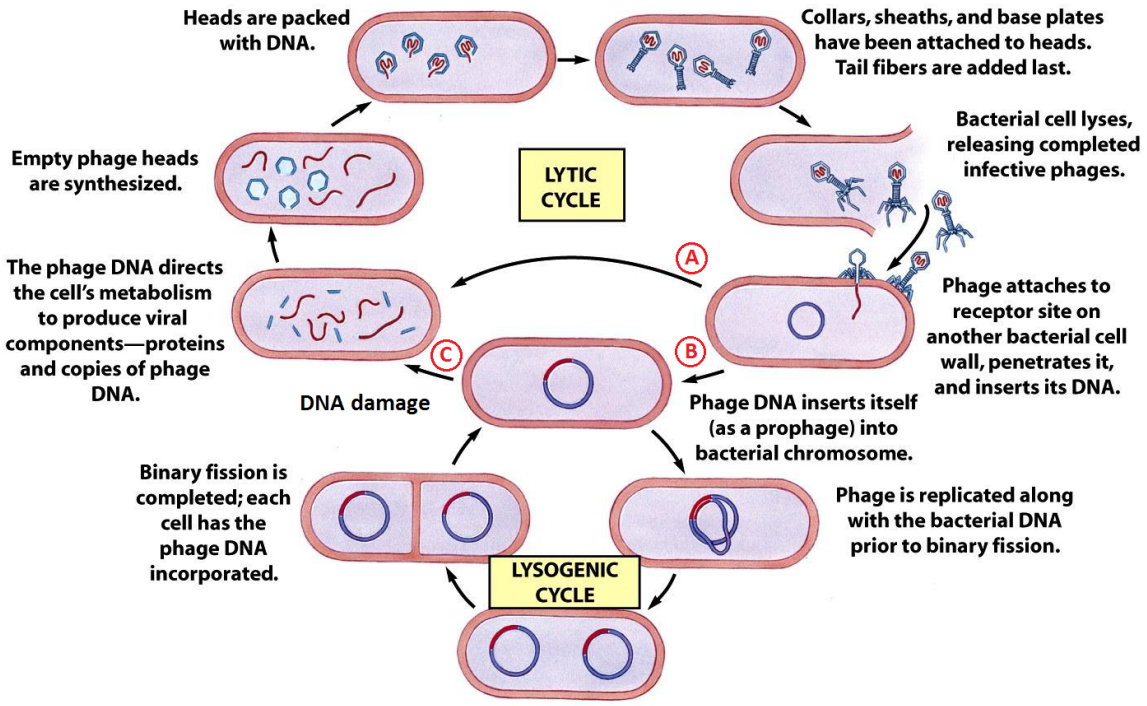


Figure 1-3 Lytic and lysogenic life cycles of phages

(A) In a lytic cycle, an infective phage immediately initiates reproduction and lyses host cells to release its progeny. **(B)** In a lysogenic cycle, a phage inserts its genome into the bacterial chromosome to become a prophage, which is replicated along with the bacterial DNA. **(C)** Under certain conditions such as DNA damage, the lysogenic cycle can convert to the lytic cycle as shown in **(A)**. This figure is adapted from *Microbiology*, 7/e © 2008 John Wiley & Sons.

Generally, a phage is composed of a protein capsid that encapsulates a DNA or RNA genome (**Figure 1-1B**). In many cases, the capsid, also known as the head, has a tail structure attached to it. The tail structure is a hollow tube made of proteins that allows the phage to penetrate bacterial membrane and cell wall and inject its genetic material into the cell. Upon infection, phages typically sustain two distinct life cycles – lytic and lysogenic (**Figure 1-3**) (Herskowitz and

Hagen 1980). Lytic phages immediately enter a reproductive phase, in which their genetic material is replicated and packaged into a few hundred of progeny phage particles. The lytic cycle concludes with lysis of the host bacteria and release of phage progeny. By contrast, temperate phages can adopt either of the two life cycles, depending on their genetic circuits and environmental factors. If the lysogenic cycle is chosen, a single copy of the phage genome integrates into the bacterial chromosome to become a prophage, which is replicated together with the host chromosome. A switch to the lytic cycle can initiate under certain stress conditions such as DNA damage (Herskowitz and Hagen 1980). During this, the once-dormant prophage excises from the bacterial genome and follows the program similar to that of a lytic phage. The lysis-lysogeny decision has been a paradigm for gene regulatory network, and has been extensively characterized in the temperate phage λ (Oppenheim et al. 2005).

For many large double-stranded DNA phages such as T4 and λ, expression of phage genes is temporally regulated (Miller et al. 2003; Oppenheim et al. 2005). Immediately transcribed upon infection, most early genes encode enzymes involved in DNA replication and phage-encoded RNA polymerases that turn on late phage genes. These genes usually encode various phage parts such as the head and tail structures, packaging machinery and lytic enzymes that degrade bacterial cell walls. Temporal gene regulation is a “clever” way to ensure efficient utilization of host resources, and to control the timing of cell lysis so that an optimum number of phage progeny can be produced.

1.2 CRISPR-Cas is a small-RNA-based adaptive immune system in prokaryotes

Bacteria have evolved different ways to protect themselves against phage infections. Some of the best-known types of phage defense mechanisms are restriction-modification (Wilson 1991; Tock and Dryden 2005), abortive infection (Chopin et al. 2005) and CRISPR-Cas.

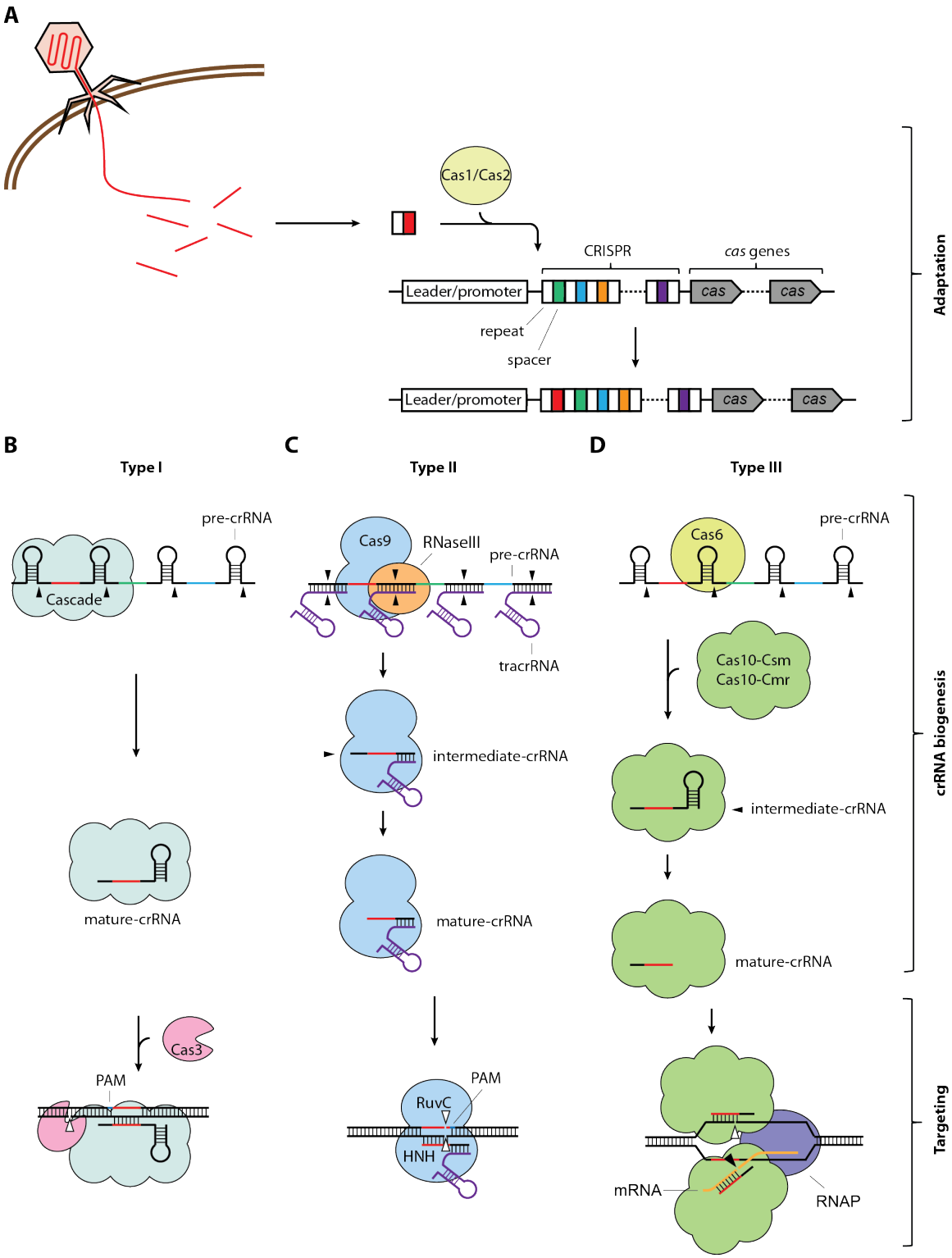
CRISPR-Cas systems encode an adaptive, small-RNA-based immune system that protects prokaryotes from infectious phages and plasmids (Deveau et al. 2010; Horvath and Barrangou 2010; Terns and Terns 2011; Wiedenheft et al. 2012; Barrangou and Marraffini 2014). A typical CRISPR-Cas locus in bacteria and archaea is composed of the CRISPR array and CRISPR-associated genes (*cas*) (**Figure 1-4A**).The CRISPR array is a cluster of short repetitive sequences (30-40 bp long) separated by equally short “spacer” sequences. Many spacer sequences match the genomes of phages and plasmids of bacteria and archaea (Bolotin et al. 2005; Mojica et al. 2005; Pourcel et al. 2005). This observation led to the hypothesis that CRISPR systems protect prokaryotes from infection by these genetic elements (Bolotin et al. 2005; Mojica et al. 2005; Pourcel et al. 2005; Makarova et al. 2006). The bioinformatics predictions were first demonstrated by two experimental studies showing that CRISPR loci prevent viral (Barrangou et al. 2007) and plasmid (Marraffini and Sontheimer 2008) infection.

CRISPR-Cas immunity is divided into three stages (**Figure 1-4**). In the first stage, known as the adaptation phase, Cas proteins integrate short regions of

the invader's genome into the CRISPR array as new spacers (**Figure 1-4A**). Reviews about this phenomenon are published elsewhere (Heler et al. 2014). In the second stage, the CRISPR array is transcribed and processed by Cas proteins to generate small CRISPR RNAs (crRNAs) that contain a full or partial spacer sequence (**Figure 1-4B-D**). During the third stage, or the targeting phase, processed crRNAs associate with Cas nucleases to guide the ribonucleoprotein complex to the target sequence (**Figure 1-4B-D**). Cleavage of the target sequence, also known as protospacer, results in both the destruction of the invader's genome and immunity. On the basis of *cas* gene conservation and operon organization, CRISPR prokaryotic immune systems can be divided into three major types (Makarova et al. 2011), each of which presents variations of the general immunity mechanism described above (**Figure 1-4B-D**).

Figure 1-4 Three stages of CRISPR-Cas immunity

(A) A typical CRISPR-Cas locus in bacteria and archaea is composed of a CRISPR array and CRISPR-associated genes (*cas*). In the first stage, known as the adaptation phase, Cas proteins acquire a short sequence of the genetic material of viral and plasmid invaders, known as a "spacer", and integrate it into the first repeat of the CRISPR array, establishing a memory of infection. (**B-D**) In the second stage, known as the crRNA biogenesis phase, the CRISPR array is transcribed as a precursor crRNA (pre-crRNA), which is processed by Cas proteins into a short crRNA (mature crRNA). In the third stage, known as the targeting phase, the crRNA acts as a guide to specify the target of cleavage by Cas nucleases. Based on the *cas* gene content, there are three different types of CRISPR-Cas systems. They differ in the mechanisms of crRNA biogenesis and targeting, but also possibly in the mechanisms of adaptation as well. Closed and open arrowheads indicate RNA and DNA cleavage, respectively. Abbreviations: crRNA, CRISPR RNA; PAM, protospacer-adjacent motif; RNAP, RNA polymerase; tracrRNA, trans-encoded crRNA.


A
Cas1/Cas2
CRISPR
cas genes
Leader/promoter
cas
cas
repeat
spacer
Leader/promoter
cas
cas
Adaptation
B
Type I
pre-crRNA
Cascade
mature-crRNA
Cas3
PAM
C
Type II
Cas9
RNaseIII
pre-crRNA
tracrRNA
intermediate-crRNA
mature-crRNA
PAM
RuvC
HNH
D
Type III
pre-crRNA
Cas6
Cas10-Csm
Cas10-Cmr
intermediate-crRNA
mature-crRNA
mRNA
RNAP
crRNA biogenesis
Targeting

1.2.1 Type I CRISPR-Cas systems

Type I CRISPR-Cas systems encode proteins that form CRISPR-associated complex for antiviral defense (Cascade) and Cas3 (**Figure 1-4B**). The CRISPR array is transcribed as a long precursor crRNA that is processed at the repeat sequences to liberate short, mature crRNAs (Brouns et al. 2008). Cleavage is achieved by the Cas6 endoribonuclease, a subunit of the Cascade complex with repeat-specific activity. Cas6 cleaves the repeat sequence eight nucleotides upstream of the spacer sequence, liberating a small crRNA containing a full spacer flanked by partial repeats. The crRNA remains associated to Cascade and guides the complex to its target.

Efficient immunity requires the interaction between the first eight nucleotides of the target (a region known as the target “seed”) and the complementary sequence of the crRNA guide (Semenova et al. 2011; Wiedenheft et al. 2011) at the 5’ end of the DNA:RNA duplex. Viruses containing mutations in this region can indeed escape type I CRISPR immunity in *E. coli* (Semenova et al. 2011). Mutations in the sixth nucleotide of the seed are exceptions – they do not affect CRISPR immunity. This is due to the noncanonical ribbon structure formed by a guide crRNA and its cognate ssDNA target within the Cascade complex. This conformation requires outward rotation of every sixth nucleotide in the crRNA:ssDNA hybrid (Jackson et al. 2014; Mulepati et al. 2014; Zhao et al. 2014). As such, the sixth nucleotide of the seed can remain unpaired without compromising the unusual structure, and a mutation at this location does not affect immunity.

A second requirement for type I immunity is the presence of a nucleotide motif immediately upstream of the target sequence, the protospacer-adjacent motif (PAM). For the type I-E system of *E. coli* the PAM is an AWG trinucleotide and it is recognized by CasA, a member of the Cascade complex (Sashital et al. 2012). It is believed that the PAM requirement for immunity allows “tolerance to self”: it prevents the autoimmunity against the spacer DNA sequences, which are complementary to the crRNAs they encode. Absence of a PAM in the 5’ end of the target prevents spacer targeting. Similar to the alteration to the seed sequence, mutations in the PAM promote the escape of bacteriophages from CRISPR immunity (Semenova et al. 2011).

If both conditions are met, i.e. base-pairing within the seed and the presence of a PAM, recruitment of the Cas3 ssDNA nuclease cleaves and degrades the displaced DNA strand within the target sequence (Westra et al. 2012) and degrades it with 3’-5’ processivity (Gong et al. 2014; Huo et al. 2014).

1.2.2 Type II CRISPR-Cas systems

Type II CRISPR-Cas systems are defined by the presence of the Cas9 RNA-guided endonuclease and are the simplest of all CRISPR types (**Figure 1-4C**). Processing of the precursor crRNA does not require a dedicated, repeat-specific endoribonuclease. Instead, type II CRISPR loci produce a small trans-encoded crRNA (tracrRNA) with a region of complementarity to the repeat sequence (Deltcheva et al. 2011). The tracrRNA is bound by the Cas9 nuclease and forms a dsRNA interaction with the repeat sequences of the precursor crRNA that is

cleaved by RNase III. The result of this reaction is the generation of a Cas9 enzyme loaded with both the tracrRNA and the crRNA guide, a ribonucleoprotein complex that is necessary and sufficient for CRISPR immunity (Sapranauskas et al. 2011).

Similar to type I, type II CRISPR immunity requires six to eight nucleotides of seed sequence (Deveau et al. 2008; Jiang et al. 2013a) as well as a PAM (Deveau et al. 2008), which for the *Streptococcus pyogenes* SF370 type II-A CRISPR-Cas system is NGG (Mojica et al. 2009; Jiang et al. 2013a). As opposed to type I, however, the seed and PAM sequences are located at the 3' end of the target. As a consequence of these requirements, viruses can escape CRISPR immunity by mutating any of these sequences (Barrangou et al. 2007; Deveau et al. 2008). Upon infection, the tracrRNA/crRNA/Cas9 ribonucleoprotein complex scans the viral genome for the presence of PAM sequences. Binding of Cas9 to the PAM favors the unwinding of the target sequence immediately upstream of the motif, allowing the crRNA to probe for the presence of a matching sequence (Sternberg et al. 2014). A productive annealing results in the formation of a crRNA:target R-loop that triggers cleavage by Cas9. The enzyme harbors two nuclease domains (HNH and RuvC (Sapranauskas et al. 2011; Jinek et al. 2014; Nishimasu et al. 2014)), each of which cuts one target DNA strand. The tracrRNA is a co-factor of Cas9 and it is also required for DNA cleavage (Gasiunas et al. 2012; Jinek et al. 2012).

1.2.3 Type III CRISPR-Cas systems

Type III CRISPR-Cas systems are possibly the most complex of all CRISPR types. These loci are defined by the presence of genes encoding Cas10 and the repeat-associated mysterious protein (RAMP) modules Csm or Cmr for type III-A or III-B, respectively (Makarova et al. 2011), which together form the Cas10-Csm or Cas10-Cmr complexes (Hale et al. 2009; Zhang et al. 2012; Hatoum-Aslan et al. 2013). As in type I systems, crRNA precursor processing is achieved by the Cas6 repeat-specific endoribonuclease (Carte et al. 2008) (**Figure 1-4D**). In contrast to type I systems, however, Cas6 is not part of the Cas10 complex (Hale et al. 2009; Zhang et al. 2012; Hatoum-Aslan et al. 2013), and recent experiments suggested that repeat sequences are important to incorporate the Cas6 products into the complex prior to targeting (Sokolowski et al. 2014). Also in contrast to type I, the crRNAs generated by Cas6 cleavage are further trimmed at the 3' end in a process known as crRNA maturation. Type III crRNA maturation eliminates the 3'-end repeat sequences that remain after Cas6 cleavage and generates a heterogeneous population of mature crRNA guides which differ by increments of six nucleotides at the maturation end (Hale et al. 2008; Hatoum-Aslan et al. 2011).

As opposed to the other CRISPR types, a seed sequence has not been identified. Moreover, multiple mutations within the target do not abrogate immunity (Manica et al. 2013). No PAM requirements have been reported for type III targeting either. Instead, the lack of base pairing between the target 5' flanking sequences and the crRNA tag is essential for type III CRISPR immunity

(Marraffini 2010). Since there is perfect homology between the upstream repeat and the crRNA tag (this is transcribed from the repeat DNA), this property enables the discrimination between bona fide targets and the CRISPR array itself to avoid autoimmunity.

Two unique features of type III CRISPR-Cas systems are (i) that transcription across the target is required for immunity (Deng et al. 2013; Goldberg et al. 2014), and (ii) both DNA (Marraffini and Sontheimer 2008; Manica et al. 2011; Goldberg et al. 2014; Samai et al. 2015) and RNA (Hale et al. 2009; Zhang et al. 2012; Staals et al. 2014; Tamulaitis et al. 2014; Zebec et al. 2014; Samai et al. 2015) targets are cleaved. In addition, only crRNA guides complementary to the coding, but not the template, DNA strand provide effective immunity (Goldberg et al. 2014). Because these guides are also complementary to, and cleave, the target transcript, type III CRISPR-Cas immunity leads to the co-transcriptional destruction of both the target DNA and its transcripts (Peng et al. 2015; Samai et al. 2015). Whereas DNA cleavage is essential for immunity, the function of RNA targeting is unknown, and has been one of the focuses of my thesis.

The transcription requirement allows tolerance of foreign elements that can silence the region targeted by type III CRISPR-Cas systems. For example, this permits discrimination between the lytic and lysogenic cycles of temperate phages. Temperate phages with targets for type III CRISPR immunity within the lytic genes, which are expressed only during the lytic cycle but not during the prophage stage, can be maintained as prophages inside the cell even if they

contain a target that perfectly matches a crRNA (Goldberg et al. 2014). Induction of the lytic cycle, however, re-initiates transcription and targeting, and thus type III immunity can prevent the lethal reactivation of prophages.

CHAPTER 2 APPLICATION OF TYPE II CRISPR-CAS SYSTEM IN BACTERIAL GENOME EDITING AND GENE REGULATION

In this chapter, I will introduce existing genome editing and gene regulation techniques in bacteria and some drawbacks associated with them. I will describe the development of type II CRISPR-Cas as a novel sequence-specific counter-selection tool, and its application in bacterial genome editing and gene regulation. Most of this work (Bikard et al. 2013; Jiang et al. 2013a) was done in close collaboration with Dr. David Bikard, a previous post-doc in our lab, who's also a great mentor and friend of mine.

Since our initial proof-of-principle demonstration of CRISPR-mediated genome editing in *E.coli* and *Streptococcus pneumoniae* (Jiang et al. 2013a), the approach has been successfully applied for genetic manipulation in many more bacterial species, including *Lactobacillus* (Oh and van Pijkeren 2014), *Clostridium* (Wang et al. 2015; Xu et al. 2015), *Streptomyces* (Cobb et al. 2015; Huang et al. 2015), *Actinomycetes* (Tong et al. 2015), *Staphylococcus* (Chapter 2.3.6), *Salmonella* (in collaboration with Zhenrun J. Zhang, unpublished data) as well as bacteriophages (Kiro et al. 2014; Martel and Moineau 2014). CRISPR-mediated gene regulation has been demonstrated in *E.coli* (Bikard et al. 2013; Qi et al. 2013), *Mycobacterium* (Choudhary et al. 2015), *Bacteroides* (Mimee et al. 2015) and *Actinomycetes* (Tong et al. 2015). The easy programmability of the CRISPR-Cas system greatly facilitated genome editing and modulation of gene expression, and is likely to substantially advance our capability to both decipher gene function in prokaryotes and manipulate them for biotechnological purposes.

2.1 Existing genome editing technology in bacteria

Reverse genetics is a powerful approach for identification and understanding of gene function. The general methodology consists of first mutation of the gene of interest and then analysis of the resulting phenotype. In bacteria, genome modification is typically achieved by using allelic replacement (Hamilton et al. 1989) or λ-Red recombinase-mediated recombination (Court et al. 2002).

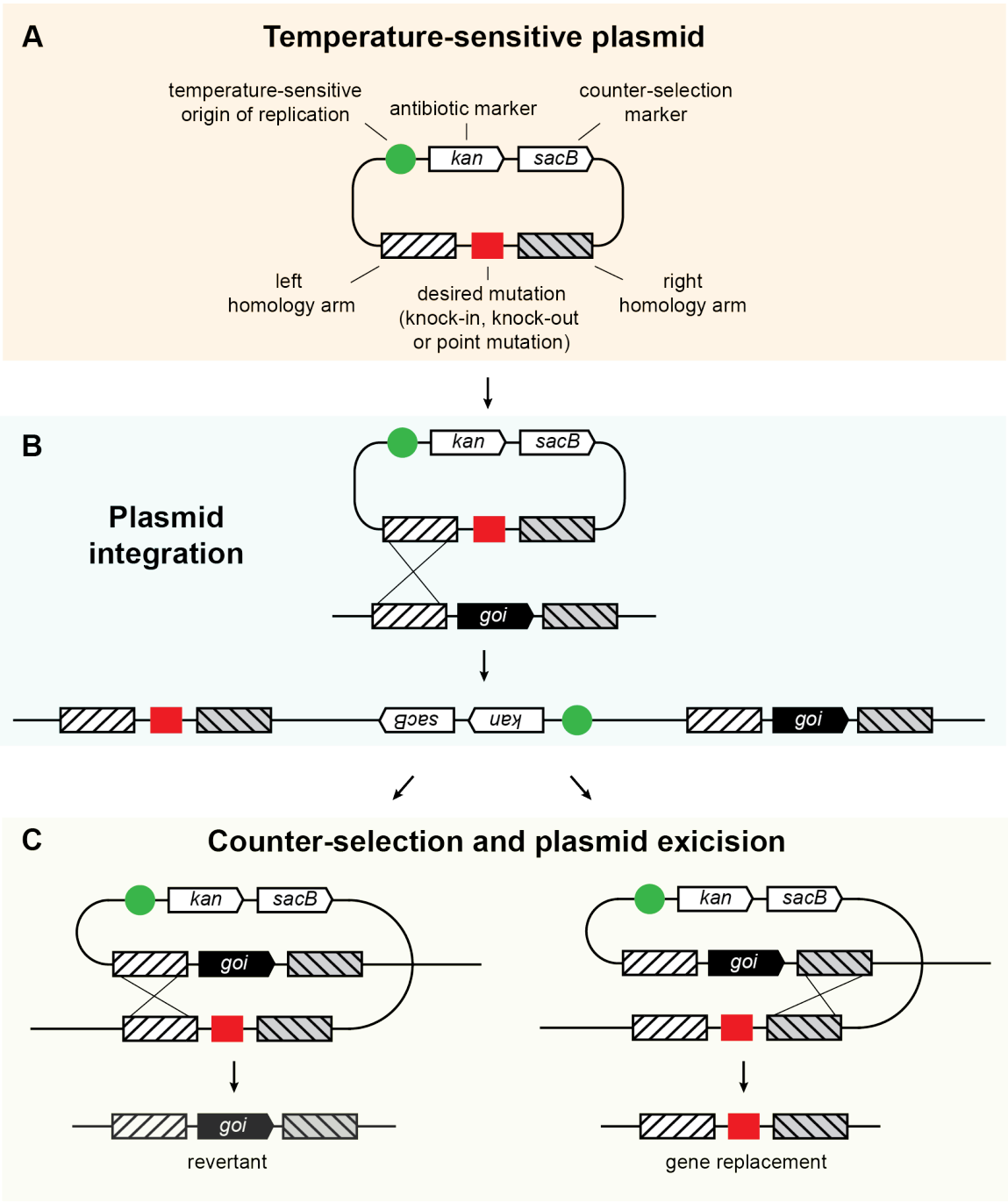
2.1.1 Allelic replacement

Allelic replacement (**Figure 2-1**) is a classical genetic engineering technique (Hamilton et al. 1989), and has been extensively used to generate knock-out, knock-in or point mutations in chromosomal genes. The method typically makes use of temperature-sensitive plasmids, on which a replacing gene fragment with homologous sequences flanking the chromosomal gene of interest is cloned. Upon introduction to a host such as *E.coli* and selection of appropriate antibiotic, the plasmid can replicate and be maintained at a permissive temperature (e.g., 30°C). When shifted to a nonpermissive temperature (e.g., 43°C), the plasmid can no longer be maintained episomally, and in the presence of the appropriate selective antibiotic, homologous recombination between the plasmid-borne homology arms and the chromosome results in plasmid integration (**Figure 2-1B** and **Figure 2-2A**). Since the integrant harbors two sets of homology arms, a second recombination event can happen. Because this is usually a low-frequency event, the plasmid typically carries a counter-selection marker such as the *sacB* gene derived from *Bacillus subtilis*. Expression of the gene in *E.coli* is lethal in the presence of sucrose (Blomfield et al. 1991), and leads to plasmid

excision. Depending on where the second recombination event takes place, the chromosome either undergoes a gene replacement or reverts back to the original genotype at a roughly 1:1 ratio (**Figure 2-1C**).

Figure 2-1 Genome editing through allelic replacement

(A) First, a temperature-sensitive plasmid harboring a desired mutation along with two homology arms are constructed. The plasmid also carries an antibiotic marker (e.g. *kan*) and a counter-selection marker (e.g. *sacB*). Upon introduction in host cells such as *E.coli* and selection using kanamycin, the plasmid can be maintained at a permissive temperature (e.g. 30°C). **(B)** When shifted to a nonpermissive temperature (e.g., 43°C), the plasmid is forced to integrate into the region of the host chromosome carrying the gene of interest (*goi*) through homologous recombination. This low-frequency event can be selected using kanamycin. **(C)** Using a counter-selection marker such as the *sacB* gene, a second recombination event resulting in gene replacement or reversion can take place at a roughly 1:1 ratio. Abbreviations: kan, kanamycin-resistant marker; goi, gene of interest.


A
Temperature-sensitive plasmid
temperature-sensitive origin of replication
antibiotic marker
counter-selection marker
kan
sacB
left homology arm
desired mutation (knock-in, knock-out or point mutation)
right homology arm
B
Plasmid integration
kan
sacB
goi
goi
C
Counter-selection and plasmid excision
kan
sacB
goi
goi
revertant
kan
sacB
goi
gene replacement

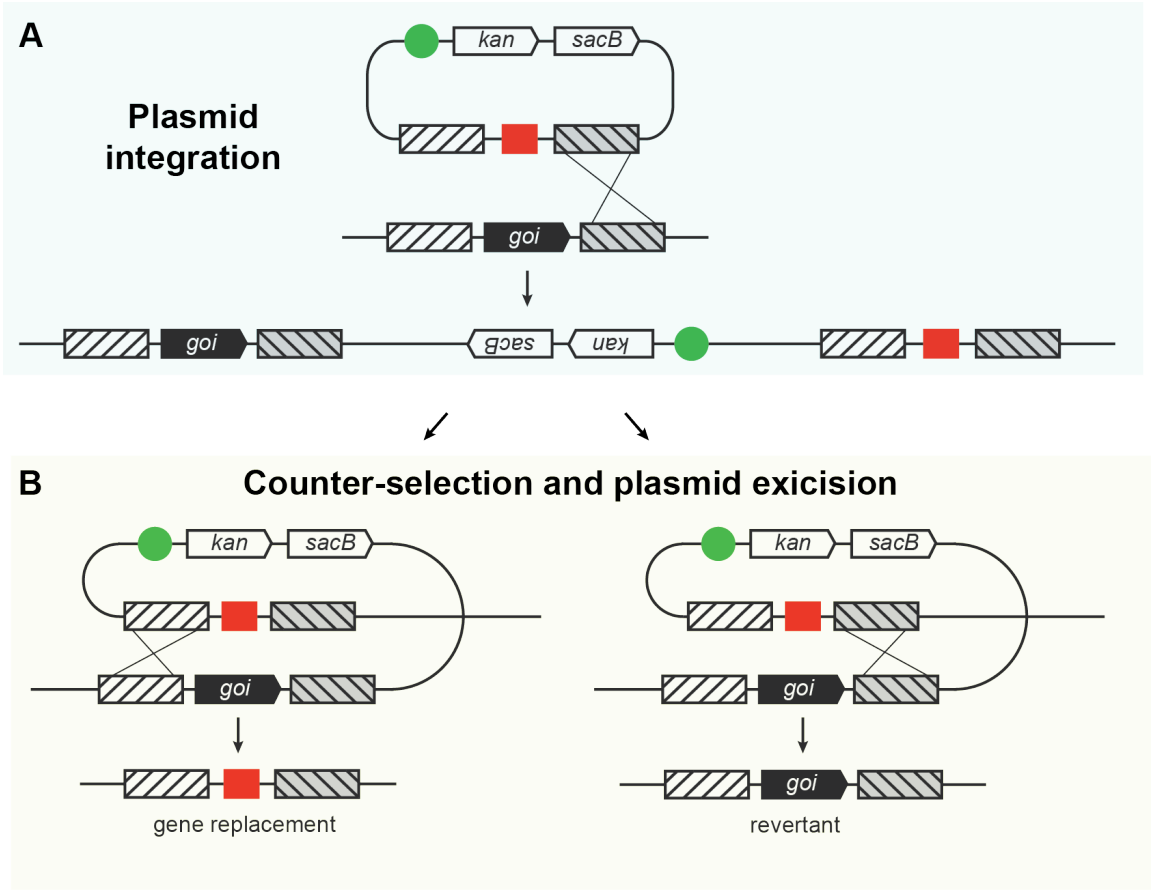


Figure 2-2 Genome editing through allelic replacement continued.

Same as **Figure 2-1** except that the initial plasmid integration step can also occur through recombination of the right homology arm, as opposed to the left homology arm shown in **Figure 2-1B**.

Allelic replacement suffers from some drawbacks. First, it is a two-step method that requires cloning, which could be lengthy and labor-intensive. Second, the success of the method depends crucially on counter-selection makers. Although the *sacB* sucrose sensitivity system remains the most effective bacterial counter-selection tool to date, expression of the gene is harmless in most gram-positive bacteria and thus cannot support counter-selection in these species (Reyrat et al. 1998).

2.1.2 λ-Red recombination

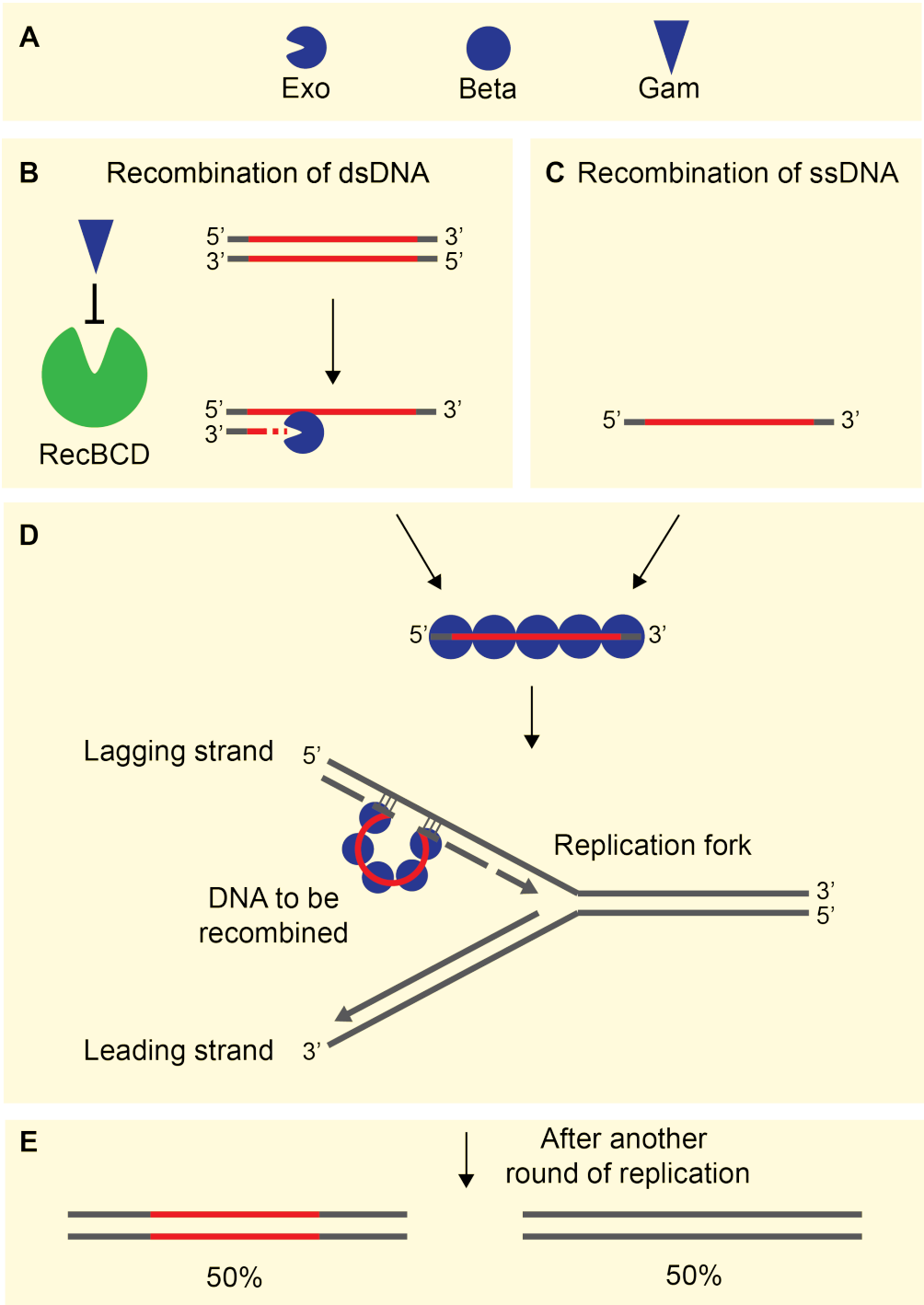
Bacterial recombineering, first developed in *E.coli* by Kenan Murphy in 1998 (Murphy 1998), is a genetic engineering technique that uses phage-encoded homologous recombination systems to recombine linear DNA fragments into bacterial chromosomes (Court et al. 2002). In contrast to classical techniques such as allelic replacement, recombineering is a single-step procedure that does not require laborious construction of plasmids containing homology arms (**Figure 2-1**). Instead, all that is required is synthesis of standard oligonucleotides (oligos) that provide the homology. These oligos can be used directly or for the making of PCR products that are used for recombineering.

Recombination systems encoded by either *redαβδ* from phage λ (Murphy 1998; Yu et al. 2000) or *recE*/*recT* from Rac prophage (Zhang et al. 1998) can mediate efficient recombination of linear DNA fragments. Now widely used in *E.coli*, the λ-Red system is composed of three λ genes: *exo*, *bet* and *gam*, which encode the Exo, Beta and Gam proteins, respectively (**Figure 2-3A**). When a linear dsDNA fragment carrying short sequences of homology on each end is introduced to host cells, λ Exo, an exonuclease, degrades the dsDNA in a 5' to 3' direction (Little 1967; Carter and Radding 1971). As Exo is highly processive (Matsuura et al. 2001), the current model suggests that the protein binds one of the two dsDNA strands and degrades that strand completely (**Figure 2-3B**) (Mosberg et al. 2010; Pines et al. 2015). The ssDNA generated is then stably bound by λ Beta (**Figure 2-3D**). The function of λ Beta is to protect the DNA from single-strand nuclease attack and promote annealing of the exogenous DNA to

the complementary ssDNA sequences in the chromosome (Kmiec and Holloman 1981; Muniyappa and Radding 1986; Karakousis et al. 1998). Specifically, the recombination of the ssDNA intermediate preferentially occurs at the lagging strand of the replication fork during DNA replication (**Figure 2-3D**) (Ellis et al. 2001). The third protein, λ Gam, confers the full recombination potential to the system. Gam inhibits RecBCD (Karu et al. 1975; Murphy 1991), host nucleases that would otherwise rapidly degrade cellular linear dsDNA (**Figure 2-3B**). ssDNAs on the other hand, are not substrates for RecBCD. Therefore, ssDNA fragments can be recombined to the host genome with the presence of only λ Beta (**Figure 2-3C**) (Ellis et al. 2001). Recombineering with ssDNA is more efficient than with dsDNA and is the method of choice to create point mutations.

Figure 2-3 Genome editing through λ-Red recombination

(A) The λ-Red recombination system comprises of the Exo, Beta and Gam proteins from phage λ. Both linear ssDNAs and dsDNAs can be used as substrates for recombination. **(B)** When a linear dsDNA fragment is introduced to host cells, λ Exo degrades the dsDNA in a 5' to 3' direction and generates a ssDNA. In the meantime, λ Gam protects the dsDNA from degradation by inhibiting host nucleases RecBCD. Red segments on DNAs denote new sequences to be incorporated and gray segments denote short sequences homologous to host chromosomes. **(C)** Recombination of ssDNA fragments does not require λ Exo and Gam. **(D)** λ Beta binds the ssDNA fragments, protects them from single-strand nuclease attack and promotes annealing of the DNA preferentially to the lagging strand of the replication fork. **(E)** Among recombined cells, another round of cellular replication is required to complete the synthesis of the newly introduced modification on the opposite strand. This results in 50% of the recombined cells inheriting the change. Abbreviations: dsDNA, double-stranded DNA; ssDNA, single-stranded DNA.


A
Exo
Beta
Gam
B
Recombination of dsDNA
5'
3'
3'
5'
5'
3'
3'
RecBCD
C
Recombination of ssDNA
5'
3'
D
5'
3'
Lagging strand
5'
Replication fork
DNA to be
recombined
3'
5'
Leading strand
3'
E
After another
round of replication
50%
50%

The single-step λ-Red recombination system integrates linear DNA fragments into the host chromosomes, which circumvents the need for the extensive in-vitro cloning that conventional genetic engineering techniques entail. Although the system substantially simplifies genome editing, currently it has only been shown to work in a limited number of species. These species include *E.coli* (Datsenko and Wanner 2000), *Salmonella* (Karlinsey 2007), *Streptomyces* (Gust et al. 2004) and *Mycobacterium* (van Kessel and Hatfull 2008). Additionally, when the intention is to create point mutations, the recombineering technique suffers from a big downside – lack of selection. Typically recombinants are found in only 0.01% of the treated cells if dsDNA is used or 0.1% if cells are treated with ssDNA (http://redrecombineering.ncifcrf.gov/) (Yu et al. 2000; Ellis et al. 2001), and thus screening of a large number of colonies is required.

2.2 Existing gene modulation tools in bacteria

Unlike eukaryotic cells that can be modulated by RNAi (Hannon and Rossi 2004), methods for gene regulation are limited in bacteria. Two existing tools for gene repression and activation are small antisense RNAs (sRNAs) (Storz et al. 2011) and DNA-binding transcription activators (Ptashne and Gann 1997).

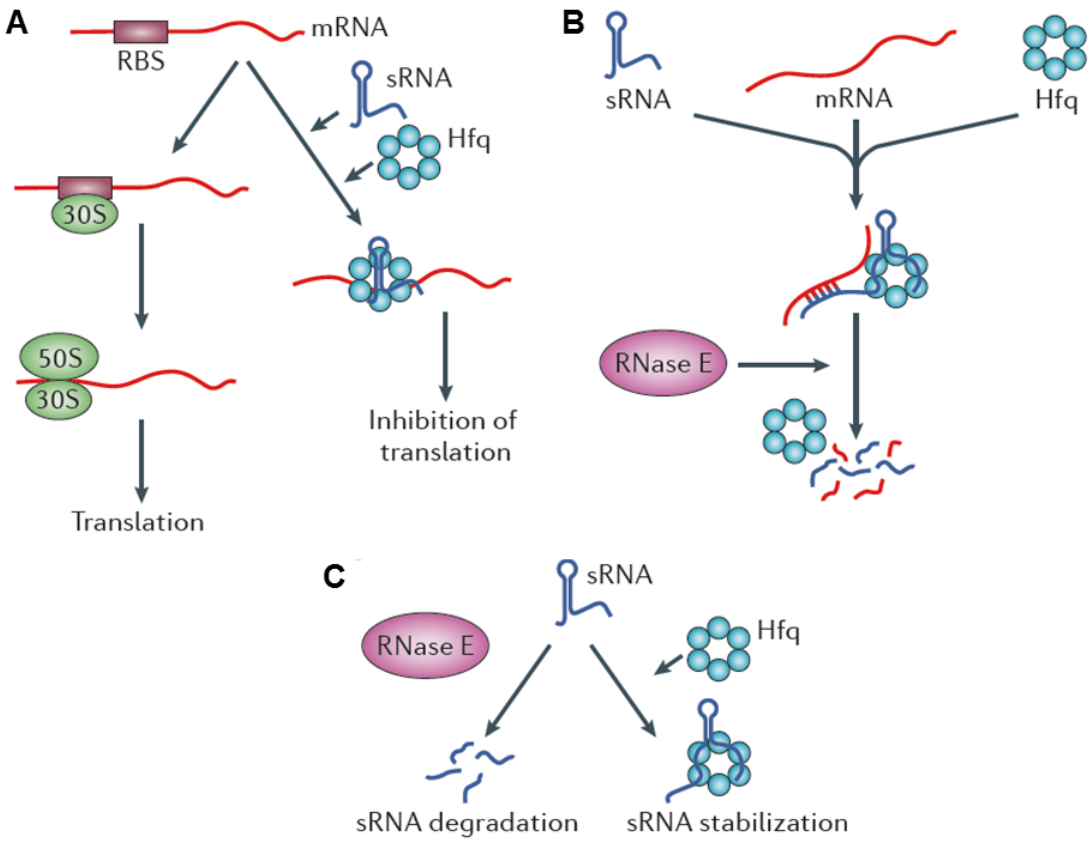
2.2.1 Small-RNA-mediated gene repression in bacteria

RNA elements are important regulators for gene expression in all domains of life. In bacteria, one such regulator comprises riboswitches, which are cis-elements that modulate the translation of the mRNAs harboring them (Waters and Storz 2009). Another class of regulators naturally found in bacteria is small

RNAs (sRNA). Typically 50-300 nucleotides long and often trans-encoded, these sRNAs usually negatively control translation and stability of target mRNAs by base-pairing (Waters and Storz 2009; Storz et al. 2011). Inhibition of translation is achieved by base-pairing at or near ribosome binding sites (RBS) and occlusion of ribosomes (**Figure 2-4A**). Base-pairing between the sRNAs and the coding region of mRNA recruits RNase E that degrades the target mRNA, leading to downregulation of the gene (**Figure 2-4B**). In many cases, the RNA chaperone Hfq is required for the function and/or stability of the trans-encoded sRNAs (**Figure 2-4A**, **B** and **C**) (Vogel and Luisi 2011).

Figure 2-4 Small RNAs mediate gene repression

(A) Small RNA (sRNA) in association with the RNA chaperone, Hfq can repress gene expression by base-pairing at or near ribosome binding sites (RBS) of its target mRNA, which occludes ribosomes from binding and initiating translation. **(B)** sRNA/Hfq complex base-pairs with the coding region of mRNA and recruits RNase E, which degrades the target mRNA and leads to downregulation of the gene. **(C)** Hfq may protect some sRNAs from degradation by ribonucleases such as RNase E. This figure is adapted from (Vogel and Luisi 2011).



Though the programmable nature of trans-encoded-sRNA offers a promising possibility to repress gene expression at will, the efficacy of different sRNAs in repression has been variable (Man et al. 2011; Sharma et al. 2012). Conceivably, the efficacy can be influenced by factors such as the sequence of sRNAs, their interaction with Hfq and thus stability in vivo, as well as sequence of target mRNAs and their secondary structures. Since no robust design rule currently exists, identification of functional sRNAs often requires screening of a library of sRNAs (Man et al. 2011; Sharma et al. 2012). This process is laborious and severely limits the application of sRNAs as a tool for gene repression.

2.2.2 Transcription activation in bacteria

In both prokaryotes and eukaryotes, transcription activation is generally achieved by DNA-binding activators (Ptashne and Gann 1997). Activators are proteins composed of a DNA-binding domain that recognizes and binds specific sites on DNA, and an activation domain that recruits transcriptional machinery to the DNA (**Figure 2-5A**). Early work in yeast showed that fusing a DNA-binding domain to a subunit of the RNA polymerase II holoenzyme activates transcription from a promoter bearing sequences recognizable by the DNA-binding domain (Barberis et al. 1995; Farrell et al. 1996; Gaudreau et al. 1997).

In *E.coli*, the RNAP core enzyme consists of subunits of α, β and β' in the stoichiometry $\alpha_2\beta\beta'$ (Burgess 1969) and one of several alternative σ subunits that direct the enzyme to specific promoters (Helmann and Chamberlin 1988). An additional protein of unknown function, ω, has been categorized as a subunit of RNAP based on its co-purification with RNAP (Burgess 1969). On the other hand, the cI protein from phage λ is a DNA-binding protein that recognizes the operator sequence located on λ DNA (Li et al. 1994). Early studies showed that by tethering λcI to α or ω subunit of RNAP, the chimeric protein effectively mediates transcription activation of genes bearing the λ operator (**Figure 2-5B**) (Dove et al. 1997; Dove and Hochschild 1998). However, one big downside of transcription activation is that the recognition sequence on the DNA is determined by the activator and fixed. Therefore it requires engineering of the promoter sequence of the gene before it can be activated. Overall, this makes the method cumbersome and time-consuming.

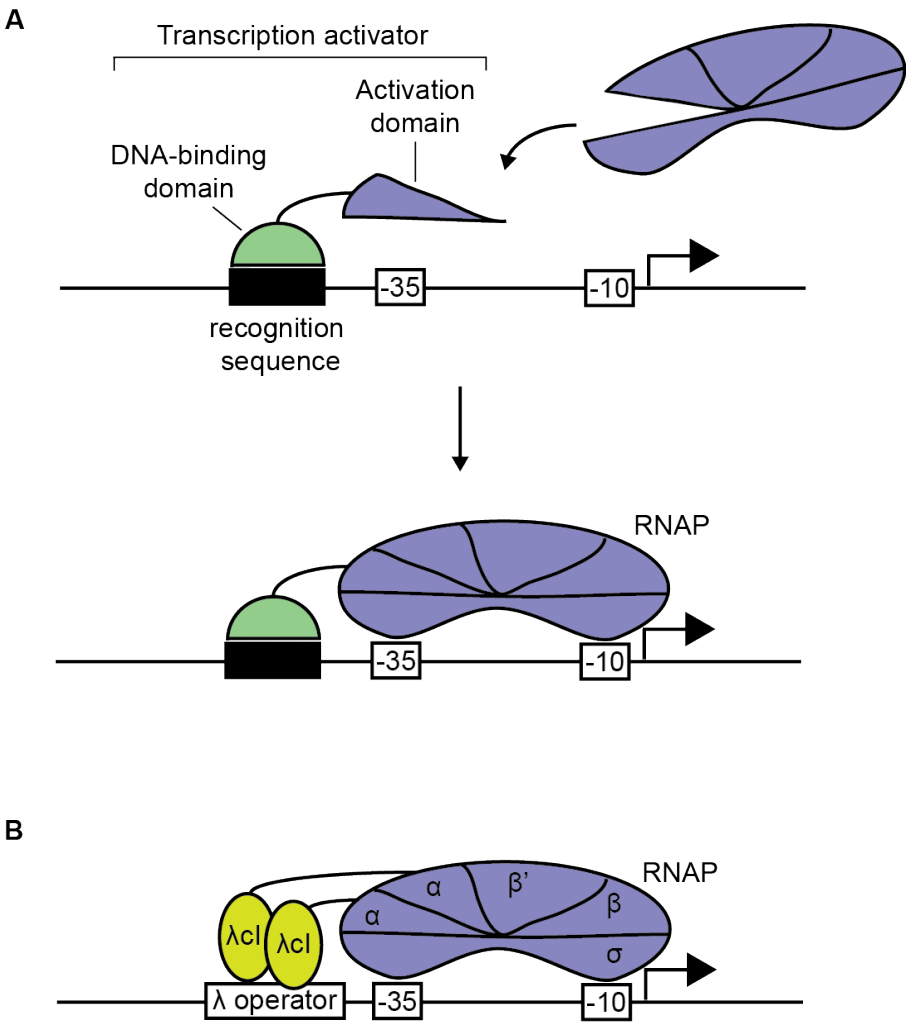


Figure 2-5 Transcription activation in bacteria

(A) A transcription activator comprises of a DNA-binding domain that recognizes and binds specific sites on DNA, and an activation domain, in many cases a subunit of the RNA polymerase (RNAP) that recruits RNAP to the DNA. **(B)** In *E.coli*, by tethering the cI protein of phage λ to the α subunit of RNAP, the transcription activator can recruit RNAP to genes bearing the λ operator, thus activating transcription.

2.3 CRISPR-Cas-mediated counter-selection and genome editing

As introduced in Chapter 2.1, current bacterial genome editing methods such as allelic replacement and λ-Red recombination suffer from a common

disadvantage – the limited utility or complete lack of counter-selection markers. Therefore there is a need to develop new and easy counter-selection tools that can function in a wide range of bacterial species.

To date, all three types of CRISPR-Cas systems are reported to cleave DNA. Importantly, our lab discovered that when programmed to target the chromosome of *Streptococcus pneumoniae*, CRISPR-Cas causes autoimmunity and kills the cells (Bikard et al. 2012). The lethal consequence upon targeting, presumably due to inability of the cellular machinery to repair the cleaved DNA, opened up a possibility to adopt CRISPR-Cas as a sequence-specific counter-selection tool in bacteria, and potentially facilitate genome editing when combined with existing methods.

Not all types of CRISPR-Cas systems are suitable for repurposing as genetic engineering tools. For example, the DNA cleavage activity of type I and type III crRNA:Cas ribonucleoprotein complexes is not well characterized and the biogenesis of crRNA guides requires a specialized Cas endoribonuclease, Cas6. In addition, these complexes are composed of multiple protein subunits (**Figure 1-4B** and **D**), increasing the engineering required to make them work in heterologous hosts. In contrast, type II CRISPR-Cas systems present a simpler mechanism which relies on the crRNA-guided nuclease Cas9, its tracrRNA cofactor, and the house-keeping RNase III (Chapter 1.2.2). To further simplify the system, the crRNA and tracrRNA can be fused to form a single guide RNA (sgRNA) (Jinek et al. 2012). Creation of sgRNA obviates the need for RNase III,

which is required for canonical crRNA processing canonically (Deltcheva et al. 2011). In the work presented below, we used the canonical tripartite system consisting of the Cas9 nuclease, its tracrRNA cofactor and its crRNA guide that specify the target. In all studied bacterial species (including *E.coli*, *Streptococcus pneumoniae*, *Staphylococcus aureus* and *Salmonella*) we worked with, heterologous expression of the tripartite system had programmable DNA targeting activity, suggesting that host RNase III or homologous enzymes in those species can support crRNA processing and thus CRISPR functionality. It remains possible that host RNase III in other species might not support crRNA processing, in which case the sgRNA-mediated version of engineering would be needed in those species.

2.3.1 CRISPR-Cas as a sequence-specific counter-selection tool

In order to test the feasibility of adapting the type II CRISPR-Cas as a counter-selection tool, we chose to work with *S. pneumoniae* because the organism is naturally competent for DNA uptake and the genome can be easily modified by linear PCR products provided exogenously (Claverys et al. 2009). Using the parental strain, R6, we first constructed strain crR6, in which the non-essential gene *IS1167* is replaced by the CRISPR01 locus of *Streptococcus pyogenes* SF370 along with a kanamycin-resistant marker, *aphA-3* (**Figure 2-6A**). This CRISPR-Cas locus encodes tracrRNA, Cas9 and a CRISPR array with six spacers. The locus also contains *cas1*, *cas2* and *csn2*, genes that are involved in the adaptation phase (Heler et al. 2014) but are dispensable for targeting. Notably, the SF370 Cas9 was later widely repurposed as a robust genome

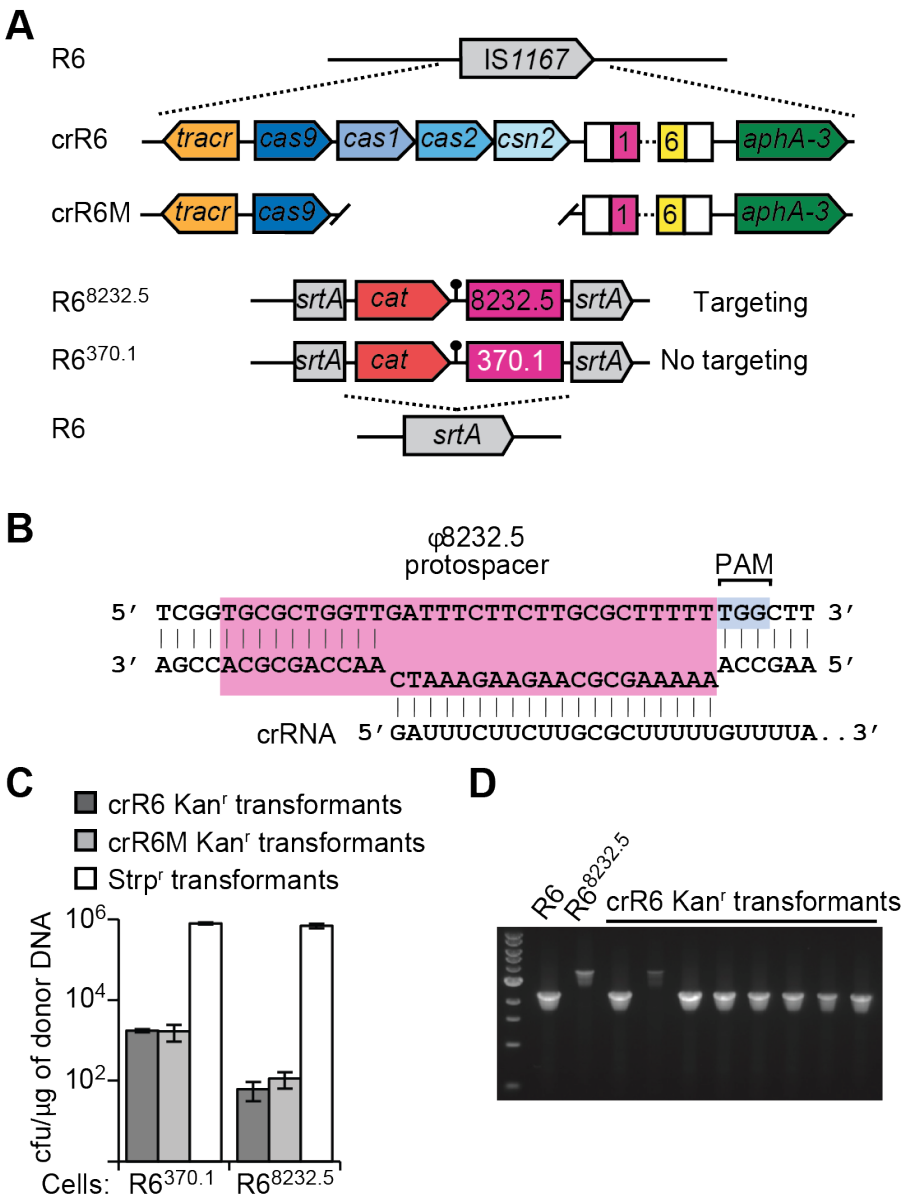
editing tool in both bacteria (Jiang and Marraffini 2015) and eukaryotic cells (Hsu et al. 2014). We also constructed two other strains, R6$^{8232.5}$, and R6$^{370.1}$, in which the non-essential *srtA* gene in R6 is disrupted by a chloramphenicol-resistant marker, *cat*, along with a 0.4 kB sequence derived from streptococcal bacteriophages φ8232.5 or φ370.1, respectively (**Figure 2-6A**). Crucially, the first spacer of the CRISPR array matches perfectly to a region on the 0.4 kB sequence cloned from both phages (**Figure 2-6B**). However, CRISPR can only target φ8232.5 but not φ370.1 as the protospacer on φ370.1 lacks a functional PAM, which is required for Cas9 targeting (see Chapter 1.2.2).

Next, we transformed R6$^{8232.5}$, and R6$^{370.1}$ cells with genomic DNA (gDNA) extracted from crR6 cells. If targeting of the chromosome by CRISPR leads to cell death, we should expect to recover only R6$^{370.1}$ transformants. Contrary to this expectation, we isolated R6$^{8232.5}$ transformants, albeit with approximately 10-fold less efficiency than R6$^{370.1}$ transformants (**Figure 2-6C**). Transformation of genomic DNA from crR6M, a strain lacking *cas1*, *cas2* and *csn2* yielded similar result (**Figure 2-6C**), consistent with the notion that these genes are dispensable for targeting. Genetic analysis of eight R6$^{8232.5}$ transformants revealed that the great majority of them had undergone a double recombination event (**Figure 2-6D**). Since the crR6 gDNA contains the wild-type *srtA* locus, it can recombine and replace the φ8232.5 target that would be otherwise recognized by Cas9 and lead to cell death. This experiment provides the initial proof that the concurrent introduction of a targeting construct (i.e. a type II CRISPR-Cas system composed of tracrRNA, crRNA and Cas9 (hereafter referred to as the Cas9 complex) that

targets a genomic locus) together with an editing template for recombination into the targeted locus, can lead to targeted genome editing.

Figure 2-6 CRISPR-Cas as a counter-selection tool: initial proof of principle in *Streptococcus pneumoniae*

(A) Using *Streptococcus pneumoniae* R6 as a parental strain, crR6 was constructed by replacing the non-essential *IS1167* by the CRISPR01 locus of *Streptococcus pyogenes* SF370 along with a kanamycin-resistant marker, *aphA-3*. crR6M is a minimal type II CRISPR-Cas system that had *cas1*, *cas2* and *csn2* deleted from crR6. Using R6 as a parental strain, $R6^{8232.5}$ and $R6^{370.1}$ were constructed by integrating a 0.4 kB region derived from the streptococcal bacteriophages φ8232.5 and φ370.1 in the *srtA* gene, respectively. This region was also fused to a chloramphenicol-resistant marker, *cat*. **(B)** The φ8232.5 protospacer matches perfectly to the crRNA derived from *spc1* of the CRISPR locus. The PAM is TGG. **(C)** Transformation of donor crR6 and crR6M genomic DNA in recipient cells, $R6^{8232.5}$ and $R6^{370.1}$. As a control of cell competence a streptomycin-resistant gene was also transformed. Transformation efficiency was measured as the number of colony-forming units (CFU) per μg of donor DNA (mean ± SD of three replicas). **(D)** PCR analysis of eight $R6^{8232.5}$ transformants with crR6 genomic DNA. Primers that amplify the *srtA* locus were used for PCR. 7/8 genotyped colonies replaced the $R6^{8232.5}$ *srtA* locus (~3.5 kB) by the wild-type locus (~2.2 kB) from the crR6 genomic DNA. Abbreviations: PAM, protospacer adjacent motif; Kan, kanamycin; Strp, streptomycin.

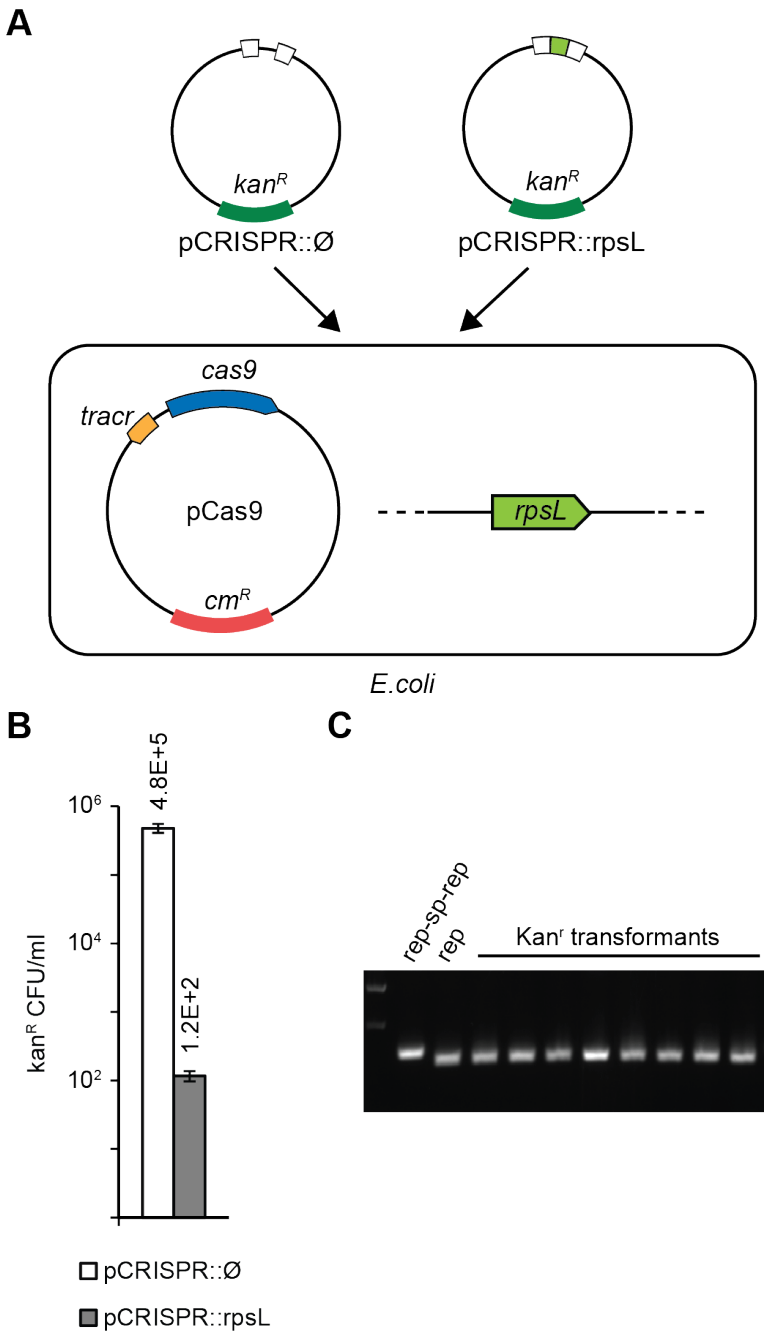


CRISPR-mediated killing also works in *E.coli* (**Figure 2-7A**). We constructed a plasmid, pCas9, which contained *tracr* and *cas9*. We also cloned pCRISPR::Ø and pCRISPR::rpsL, which were plasmids harboring either an empty spacer, or a spacer that matched the *rpsL* gene on the genome, respectively. When these two plasmids were transformed into *E.coli* cells that already carried pCas9, the number of pCRISPR::Ø transformants was over three orders of magnitude more

than that of pCRISPR::rpsL (**Figure 2-7B**). In addition, genetic analysis of eight pCRISPR::rpsL transformants revealed that all had deleted the spacer (**Figure 2-7C**). These data suggests that CRISPR targeting of host chromosome leads to cell death, a property that can be exploited as a powerful tool for counter-selection.

Figure 2-7 CRISPR-Cas as a counter-selection tool in *E.coli*

(A) pCRISPR::Ø and pCRISPR::rpsL are kanamycin-resistant plasmids that carried either an empty spacer or a spacer that matched the *rpsL* gene on the *E.coli* genome. These plasmids were transformed into an *E.coli* strain HME63 that already harbored pCas9, which contained *tracr* and *cas9*. **(B)** Transformation of the pCRISPR::Ø or pCRISPR::rpsL plasmids into HME63 competent cells. Transformation efficiency was measured as the number of colony-forming units (CFU) per mL of competent cells (mean ± SD of three replicas). **(C)** PCR analysis of eight pCRISPR::rpsL transformants using primers that amplified the CRISPR array.



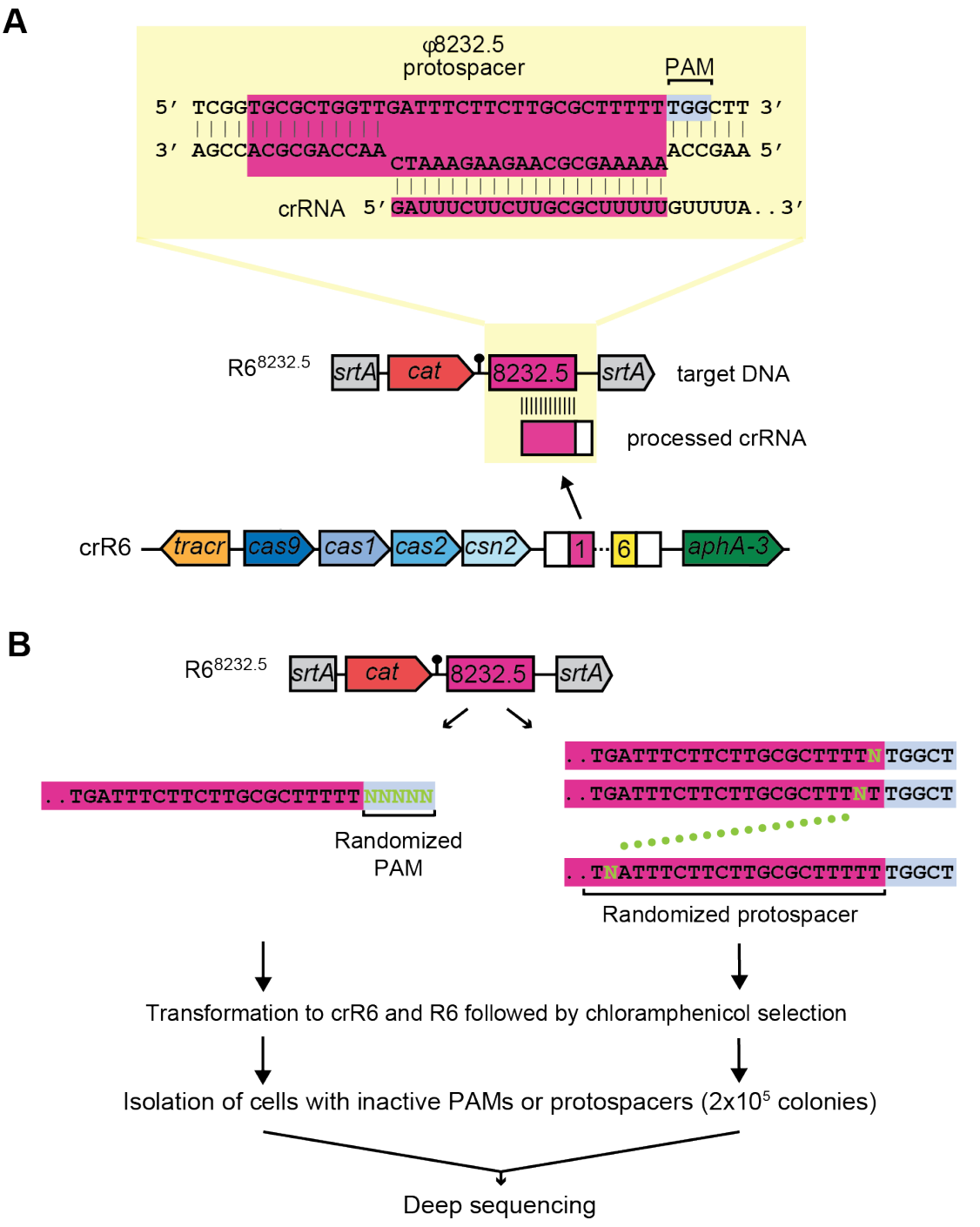
2.3.2 Analysis of targeting requirement by Cas9 complex

To use the Cas9 complex as a counter-selection tool for genome editing, one must design an editing template carrying mutations that abolish cleavage by it, thereby preventing cell death. This is easy to achieve when the target to be

recognized by CRISPR is deleted or replaced by another sequence. When the goal is to generate gene fusions or single-nucleotide mutations, the prevention of cleavage by Cas9 is possible only by introducing mutations in the editing template that alter either the PAM or the protospacer sequences. To determine the constraints imposed by these sequences, we performed a thorough analysis of PAM and protospacer mutations that abrogate targeting by Cas9 complex.

Figure 2-8 Analysis of targeting requirement by Cas9 complex

(A) The crRNA derived from *spc1* of the crR6 CRISPR locus targets the φ8232.5 protospacer containing a perfect match and a functional PAM, TGG. **(B)** Genomic DNA from $R6^{8232.5}$ was used as a template to obtain a randomized PAM or protospacer library. This was achieved by PCR with primers containing these randomized sequences. The PCR products were then transformed into crR6 or R6 cells followed by chloramphenicol selection. More than 2×10^5 chloramphenicol-resistant transformants, carrying inactive PAM or protospacer sequences, were combined for amplification and deep sequencing of the target region. In **(B)**, single-stranded DNAs were used for illustration but in real experiments double-stranded PCR products were used.



Previous studies proposed that *S. pyogenes* Cas9 requires an NGG PAM immediately downstream of the protospacer (Deltcheva et al. 2011; Gasiunas et al. 2012; Jinek et al. 2012). However, because only a very limited number of

PAM-inactivating mutations have been described thus far (Deltcheva et al. 2011; Bikard et al. 2012; Gasiunas et al. 2012; Jinek et al. 2012), we conducted a systematic analysis to find all 5-nucleotide sequences following the protospacer that eliminate Cas9 cleavage (**Figure 2-8**). We used randomized oligonucleotides to generate all possible 1,024 5-nucleotide PAM sequences in a heterogeneous PCR product that was transformed into crR6 or R6 cells. Constructs carrying functional PAMs are expected to be recognized and destroyed by crR6 but not R6 cells. More than 2×10^5 colonies were pooled to extract DNA for use as templates for the co-amplification of all targets. PCR products were deep sequenced and found to contain all 1,024 sequences, with coverage ranging from 5 to 42,472 reads (see supplementary materials of (Jiang et al. 2013a)). The functionality of each PAM was estimated by the relative proportion of its reads in the crR6 sample over the R6 sample. Analysis of the first three bases of the PAM clearly shows that the NGG pattern is under-represented in crR6 transformants (**Figure 2-9A**). Furthermore, the next two bases have no detectable effect on the NGG PAM (see supplementary materials of (Jiang et al. 2013a)), demonstrating that the NGGNN sequence is sufficient to license Cas9 activity. Partial targeting was observed for NAG PAM sequences (**Figure 2-9A**). Also the NNGGN pattern partially inactivates CRISPR targeting (**Figure 2-9B**, orange and blue sequences), indicating that the NGG motif can still be recognized by the Cas9 complex with reduced efficiency when shifted by 1 bp. These data shed light on the molecular mechanism of Cas9 target recognition, and they revealed that NGG (or CCN on the complementary strand)

sequences were sufficient for targeting. As an editing strategy, mutation of the two “G”s can abolish targeting, though NGG to NAG or NNGGN mutations should be avoided. In theory, a given sequence of any di-nucleotide (i.e. the GG in the PAM) occurs once every 8 bp in a stretch of perfectly random sequences. This means that almost any position of the genome can be targeted by the Cas9 complex.

A

1st PAM position	2nd PAM position: A		2nd PAM position: C		2nd PAM position: G		2nd PAM position: T		3rd PAM position
A	AAA	1.02	ACA	1.06	AGA	0.95	ATA	1.02	A
C	CAA	1.03	CCA	1.03	CGA	0.97	CTA	1.00	
G	GAA	1.10	GCA	1.09	GGA	0.93	GTA	1.01	
T	TAA	1.02	TCA	1.11	TGA	0.91	TTA	1.11	
A	AAC	1.07	ACC	1.02	AGC	1.00	ATC	1.09	C
C	CAC	1.06	CCC	1.04	CGC	1.00	CTC	1.09	
G	GAC	1.05	GCC	1.02	GGC	1.02	GTC	1.22	
T	TAC	1.06	TCC	1.05	TGC	1.03	TTC	1.06	
A	AAG	0.49	ACG	0.97	AGG	0.06	ATG	0.96	G
C	CAG	0.78	CCG	1.08	CGG	0.06	CTG	1.03	
G	GAG	0.50	GCG	1.00	GGG	0.06	GTG	1.20	
T	TAG	0.86	TCG	0.91	TGG	0.07	TTG	1.06	
A	AAT	1.01	ACT	1.09	AGT	1.04	ATT	1.03	T
C	CAT	1.03	CCT	1.03	CGT	1.07	CTT	1.08	
G	GAT	1.04	GCT	1.08	GGT	1.08	GTT	1.12	
T	TAT	1.06	TCT	1.03	TGT	1.06	TTT	1.01	

B

2nd PAM position	3rd PAM position: A		3rd PAM position: C		3rd PAM position: G		3rd PAM position: T		4th PAM position
A	AAA	1.04	ACA	1.12	AGA	0.73	ATA	1.10	A
	AAC	1.07	ACC	1.04	AGC	0.64	ATC	0.97	C
	AAG	1.00	ACG	1.09	AGG	0.61	ATG	1.07	G
	AAT	0.98	ACT	1.02	AGT	0.65	ATT	1.01	T
C	CAA	1.05	CCA	1.05	CGA	0.99	CTA	1.07	A
	CAC	1.04	CCC	1.02	CGC	1.08	CTC	1.04	C
	CAG	1.08	CCG	1.08	CGG	0.81	CTG	1.05	G
	CAT	1.13	CCT	1.05	CGT	1.07	CTT	1.08	T
G	GAA	0.97	GCA	1.05	**GGA**	**0.08**	GTA	0.99	A
	GAC	0.92	GCC	1.00	**GGC**	**0.05**	GTC	1.15	C
	GAG	0.96	GCG	0.98	**GGG**	**0.07**	GTG	0.98	G
	GAT	0.98	GCT	0.99	**GGT**	**0.06**	GTT	1.05	T
T	TAA	1.08	TCA	1.16	TGA	1.05	TTA	1.14	A
	TAC	1.00	TCC	1.08	TGC	1.08	TTC	1.05	C
	TAG	1.02	TCG	1.11	TGG	0.77	TTG	1.01	G
	TAT	1.01	TCT	1.12	TGT	1.21	TTT	1.02	T

Figure 2-9 PAM analysis

(A) Relative proportion of number of reads after transformation of the random PAM constructs in crR6 cells (compared to number of reads in R6 transformants). The relative abundance of each 3-nucleotide PAM sequence is shown. Severely underrepresented sequences (NGG) are shown in red; partially underrepresented one, in orange (NAG). **(B)** Same as **(A)** except showing the relative abundance of the 2nd, 3rd and 4th PAM sequences. Partially underrepresented sequences are also shown in blue.

Another way to disrupt the Cas9-complex-mediated cleavage is to introduce mutations in the protospacer region of the editing template. It is known that point mutations within the “seed sequence” (the 8 to 10 protospacer nucleotides immediately adjacent to the PAM) can abolish cleavage by CRISPR nucleases (Semenova et al. 2011; Wiedenheft et al. 2011; Jinek et al. 2012). However, the exact length of this region is not known, and it is unclear whether mutations to any nucleotide in the seed can disrupt Cas9 target recognition. We followed the same deep sequencing approach described above to randomize the entire protospacer sequence involved in base pair contacts with the crRNA and to determine all sequences that disrupt targeting. Each position of the 20 matching nucleotides (Deltcheva et al. 2011) in the *spc1* target present in $R6^{8232.5}$ cells (**Figure 2-8A**) was randomized, and a library containing the resulting sequences was transformed into crR6 and R6 cells (**Figure 2-8B**). Consistent with the presence of a seed sequence, only mutations in the 12 nucleotides immediately upstream of the PAM abrogated cleavage by Cas9 (**Figure 2-10**). However, different mutations displayed markedly different effects. The distal (from the PAM) positions of the seed (12 to 7) tolerated most mutations and only one particular base substitution abrogated targeting. In contrast, mutations to any nucleotide in the proximal positions (6 to 1, except 3) prevented cleavage, although at different levels for each particular substitution. At position 3, only two substitutions affected cleavage activity with different strength. We conclude that, although seed sequence mutations can prevent targeting by the Cas9 complex, there are restrictions regarding the nucleotide changes that can be made in each position

of the seed. Moreover, these restrictions can most likely vary for different spacer sequences.

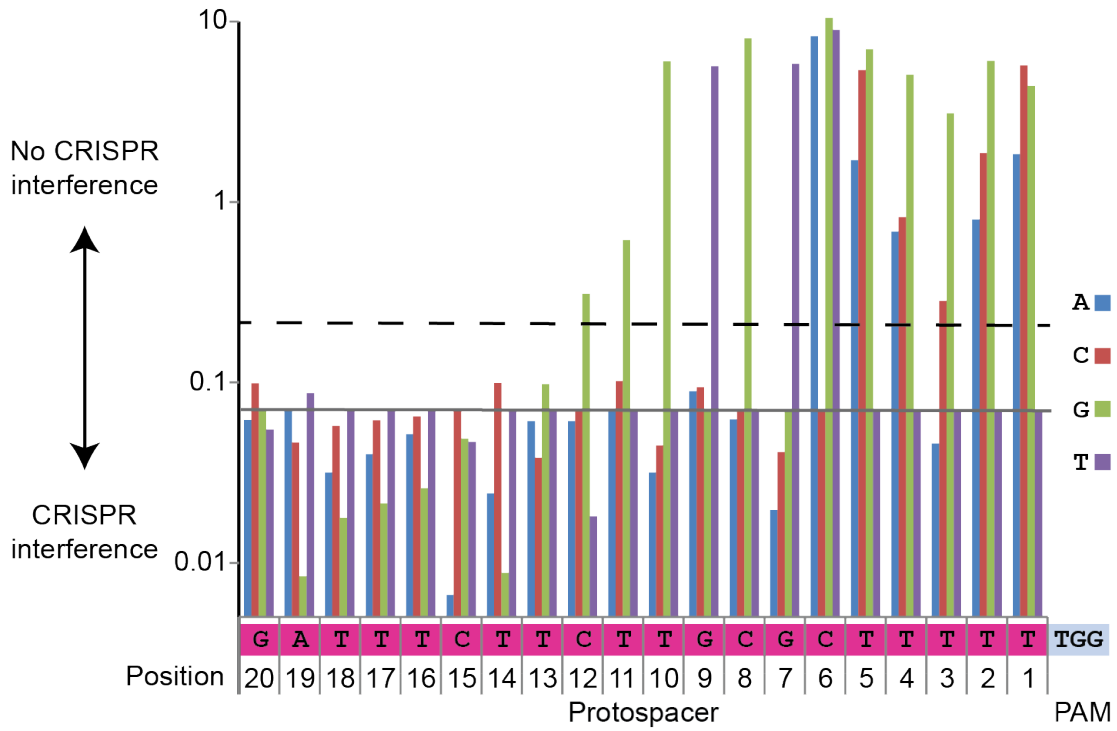


Figure 2-10 Analysis of the protospacer sequences

Relative proportion of the number of reads after transformation of the random protospacer constructs in crR6 cells (compared to number of reads in R6 transformants). The relative abundance of each nucleotide for each position of the first 20 nucleotides of the protospacer sequence is shown. High abundance indicates lack of cleavage by Cas9 complex, that is, an inactivating mutation. The gray line shows the level of the wild-type sequence. The dashed line represents the level above which a mutation significantly disrupts cleavage (see supplementary materials of (Jiang et al. 2013a)).

2.3.3 General scheme for the Cas9-complex-mediated counter-selection for genome editing

Based on the analysis of targeting requirement by the Cas9 complex, we describe a Cas9-complex-mediated counter-selection scheme for genome editing (**Figure 2-11**). First, a targeting construct containing a minimal type II CRISPR-Cas system (i.e. *tracr*, *cas9* and a spacer that matches the genomic region to be modified) is designed. An editing template is designed to carry mutations that abolish recognition and cleavage by the Cas9 complex. Preferably, the mutation should be deletion or gene replacement that entirely removes the target sequence recognized by the crRNA. If this is not feasible, a mutation in the PAM or the seed sequence on the protospacer is generally considered. Based on our analysis, mutations in the PAM clearly inactivates targeting by the Cas9 complex, whereas the effect of single-nucleotide mutations in the protospacer is highly dependent on the sequence context and can be difficult to predict. Therefore we believe that mutations in the PAM, if possible, should be the preferred editing strategy. Alternatively, multiple mutations in the seed sequence are likely to abolish cleavage by the Cas9 complex.

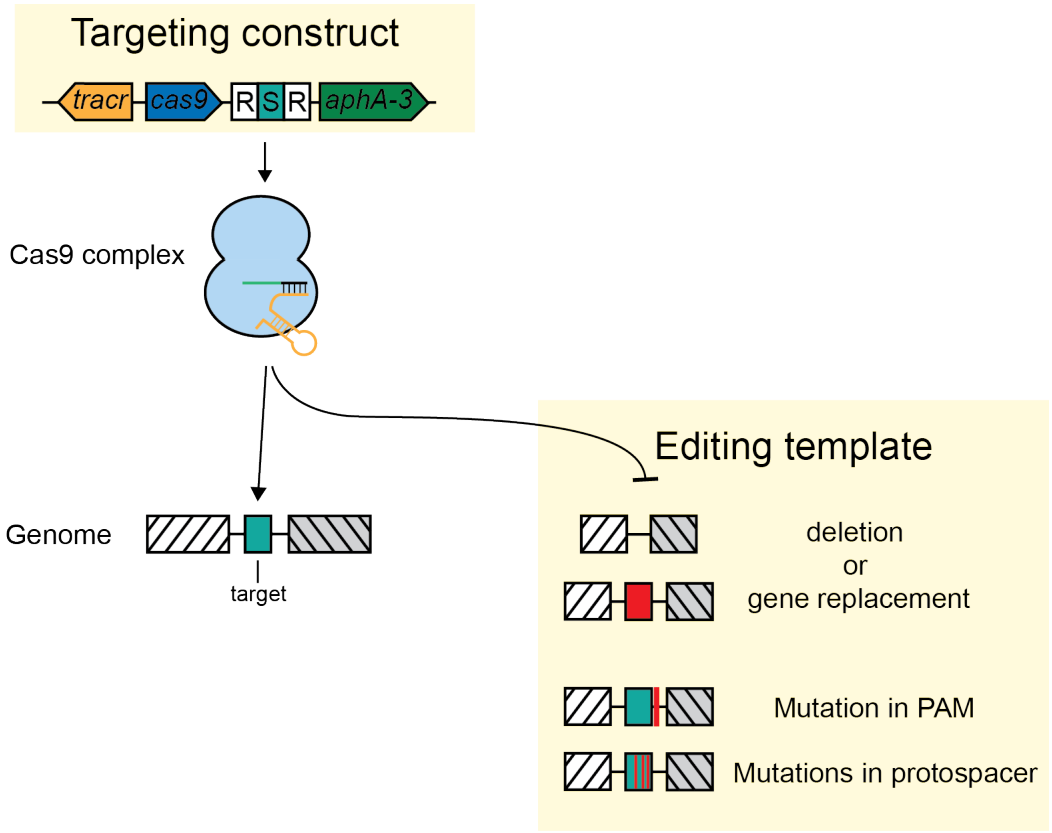


Figure 2-11 A general scheme for genome editing mediated by the Cas9 complex

A targeting construct encoding the Cas9 complex contains a spacer that matches a genomic region to be modified. An editing template should contain either deletion, replacement, or mutations within the PAM or protospacer of the target sequence. Hence, targeting by the Cas9 complex eliminates cells containing the wild-type sequence and selects for the newly modified sequence that abolishes targeting. Abbreviations: R, repeat; S, spacer; PAM, protospacer adjacent motif.

2.3.4 Genome editing in *Streptococcus pneumoniae*

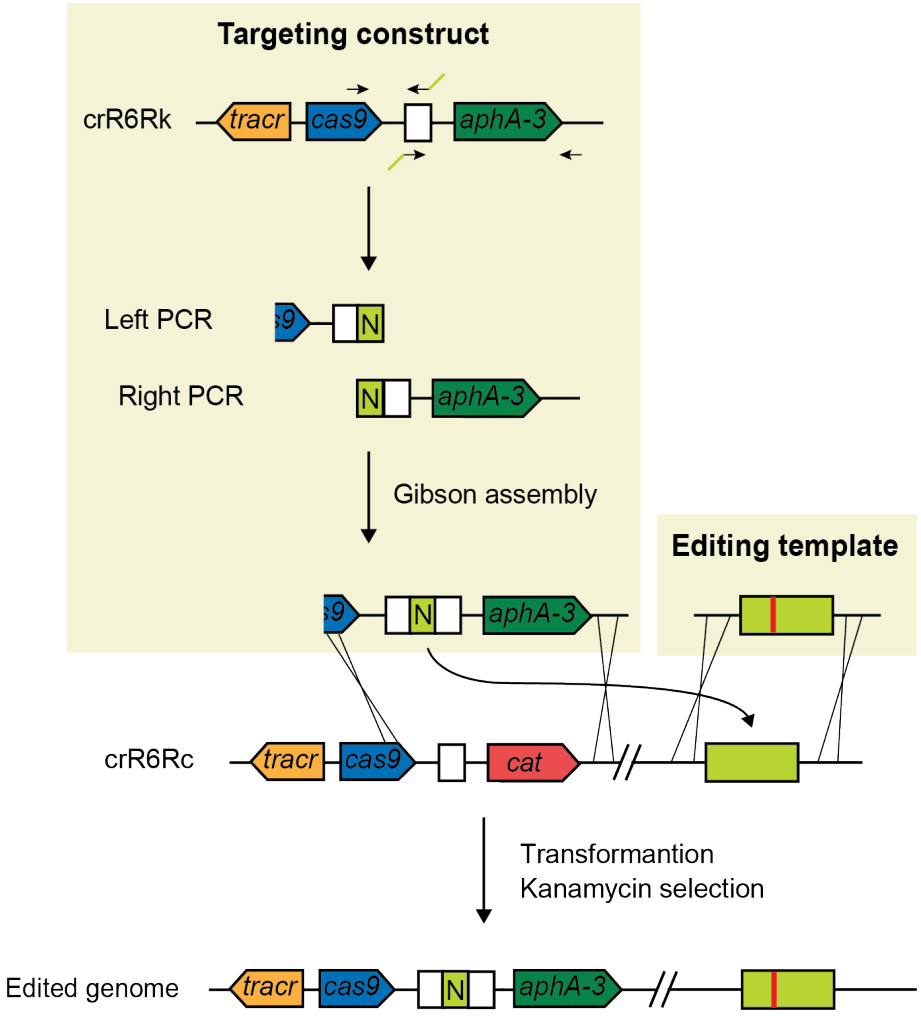


Figure 2-12 Methodology for genome editing in *S. pneumoniae*

S. pneumoniae crR6Rk contains *tracr*, *cas9*, a single repeat (white box) and a kanamycin-resistant marker (*aphA-3*). Genomic DNA from this strain was used as a template for PCR with primers designed to introduce a new spacer (green box designated with N). The left and right PCRs are ligated using Gibson assembly to make the targeting construct. The targeting construct and an editing template carrying a desired mutation (red bar), both in the form of dsDNA, were co-transformed into strain crR6Rc, which is a strain equivalent to crR6Rk but has the kanamycin-resistant marker replaced by a chloramphenicol-resistant marker (*cat*). Selection of kanamycin results in successful editing.

To develop a rapid and efficient method for targeted genome editing in *S. pneumoniae*, we generated crR6Rk and crR6Rc, strains in which spacers can be easily introduced by PCR (**Figure 2-12**). We decided to edit the β-galactosidase (*bgaA*) gene of *S. pneumoniae*, whose activity can be easily measured (Zahner and Hakenbeck 2000). We set out to introduce alanine substitutions of amino acids in the active site of this enzyme: R481A (R→A) and N563A,E564A (NE→AA) mutations. To illustrate different editing strategies, we designed mutations in both the PAM sequence and the protospacer seed. In both cases we used the same targeting construct with a crRNA complementary to a region of the β-galactosidase gene that is adjacent to a TGG PAM sequence (CCA in the complementary strand) (**Figure 2-13A**). The R→A editing template (**Figure 2-13A**) created a three-nucleotide mismatch on the protospacer seed sequence (CGT to GCA, also introducing a BtgZI restriction site). In the NE→AA editing template (**Figure 2-13A**) we simultaneously introduced a synonymous mutation that creates an inactive PAM (TGG to TTG) along with mutations that were 218 nt downstream of the protospacer region (AAT GAA to GCT GCA, also generating a TseI restriction site). This last editing strategy demonstrated the possibility of using a remote PAM to make mutations in places where a proper target might be hard to choose. For instance, although the *S. pneumoniae* R6 genome (with a 39.7% GC content) contains on average one PAM every 12 bp, some PAMs are separated by up to 194 bp (**Figure 2-14**). Finally we designed a Δ*bgaA* in-frame deletion template of 6,664 bp. In all three cases, co-transformation of the targeting construct and editing template produced 10-times

more kanamycin-resistant cells than co-transformation with a control editing template containing wild-type *bgaA* sequences (**Figure 2-13B**). We genotyped 24 transformants (8 for each editing experiment) by PCR, restriction digest and DNA sequencing and found that all but one incorporated the desired change (**Figure 2-13C, D** and **E**). These data indicate that the editing efficiency in *S. pneumoniae* is approximately 90%. The 10% that were not edited were mostly CRISPR mutants, primarily caused by the error-prone Gibson assembly of two DNA fragments carrying repeat sequences (data not shown). Finally, we measured β-galactosidase activity (Zahner and Hakenbeck 2000) to confirm that all edited cells displayed the expected phenotype (**Figure 2-13F**).

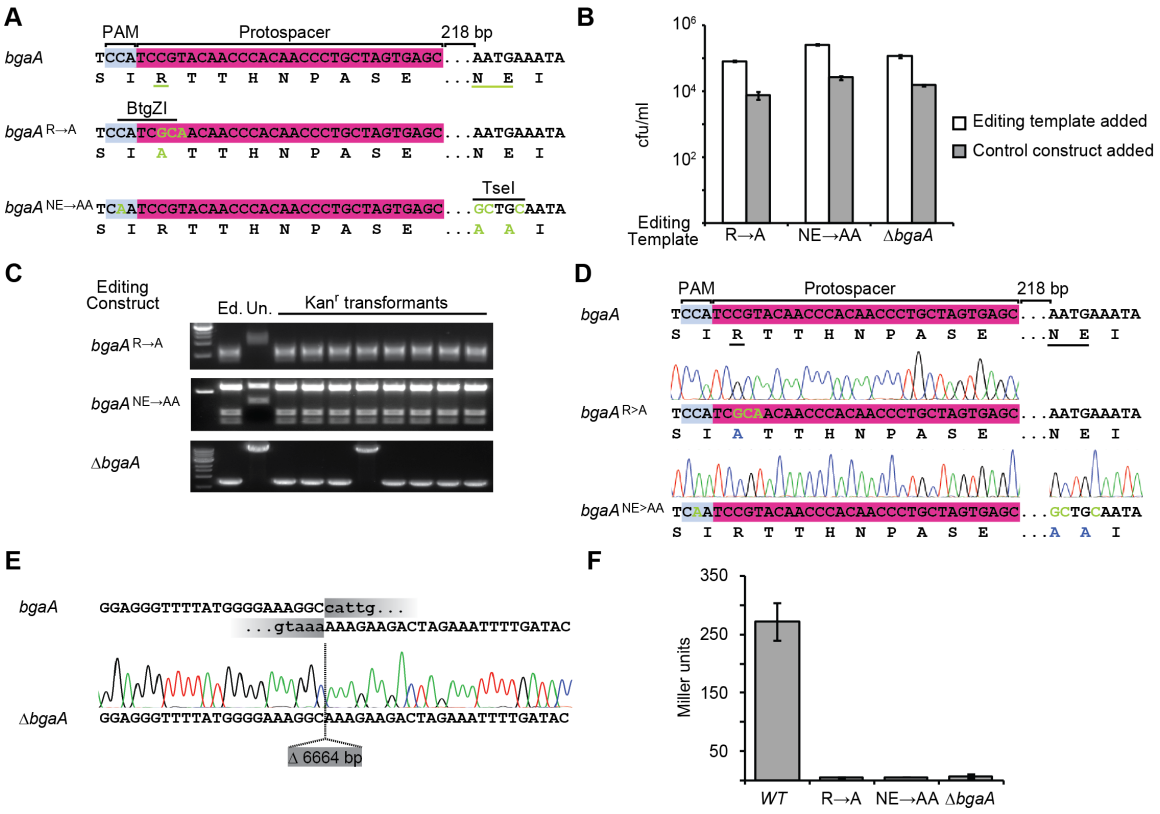


Figure 2-13 Genome editing in *S. pneumoniae*

(A) Nucleotide and amino acid sequences of the wild-type and edited (green nucleotides; underlined amino acid residues) *bgaA*. The protospacer, PAM and restriction sites are shown. **(B)** Transformation efficiency (counted as CFU/mL) of cells transformed with targeting constructs in the presence of an editing template or control. Error bars, mean ± SD for three replicas. **(C)** PCR analysis for eight transformants of each editing experiment followed by digestion with BtgZI (R→A) and TseI (NE→AA). Deletion of *bgaA* was revealed as a smaller PCR product. **(D** and **E)** Editing was confirmed by DNA sequencing. **(F)** Miller assay to measure the β-galactosidase activity of wild-type and edited strains. Error bars, mean ± SD for three replicas.

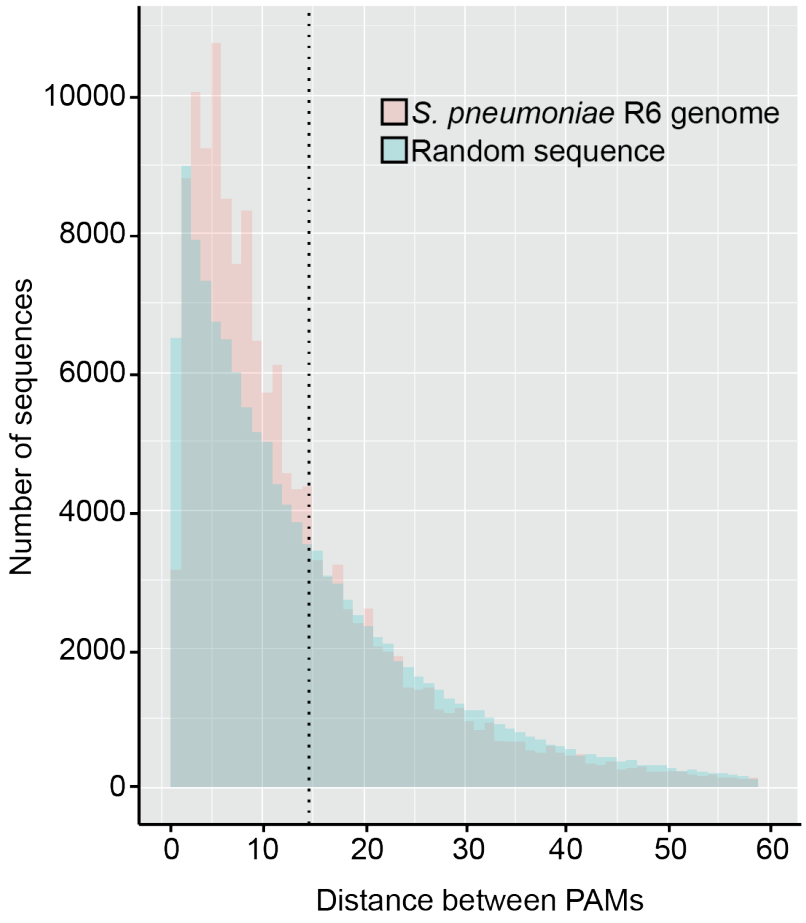


Figure 2-14 Distribution of distances between PAMs

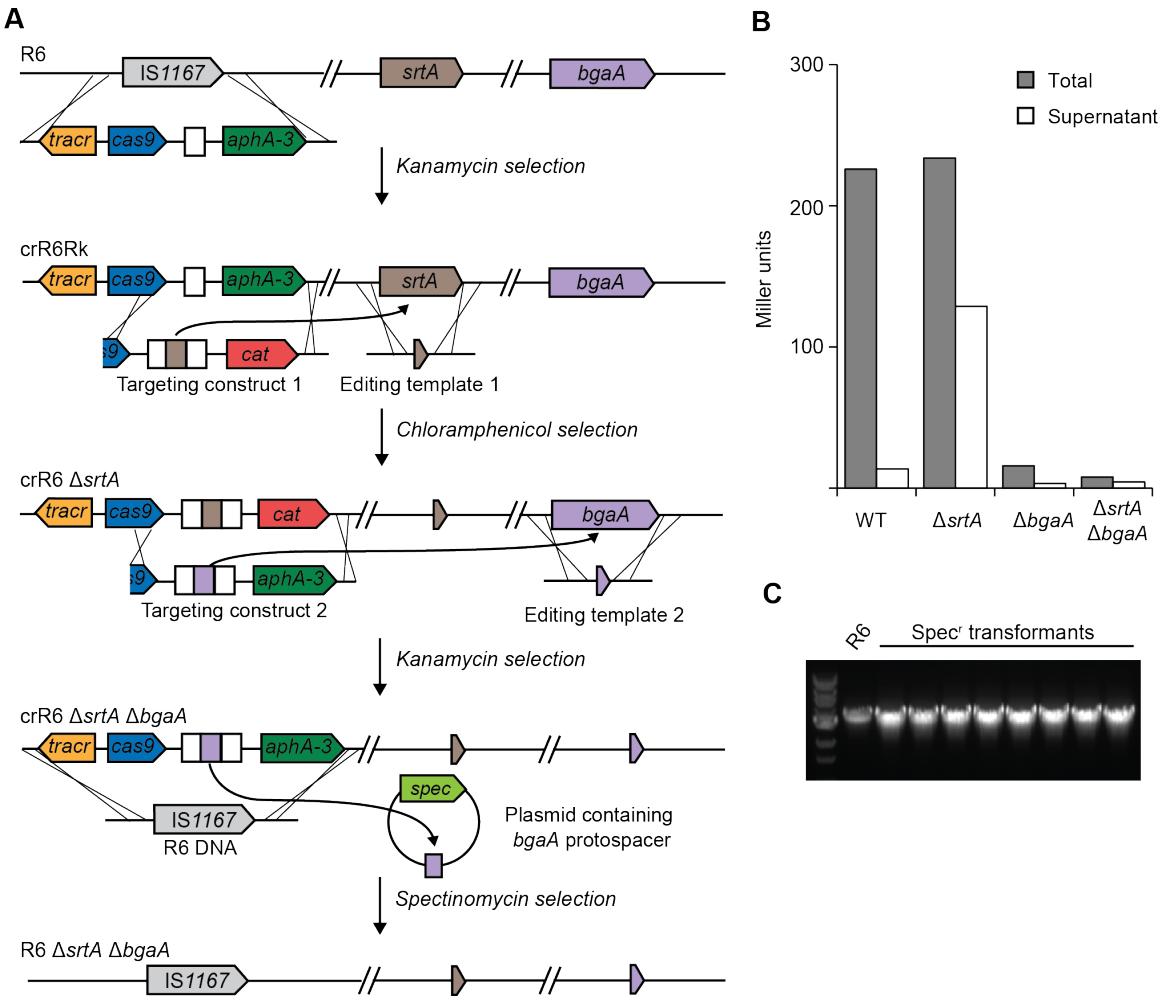
NGG and CCN are considered to be valid PAMs. Data is shown for the *S. pneumoniae* R6 genome as well as for a random sequence of the same length and with the same GC-content (39.7 %). The dotted line represents the average distance (12 nt) between PAMs in the R6 genome.

Cas9-complex-mediated editing can also be used to generate multiple mutations for the study of biological pathways. We decided to illustrate this for the sortase-dependent pathway that anchors surface proteins to the envelope of Gram-positive bacteria (Marraffini et al. 2006). We introduced a sortase deletion by co-transformation of a chloramphenicol-resistant targeting construct and a Δ*srtA* editing template (**Figure 2-15A**). Next we introduced a β-galactosidase

deletion by co-transformation of a kanamycin-resistant targeting construct that replaced the previous one and a Δ*bgaA* editing template. In *S. pneumoniae*, β-galactosidase is covalently linked to the cell wall by sortase (Zahner and Hakenbeck 2000). Therefore, deletion of *srtA* resulted in the release of the surface protein into the supernatant, whereas the double deletion had no detectable β-galactosidase activity (**Figure 2-15B**). Such a sequential selection can be iterated as many times as required to generate multiple mutations. Finally, to eliminate the CRISPR-Cas locus, we introduced a plasmid containing the *bgaA* target and a spectinomycin resistance gene along with genomic DNA from the wild-type strain R6. Spectinomycin-resistant transformants that retain the plasmid eliminated the CRISPR sequences (**Figure 2-15C**).

Figure 2-15 Sequential introduction of mutations by CRISPR-mediated genome editing in *S. pneumoniae*

(A) A schematic for sequential introduction of mutations by CRISPR-mediated genome editing. First, R6 is engineered to generate crR6Rk. crR6Rk is co-transformed with a *srtA*-targeting construct fused to *cat* for chloramphenicol selection of edited cells, along with an editing construct containing a Δ*srtA* in-frame deletion. Strain crR6 Δ*srtA* is generated by selection on chloramphenicol. Subsequently, the Δ*srtA* strain is co-transformed with a *bgaA*-targeting construct fused to *aphA-3* for kanamycin selection of edited cells, along with an editing construct containing a Δ*bgaA* in-frame deletion. Finally, the engineered CRISPR locus can be erased from the chromosome by first co-transforming R6 DNA containing the wild-type *IS1167* locus and a plasmid carrying a *bgaA* protospacer (pDB97), and selection on spectinomycin. **(B)** β-galactosidase activity as measured by Miller assay. In *S. pneumoniae*, this enzyme is anchored to the cell wall by sortase A. Deletion of the *srtA* gene results in the release of β-galactosidase into the supernatant. Δ*bgaA* mutants show no activity. **(C)** PCR analysis for eight spectinomycin (spec)-resistant transformants to detect the replacement of the CRISPR locus by wild-type *IS1167.*



These two mutations can also be introduced simultaneously. We designed a targeting construct containing two spacers, one matching *srtA* and the other matching *bgaA*, and co-transformed it with both editing templates at the same time (**Figure 2-16A**). Genetic analysis of transformants showed that editing occurred in 6/8 cases (**Figure 2-16B**). Notably, the remaining two clones each contained either a Δ*srtA* or a Δ*bgaA* deletion, suggesting the possibility of performing combinatorial mutagenesis using this strategy.

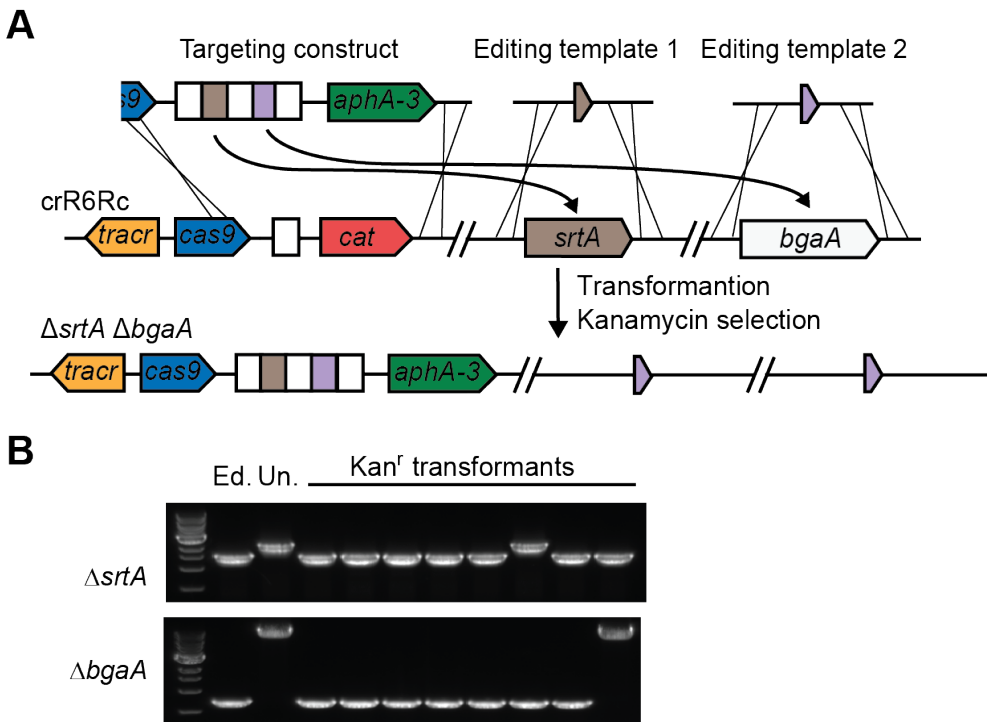


Figure 2-16 Multiplex editing in *S. pneumoniae*

(A) The Δ*srtA* and Δ*bgaA* in-frame deletion can be performed simultaneously by co-transforming a targeting construct that contains two spacers, each targeting *srtA* and *bgaA*, into strain crR6Rc followed by kanamycin selection. **(B)** PCR analysis of eight transformants to detect deletions in *srtA* and *bgaA* loci.

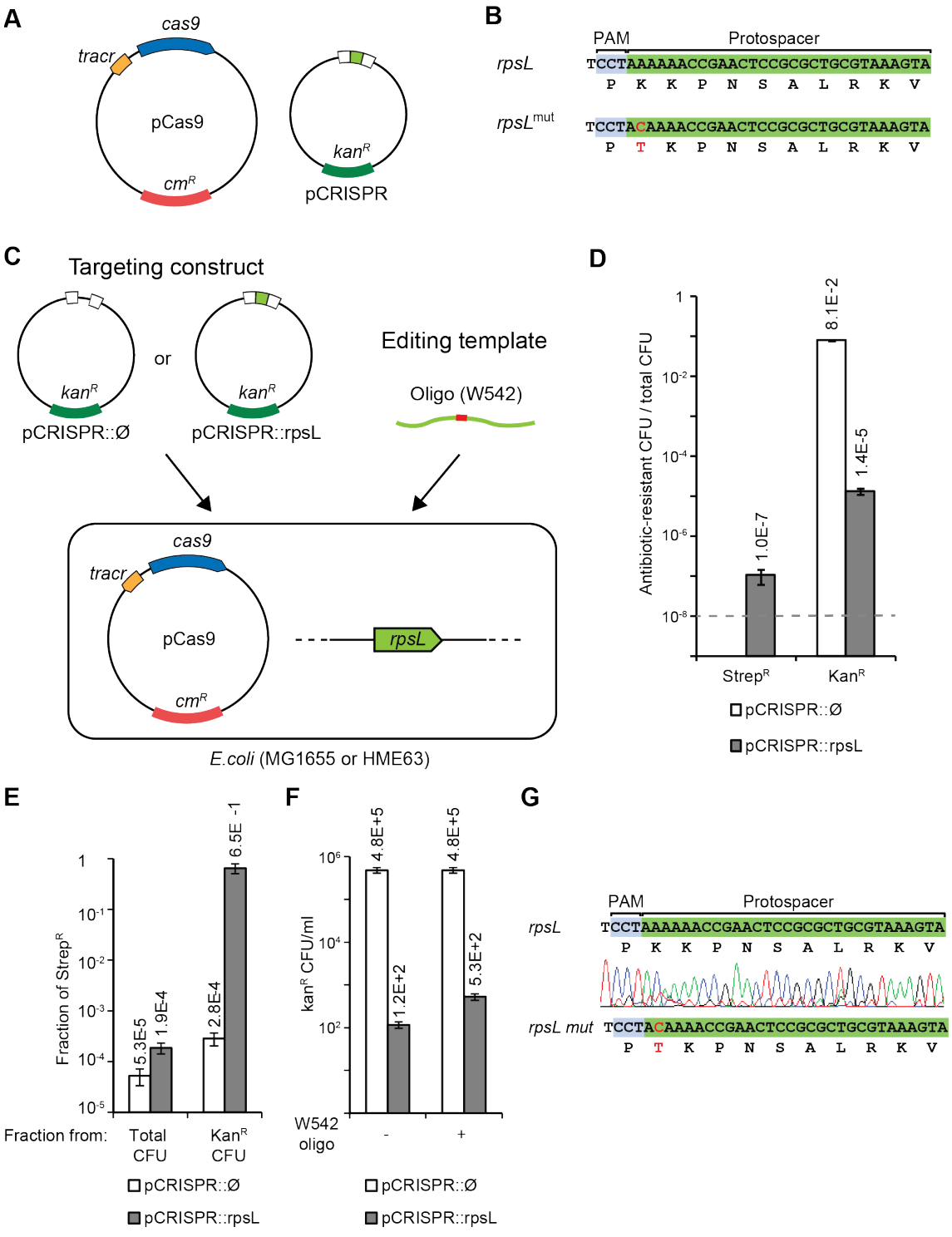
2.3.5 Genome editing in *E.coli*

We also repurposed the type II CRISPR-Cas system as a counter-selection tool to facilitate genome editing in *E.coli*. Two plasmids were constructed: a pCas9 plasmid carrying the tracrRNA, Cas9 and a chloramphenicol-resistant marker, and a pCRISPR plasmid carrying the CRISPR array and a kanamycin-resistant marker (**Figure 2-17A**). To be able to measure the efficiency of editing independently of CRISPR counter-selection, we sought to introduce an A to C transversion in the *rpsL* gene (**Figure 2-17B**) that confers streptomycin

resistance (Hosaka et al. 2004). We constructed a pCRISPR::rpsL plasmid harboring a spacer that would guide Cas9 to cleave the wild-type, but not the mutant *rpsL* allele (**Figure 2-17C**). The pCas9 plasmid was first introduced into *E. coli* wild-type strain MG1655 and the resulting strain was transformed with either pCRISPR::rpsL, the targeting construct, or pCRISPR::Ø, which contained a non-matching spacer serving as a negative control. W542, an editing oligonucleotide containing the A →C mutation was also co-transformed (**Figure 2-17C**). As a result, we were able to recover streptomycin-resistant colonies after transformation of only pCRISPR::rpsL but not pCRISPR::Ø, suggesting that Cas9 cleavage induced recombination of the oligonucleotide (**Figure 2-17D**). However, when co-transformed with pCRISPR::rpsL, the frequency of the streptomycin-resistant colonies (1.0×10^{-7}) was two orders of magnitude lower than that of the kanamycin-resistant colonies (1.4×10^{-5}), which were presumably cells that escaped cleavage by Cas9, similar to those found in **Figure 2-7C**. Therefore, in these conditions, cleavage by the Cas9 complex facilitated the recombination of the editing template, but with an efficiency that was not enough to select the edited cells above the background of “escapers”.

Figure 2-17 CRISPR-mediated genome editing in *E.coli*

(A) pCas9 plasmid carries *tracr*, *cas9* and a chloramphenicol-resistant marker; pCRISPR plasmid carries a CRISPR array and a kanamycin-resistant marker. **(B)** A K42T mutation (A→C on the nucleotide level) conferring streptomycin resistance was introduced in *rpsL*. **(C)** The targeting construct, pCRISPR::rpsL contained a spacer that matched the wild-type, but not mutant *rpsL*. To perform genome editing, this targeting construct or a control construct, pCRISPR::Ø (which contains a non-matching spacer) was co-transformed with W542, an editing oligonucleotide that contained the A→C mutation, into *E.coli* MG1655 that already harbored pCas9. **(D)** Fraction of cells that became streptomycin-resistant (*Strep*^{R}) or kanamycin-resistant (*Kan*^{R}) after co-transformation performed in **(C)**. Dashed line indicates limit of detection of the assay. **(E)** Fraction of *Strep*^{R} colony-forming units (CFUs) calculated from total or *Kan*^{R} CFUs. **(F)** Transformation of the pCRISPR::Ø or pCRISPR::rpsL plasmids into HME63 competent cells in the absence or presence of the W542 oligo. Transformation efficiency was measured as the number of CFU per mL of competent cells. **(G)** Editing of the *rpsL* gene was confirmed by DNA sequencing. In **(D)**, **(E)** and **(F)**, error bars represent mean ± SD of three replicas.


A
cas9
tracr
pCas9
cmR
kanR
pCRISPR
B
PAM
Protospacer
rpsL
TCCTAAAAAACCGAACTCCGCGCTGCGTAAAGTA
P K K P N S A L R K V
rpsLmut
TCCTACAAAACCGAACTCCGCGCTGCGTAAAGTA
P T K P N S A L R K V
C
Targeting construct
kanR
pCRISPR::Ø
or
kanR
pCRISPR::rpsL
Editing template
Oligo (W542)
cas9
tracr
pCas9
cmR
rpsL
E.coli (MG1655 or HME63)
D
Antibiotic-resistant CFU / total CFU
1
10-2
10-4
10-6
10-8
8.1E-2
1.4E-5
1.0E-7
StrepR
KanR
pCRISPR::Ø
pCRISPR::rpsL
E
Fraction of StrepR
1
10-1
10-2
10-3
10-4
10-5
5.3E-5
1.9E-4
2.8E-4
6.5E -1
Fraction from:
Total CFU
KanR CFU
pCRISPR::Ø
pCRISPR::rpsL
F
kanR CFU/ml
10⁶
10⁴
10²
4.8E+5
1.2E+2
4.8E+5
5.3E+2
W542 oligo
-
+
pCRISPR::Ø
pCRISPR::rpsL
G
PAM
Protospacer
rpsL
TCCTAAAAAACCGAACTCCGCGCTGCGTAAAGTA
P K K P N S A L R K V
rpsL mut
TCCTACAAAACCGAACTCCGCGCTGCGTAAAGTA
P T K P N S A L R K V

To improve the efficiency of genome editing in *E. coli*, we combined our CRISPR-Cas system with the λ-Red recombination technology (see Chapter 2.1.2). The pCas9 plasmid was introduced into the *E.coli* recombineering strain HME63 (Costantino and Court 2003), which contains the Exo, Beta and Gam functions of the λ phage. The resulting strain was co-transformed with the pCRISPR::rpsL plasmid (or the pCRISPR::*Ø* control) and the W542 oligonucleotide (**Figure 2-17C**). The recombineering efficiency was 5.3×10^{-5}, calculated as the fraction of total cells that became streptomycin-resistant when the control plasmid was used (**Figure 2-17E**). In contrast, transformation with the pCRISPR::rpsL plasmid and selection with kanamycin increased the percentage of edited cells to 65 ± 14 % (**Figure 2-17E**). Editing efficiency can be also calculated in another way. When transformation was performed with the targeting construct alone, we found pCRISPR::rpsL transformants were 1.2×10^{2} CFU/mL, as opposed to 5.3×10^{2} CFU/mL for transformation performed with both targeting construct and editing template (**Figure 2-17F**). This rendered the editing efficiency 77% ((5.3×10^{2} -1.2×10^{2}) / 5.3×10^{2}), consistent with the efficiency calculated from **Figure 2-17E**. Streptomycin-resistant colonies were confirmed for editing by DNA sequencing of the *rpsL* allele (**Figure 2-17G**).In sum, these data show that CRISPR-Cas can act as a robust counter-selection tool when combined with λ-Red recombination. The editing efficiency is drastically boosted from 0.1% (without CRISPR-Cas, see Chapter 2.1.2) to 65 ± 14 % (with CRISPR-Cas).

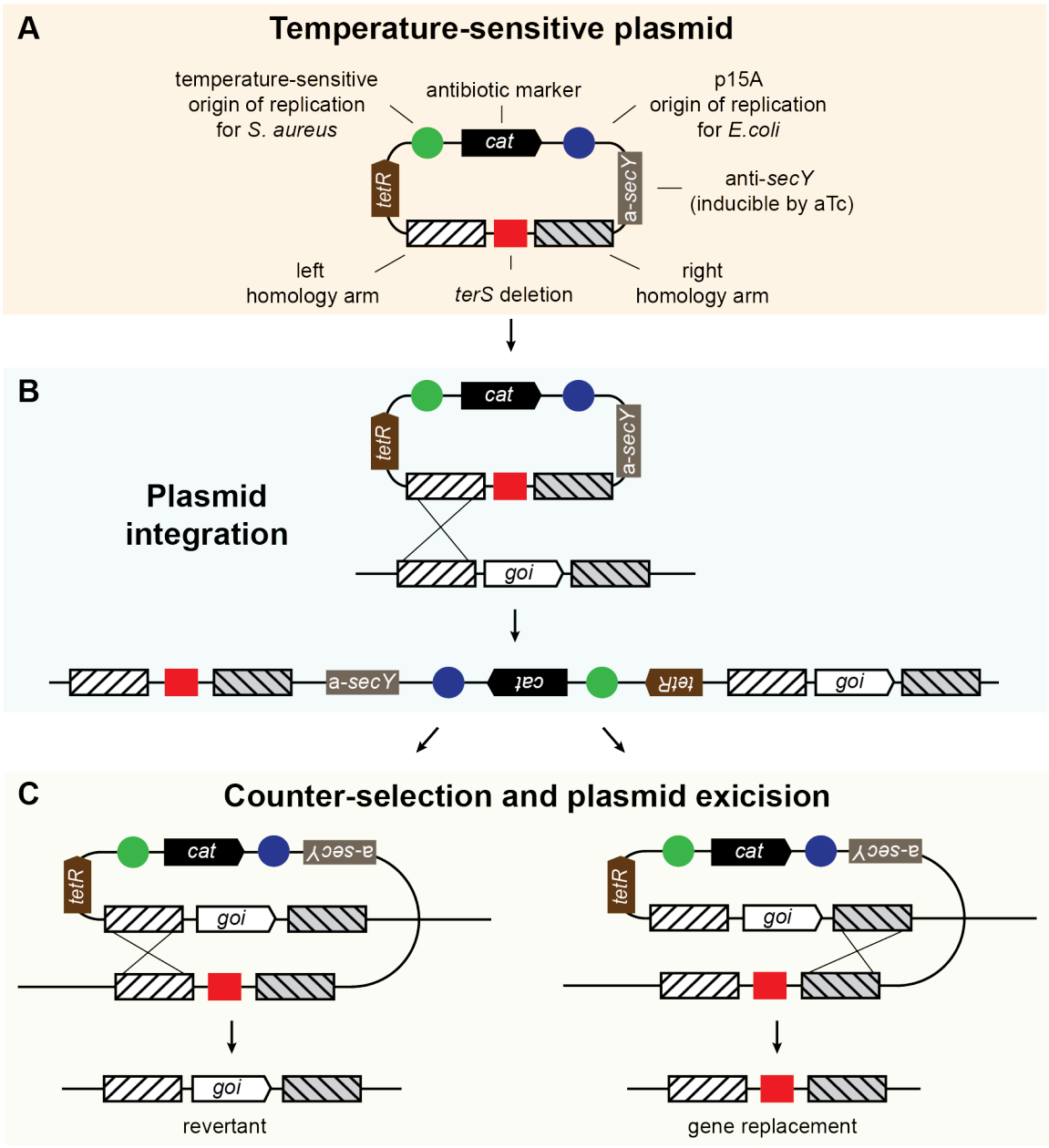
2.3.6 Genome editing in *Staphylococcus aureus*

In both *S. pneumoniae* and *E.coli*, linear DNA fragments can be used as editing templates due to either natural competence or the presence of phage recombineering machineries. However, many other bacteria are not naturally competent and recombineering typically do not work in them. In these species, introduction of linear DNA fragments are rapidly degraded by the RecBCD or homologue complexes and cannot be used as template for genome editing. To demonstrate that it's possible to use the same CRISPR-mediated counter-selection scheme to facilitate genome editing in these species, I developed a system in which the editing template was placed on a temperature-sensitive plasmid. It needs to be mentioned since the following work is still under progress, the results I showed are qualitative rather than quantitative.

I chose to develop a CRISPR-mediated counter-selection system in *Staphylococcus aureus*, the most studied bacterium in our lab yet for which the genome editing tool had not been optimized. The existing technology relies on allelic replacement of temperature-sensitive plasmids and subsequent counter-selection (Bae and Schneewind 2006; Monk et al. 2012). As depicted in **Figure 2-18**, the counter-selection marker is a *secY* antisense RNA cloned under a pTet promoter inducible by anhydrotetracycline (aTc). Overexpression of the antisense RNA represses *secY*, which encodes a membrane protein that makes up the SecYEG translocase essential for bacterial growth and survival (Manting and Driessen 2000).

Figure 2-18 Allelic replacement in *S. aureus*

(A) First, a temperature-sensitive plasmid harboring a desired mutation along with two homology arms are constructed. The plasmid also carries a chloramphenicol-resistant marker (*cat*) and a counter-selection marker, an inducible *secY* antisense RNA. Upon introduction to *S. aureus* and selection of chloramphenicol, the plasmid can be maintained at a permissive temperature (i.e. 30°C). **(B)** When shifted to a nonpermissive temperature (i.e., 43°C), the plasmid is forced to integrate into the region of the host chromosome carrying the gene of interest (*goi*) through homologous recombination. This low-frequency event can be selected using chloramphenicol. **(C)** When applying anhydrotetracylcline (aTc), overexpression of the *secY* antisense RNA (a counter-selection marker) forces plasmid excision, resulting in either gene replacement or reversion at a roughly 1:1 ratio.

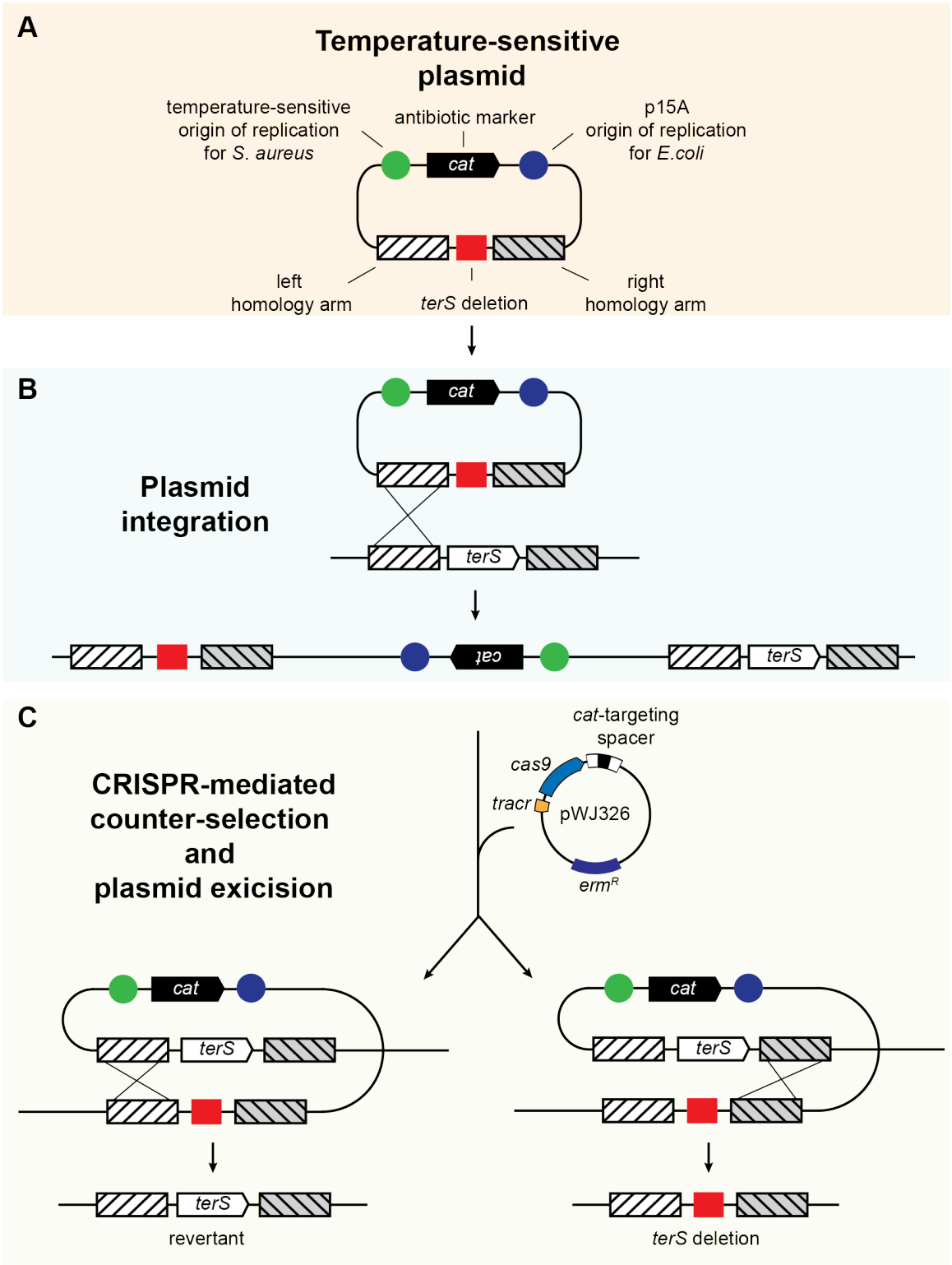


We and others attempted to perform genome editing in *S. aureus* following these established protocols. However, we encountered hindrance during the counter-selection step and were unable obtain edited cells in many cases. In order to circumvent this problem, I decided to apply the type II CRISPR-Cas as a counter-selection tool. First, I constructed pWJ40, a plasmid carrying the full type

II CRISPR-Cas from *S. pyogenes* and demonstrated heterologous expression in *S. aureus* constituted a functional immune system (Goldberg et al. 2014). Next, *cas1*, *cas2* and *csn2* were removed from the plasmid, again creating a minimal type II system. When programmed to target the *S. aureus* genome, this minimal system effectively killed the cells (Bikard et al. 2014) just as in *S. pneumoniae* and *E.coli*. After confirming that Cas9 can be used as a counter-selection tool in *S. aureus*, I removed the overexpression element from the original temperature-sensitive vector. In the new scheme, I introduced pWJ326, a plasmid that carried tracrRNA, Cas9 and a crRNA that targets the chloramphenicol-resistant marker (*cat*) in order to achieve counter-selection. As a proof-of-concept (**Figure 2-19**), I managed to delete *terS* from the ΦNM1 prophage genome in two weeks.

Figure 2-19 CRISPR-mediated genome editing in *S. aureus*

(A) First, a temperature-sensitive plasmid harboring the *terS* deletion along with two homology arms are constructed. The plasmid also carries a chloramphenicol-resistant marker (*cat*). Upon introduction to *S.aureus* and selection of chloramphenicol, the plasmid can be maintained at a permissive temperature (i.e. 28°C). **(B)** When shifted to a nonpermissive temperature (i.e., 37°C), the plasmid is forced to integrate into the *terS* region of the host chromosome through homologous recombination. This low-frequency event can be selected using chloramphenicol. **(C)** In order to counter-select, pWJ326, a temperature-sensitive plasmid containing *tracr*, *cas9* and a *cat*-targeting spacer was introduced to integrant cells, leading to plasmid excision. Gene replacement or reversion occurred at a roughly 1:1 ratio. Following successful *terS* deletion, pWJ326 can be readily lost by growing cells at 37°C.


A
Temperature-sensitive plasmid
temperature-sensitive origin of replication for S. aureus
antibiotic marker
p15A origin of replication for E.coli
cat
left homology arm
terS deletion
right homology arm
B
Plasmid integration
cat
terS
cat
terS
C
CRISPR-mediated counter-selection and plasmid excision
cat-targeting spacer
cas9
tracr
pWJ326
ermR
cat
terS
terS
revertant
cat
terS
terS deletion

Interestingly, since plasmid excision is normally a low-frequency event, it was expected that counter-selection with pWJ326 should yield only a few colonies. In contrast, my preliminary results showed that the transformation efficiency of pWJ326 was very high, suggesting that CRISPR-induced recombination may be higher in *S. aureus* than it is in *E.coli* and *S. pneumoniae* (Chapter 2.3.7). Induction of recombination makes CRISPR a superior counter-selection tool in the case of genome editing, and it undeniably deserves a close investigation in the future.

2.3.7 Cas9-complex-mediated cleavage actively induces recombination

Continuing from Chapter 2.3.5, the fact that we were able to recover streptomycin-resistant colonies after transformation of pCRISPR::rpsL but not pCRISPR::Ø in *E.coli* suggests that Cas9-complex-mediated cleavage induced recombination of the editing template (**Figure 2-17D**). However, since this recombination efficiency in wild-type *E.coli* (i.e. MG1655) was too low, we sought to better quantify the fold-induction of recombination by CRISPR in the recombineering strain HME63. The experimental set-up is the same as schematized in **Figure 2-17C** in which the targeting construct and the editing template were co-transformed. When we measured the number of edited cells ($Strep^R$ CFU) that simultaneously received the targeting constructs (Kan^R CFU), we found that the pCRISPR::rpsL transformants that were resistant to both antibiotics were isolated at a frequency of 2.0×10^{-4}, while for pCRISPR::Ø the frequency was 3.0×10^{-5} under the same condition (**Figure 2-20A**). This means Cas9-complex-mediated cleavage induced recombination by 6.7-fold

($2.0\times10^{-4}/3.0\times10^{-5}$). A milder induction (2.2-fold) was found in *S. pneumoniae* using an erythromycin-resistant marker (**Figure 2-20B**). Altogether, the Cas9 complex not only provided counter-selection, the cleavage by it induced modest recombination which aided in genome editing.

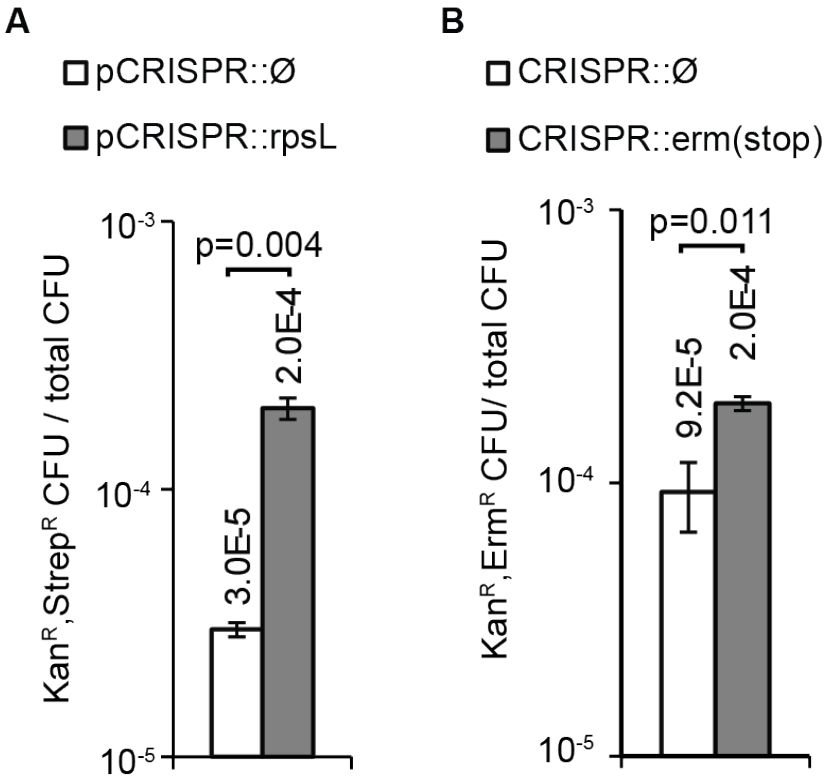


Figure 2-20 CRISPR-Cas induces recombination

(A) Fraction of total *E.coli* cells that became both kanamycin-resistant (*Kan*^{R}) and streptomycin-resistant (*Strep*^{R}) after co-transformation of the targeting construct (pCRISPR::rpsL or the pCRISPR::Ø control) and editing template (W542) shown in **Figure 2-17C** . **(B)** Fraction of total *S. pneumoniae* cells that became both kanamycin-resistant (*Kan*^{R}) and erythromycin-resistant (*Erm*^{R}) after co-transformation of the targeting construct (CRISPR::erm(stop) or the CRISPR::Ø control) and an editing template that confers erythromycin resistance. Error bars, mean ± SD of three replicas.

2.3.8 Discussion

2.3.8.1 CRISPR may be a universal counter-selection tool

As mentioned in Chapter 2.1, the biggest drawback for existing genome editing tools such as allelic replacement and λ-Red recombination is limited utility or complete lack of counter-selection markers. Here we developed a system consisting of a minimal type II CRISPR-Cas components, i.e. Cas9, tracrRNA and crRNA derived from *S. pyogenes*. Cas9 is a nuclease whose specificity can be programmed by a crRNA. Crucially, we showed that the Cas9 complex can cleave host chromosome and lead to cell death, hence it can be used as a counter-selection tool. We expressed this system heterologously in *S. pneumoniae*, *E.coli* and *S. aureus* without codon optimization, and showed by programming with different crRNAs, the system can act as a sequence-specific counter-selection tool and facilitate genome editing. Furthermore, we and others successfully expanded this tool to many more bacterial species, including *Lactobacillus* (Oh and van Pijkeren 2014), *Clostridium* (Wang et al. 2015; Xu et al. 2015), *Streptomyces* (Cobb et al. 2015; Huang et al. 2015), *Actinomycetes* (Tong et al. 2015), *Salmonella* (in collaboration with Zhenrun J. Zhang, unpublished data) as well as bacteriophages (Kiro et al. 2014; Martel and Moineau 2014). These works suggest that the type II CRISPR-Cas system is likely to serve as a universal counter-selection tool in a wide variety of bacteria. What if heterologous expression of the Cas9 complex does not function in some bacteria? As an immune system, CRISPR-Cas is present in 40% of eubacteria and 90% of archaea (Grissa et al. 2007). Therefore it is possible to exploit bacteria's

endogenous CRISPR-Cas systems for counter-selection. To this end, the more naturally-abundant type I and type III CRISPR-Cas may be harnessed for this purpose (Li et al. 2016).

2.3.8.2 The mutation frequency of CRISPR-Cas in the targeting construct and its effect on counter-selection

In both *S. pneumoniae* and *E. coli*, we found that although genome editing is facilitated by a co-selection of transformant cells and a modest induction of recombination at the target site by Cas9-complex-mediated cleavage, the mechanism that contributes the most to editing is the selection against non-edited cells. Therefore the major limitation of the method is the presence of a background of cells that escape CRISPR-induced cell death and lack the desired genomic modification. These “escapers” emerged primarily through the deletion of the targeting spacer, thus inactivating CRISPR-Cas. In *E.coli* HME63, the frequency of CRISPR mutation occurred at 2.5×10^{-4}, (**Figure 2-7B**), as calculated by taking the ratio of the number of pCRISPR::rpsL transformants (1.2×10^{2} CFU/mL) and the pCRISPR::Ø transformants (4.8×10^{5} CFU/mL). This means if editing templates are recombined to the host chromosome at a lower frequency, CRISPR-Cas no longer effectively counter-selects. For instance, if the effective recombination frequency (i.e. basal recombination frequency + induction by CRISPR-Cas) is the same as the mutation frequency, then the theoretical editing efficiency is 50%; if the effective recombination frequency is ten times less, the editing efficiency drops to 9%, etc. The recombination frequency in the

recombineering strain of *E.coli* was 5.3×10^{-5} and the fold-induction by Cas9 cleavage was 6.7-fold (**Figure 2-20A**), resulting in an effective recombination frequency of 3.6×10^{-4}. This makes the theoretical editing efficiency 59% ($3.6\times10^{-4}/(3.6\times10^{-4} + 2.5\times10^{-4})$), which is consistent with experimental efficiency we obtained: 65 ± 14 % (**Figure 2-17E**).

2.3.8.3 Off-target effect

A fundamental aspect of any targeted genome engineering technology is target specificity. This is extremely important in eukaryotes since recognition and cleavage of off-target sites will lead to the introduction of undesired mutations by NHEJ (non-homologous end joining) repair. We think off-target is a much less pronounced issue in bacteria. First, the genomes are smaller. Our analysis of targeting requirement (Chapter 2.3.2) revealed that the PAM (NGG) and sequence of the 12 nucleotides immediately upstream of the PAM were important for recognition and cleavage by the Cas9 complex. This means that off-target sequence occurs randomly once every 268 Mb (4^{14} Mb) in theory, and thus is unlikely to be found in smaller genomes such as the bacterial ones, which range from 130 Kb to 14 Mb (Koonin 2012).

Second, we demonstrated that cells were killed by CRISPR-Cas when programmed with crRNAs targeting the host chromosome, hence counter-selection. This means if off-target sites exist, it would not be possible to obtain even the on-target modification, since the Cas9 complex would cleave the off-target sites and cause cell death. In other words, if certain genomic modifications

are proved to be difficult to obtain through CRISPR counter-selection, one should consider the presence of off-target sites as a possible explanation and change the design of crRNAs accordingly. In our hands, we found that crRNAs matching off-target sites were rare, and can be usually avoided by using the BLAST tool provided by NCBI (http://www.ncbi.nlm.nih.gov/). Additionally, we were only able to isolate CRISPR mutants but not target mutant after applying CRISPR counter-selection without the editing template. This means target mutations are indeed very rare events. This observation is also supported by the fact that the error-prone DNA repair pathway, NHEJ, is absent in most bacteria (Pitcher et al. 2007).

2.4 Dead Cas9 (dCas9) mediates sequence-specific genetic repression and activation in bacteria

Cas9 has two nuclease domains, RuvC and HNH, each of which cleaves one strand within the target DNA (Gasiunas et al. 2012; Jinek et al. 2012). Mutation of the active sites of the two domains, D10A and H840A, abolish cleavage but do not impair DNA binding. We therefore thought that it may be possible to exploit the DNA-binding activity of this Cas9 catalytic site mutant, hereafter referred to as “dead” Cas9 (dCas9), to engineer a programmable transcription regulator. As a proof-of-concept, we demonstrated in *E.coli* that tracrRNA and dCas9 when guided with a proper crRNA (hereafter referred as the dCas9 complex), act as a transcription repressor by preventing the binding of the RNA polymerase (RNAP) to promoter sequences or as a transcription terminator by blocking the running

RNAP (**Figure 2-21**). In addition, a fusion between the ω subunit of the RNAP and dCas9 allows for programmable transcription activation (**Figure 2-21**).

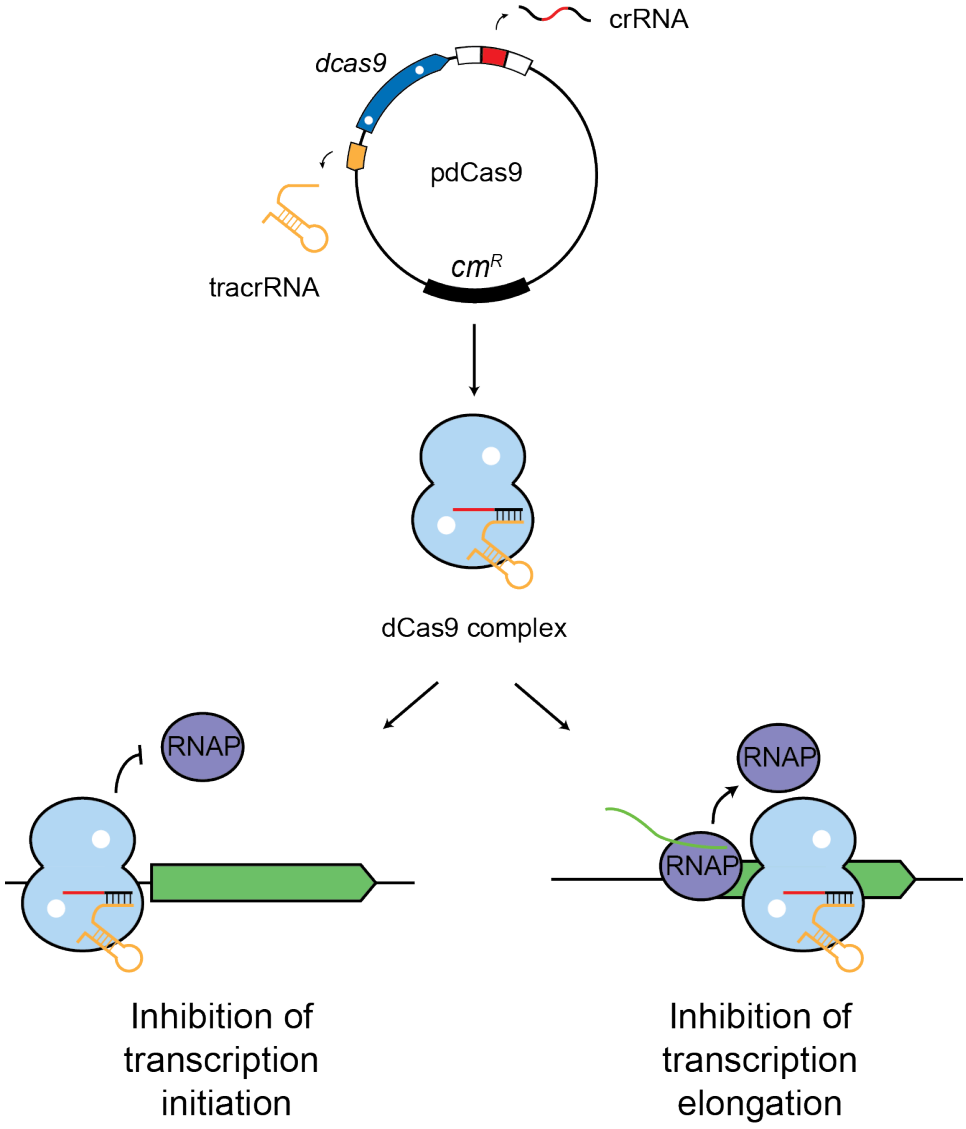


Figure 2-21 A schematic for dCas9-mediated transcription repression

Plasmid pdCas9 encodes a *cas9* mutant containing D10A and H840A substitutions (white holes) that abrogate nuclease activity. This catalytically dead Cas9 (dCas9), along with tracrRNA and crRNA form a complex. The crRNA directs binding of the dCas9 complex to promoter or open reading frame regions to prevent RNAP (RNA polymerase) binding or elongation, respectively.

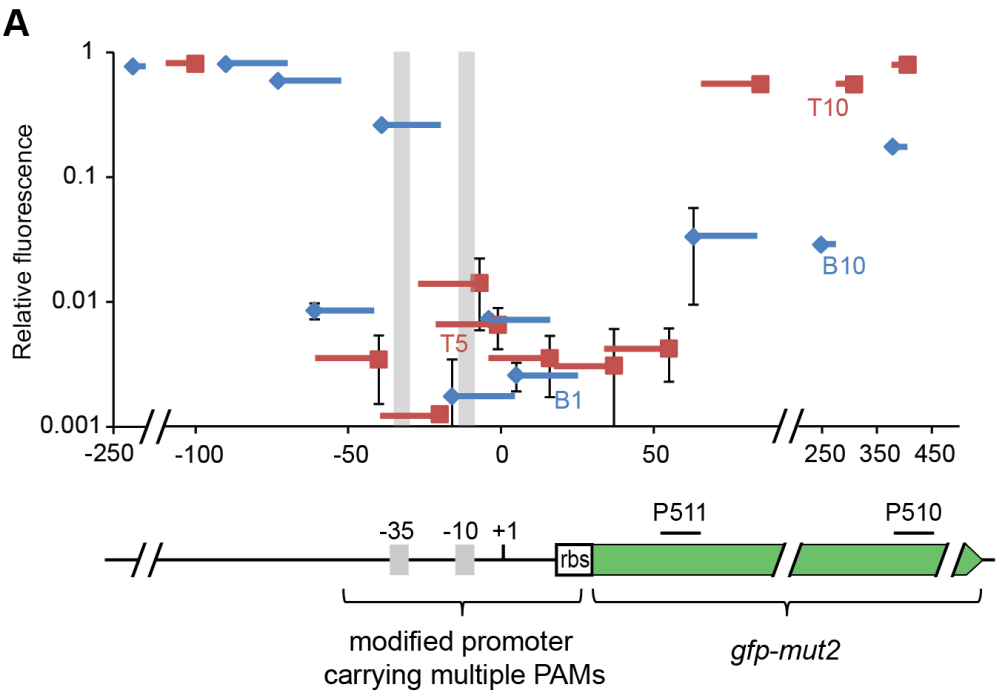
2.4.1 dCas9 mediates transcription repression

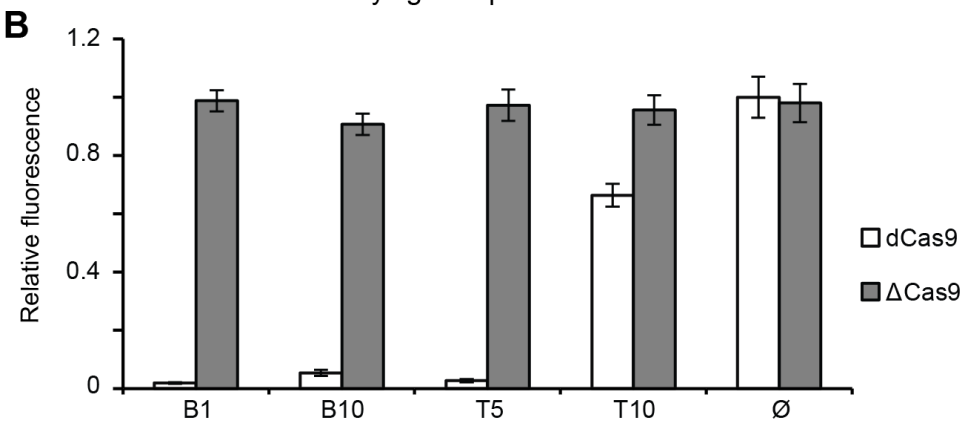
We converted Cas9 into a programmable DNA-binding protein by introducing the D10A and H840A mutations in the RuvC and HNH nuclease domains, respectively (Jinek et al. 2012). This catalytically “dead” version of Cas9, dCas9, was introduced into the pdCas9 plasmid along with the tracrRNA and a minimal CRISPR array designed for the easy cloning of new spacers and expression of crRNA guides (**Figure 2-21**) (Jiang et al. 2013a). To evaluate how dCas9 binding to the promoter can affect gene expression in *E. coli*, we constructed a green fluorescence protein reporter plasmid (pDB127) carrying the *gfp-mut2* gene (Cormack et al. 1996) under the control of a promoter designed to carry several NGG PAM sequences on both strands.

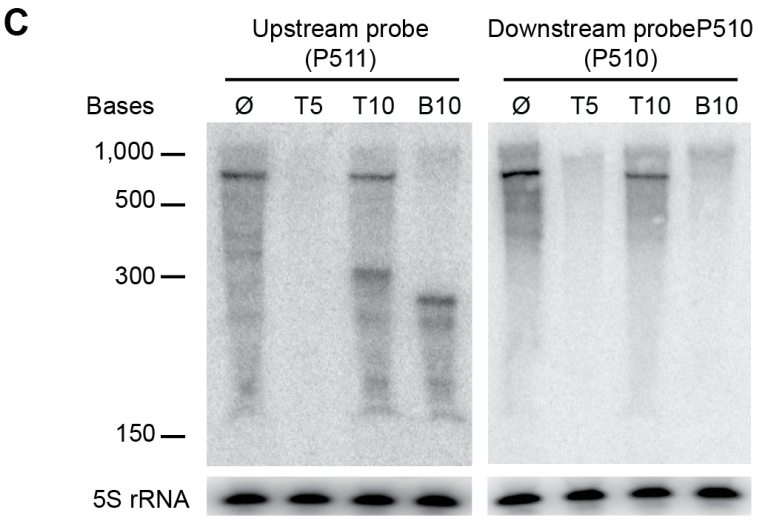
Twenty-two different spacers were engineered to express crRNAs guiding dCas9 to different regions of the *gfp-mut2* promoter and open reading frame. A greater than 100-fold reduction in fluorescence was observed upon targeting of regions overlapping or adjacent to the -35 and -10 promoter elements and to the Shine-Dalgarno sequence (**Figure 2-22A**). Targets on both strands showed similar repression levels. These experiments suggest that the binding of the dCas9 complex to any position within the promoter region prevents transcription initiation, presumably through steric inhibition of RNAP binding. In order to confirm that repression was due to dCas9 binding to the promoter DNA and not an effect of the antisense guide crRNA by itself, we repeated experiments in the absence of dCas9. In all cases tested, the fluorescence levels were identical to a non-targeting dCas9 control (**Figure 2-22B**).

Figure 2-22 dCas9 mediates transcription repression

(A) GFP fluorescence of cells expressing dCas9 guided to different regions of the *gfp-mut2* gene, relative to the fluorescence of cells expressing a non-targeting dCas9, was plotted as a function of the position of the target sequence within the gene (+1, transcription start). Squares indicate the PAM position, lines the extension of complementarity between the crRNA guide and the reporter gene. Red and blue lines indicate crRNAs sequences identical to top or bottom DNA strand, respectively. The *gfp-mut2* gene (green), its promoter, including the -35 and -10 elements (gray shade), the ribosome binding site (rbs) and the two probes (P511 and P510) used for Northern blot in **(C)** are shown. Error bars, mean ± SD of three replicas. **(B)** Transcription repression in the absence of dCas9. GFP fluorescence of cells expressing crRNAs alone (ΔCas9), relative to the fluorescence of control cells lacking a targeting crRNA (Ø) was measured. Error bars, mean ± SD of three replicas. **(C)** Northern blot with probes annealing either upstream (P511) or downstream (P510) of the T10 and B10 target sites using RNA extracted from cells expressing T5-, T10-, B10-guided dCas9 or a control strain without a target (Ø). Detection of 5S RNA serves as a control.


A
Relative fluorescence
1
0.1
0.01
0.001
-250
-100
-50
0
50
250
350
450
T10
B10
T5
B1
-35
-10
+1
rbs
P511
P510
modified promoter
carrying multiple PAMs
gfp-mut2


B
Relative fluorescence
1.2
0.8
0.4
0
B1
B10
T5
T10
Ø
dCas9
ΔCas9


C
Upstream probe
(P511)
Downstream probeP510
(P510)
Bases
Ø
T5
T10
B10
1,000
500
300
150
5S rRNA

To determine whether DNA binding of the dCas9 complex could prevent transcription elongation, we directed it to the open-reading frame of *gfp-mut2*. A reduction in fluorescence was observed when both the coding and non-coding strands were targeted, suggesting that dCas9 binding could block the elongating RNAP (**Figure 2-22A**). A range of 2.5- to 5-fold reduction in expression was observed when the coding strand was targeted (red spacers), while a range of 6- to 35-fold reduction was observed when the dCas9 complex was directed to the non-coding strand (blue spacers). To directly determine the effects of dCas9 binding on transcription, we extracted RNA from strains expressing the T5, T10 or B10 crRNA guides or a non-targeting dCas9 and subjected it to Northern blot analysis using probes binding before (P511) or after (P510) the B10 and T10 target sites (**Figure 2-22C**). Consistent with our fluorescence measurement, no *gfp-mut2* transcription was detected when the dCas9 complex was directed to the promoter region (T5 target) and lower levels of transcription were observed after the targeting of the T10 region. Interestingly, a smaller transcript was observed with the P511 probe in cells where the complex binds to the T10 or B10 target. These species corresponded to the expected size of a transcript that would be interrupted by the dCas9 complex (calculated as ~250 or 300 nt between the transcription start and the B10 or T10 target sites, respectively). This result is a direct indication that the dCas9 complex caused transcription termination. In accordance with the pronounced decrease in fluorescence caused by B10-bound dCas9 complex, only the truncated *gfp* transcript, but no full-length transcript, was detected with the P511 probe. Altogether, these results

demonstrate that directing the dCas9 complex to different gene regions can prevent both the initiation and elongation of transcription. Targeting of the non-coding strand blocks transcription more efficiently than targeting of the coding strand, suggesting a more efficient displacement of the elongating RNAP by the complex in this configuration.

2.4.2 dCas9 mediates transcription activation

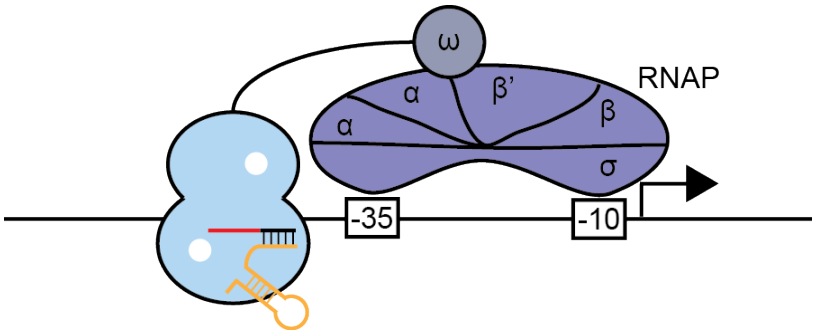


Figure 2-23 A schematic for dCas9-mediated transcription activation.

Transcription activation in *E. coli* using dCas9 fused to the ω subunit of RNA polymerase (RNAP). The dCas9-ω fusion protein is directed to promoter regions, and the ω subunit recruits the RNAP by interacting with the β' subunit. A host with a deletion of *rpoZ*, encoding ω, is used.

The dCas9 complex can be also converted into a transcription activator. Previous work demonstrated that a fusion between the λcI protein and the RNAP omega subunit (ω) can activate transcription by stabilizing the binding of RNAP to a promoter bearing an upstream λ operator (Dove and Hochschild 1998). Therefore we made both C- and N-terminal fusions between the ω subunit and dCas9 (**Figure 2-23**), generating plasmids pWJ66 and pWJ68, respectively. We also expressed different crRNAs to program the binding of both chimeric proteins

to four different positions in a constitutive synthetic promoter controlling the *lacZ* gene (**Figure 2-24A**) in an *E. coli* strain lacking the gene encoding the ω subunit (*rpoZ*). Using β-galactosidase activity as a reporter, we investigated the effect of both chimeras on gene expression. When the binding site was too close to the promoter (e.g. Z1 position), we found binding of the chimeras could still lead to repression (**Figure 2-24B**). With binding sites more distant from the promoter, we observed a modest increase in β-galactosidase activity; in the best case we observed 2.8-fold activation for the dCas9-ω complex (i.e. when ω is fused to the C-terminal of dCas9).

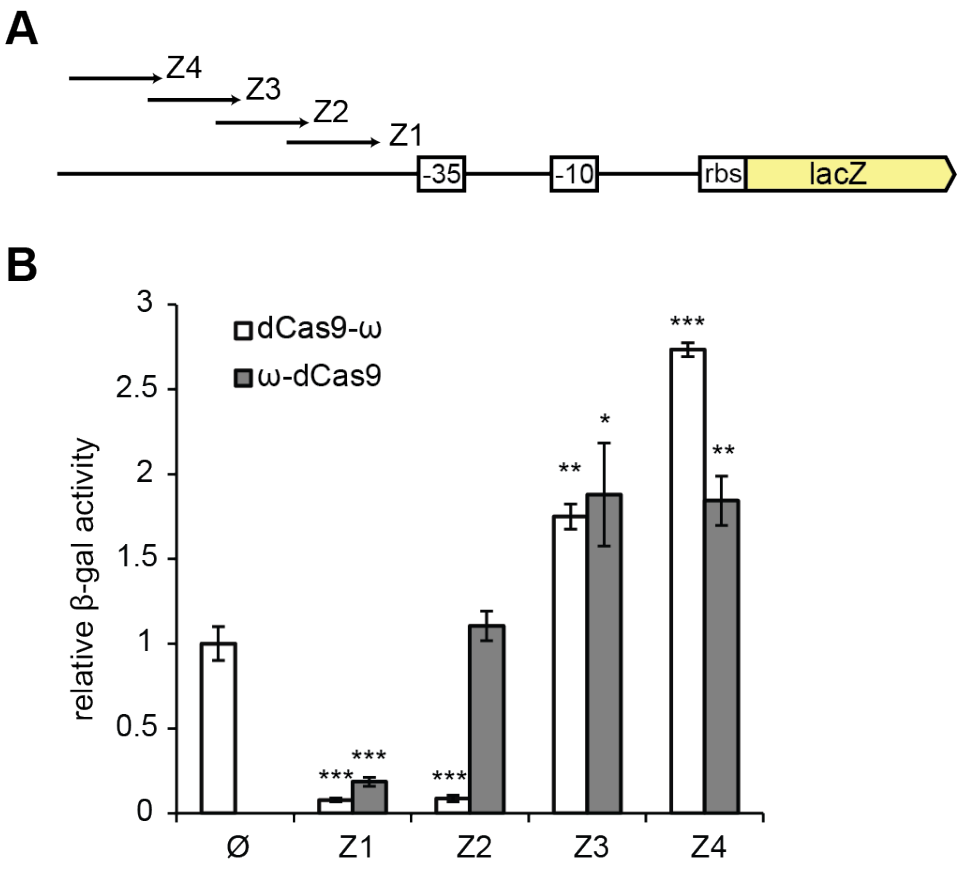


Figure 2-24 dCas9-ω mediates transcription activation

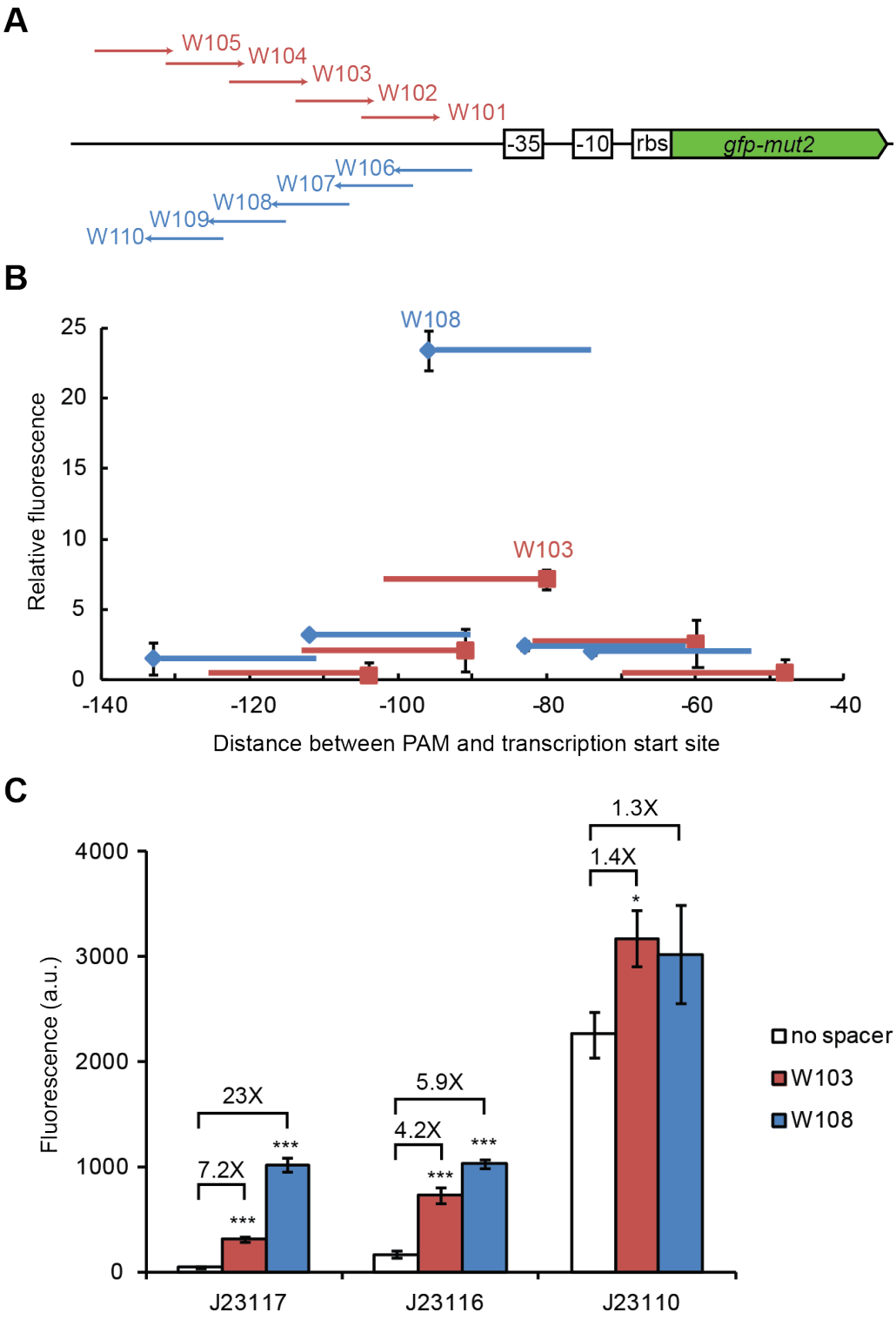
(A) Either N- or C-terminal fusions of the ω subunit to dCas9 were directed to four positions (Z1-Z4) of the top strand upstream of the -35 element of the *lacZ* gene. **(B)** *lacZ* gene expression levels in the different strains were measured as β -galactosidase activity (Miller units). Activation is reported as the relative Miller units normalized against the units obtained with cells expressing a C-terminal dCas9-ω fusion but no crRNA guide (Ø). Error bars indicates mean ± SD of three replicas. Asterisks indicate the P-values associated with each measurement compared with cells expressing a C-terminal dCas9-ω fusion but no crRNA guide (Ø). **P*≤0.05; ***P*≤0.005; ****P*≤0.001.

We thus decided to further investigate the activation capabilities of dCas9-ω when targeted to regions increasingly distant from the -35 promoter element as

well as to both DNA strands (**Figure 2-25A**). To facilitate measurements we used a GFP reporter plasmid, pWJ89, with the *gfp-mut2* gene under the control of a weak biobrick promoter (BBa_J23117) that is preceded by a sequence rich in NGG PAM sequences on both strands. Among ten tested binding sites, two (specified by W103 and W108 crRNAs, each targeting a different strand) were found to strongly activate *gfp-mut2* (**Figure 2-25B**). These were located 45 and 56 nt away from the -35 element (80 and 96 nt upstream of the transcription start site), and provided a 7.2- and 23-fold induction, respectively. A similar level of induction was observed when fluorescence is measured at different growth phases (data not shown). These data suggest that the dCas9-ω complex can induce gene expression when it binds at an optimal distance from the promoter.

Figure 2-25 dCas9-ω mediates transcription activation (continued)

(A) dCas9-ω was directed to ten positions (W101-W110) of both strands upstream of the -35 element of the *gfp-mut2* gene. **(B)** GFP fluorescence of cells expressing dCas9-ω guided to different positions shown in **(A)**, relative to the fluorescence of cells expressing a non-targeting dCas9-ω, was plotted as a function of the position of the target sequence within the *gfp-mut2* upstream region (+1, transcription start). Squares indicate the PAM position, lines the extension of complementarity between the crRNA guide and the reporter gene; red lines, top strand targets; blue, bottom strand. Error bars, mean ± SD of three replicas. **(C)** Transcription activation of *gfp-mut2* containing promoters of three different strengths (J23117, J23116 and J23110) by W103- or W108-guided dCas9-ω. The relative induction, compared with the fluorescence of cells expressing a non-targeting dCas9-ω, is shown. Error bars indicate mean ± SD of three replicas. Asterisks indicate the P-values associated with each measurement compared with the “no spacer” control. *$P \leq 0.05$; **$P \leq 0.005$; ***$P \leq 0.001$.


A
W105
W104
W103
W102
W101
-35
-10
rbs
gfp-mut2
W106
W107
W108
W109
W110
B
W108
W103
Relative fluorescence
25
20
15
10
5
0
-140
-120
-100
-80
-60
-40
Distance between PAM and transcription start site
C
Fluorescence (a.u.)
4000
3000
2000
1000
0
23X
7.2X
5.9X
4.2X
1.3X
1.4X
***
*
J23117
J23116
J23110
no spacer
W103
W108

To test whether dCas9-ω can induce expression of *gfp-mut2* under the control of stronger promoters, we replaced the weak promoter in pWJ89 (BBa_J23117) with promoters of intermediate (BBa_J23116) and high (BBa_J23110) strength (**Figure 2-25C**), generating plasmids pWJ96 and pWJ97, respectively. We compared dCas9-ω-mediated activation of *gfp* expression from the three constructs using the W103 and W108 crRNA guides. For both binding sites, all three promoters could be induced by the dCas9-ω complex; however the relative induction diminished as the promoter became stronger. Altogether, these results show that dCas9 can be used to activate gene expression, with the possibility of achieving different levels of activation depending on the strength of the targeted promoter (the best induction being obtained with weak promoters).

2.4.3 Discussion

Here I described the use of an RNA-guided DNA binding protein, dCas9, to either repress or activate genes in *E. coli*. In this method dCas9 can be directed to any region of the bacterial chromosome that is specified by the base-pair complementarity between the crRNA guide and the cognate genomic sequence.

Repression was achieved by directing the dCas9 complex to either promoter or open reading frame regions. While binding of the dCas9 complex to promoters prevented transcription initiation, binding to the open reading frame prevented elongation, especially when the non-coding strand is targeted. In general, dCas9 complex mediates stronger repression when directed to the promoter region than to the open reading frame.

Compared to sRNA-mediated transcription repression (Chapter 2.2.1), in which repression occurs at the mRNA level, CRISPR-mediated repression occurs at the DNA level, by either blocking transcription initiation or elongation. Being mechanistically different, the CRISPR method can complement the sRNA method, and possibly allow researchers to explore new avenues of research that were previously untenable. In addition, since many randomly designed crRNA guides that target the promoter region can effectively repress gene expression, it seems CRISPR-mediated repression can be more reliable and predictable than sRNAs (see Chapter 2.2.1). Future studies should thoroughly compare the efficacy and predictability of the two methods in parallel.

We also converted the dCas9 complex to a transcription activator by fusing dCas9 to the ω subunit of RNAP, which was previously shown to provide effective recruitment of RNAP (Dove and Hochschild 1998). Contrary to dCas9-complex-mediated repression, in which directing of the complex by many crRNAs to a large region surrounding the promoter (-60 ~ +40) led to strong repression (**Figure 2-22A**), strong activation was only achieved when guiding the dCas9-ω-complex to a much narrower region spanning from -100 to -75 (**Figure 2-25B**). Among ten tested crRNA guides, only two showed strong activation. This shows that activation of a gene of interest will be highly contingent upon the availability of PAM sequences in that defined region. Given the strict position requirement for gene activation, it may be necessary to harness the power of Cas9 orthologs that can recognize different PAM sequences. In a recent study, the Cas9 proteins derived from *Streptococcus thermophilus*, *Neisseria meningitidis* and *Treponema*

denticola, each recognizing a different PAM, were converted to dCas9. The study showed that these dCas9 proteins are orthogonal, and can mediate simultaneous targeted gene regulation in bacterial and human cells (Esvelt et al. 2013). Another study reported that by applying selection-based directed evolution, the *S. pyogenes* Cas9 can be modified to recognize alternative PAM sequences (Kleinstiver et al. 2015). These natural and engineered Cas9 proteins expand targetable sequences that are useful for genome editing and gene regulation.

A maximum of a 23-fold induction was achieved when directing a crRNA to target the bottom strand with a PAM positioned 56 nt upstream of the -35 element (**Figure 2-25B**), suggesting this distance may be optimal for activation. Still, the level of activation is lower than the reported 70-fold induction by using an engineered λ operator and λcI-ω fusion protein (Dove and Hochschild 1998), though a side-by-side comparison is lacking. We think that activation can be further optimized by changing the protein linker between dCas9 and ω and/or testing activation domains other than ω. Importantly, the main advantage of CRISPR-mediated gene activation is that it circumvents the need to modify the promoter sequence of the gene of interest, as was required for conventional transcription activation by a recruitment method such as the λcI-ω fusion.

CHAPTER 3 TYPE III CRISPR-CAS SYSTEM

Type III CRISPR-Cas systems are defined by the presence of genes encoding Cas10 and repeat-associated mysterious protein (RAMP) modules Csm or Cmr for type III-A or III-B sytems, respectively (Makarova et al. 2011), which together form the Cas10-Csm or Cas10-Cmr complexes (Hale et al. 2009; Zhang et al. 2012; Hatoum-Aslan et al. 2013). Among all CRISPR types, the type III systems are possibly the most complex. In contrast to type I and type II systems, type III CRISPR-Cas displays at least two distinct targeting mechanisms. First, transcription across the target is required for immunity. Second, the system performs co-transcriptional cleavage of both DNA and RNA targets (see Chapter 1.2.3).

In this chapter, I will describe my published work (Jiang et al. 2016) that studies the type III-A system of *Staphylococcus epidermidis* RP62a (Gill et al. 2005). Previous work conducted by Gregory Goldberg and Poulami Samai from our lab have elucidated key mechanistic questions with regard to co-transcriptional DNA and RNA cleavage. However, while it is well known that DNA targeting is essential for immunity in all types of CRISPR-Cas, the function of the RNA-targeting module in type III systems has been elusive. In my research, I identified a new CRISPR-associated RNase – Csm6. Along with Csm3, a previously characterized RNase, Csm6 plays an important role in type III-A CRISPR immunity. In particular, I demonstrated that RNA targeting by one or both proteins provides a second line of defense – it is essential for cell survival

when target sequences recognized by CRISPR-Cas reside in late-expressed phage genes or regions that bear mismatches to the crRNAs.

3.1 Functionality of type III-A CRISPR-Cas of *Staphylococcus epidermidis*

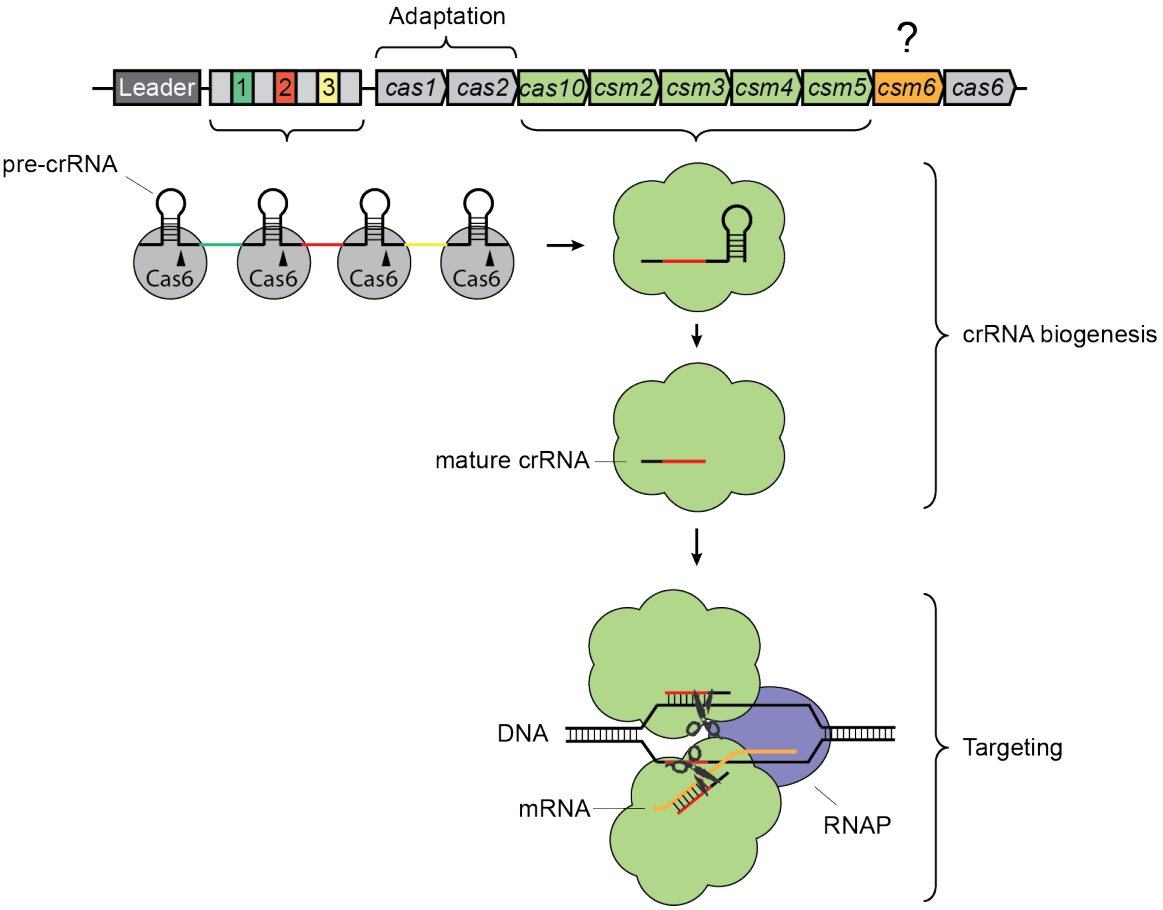


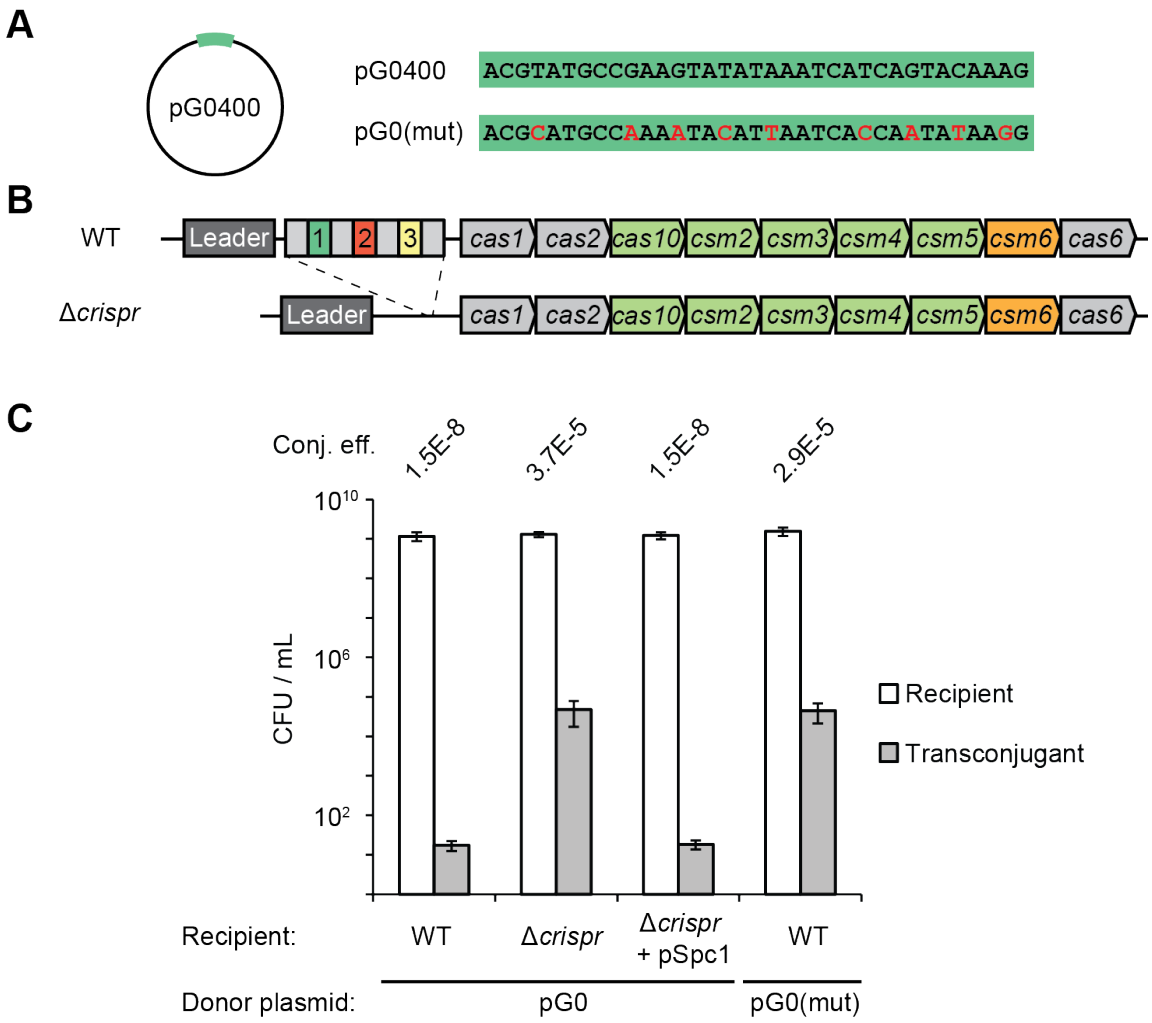
Figure 3-1 The type III-A CRISPR-Cas locus of *Staphylococcus epidermidis* RP62a

The type III-A CRISPR-Cas locus of *Staphylococcus epidermidis* RP62a harbors three spacers and nine *cas* genes. *cas1* and *cas2* are present in most CRISPR-Cas systems and participate in the adaptation phase. *cas6* encodes an endoribonuclease that processes the crRNA precursor (pre-crRNA) into small crRNA guides. *cas10*, along with *csm2*, *csm3*, *csm4* and *csm5* encode a protein complex responsible for crRNA maturation and targeting. The type III-A system cleaves both DNA and RNA when transcription happens across the target. *csm6* is the only gene that has not been characterized in detail.

The type III-A CRISPR-Cas of *Staphylococcus epidermidis* RP62a harbors three spacers and nine *cas* genes (**Figure 3-1**) (Gill et al. 2005; Marraffini and Sontheimer 2008). The first spacer (*spc1*) matches a region of the *nickase* (*nes*) gene found in a staphylococcal conjugative plasmid, pG0400 (**Figure 3-2A**) (Morton et al. 1995). To demonstrate that CRISPR-Cas can act as a defense system and that *spc1* can prevent plasmid conjugation into *S. epidermidis* RP62a, the CRISPR array was deleted from the chromosome, generating strain *Δcrispr* (**Figure 3-2B**). As shown in **Figure 3-2C** and a seminal study by my advisor (Marraffini and Sontheimer 2008), wild-type cells were refractory to pG0400 transfer. The few transconjugants recovered were subjected to thorough genotyping and all contained mutation in the CRISPR array or *cas* genes (Jiang et al. 2013b). In contrast, conjugation efficiency was at least three orders of magnitude higher in *Δcrispr* cells (**Figure 3-2C**). Complementation of pSpc1, a plasmid carrying *spc1* restored the CRISPR immunity (**Figure 3-2C**). These data indicates that *spc1* and the *cas* genes are involved in preventing plasmid conjugation. To further confirmed that the protospacer sequence on pG0400 was important for recognition by *spc1*, the target was introduced with nine silent mutations, generating plasmid pG0(mut) (**Figure 3-2A**) (Marraffini and Sontheimer 2008). Indeed, transfer of pG0(mut) was no longer blocked by wild-type cells, and the conjugation efficiency was as high as that of pG0400 into *Δcrispr* cells (**Figure 3-2C**).

Figure 3-2 Type III-A CRISPR-Cas has anti-plasmid activity

(A) A staphylococcal conjugative plasmid, pG0400 harbors a protospacer sequence that matches Spacer 1 (*spc1*) of the CRISPR array of *Staphylococcus epidermidis* RP62a. The protospacer sequence was altered to contain nine synonymous mutations shown in red, generating plasmid pG0(mut). **(B)** Schematics of the wild-type CRISPR-Cas or Δ*crispr* locus (harboring deletion of the CRISPR array) of *Staphylococcus epidermidis* RP62a. **(C)** Conjugation experiments were performed by mixing recipient cells carrying different genetic background (WT: wild-type; Δ*crispr*: deletion of CRISPR array; Δ*crispr* + pSpc1: deletion of CRISPR array complemented with *spc1* on a plasmid) and donors cells carrying pG0400 or pG0(mut). Colony-forming units (CFU) per milliliter (mL) values (mean ± SD of three replicas) obtained for recipients and transconjugants are shown. Conjugation efficiency (Conj. Eff.) was calculated as the mean transconjugants / mean recipients ratio.

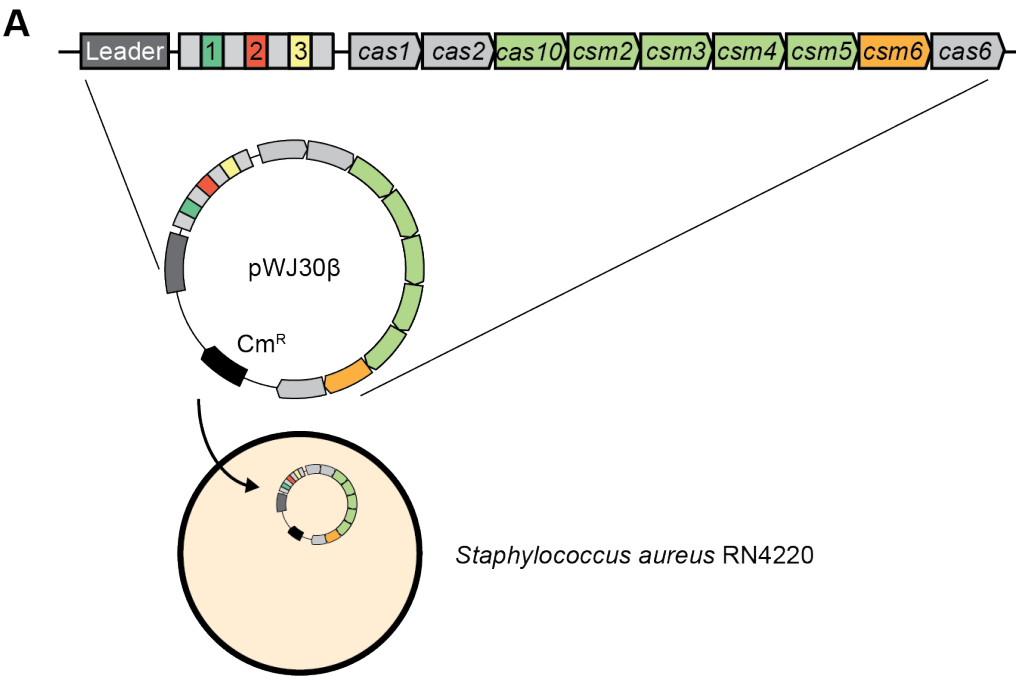


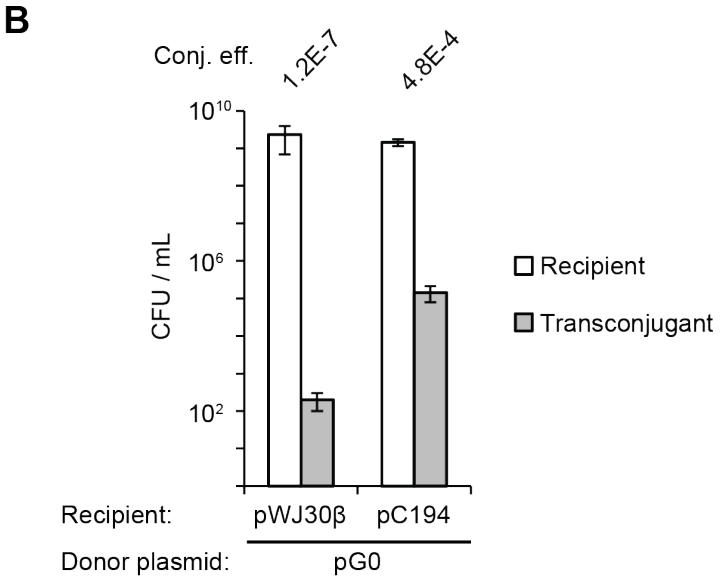
Genetic approaches are required to understand the different function of *cas* genes in CRISPR immunity. However, reagents for genetic manipulation such as availability of antibiotic markers and efficient transformation/transduction methods had been lacking in *S. epidermidis*. In addition, we were unable to identify any bacteriophages, natural immunogens of the CRISPR-Cas system, to infect our strain. In order to facilitate future studies of the CRISPR immune response, I cloned the entire type III CRISPR-Cas system onto a plasmid, generating pWJ30β (**Figure 3-3A**). I decided to study the plasmid-borne CRISPR-Cas in *Staphylococcus aureus*, a close relative of *S. epidermidis* that is

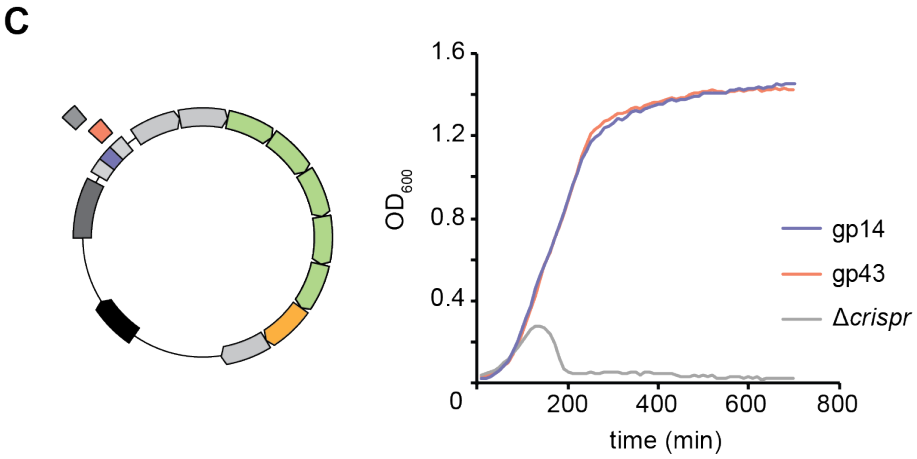
much more amenable to genetic manipulation. Importantly, a few phages capable of infecting *S. aureus* had been characterized (Bae et al. 2006). The plasmid-borne CRISPR-Cas system was functional as an immune system and had both anti-plasmid and anti-phage activity (**Figure 3-3B and C**).

Figure 3-3 A plasmid-borne type III-A CRISPR-Cas system is functional in *Staphylococcus aureus*

(A) The type III-A CRISPR-Cas system of *Staphylococcus epidermidis* RP62a is cloned onto plasmid pC194 (Horinouchi and Weisblum 1982), generating pWJ30β, which is subsequently introduced to *Staphylococcus aureus* RN4220. **(B)** To test functionality of the plasmid-borne CRISPR-Cas system, conjugation experiments were performed by mixing recipient cells (RN4220 cells harboring pWJ30β or parental plasmid pC194) and donor cells carrying pG0400. Colony-forming units (CFU) per milliliter (mL) values (mean ± SD of three replicas) obtained for recipients and transconjugants are shown. Conjugation efficiency (Conj. Eff.) was calculated as the mean transconjugants / mean recipients ratio. **(C)** Functionality of the plasmid-borne CRISPR-Cas system was subjected to phage infection assay. First, the CRISPR array of pWJ30β was replaced by spacers matching the *gp14*, *gp43* genes of lytic phage ΦNM1γ6 (Goldberg et al. 2014), and a non-matching spacer (Δ*crispr*), generating pWJ245, pWJ191 and pGG-BsaI-R (Goldberg et al. 2014), respectively. Liquid cultures of RN4220 cells harboring these individual plasmids were infected with ΦNM1γ6 (at 0hr). Optical density at 600 nm (OD_{600}) was measured for the following 12 hr to monitor cell survival. Representative growth curves of at least three independent experiments are shown.


A
Leader
1
2
3
cas1
cas2
cas10
csm2
csm3
csm4
csm5
csm6
cas6
pWJ30β
Cm^R
Staphylococcus aureus RN4220


B
Conj. eff.
1.2E-7
4.8E-4
10^10
10^6
10^2
CFU / mL
Recipient
Transconjugant
Recipient:
pWJ30β
pC194
Donor plasmid:
pG0


C
1.6
1.2
0.8
0.4
0
OD600
200
400
600
800
time (min)
gp14
gp43
Δcrispr

The plasmid system I established allowed our lab to investigate the type III CRISPR-Cas system more easily. First, a biochemical study performed by a former post-doc, Asma Hatoum-Aslan found that Cas10, Csm2, Csm3, Csm4 and Csm5 can be co-purified with each other in *Staphylococcus*, suggesting these proteins form a complex (Hatoum-Aslan et al. 2013). Next, mature crRNAs of various lengths were found to be associated with the complex (Hatoum-Aslan et al. 2013). Individual deletion of *cas10*, *csm3* or *csm4*, but not *csm2* or *csm5* abolished complex formation and thus crRNA maturation; deletion of *csm2* or *csm5* individually compromised crRNA maturation (Hatoum-Aslan et al. 2014). Moreover, individual deletion of all these genes abrogated anti-plasmid activity, suggesting complex formation and crRNA maturation were required for immunity (Hatoum-Aslan et al. 2014). The Cas10-Csm::crRNA ribonucleoprotein complex (hereafter referred to as Cas10-Csm complex) could be also purified from *E.coli*, and was shown to possess DNA- and RNA-targeting activity specified by the sequence of the crRNA in vitro (Samai et al. 2015). These data indicate that Cas10, Csm2, Csm3, Csm4 and Csm5 proteins form a complex that is necessary for crRNA maturation and targeting (**Figure 3-1**).

Supplementary to the characterization of the Cas10 – Csm5 proteins, it is established that Cas1 and Cas2 are involved in the adaptation phase, and are dispensable for crRNA biogenesis and targeting (Heler et al. 2014). Cas6 is an endoribonuclease that participates in the primary processing of the crRNAs (**Figure 3-1**) (Carte et al. 2008). Since these crRNAs then become the substrates for maturation, deletion of Cas6 results in loss of maturation and thus loss of

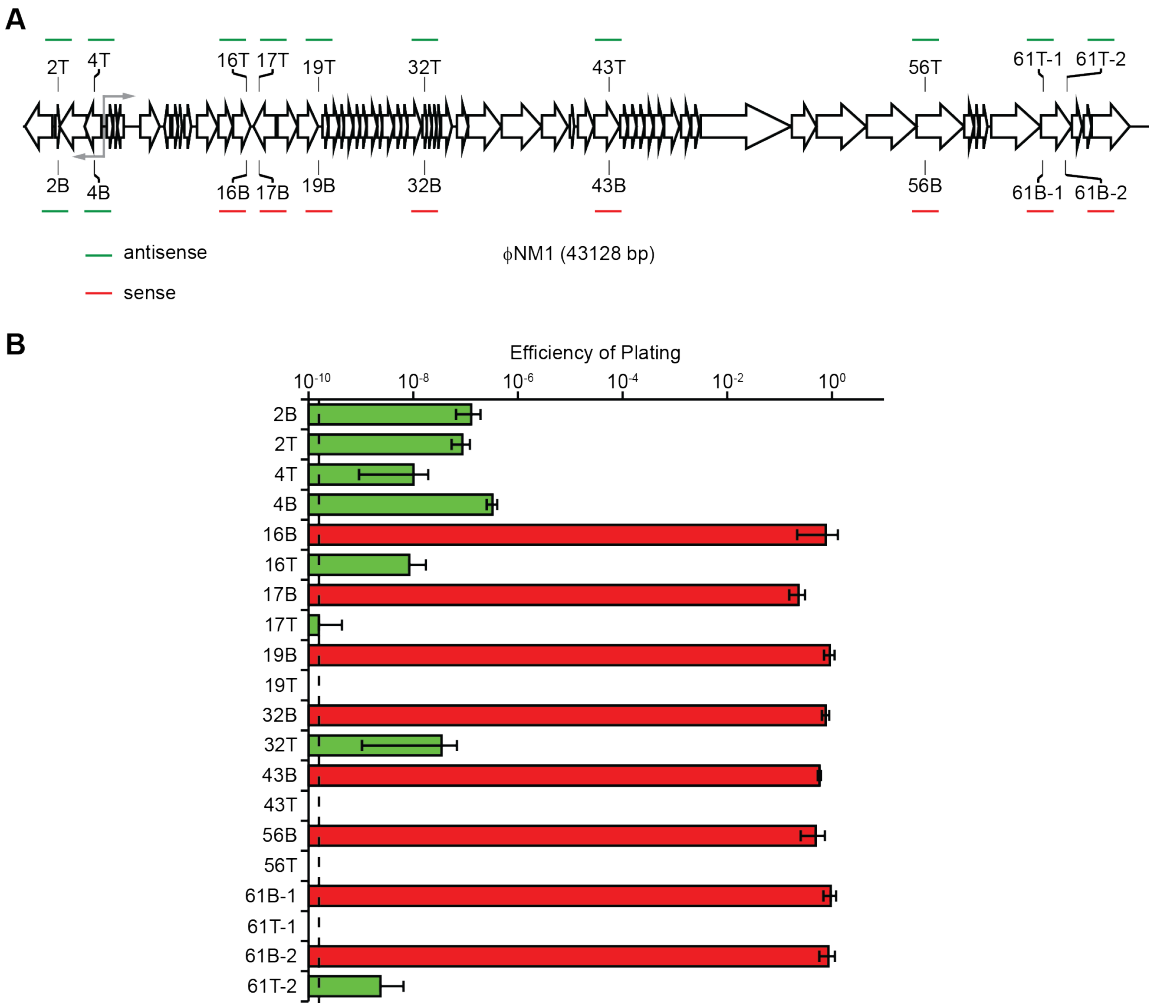
CRISPR immunity (Hatoum-Aslan et al. 2014). The function of *csm6* was unknown. A genetic study from our lab showed that deletion of the gene abolishes anti-plasmid immunity (Hatoum-Aslan et al. 2014). Our lab also showed that Csm6 does not co-purify with the Cas10-Csm complex (Hatoum-Aslan et al. 2013) and that deletion of *csm6* does not disrupt crRNA biogenesis (Hatoum-Aslan et al. 2011). Therefore it has been proposed that *csm6* may be involved in the targeting stage of CRISPR immunity.

3.2 Type III CRISPR-Cas immunity requires transcription

The transcription-dependent requirement for type III CRISPR-Cas was first discovered by Gregory Goldberg in our lab. Using the plasmid-borne version of the *S. epidermidis* type III CRISPR-Cas (see Chapter 3.1), Gregory designed multiple spacers that matched the genome of a dsDNA phage, ΦNM1 (Bae et al. 2006) (**Figure 3-4A**). To his surprise, half of the spacers he designed did not protect host cells from phage infection (**Figure 3-4B**). A close examination of the orientation of spacers and the phage genome revealed that spacers that confer immunity were all antisense to the phage transcripts, while the inactive spacers all represented the sense strand. These experiments suggested that successful CRISPR immunity may depend on target transcription (Goldberg et al. 2014).

Figure 3-4 Transcription is required for the type III-A CRISPR-Cas to target phage

(A) A schematic of the genome of a dsDNA phage, ΦNM1. Twenty spacers that matched the top strand (T) or the bottom strand (B) of ΦNM1 were designed and the positions of their targets are shown. Green spacers are antisense to the phage transcripts, and red spacers have orientations matching the phage sense strand. **(B)** *Staphylococcus aureus* RN4220 cells harboring plasmid-borne CRISPR-Cas are challenged by ΦNM1. The efficiency of plating was used to determine immunity against the phage provided by spacers targeting the phage regions shown in **(A)**. Dotted line indicates the limit of detection of the assay. Error bars indicates mean ± SD of three replicas. Figure is adapted from (Goldberg et al. 2014) with modifications.



One of the key experiments that demonstrated transcription-dependent targeting is a plasmid-curing assay I developed (Goldberg et al. 2014). In this assay, host cells harbored two plasmids (**Figure 3-5A**). pCRISPR-Cas carried the canonical *spc1* and all the *cas* genes, and pTarget contained a target sequence cloned after a tetracycline-inducible promoter, P_{tet}. The two plasmids were compatible in the absence of the inducer, anhydrotetracycline (aTc); however, when aTc was added, induction of transcription across the target led to CRISPR targeting and elimination of the target plasmid (**Figure 3-5B**). Additional evidence such as tolerance of lysogenization, phage escaper analysis and

transcriptome analysis (Goldberg et al. 2014) supported the conclusion that transcription is required for type III CRISPR-Cas immunity. Moreover, biochemical work performed by Poulami Samai in the lab showed that the purified Cas10-Csm complex cleaves double-stranded DNA targets only when transcription occurs across the target, further corroborating the transcription-dependent targeting model in vitro (Samai et al. 2015).

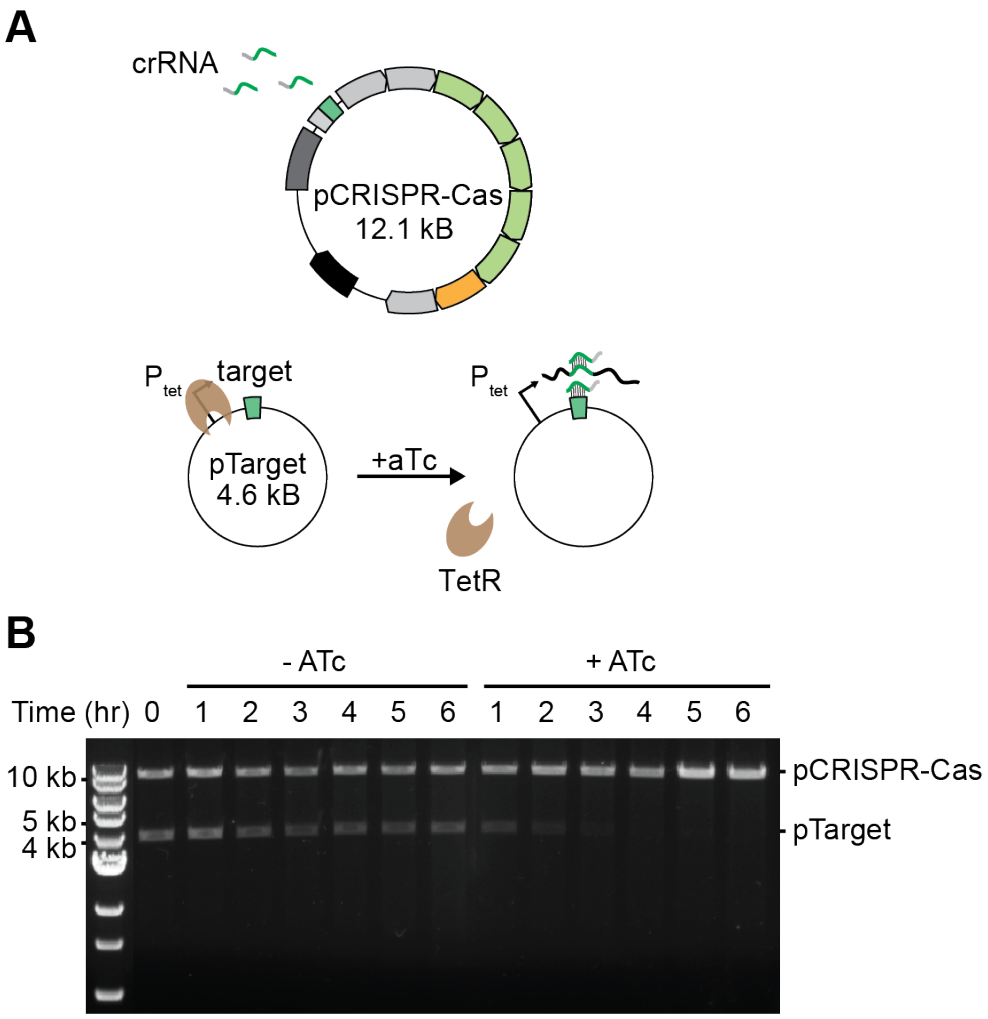


Figure 3-5 Transcription is required for the type III-A CRISPR-Cas to target plasmids

(A) The plasmid-curing assay used *Staphylococcus aureus* RN4220 harboring two plasmids. pCRISPR-Cas carried the conical *spc1* and all *cas* genes of *Staphylococcus epidermidis* RP62a. pTarget (pWJ153) carried a target sequence inserted after a tetracycline-inducible promoter (P_{tet}). Addition of anhydrotetracycline (aTc) induces transcription across the target sequence by derepressing the Tet repressor protein (TetR). **(B)** Agarose gel of linearized plasmid DNA purified from aTc-treated (+aTc) and untreated (-aTc) liquid cultures at the indicated time points.

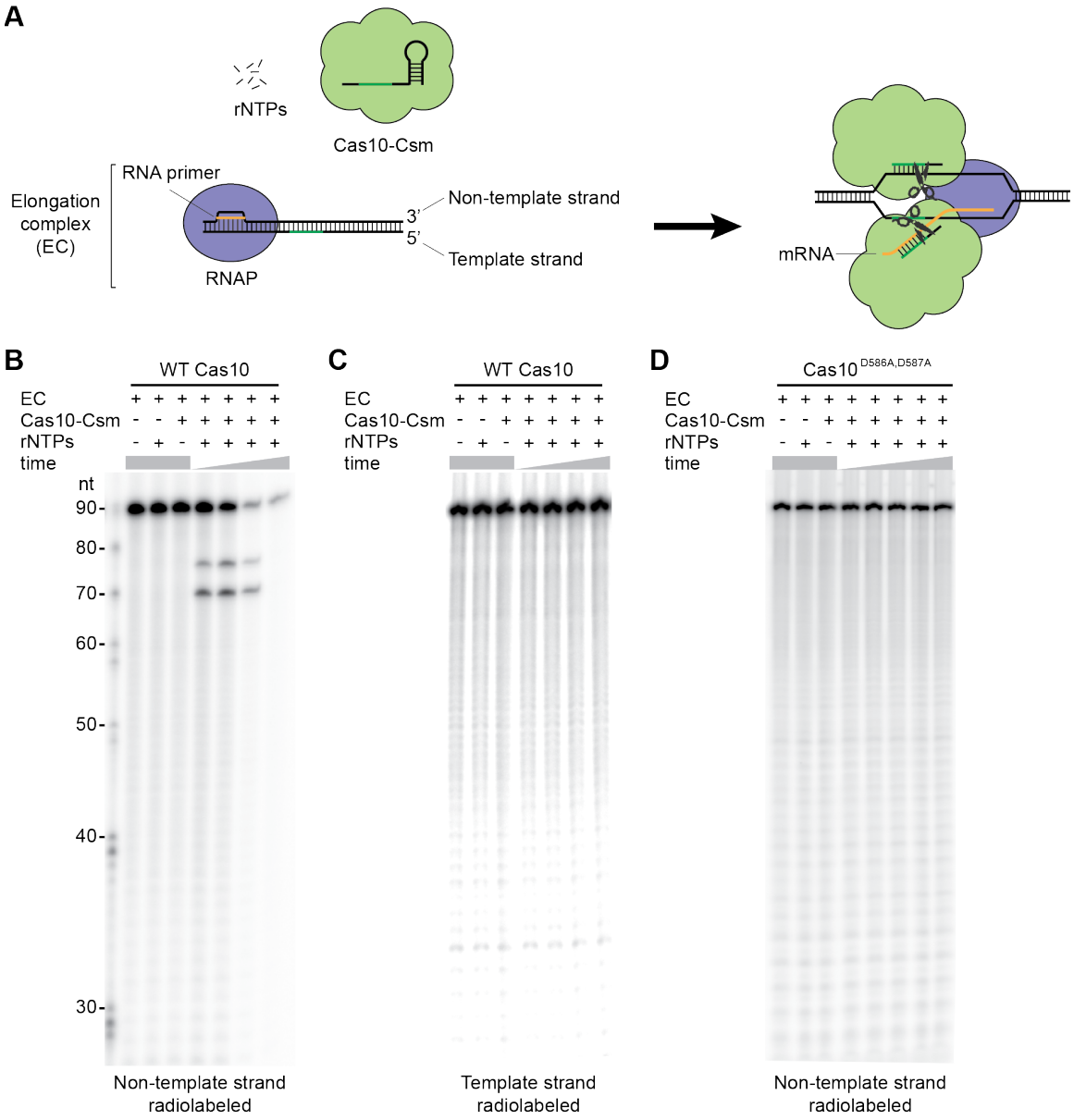
3.3 Type III CRISPR-Cas targets both DNA and RNA

The identity of the target nucleic acid for type III CRISPR-Cas systems was controversial at a time. In vivo genetic studies done by our lab showed the type III-A system of *S. epidermidis* targets DNA (Marraffini and Sontheimer 2008; Goldberg et al. 2014; Hatoum-Aslan et al. 2014). However, various other groups showed that type III-B systems of *Pyrococcus furiosus* (Hale et al. 2009), *Sulfolobus solfataricus* (Zhang et al. 2012) and *Streptococcus thermphilus* (Zebec et al. 2014) all target RNA in vitro. Therefore, it was thought that the III-A and III-B subtypes were functionally different, as the nucleic acids recognized and cleaved by the systems are DNA and RNA, respectively. Two subsequent reports found that the III-A systems of *Streptococcus thermphilus* (Tamulaitis et al. 2014) and *Thermus thermophilus* (Staals et al. 2014) also target RNA in vitro, further fueling the controversy. However, the aforementioned work conducted by Poulami settled the discussion by demonstrating that the type III-A complex of *S. epidermidis* can cleave both DNA and RNA targets, in vitro and in vivo (Samai et al. 2015). This finding is reinforced by two subsequent studies, both confirming the dual DNA- and RNA-targeting activity of the III-B systems (Elmore et al. 2016; Estrella et al. 2016). Poulami's work also identified active sites within the staphylococcal Cas10-Csm complex responsible for DNA and RNA targeting: the D586, D587 residues of Cas10 (**Figure 3-6**) and D32 residue of Csm3 (**Figure 3-7**), respectively. Furthermore, mutations that abrogate DNA cleavage did not affect RNA cleavage and vice versa, strongly suggesting that DNA- and RNA-

targeting functions of the Cas10-Csm complex are independent of each other (Samai et al. 2015).

Figure 3-6 Co-transcriptional DNA cleavage by purified Cas10-Csm complex

(A) Schematic of the co-transcriptional DNA cleavage assay. First, an elongation complex (EC) is assembled in the following order: template strand, RNA primer, RNA polymerase (RNAP) and non-template strand. Next the purified Cas10-Csm complex containing a matching crRNA (derived from *spc1*) and ribonucleoside tri-phosphates (rNTPs) are added to the EC to reconstitute the co-transcriptional DNA cleavage in vitro. **(B, C)** Denaturing PAGE and autoradiography of the products of two co-transcriptional dsDNA cleavage assays differing in the location of the radioactive label: **(B)** non-template strand; **(C)** template strand. **(D)** Same as **(B)** but using a mutant complex containing $Cas10^{D586, D587}$. Panels **(B-D)** are adapted from (Samai et al. 2015) with modifications.


A
rNTPs
Cas10-Csm
RNA primer
Elongation complex (EC)
RNAP
3'
5'
Non-template strand
Template strand
mRNA
B
WT Cas10
EC
Cas10-Csm
rNTPs
time
nt
90
80
70
60
50
40
30
Non-template strand radiolabeled
C
WT Cas10
Template strand radiolabeled
D
Cas10 D586A,D587A
Non-template strand radiolabeled

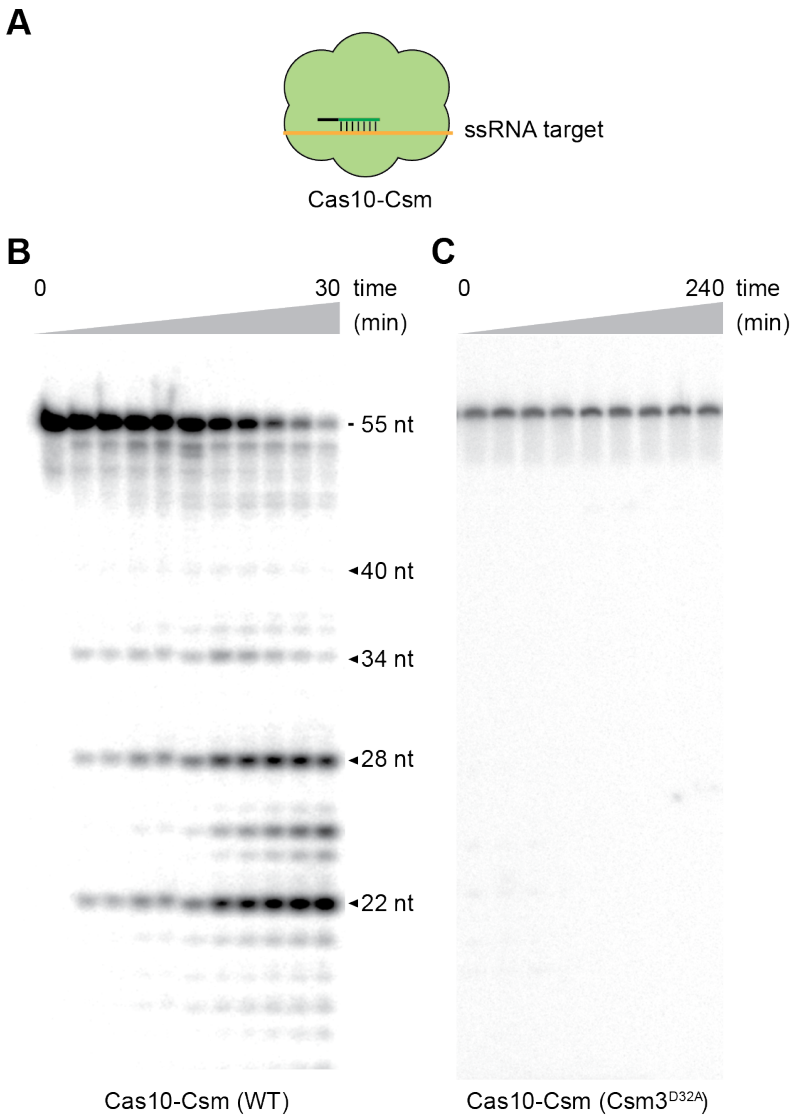


Figure 3-7 RNA cleavage by purified Cas10-Csm complex

(A) Schematic of the RNA cleavage assay. A 5' radiolabeled ssRNA substrate is mixed with purified Cas10-Csm complex containing a matching crRNA (derived from *spc1*). (B) Reaction products of the RNA cleavage assay are collected at 0, 1, 2, 3, 4, 5, 7.5, 10, 15, 20, and 30 min, separated by denaturing PAGE and visualized by phosphorimaging. **(C)** Same as **(B)** but using a mutant complex containing Csm3^{D32A}, and the reaction times are 0, 5, 10, 20, 30, 60, 120, 180 and 240 min. Panels **(B)** and **(C)** are adapted from (Samai et al. 2015) with modifications.

3.4 Discovery of Csm6 as a novel RNase associated with the type III-A CRISPR-Cas system

csm6 is the least characterized among the *cas* genes. A previous study by our lab showed that deletion of the *csm6* abolishes anti-plasmid activity, suggesting a potential role in DNA cleavage (Hatoum-Aslan et al. 2014). Contradicting to this hypothesis, the Csm6 protein does not co-purify with the Cas10-Csm complex, which is sufficient to cleave target DNA in vitro (Samai et al. 2015). A recent bioinformatics analysis of the Higher Eukaryotes and Prokaryotes Nucleotide-binding (HEPN) family reveals that Csm6 is a member of this group and may function as a metal-independent RNase (Anantharaman et al. 2013). To test this possibility, I expressed Csm6 in *E. coli* and purified it to homogeneity (**Figure 3-8A**). The putative active site double mutant R364A, H369A (Anantharaman et al. 2013) was also purified (**Figure 3-8A**). Incubation of the purified proteins with ssRNAs radiolabeled at either end resulted in degradation of the substrate by the wild-type Csm6, but not Csm6$^{R364A,\ H369A}$ or "dead" Csm6 (dCsm6) (**Figure 3-8B**). The reaction did not require any metal cation (Mg, Mn and EDTA were tested and obtained the same cleavage; data not shown). I obtained similar results with the individual active site mutants (**Figure 3-8A**), using substrates of different sequences and lengths (**Figure 3-8C**). These results confirmed that Csm6 is a metal-independent, sequence-independent RNase.

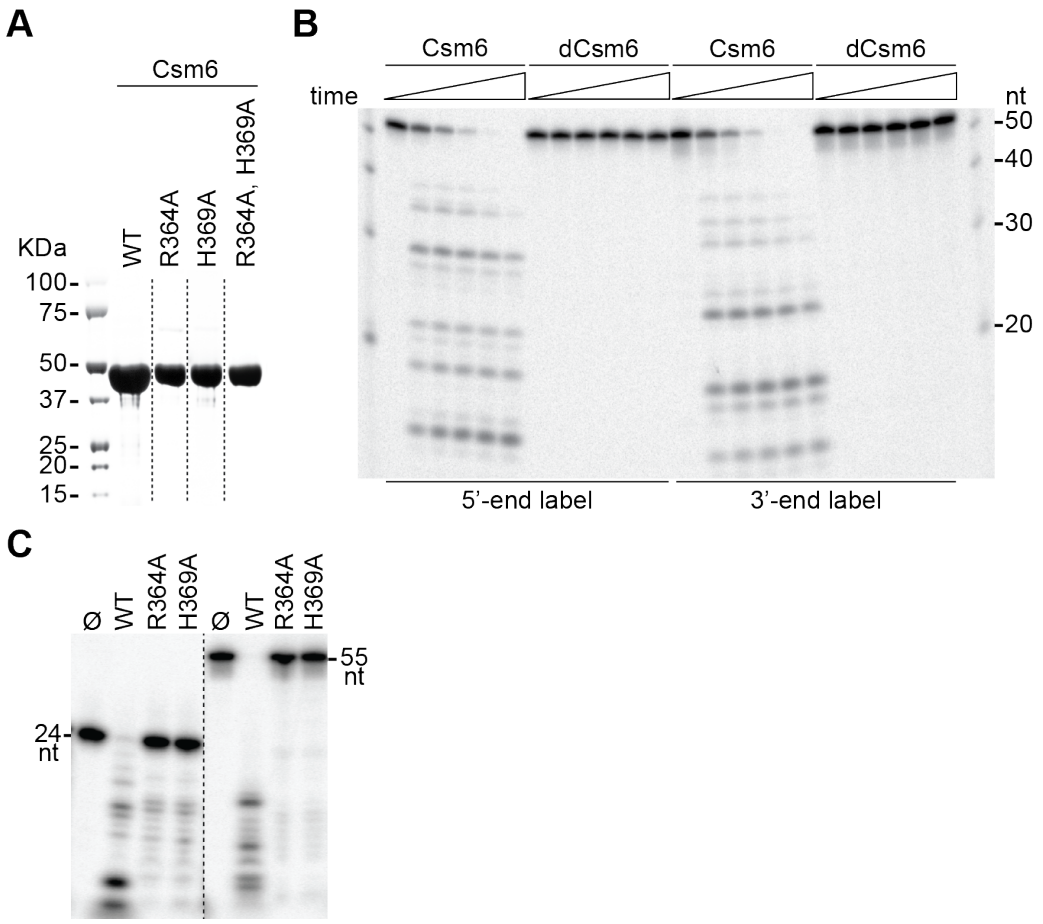


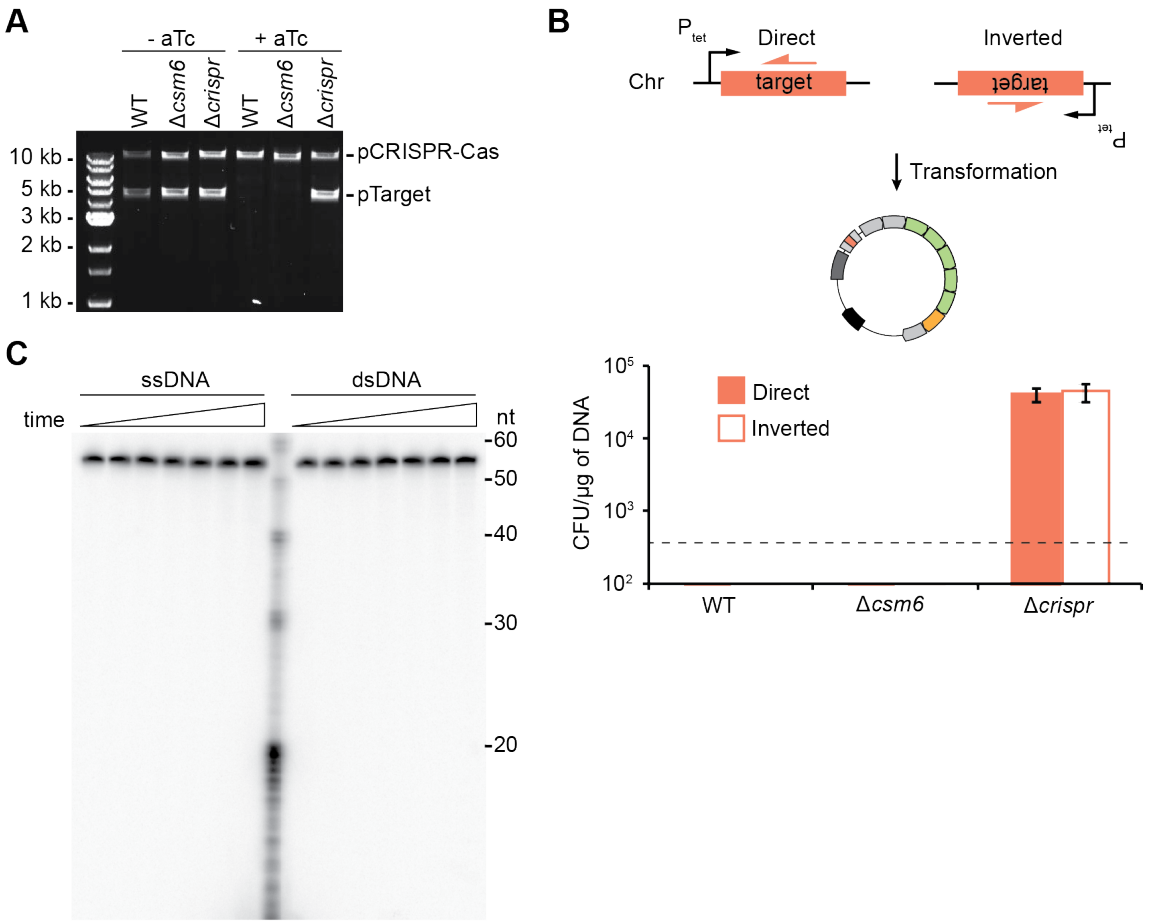
Figure 3-8 RNase activity of Csm6

(A) Csm6 purification. SDS-PAGE of *S. epidermidis* Csm6 and its putative active site mutants, Csm6^{R364A} and Csm6^{H369A} as well as the double mutant, Csm6$^{R364A, H369A}$ were purified from *E.coli*. **(B)** Purified wild-type Csm6 (WT) and Csm6$^{R364A, H369A}$, or "dead" Csm6 (dCsm6) were incubated with a ssRNA substrate (R55) radiolabeled at either end. The reaction proceeded for 1 hr and aliquots were taken at 0, 5, 15, 30, 45 and 60 min for PAGE and phosphorimager visualization. **(C)** Purified Csm6, Csm6^{R364A} and Csm6^{H369A} were incubated with two different 5' radiolabeled ssRNA substrates of 24 and 55 nucleotides in length (R24 and R55, respectively). The reaction proceeded for 1 hr and products were subjected to PAGE and visualized by phosphorimaging. "Ø" indicates a control reaction without enzyme added. Abbreviations: kDA, kilodalton; nt, nucleotide.

Since previous work proposed that Csm6 participates in DNA targeting (Hatoum-Aslan et al. 2014), I tested this idea directly using the plasmid-curing assay I previously developed (**Figure 3-5**). When transcription was induced across the target, the pTarget plasmid was equally degraded by both wild-type CRISPR-Cas and a Δ*csm6* system (**Figure 3-9A**), suggesting DNA degradation does not require *csm6*. To corroborate this finding, I determined that *csm6* is not required for targeting of chromosomal DNA (**Figure 3-9B**). Finally, I tested the activity of Csm6 for ssDNA and dsDNA substrates in vitro, but failed to obtain any cleavage products (**Figure 3-9C**). These results demonstrate that Csm6 is not involved in DNA targeting.

Figure 3-9 Csm6 is not involved in DNA targeting

(A) Plasmid-curing assay schematized in **Figure 3-5A** is using plasmids harboring a wild-type CRISPR-Cas (WT), a Δ*csm6* or a non-targeting CRISPR-Cas system (Δ*crispr*). Plasmid DNA was extracted from cells before and after 10 hr of treatment with the inducer, aTc, linearized and subjected to agarose gel electrophoresis. **(B)** Chromosomal targeting assay. A target sequence (a region of the *gp43* gene from phage ΦNM1) under the control of a tetracycline-inducible promoter, P_{tet}, was inserted to the chromosome (*geh* locus) of *S. aureus* RN4220. The target was placed in both orientations with respect to the origin of replication (direct and inverted insertions). Orange arrows represent crRNAs. Competent cells containing these targets were transformed with different pCRISPR-Cas plasmids carrying wild-type CRISPR-Cas (WT), Δ*csm6* or Δ*crispr* (non-targeting control), and plated in the presence of the aTc inducer. Co-transcriptional cleavage of the target DNA prevents colony formation, presumably through the introduction of lethal chromosomal lesions. Transformation efficiency was measured as the number of colony-forming units (CFU) per µg of plasmid DNA (mean ± SD of three replicas). The dotted line indicates the limit of detection of the assay. **(C)** DNase activity assay for wild-type Csm6. The purified protein was incubated either with a 5' radiolabeled ssDNA oligonucleotide (PS362) or a dsDNA substrate obtained by annealing PS362 and PS363. The reaction proceeded for 2 hr and aliquots were taken at 0, 15, 30, 45, 60, 90 and 120 min for PAGE and phosphorimager visualization. Abbreviations: aTc, anhydrotetracycline; kB, kilobase; chr, chromosome; ssDNA, single-stranded DNA; dsDNA, double-stranded DNA.



Together with previous reports, the results above show that *S. epidermidis* type III-A CRISPR-Cas system encodes two RNases: a sequence-specific, crRNA-guided endoribonuclease, Csm3, and a crRNA-independent ribonuclease, Csm6. While it is well established that DNA targeting is required for CRISPR immunity, the function of RNases has not been fully explored. One report showed that the heterologous expression of the Cas10-Csm complex in *E.coli* could provide immunity against RNA phage MS2 (Tamulaitis et al. 2014). However, since RNA phages are rare and a great majority of phages are dsDNA in nature (Koonin et al. 2015), the function of RNases merited further investigation.

3.5 A challenge for transcription-dependent targeting

What is the function of the RNases encoded by the type III CRISPR-Cas systems? To answer this question, it's worth revisiting the biology of phage lytic cycles. As introduced in Chapter 1.1 and recapitulated in **Figure 3-10**, gene expression of many lytic phages is precisely programmed over time. For instance, genes that are involved in DNA replication and transcriptional regulation are transcribed immediately upon infection. Selective expression of the early genes allows phages to replicate at an optimal pace. Genes that encode various phage parts are transcribed near and/or after replication is complete, which then package the replicated DNA and assemble into hundreds of live phage progenies. Two other important late genes encode lysin and holin, which lyse cell walls and therefore are toxic to the host. Upon host cell lysis, the phage progeny is released and they go on to infect neighboring susceptible cells.

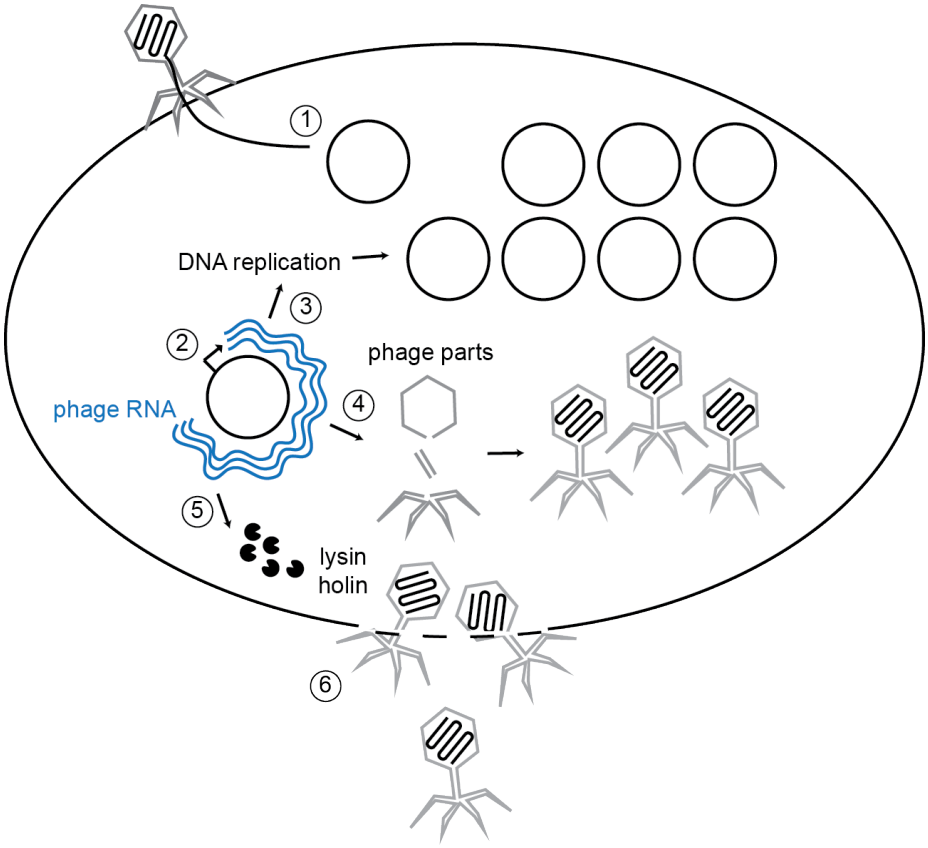


Figure 3-10 Phage lytic cycle

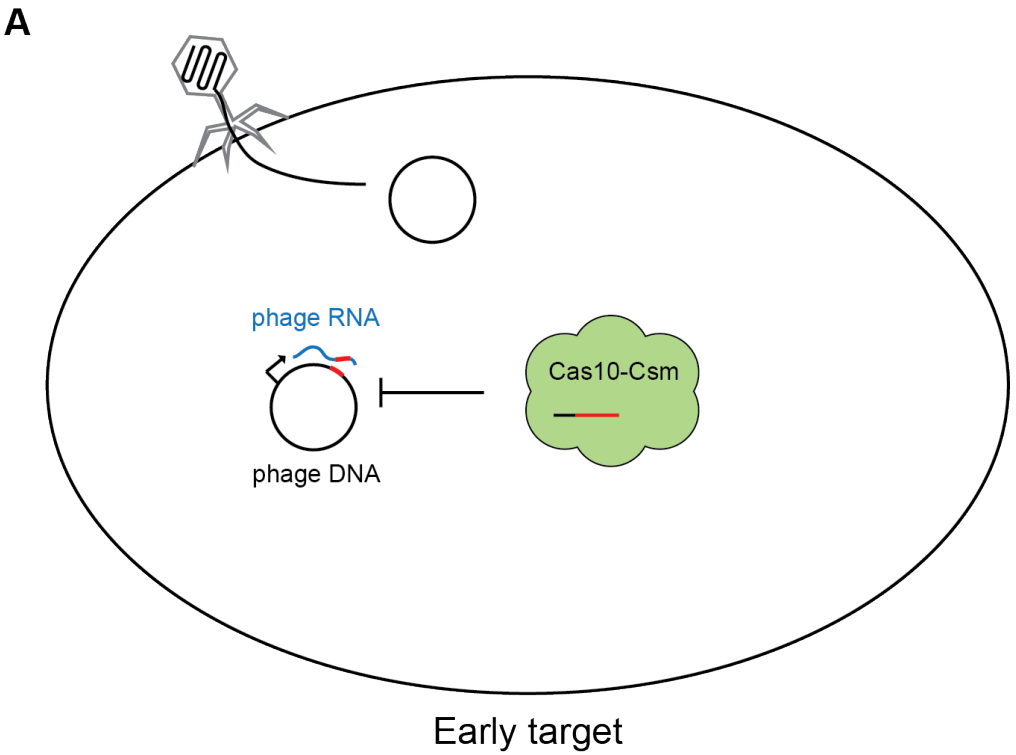
① In general, when a lytic phage infects a cell, it injects its DNA (in this case in a form of dsDNA) which is quickly circularized. ② Transcription starts. The expression of phage genes are temporally regulated. For instance, genes that are involved in DNA replication are transcribed early. ③ This allows for phage replication. ④ Late genes encode various phage parts, which package the replicated DNA and assemble into hundreds of live progenies. ⑤ Two other late genes encode lysin and holin. ⑥ These enzymes lyse host bacteria and liberate infectious phage progeny.

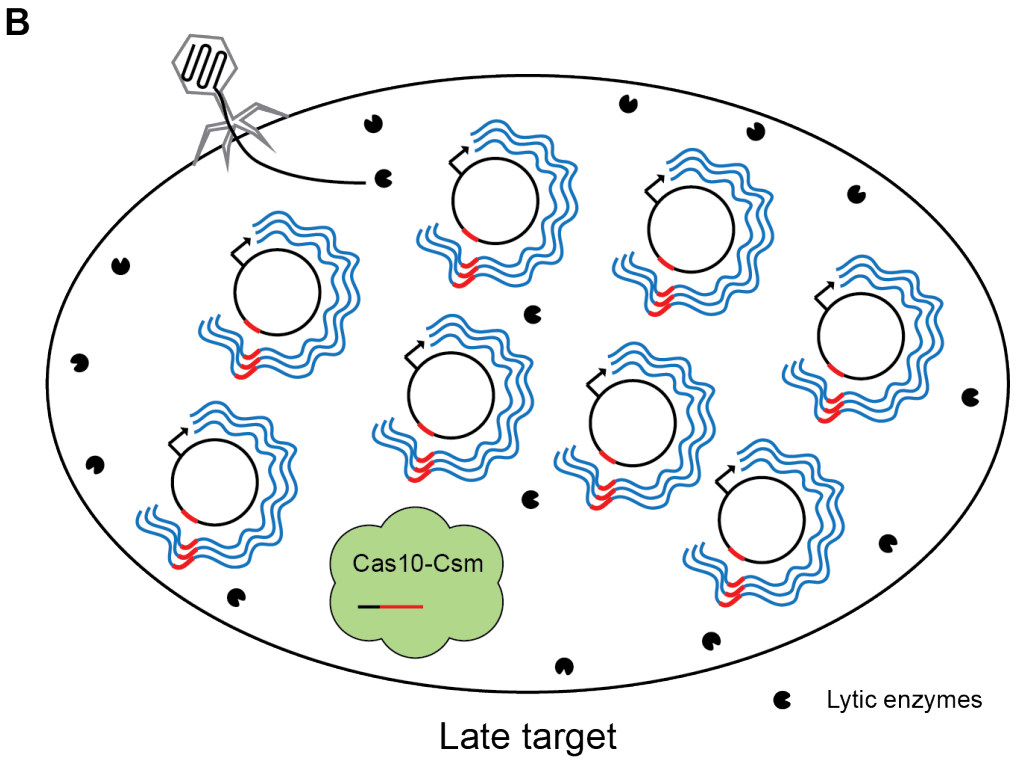
The fact that phage genes are temporally regulated creates a challenge for the transcriptional requirement of the type III CRISPR-Cas system. When CRISPR spacers target early genes, which are immediately transcribed upon

infection, CRISPR can mount an efficient attack while the phage is still at low copy number (**Figure 3-11A**). However, when CRISPR spacers target late genes, the infection cycle can proceed unchecked until the target is transcribed. At this point, the phage may have already finished replication and host cells may be loaded with phage DNA and transcripts (**Figure 3-11B**). This is a challenge because first, CRISPR-Cas may be overwhelmed by the number of target nucleic acids inside the cell. Secondly, if CRISPR-Cas is only capable of targeting DNA, phage transcripts cannot be restricted and will continue to produce toxic lytic enzymes.

Figure 3-11 Models for co-transcriptional type III CRISPR-Cas immunity when targeting early- and late-expressed phage genes

(A) When the target (red segment) is located in an early-expressed phage gene, type III CRISPR-Cas can act soon upon phage DNA injection and clear infection. **(B)** When the target is located in a late-expressed phage gene, CRISPR-Cas can only mount an attack after phage replication. At this stage, host bacteria are full of phage DNA and transcripts. Lytic enzymes such as holin and lysin are poised to lyse the cell and liberate phage progenies.


A
phage RNA
phage DNA
Cas10-Csm
Early target


B
Cas10-Csm
Lytic enzymes
Late target

Based on these properties of the phage lytic cycle, I hypothesize that (i) when type III CRISPR-Cas systems target a late-expressed phage gene, the transcriptional requirement for targeting can lead to accumulation of phage DNA inside host cells and (ii) in this situation, RNA targeting is required to eliminate phage transcripts that would otherwise compromise cell viability.

3.6 Transcriptional requirement of type III-A CRISPR-Cas targeting leads to accumulation of phage DNA

To investigate the fate of phage DNA upon infection, I constructed two type III-A CRISPR-Cas systems: one with a spacer matching an early-expressed phage gene, *gp14* (encoding a protein involved in phage DNA replication) and the other with a spacer matching a late-expressed gene, *gp43* (encoding a phage capsid subunit) (**Figure 3-12A**). Previous transcriptome analysis of *S. aureus* cells infected with lytic phage ΦNM1γ6 in the absence of CRISPR immunity confirmed that *gp14* is an early gene transcribed between 6 minutes and 15 minutes after infection while *gp43* is a late gene transcribed between 15 minutes and 30 minutes after infection (Goldberg et al. 2014). I also used a CRISPR system containing a non-matching spacer as a negative control (Δ*crispr*). In addition to these three different type III spacers, I constructed a type II CRISPR-Cas system of *Streptococcus pyogenes* (Chapter 2) with a spacer matching *gp43* as a positive control, since it is known that type II CRISPR immunity does not require transcription (Gasiunas et al. 2012; Jinek et al. 2012). Next, I infected *S. aureus* cells carrying the four different CRISPR-Cas systems with lytic phage,

ΦNM1γ6. To determine the abundance of phage DNA inside host bacteria, I extracted total DNA at various time points post infection and subjected them to qPCR with primers specific to the phage (**Figure 3-12B**). For quantification, the relative abundance of the phage DNA at 15 minutes post-infection in cells harboring the type II system was set as the reference point (i.e. a value of 1). As a result of CRISPR immunity, phage DNA was rapidly cleared by the type II system and its accumulation was minimal (**Figure 3-12B**), consistent with a previous report showing that phage DNA is immediately degraded upon injection by the type II CRISPR-Cas (Garneau et al. 2010). By contrast, in cells harboring a non-matching CRISPR-Cas system (Δ*crispr*), the abundance of phage DNA increased dramatically with time (**Figure 3-12B**), reflecting the progression of viral replication during the infectious cycle. Phage DNA clearance by type III CRISPR-Cas immunity strongly depended on the region targeted. When the early-expressed gene *gp14* was targeted, the phage DNA was barely detectable (**Figure 3-12B**). In contrast, when the late-expressed gene *gp43* was targeted, substantial phage DNA accumulated (**Figure 3-12B**). For instance, the relative abundance at 0.75 hr was 64x (134 divided by 2.1) more than that of the *gp14*-targeting cells at the same time point. Phage DNA reached a very high level for the first 1.5 hr before it started a slow decline, in contrast to near-instant clearance in the *gp14*-targeting cells. A complete clearance was only achieved when cells at 9 hr post-infection were diluted and refreshed in new culture broth and grown for an additional 9 hr (**Figure 3-12B** R9 point).

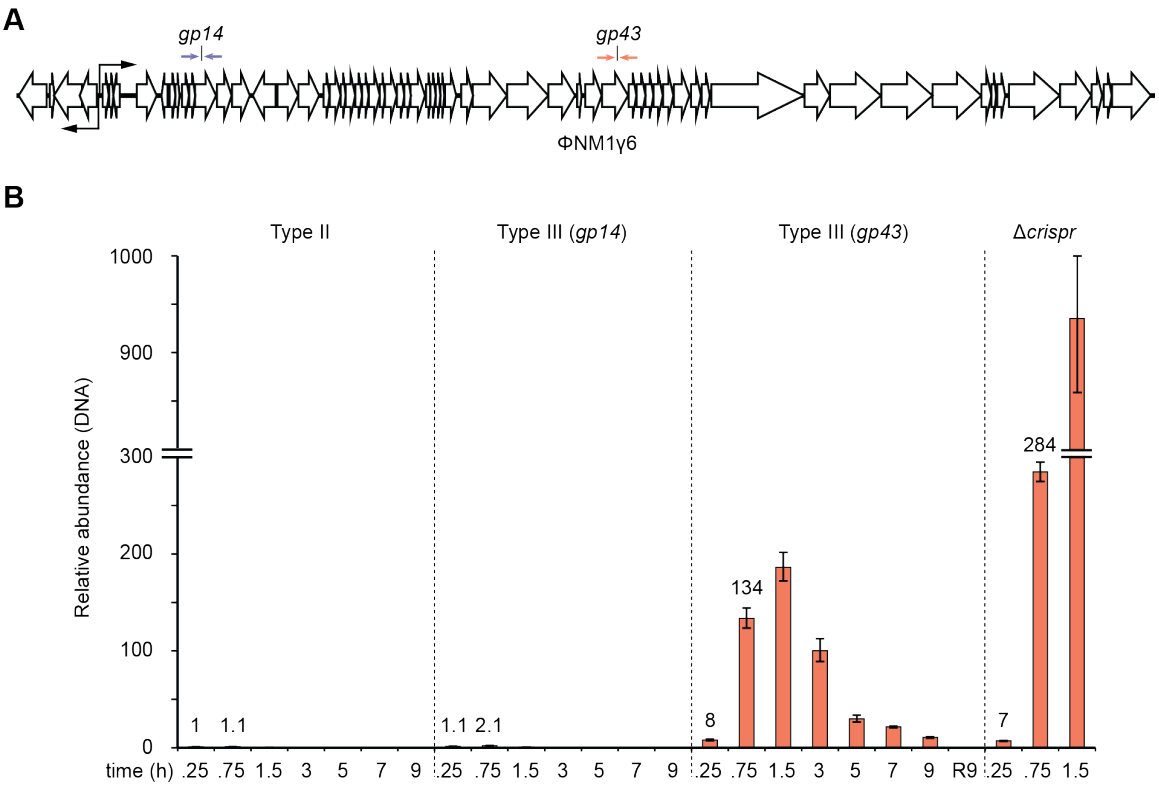


Figure 3-12 Co-transcriptional type III CRISPR-Cas targeting leads to accumulation of phage DNA

(A) A schematic of the genome of lytic phage, ΦNM1γ6. The position of two spacers that target *gp14* and *gp43* is shown. Opposed arrows indicate the primers used for qPCR experiments in **(B)** and **Figure 3-13A**. **(B)** qPCR performed on the ΦNM1γ6 *gp43* gene using total DNA collected from *S. aureus* RN4220 cells carrying different CRISPR-Cas systems at different times post-infection. Values for the *rho* gene were used for normalization. The normalized value for the measurement at 15 min in cells harboring a type II system was set to 1 to obtain the relative abundance of the phage DNA for the rest of the data points (mean ± SD of four replicas). The R9 time point indicates that cells were refreshed with new culture broth at 9 hr post-infection and were grown for an additional 9 hr before collection of DNA for qPCR.

The qPCR was performed with primers that amplify the *gp43* target. A similar result was obtained when I performed qPCR using primers that amplify the *gp14* target (**Figure 3-13A**), suggesting that DNA abundance at the target as well as at a distant locus is equally affected by CRISPR targeting. Phage DNA accumulation was also corroborated by Southern blot analysis (**Figure 3-13B, C**). Whereas phage DNA in cells carrying a type II CRISPR-Cas system was minimally detected, it accumulated over the first 90 min of infection cycle in cells carrying a type III system targeting the late *gp43* gene, with levels similar to those determined by qPCR (**Figure 3-12B**). Interestingly, no cleavage products (expected at 2.2 kB and 1.7 kB) were detected. This indicates that cleaved DNA may be rapidly degraded by the CRISPR-Cas systems or host nucleases under these experimental conditions.

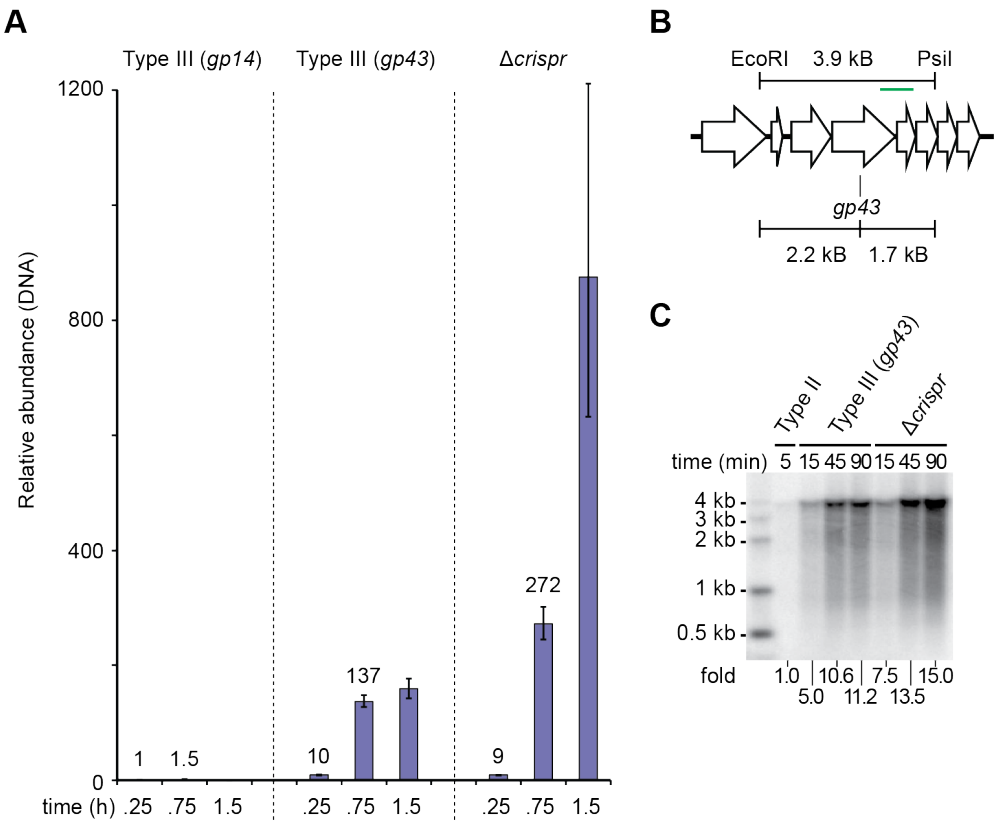


Figure 3-13 Corroboration of phage DNA accumulation in vivo

(A) qPCR performed on the ΦNM1γ6 *gp14* gene (**Figure 3-12A**) using total DNA collected from *S. aureus* RN4220 cells carrying different CRISPR-Cas systems at different times post-infection. Values for the *rho* gene were used for normalization. The normalized value for the measurement at 15 min in cells harboring the *gp14*-targeting spacer was set to 1 to obtain the relative abundance of the phage DNA for the rest of the data points (mean ± SD of four replicas). **(B)** Schematic showing the location of the EcoRI and PsiI restriction sites used to detect phage DNA via Southern blot in **(C)**. The position of the *gp43* target is located 2.2-kB from the EcoRI site and 1.7-kB from the PsiI site. The green line indicates the location of the dsDNA probe used in this assay. **(C)** Southern blot on total DNA extracted from cells treated with ΦNM1γ6 at different times after infection and digested with EcoRI and PsiI. The intensity, relative to cells harboring the type II CRISPR-Cas system, of the 3.9-kb phage fragment is reported.

Together these data revealed that the requirement of target transcription for type III CRISPR-Cas DNA cleavage results in the accumulation of phage DNA when a region that is expressed late in the infectious cycle is targeted. To support this finding in vitro, an co-transcriptional DNA cleavage assay (schematized in **Figure 3-6**) with different complex:target molar ratios were performed, and target DNA cleavage was detected at a 10:1, but not 1:1, ratio (**Figure 3-14**). This means an excess of target DNA can prevent efficient cleavage by the Cas10-Csm complex, suggesting that the accumulation of phage DNA found in cells targeting a late gene can be due to the excess of target DNA and inefficient CRISPR targeting.

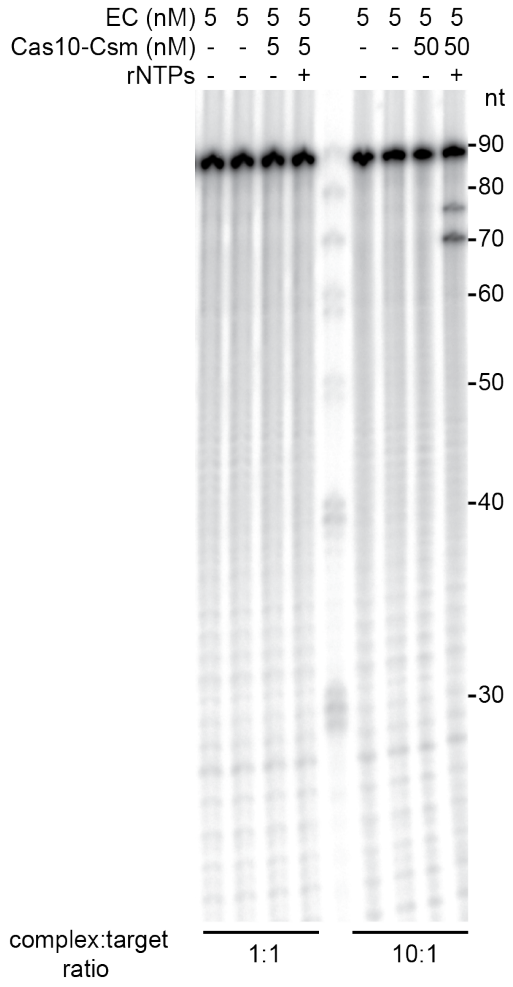


Figure 3-14 Co-transcriptional DNA cleavage using different complex:target ratios

Co-transcriptional DNA cleavage of the Cas10-Csm complex and dsDNA target at different complex:target ratio. 5 nM of elongation complex (EC) containing a radiolabeled dsDNA target, RNA primer and RNA polymerase were incubated with 5 or 50 nM of purified Cas10-Csm complex harboring a matching crRNA (derived from *spc1*). Ribonucleoside tri-phosphates (rNTPs) were added to initiate transcription and the products of the reactions (30 min after addition of rNTPs) were subject to PAGE and phosphorimager visualization.

3.7 The biological function of RNases in type III CRISPR immunity

In spite of this significant difference in phage DNA accumulation and cleavage of the target DNA (**Figure 3-12B**), both *gp14*- and *gp43*-targeting spacers protected host bacteria from lytic infection (**Figure 3-3C**). This raises the question as to how lytic phages in great abundance can be in harmony with host cells for an extended period of time. It should be pointed out that the phage DNA per se does not kill cells. However, host cells are in peril if lytic enzymes encoded by the DNA are made in sufficient amount. I therefore hypothesized that the RNases of the type III CRISPR-Cas may act as a second line of defense; they degrade phage transcripts and prevent translation of toxic proteins when DNA targeting is inefficient.

As previously mentioned, type III-A CRISPR-Cas systems encode two RNases: a sequence-specific, crRNA-guided endoribonuclease, Csm3, and another, crRNA-independent ribonuclease, Csm6. Active sites of the two proteins were also validated in vitro (**Figure 3-7B** and **Figure 3-8B** and **C**). To test if these RNases function in vivo and play a role in immunity, I generated mutant CRISPR-Cas systems containing catalytically dead Csm3 (dCsm3), catalytically dead Csm6 (dCsm6), or both (dCsm3/dCsm6), respectively. Along with cells harboring a wild-type (WT) or a non-matching spacer (Δ*crispr*), I performed a phage infection assay. Total RNAs were extracted at 15, 45 and 90 minutes post-infection and were subjected to RT-qPCR. When CRISPR targeted the early-expressed *gp14*, I did not detect any substantial accumulation of phage transcripts in the wild-type, d*csm6*, d*csm3* or d*csm3*/d*csm6* cells (**Figure 3-15A**).

In contrast, phage transcripts increased dramatically in the absence of immunity (Δ*crispr*) as expected (**Figure 3-15A**). When I conducted the same experiment but targeting the late-transcribed *gp43* gene, I observed minor differences between target transcripts abundance in wild-type, d*csm3* and d*csm6* strains (**Figure 3-15B**). However, I observed a 50-fold (351 divided by 7) increase in viral transcripts in the d*csm3*/d*csm6* double mutant, which accumulated similar level of phage mRNA as Δ*crispr* cells 45 minutes after infection (**Figure 3-15B**). I also performed RNA deep sequencing (RNA-seq) in infected cells with the different genetic backgrounds. The results confirmed the RT-qPCR data, showing that *gp14* phage transcripts did not accumulate during the immune response of the different CRISPR-Cas systems (**Figure 3-15C**), but that there was a substantial increase in *gp43* transcript levels (similar to the Δ*crispr* control) in the d*csm3*/d*csm6* double mutant at 45 minutes post-infection (**Figure 3-15D**). Together, these results indicate that the RNase activity of either Csm3 or Csm6 is required to prevent the accumulation of phage transcripts when the type III-A CRISPR-Cas system targets late-, but not early-, expressed phage genes.

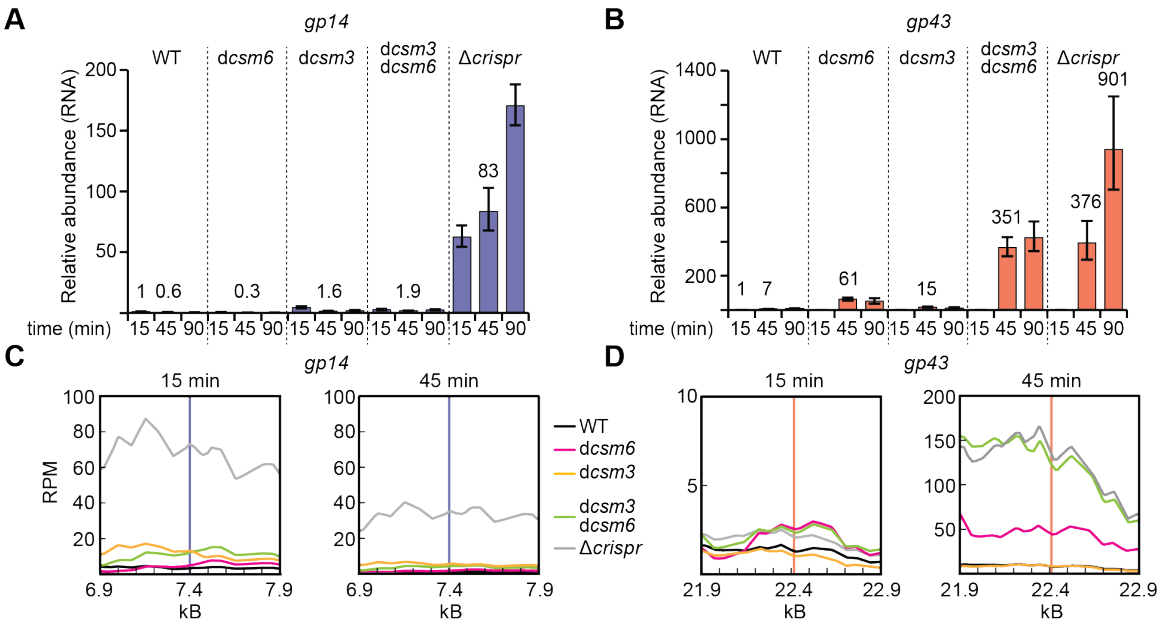


Figure 3-15 Csm3 and Csm6 are required for the degradation of phage transcripts

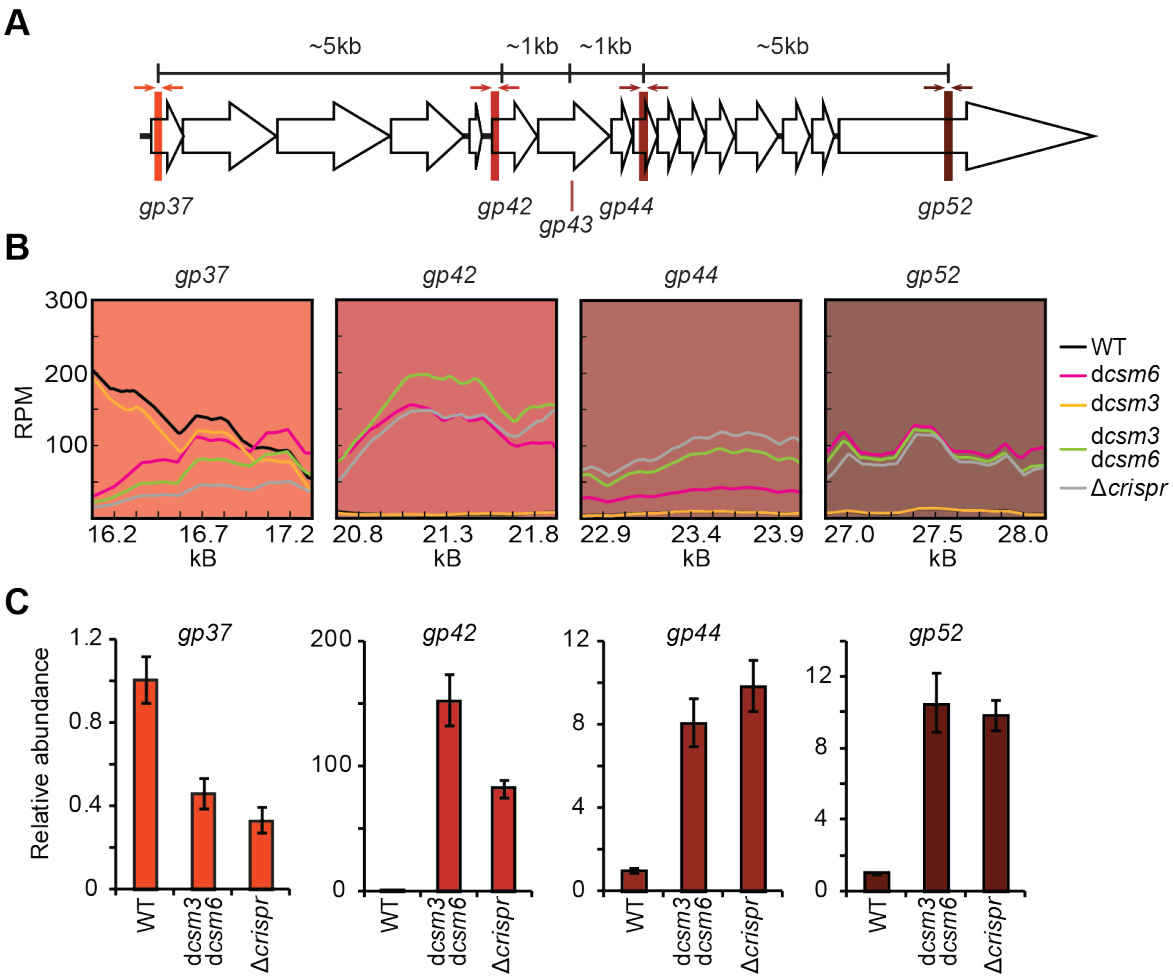
(A) RT-qPCR performed on the ΦNM1γ6 *gp14* transcript using total RNA collected from *S. aureus* RN4220 cells carrying different type III-A CRISPR-Cas systems targeting the *gp14* gene at different times post-infection. Values for the *rho* gene were used for normalization. The normalized value for the measurement at 15 min in wild-type cells was set to 1 to obtain the relative abundance of the *gp14* transcript for the rest of the data points (mean ± SD of four replicas). **(B)** Same as **(A)**, but using CRISPR-Cas systems targeting the ΦNM1γ6 *gp43* gene and measuring relative abundance of the *gp43* transcript. **(C)** RNA-seq reads (reads per 500 bases of transcript per million mapped reads, RPM) for transcripts in the vicinity of the *gp14* target at 15 and 45 min post-infection of cells harboring different type III-A CRISPR-Cas systems. Vertical purple line indicates target position. **(D)** Same as **(C)**, but showing transcription levels in the *gp43* target region. Vertical orange line indicates target position.

Since it is known that λ-like phages such as ΦNM1γ6 produce long transcripts through antitermination mechanisms (Krebs 2010), it is possible to investigate

the effect of the RNase activity of Csm3 and Csm6 on phage transcripts further away from the target site defined by the crRNA guide (**Figure 3-16A**). RNA-seq of phage transcripts in wild-type, d*csm3*, d*csm6* or d*csm3*/d*csm6* cells revealed that degradation of the phage transcripts extended for at least 1 kB at each side of the target site (**Figure 3-16B**, *gp42* and *gp44*). In particular, a very low level of phage transcripts were found in d*csm3* cells, indistinguishable to that of wild-type cells. On the other hand, transcript levels in d*csm6* cells were higher. Interestingly, this pattern extended 5 kB further downstream, but not upstream from the target site (**Figure 3-16B**). These data suggest that Csm6, and not Csm3, is responsible for much of the transcript degradation outside of the target region. Since Csm6 is not part of the Cas10-Csm complex (Hatoum-Aslan et al. 2013), the mechanism by which the RNase activity of Csm6 is first localized to the target transcript remains to be elucidated.

Figure 3-16 Csm6, and not Csm3, degrades phage transcripts in the vicinity of the region targeted by the Cas10-Csm complex

(A) Schematic of the genomic region of ΦNM1γ6 in the vicinity of the *gp43* target (up to 5 kB upstream and downstream). Vertical colored bars indicate the 100-nt window of the RNA-seq data shown in **(B)**. Opposed arrows indicate the primers used for RT-qPCR experiments in **(C)**. **(B)** Total RNA collected from *S. aureus* RN4220 cells carrying different type III-A CRISPR-Cas systems targeting the *gp43* gene at 45 min post-infection were subjected to RNA-seq. RNA-seq reads (reads per 500 bases of transcript per million mapped reads, RPM) for transcripts in the regions indicated in **(A)** were shown. **(C)** RT-qPCRs using primers shown in **(A)**. RT-qPCRs were performed on the ΦNM1γ6 *gp37*, *gp42*, *gp44* and *gp52* transcripts using total RNA collected from *S. aureus* RN4220 cells carrying wild-type, d*csm3*/d*csm6* or Δ*crispr* type III-A CRISPR-Cas systems targeting the *gp43* gene at 45 min post-infection. Values for the *rho* gene were used for normalization. The normalized value for the measurement in wild-type cells was set to 1 to obtain the relative abundance of the transcripts in other genetic backgrounds (mean ± SD of four replicas).

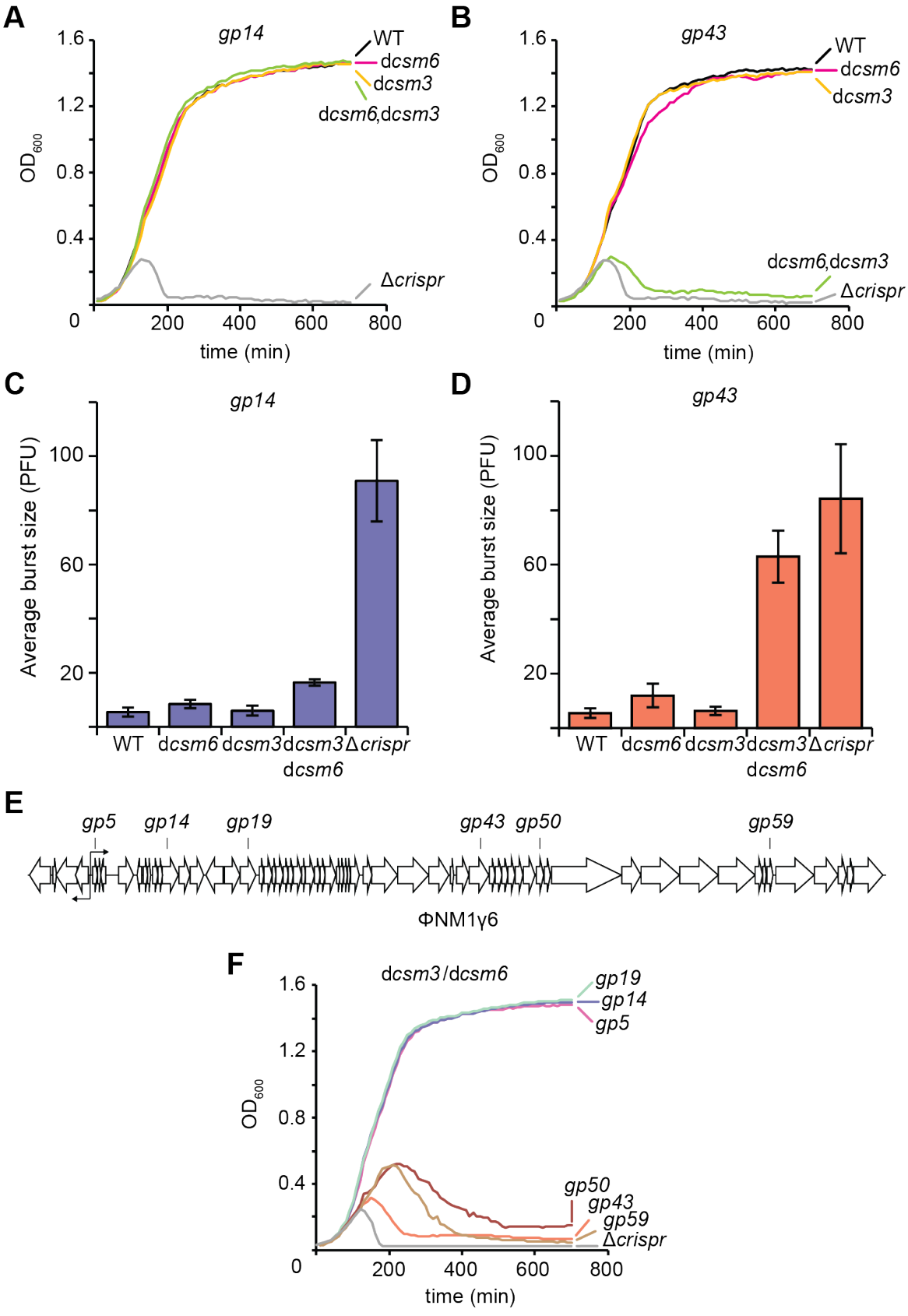


In light of the results above, I wanted to investigate whether the degradation of late-expressed phage transcripts mediated by Csm3 and Csm6 can truly act as a second line of defense, i.e., required for CRISPR-Cas immunity during the targeting of late genes. To this end, I infected *S. aureus* cells harboring wild-type and RNase-null type III CRISPR-Cas systems and monitored their growth. When CRISPR targeted *gp14*, the mutants d*csm3*, d*csm6* and d*csm3*/d*csm6* were as effective as the wild-type CRISPR-Cas to confer immunity (**Figure 3-17A**), a result that demonstrates the sufficiency of DNA cleavage for viral clearance. In contrast, when *gp43* was targeted the d*csm3*/d*csm6* double mutant failed to

provide immunity (**Figure 3-17B**), similarly to a no-targeting control (Δ*crispr*). A similar result was obtained when I measured the effect of type III-A CRISPR-Cas immunity on the propagation of the phage by determining the average burst size, i.e. the number of viral particles (counted as plaque forming units, PFU) released per infected cells (**Figure 3-17C** and **D**). Both experiments indicate that the RNase activity of either Csm3 or Csm6 is required for immunity when targeting a late-, but not early-, expressed gene. To confirm this temporal pattern I tested immunity mediated by d*csm3*/d*csm6* mutant systems targeting two other early-transcribed (*gp5* and *gp19*) and two other late-transcribed (*gp50* and *gp59*) genes (**Figure 3-17E**). Upon infection, targeting of *gp5*, *gp14* and *gp19* produced efficient immunity, whereas targeting of *gp43*, *gp50* and *gp59* resulted in the death of bacteria expressing inactive Csm3 and Csm6 RNases (**Figure 3-17F**). These data demonstrate that the RNase activities of Csm3 and Csm6 are required for type III-A CRISPR-Cas immunity when the targets specified by the crRNA guide reside within late-expressed genes.

Figure 3-17 Degradation of phage transcripts by Csm3 and Csm6 enables type III CRISPR-Cas immunity targeting late-expressed genes

(A) *S. aureus* RN4220 cells carrying different type III-A CRISPR-Cas systems targeting the *gp14* gene were grown in liquid media and infected with phage ΦNM1γ6 phage (at 0 hr) with a multiplicity of infection of five viruses per bacteria (MOI = 5). Optical density at 600 nm (OD_{600}) was measured for the following 12 hr to monitor cell survival. Representative growth curves of at least three independent assays are shown. **(B)** Same as **(A)**, but with the CRISPR-Cas systems programmed to target *gp43*. **(C)** The different infections performed in **(A)** were plated to enumerate plaque forming units (PFU) and calculate the average burst size. Mean ± SD of three replicas are reported. **(D)** Same as **(C)**, but with the CRISPR-Cas systems programmed to target *gp43*. **(E)** A schematic of the genome of ΦNM1γ6. The position of spacers that target early genes, *gp5*, *gp14* and *gp19*, and late genes, *gp43*, *gp50* and *gp59* were shown. **(F)** Survival of cells (determined by measuring growth at OD_{600}) carrying d*csm3*/d*csm6* type III-A CRISPR-Cas systems targeting the different ΦNM1γ6 genes shown in **(E)**. Representative growth curves of at least three independent assays are shown.

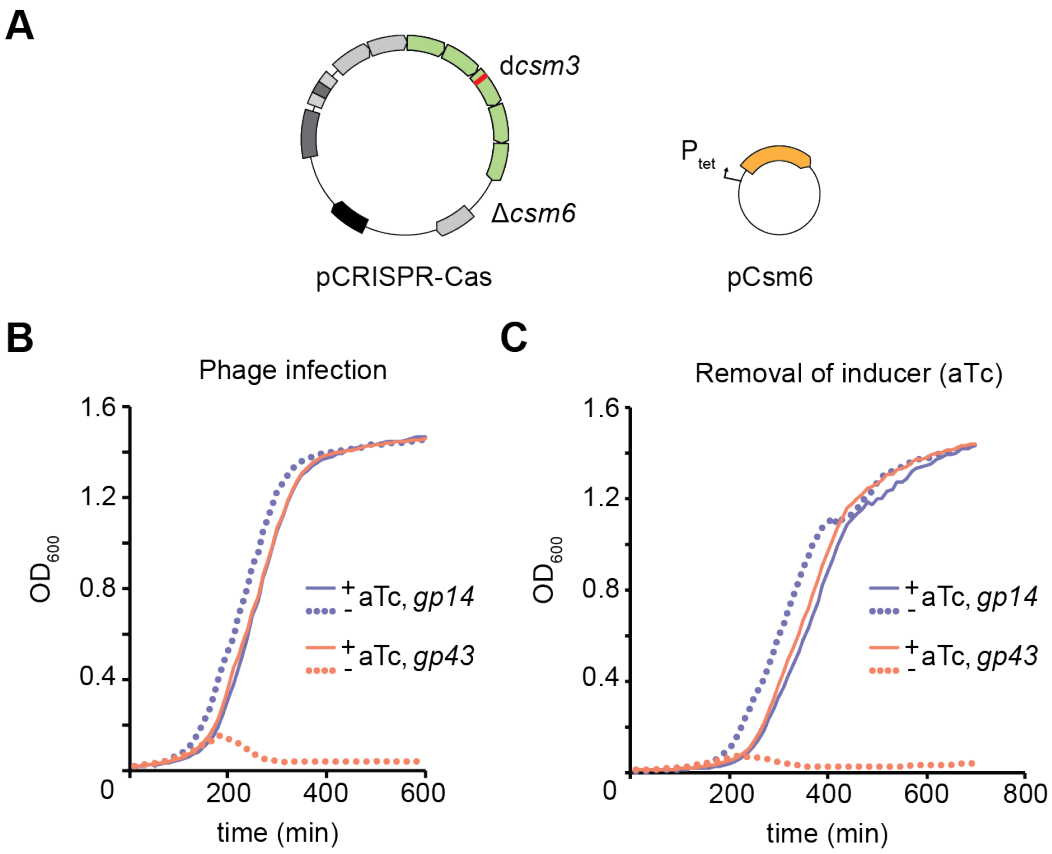

A
gp14
WT
dcsm6
dcsm3
dcsm6,dcsm3
Δcrispr
OD600
time (min)
B
gp43
WT
dcsm6
dcsm3
dcsm6,dcsm3
Δcrispr
OD600
time (min)
C
gp14
Average burst size (PFU)
WT
dcsm6
dcsm3
dcsm3 dcsm6
Δcrispr
D
gp43
Average burst size (PFU)
WT
dcsm6
dcsm3
dcsm3 dcsm6
Δcrispr
E
gp5
gp14
gp19
gp43
gp50
gp59
ΦNM1γ6
F
dcsm3/dcsm6
gp19
gp14
gp5
gp50
gp43
gp59
Δcrispr
OD600
time (min)

Why is the RNase activity required when late phage genes are targeted? I showed earlier that when CRISPR targeted a late-expressed phage gene, phage DNA accumulated inside host bacteria and persisted for a long time (**Figure 3-12B**). The experiments performed with the RNase-null systems (**Figure 3-15B** and **Figure 3-17B**) strongly indicate that degradation of phage transcripts by Csm3 and Csm6 limits further viral gene expression and the continuation of the lytic infectious cycle, thus allowing for a "peaceful" co-existence of the phage DNA and host bacteria for an extended period of time. To validate the essentiality of the RNases and demonstrate that the lingering phage DNA is capable of continuing the lytic cycle in the absence of the RNase activity, I designed an experiment to eliminate RNase activity 10 hours post-infection, a time when phage DNA is still present in host bacteria, and checked for host cell viability. First, I infected cells harboring d*csm3*/Δ*csm6* CRISPR-Cas systems (Δ*csm3* is not feasible since deletion of the gene results in loss of formation of the Cas10-Csm complex) targeting *gp14* or *gp43* and carrying a pCsm6 plasmid, which provides aTc-dependent expression of Csm6 (**Figure 3-18A**). As expected from my previous results, in the absence of the inducer the cells targeting *gp43*, but not those targeting *gp14*, succumbed to phage infection (**Figure 3-18B**). In the presence of aTc, both populations survived (**Figure 3-18B**). The cells from these two populations were washed with fresh broth to eliminate aTc, and thus Csm6 expression, 10 hours after infection. Cells were diluted in fresh broth with or without aTc and their growth was monitored (**Figure 3-18C**). While the growth of *gp14*-targeting cells was not affected by removal of Csm6, *gp43*-targeting cells

were lysed by phage. This finding corroborates that Csm6 has been keeping the accumulated phage DNA “in check” and keep host bacteria alive. Once the RNase activity is abrogated, the accumulated phage DNA can continue the lytic cycle in spite of DNA cleavage (by Cas10 within the complex), leading to the lysis of host bacteria.

Figure 3-18 Degradation of phage transcripts by Csm6 enables type III CRISPR-Cas immunity against late-expressed genes

(A) *S. aureus* RN4220 cells harboring Δ*csm6*/d*csm3* type III-A CRISPR-Cas systems targeting *gp14* or *gp43* were complemented with a pCsm6 plasmid, which carried the *csm6* gene under the control of a tetracycline-inducible promoter (P_{tet}). **(B)** These cells were infected with ΦNM1γ6 (MOI = 5) in the presence or absence of the inducer, aTc (0.008 mg/ml), i.e., induction of Csm6 expression. Bacterial growth was monitored by measuring OD_{600} for 10 hr. **(C)** The cells grown in the presence of aTc shown in **(B)** were collected after 10 hr, washed to remove residual phages in the supernatant as well as the inducer in order to eliminate further expression of Csm6. Washed cells were diluted (1:333) in fresh media without phage nor aTc. As a control washed cells were also diluted (1:333) in fresh media with aTc (0.008 mg/ml). Bacterial growth was monitored by measuring OD_{600} for 12 hr.



3.8 Requirement of Csm6 in providing immunity against phages with target mutations

So far, my data showed that CRISPR targeting of a late-expressed gene led to the accumulation of target DNA. In this scenario the RNase activities of Csm3 or Csm6 were required to clear the phage transcripts and slow down the phage lytic cycle until all DNA targets were destroyed. A similar situation could present during infection with phages harboring target mismatches. In both type I and type II CRISPR-Cas immunity mutations in the target region lead to the escape of mutant phages due to reduced crRNA-guided DNA cleavage. Type III CRISPR-Cas systems, however, seem much more tolerant of such mutations and are able

to provide immunity even in the presence of several mismatches within the crRNA:target interaction (Manica et al. 2013; Goldberg et al. 2014). I speculated that, if target mutations result in inefficient DNA cleavage also during type III CRISPR-Cas immunity, the reported tolerance to mutations could be the result of phage transcript cleavage by the Csm3 and/or Csm6 RNase activity, as it is known that RNA-RNA binding can be promiscuous. To test this I introduced 3, 4 and 5 mismatches into the spacer sequence targeting the *gp43* gene of ΦNM1γ6 (**Figure 3-19A**). I infected hosts carrying these mutations and looked for the CRISPR immune response. Consistent with previous reports, type III CRISPR-Cas immunity protected hosts even in the presence of 3 and 4 mismatches (but not 5) between the crRNA guide and its target (**Figure 3-19B**). I first tested if Csm6 was important for immunity in the presence of mismatches by performing infections in a Δ*csm6* host (**Figure 3-19C**). Consistent with **Figure 3-17B**, *csm6* was not required to provide immunity when the phage carried a target with perfect homology. However, Δ*csm6* cells were not as protected in the presence of 3 mismatches and immunity was completely abrogated with 4 mismatches in the crRNA:target interaction. Protection in the presence of 4 mismatches required the RNase activity of Csm6 but not that of Csm3 (**Figure 3-19E**).

The RNase activity of Csm3 or Csm6 was required for full immunity when DNA clearance was slow and inefficient, as was the case when CRISPR-Cas targeted late-expressed phage genes. To see if the presence of mismatches between crRNA:target can also lead to slow DNA clearance, I performed qPCR of the *gp43* target to compare phage DNA accumulation during the course of

infection of wild-type hosts carrying a perfectly matching or 4-mismatch spacer (**Figure 3-20A**). I observed that indeed the presence of mismatches led to the accumulation of target phage DNA. This was corroborated by a plasmid-curing assay similar to the one presented in **Figure 3-5A** (**Figure 3-20B**, **C** and **D**).

Figure 3-19 Csm6 is required to provide immunity against viruses with target mutations

(A) Introduction of mutations (in red) in the spacer matching the *gp43* gene of phage ΦNM1γ6 that generate three, four, or five mismatches in the crRNA:target region. **(B)** *S. aureus* RN4220 cells that harbored a wild-type III-A CRISPR-Cas system targeting the *gp43* gene in the presence of different crRNA:target mismatches were grown in liquid media and infected with ΦNM1γ6 (at 0 hr) with a multiplicity of infection of five viruses per bacteria (MOI = 5). Optical density at 600 nm (OD_{600}) was measured for the following 12 hr to monitor cell survival. Representative growth curves of at least three independent assays are shown. **(C)** Same as **(B)**, but with cells harboring a CRISPR-Cas locus without *csm6*. **(D)** *S. aureus* RN4220 cells that carried different type III-A CRISPR-Cas systems matching the *gp43* gene were grown in liquid media and infected with phage ΦNM1γ6 (at 0 hr) with a multiplicity of infection of five viruses per bacteria (MOI = 5). Optical density at 600 nm (OD_{600}) was measured for the following 12 hr to monitor cell survival. Representative growth curves of at least three independent assays are shown. **(E)** Same as **(D)**, but with cells expressing a crRNA with 4 mismatches. Abbreviation: mm, mismatch.

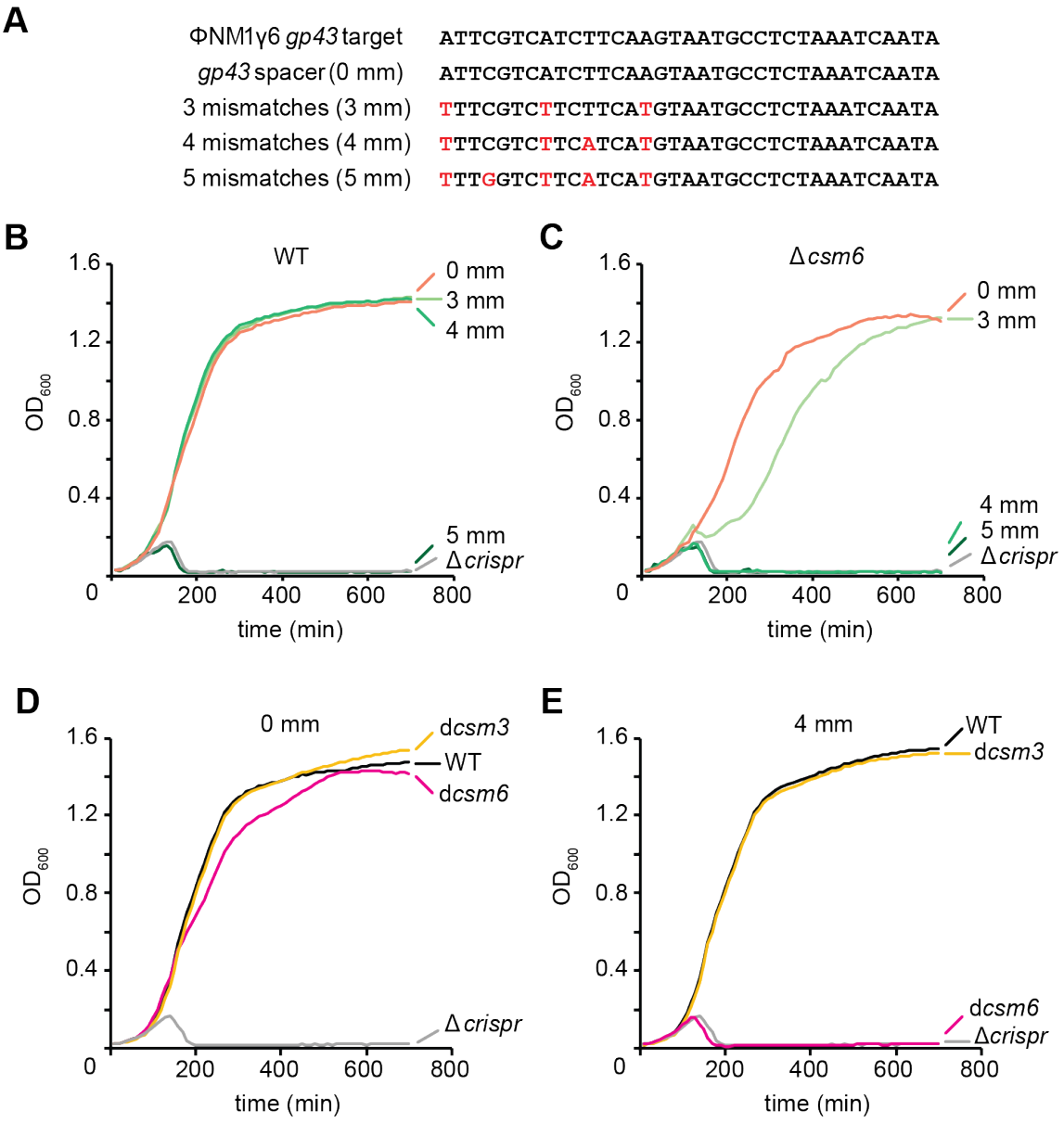
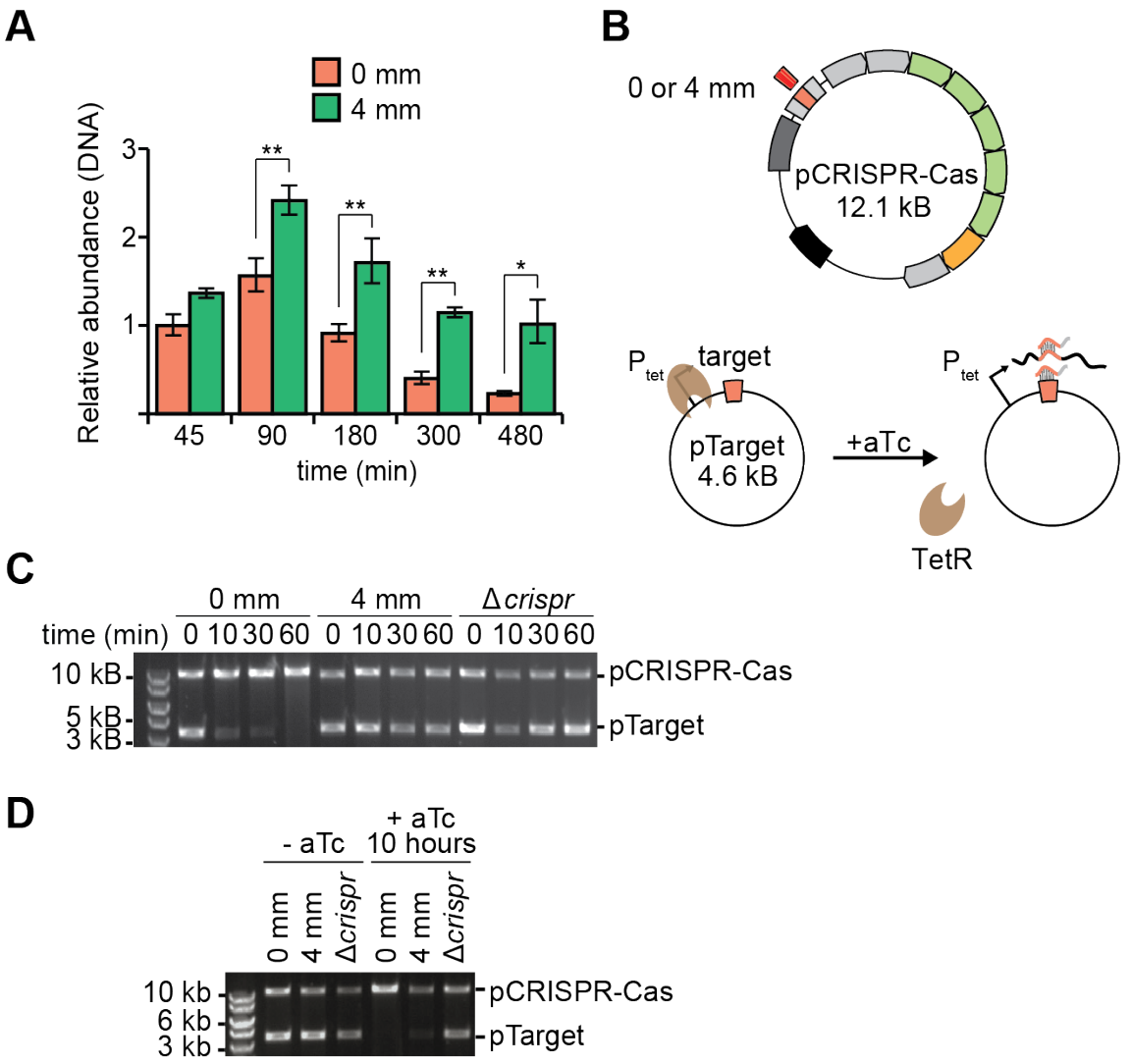

A
ΦNM1γ6 gp43 target ATTCGTCATCTTCAAGTAATGCCTCTAAATCAATA
gp43 spacer (0 mm) ATTCGTCATCTTCAAGTAATGCCTCTAAATCAATA
3 mismatches (3 mm) TTTCGTCTTCTTCATGTAATGCCTCTAAATCAATA
4 mismatches (4 mm) TTTCGTCTTCATCATGTAATGCCTCTAAATCAATA
5 mismatches (5 mm) TTTGGTCTTCATCATGTAATGCCTCTAAATCAATA
B
WT
0 mm
3 mm
4 mm
5 mm
Δcrispr
OD600
time (min)
C
Δcsm6
0 mm
3 mm
4 mm
5 mm
Δcrispr
OD600
time (min)
D
0 mm
dcsm3
WT
dcsm6
Δcrispr
OD600
time (min)
E
4 mm
WT
dcsm3
dcsm6
Δcrispr
OD600
time (min)

Figure 3-20 Mismatches in crRNA:target leads to accumulation of target DNA

(A) *S. aureus* RN4220 cells that harbored a wild-type CRISPR-Cas system expressing crRNAs with no mismatch (0 mm) or 4 mismatches (4 mm) to the target were infected with phage ΦNM1γ6. qPCR was performed on the ΦNM1γ6 *gp43* gene using total DNA collected from cells at different times post-infection. Values for the *rho* gene were used for normalization. The normalized value for the measurement at 45 min in 0 mm cells was set to 1 to obtain the relative abundance of the phage DNA for the rest of the data points (mean ± SD of four replicas). (*) $P<0.05$; (**) $P<0.01$. **(B)** Plasmid-curing assay similar to **Figure 3-5A** was performed. pCRISPR-Cas carried a *gp43*-targeting spacer containing 0 or 4 mismatches. pTarget (pWJ267) carried a *gp43* target sequence inserted after a tetracycline-inducible promoter (P_{tet}). **(C)** Agarose gel of linearized plasmid DNA purified from cells at 0, 10, 30 and 60 min after aTc induction. Δ*crispr* cells were used as a non-targeting control. **(D)** Agarose gel of linearized plasmid DNA purified from cells either before or after 10 hr of treatment with aTc. Abbreviation: mm, mismatch.



Target mismatches are not only present within a viral population but are very common between related phages. For example, our lab previously engineered a spacer matching the *gp32* gene present in phage ΦNM1γ6 (Goldberg et al. 2014) but simultaneously containing 4 mismatches to the same gene in the related phage ΦNM4γ4 (Bae et al. 2006; Heler et al. 2015) (**Figure 3-21A**). Consistent with my results, type III-A CRISPR-Cas immunity against ΦNM1γ6 mediated by this spacer does not require the RNase activity of Csm6 (**Figure 3-21A** and **B**). In contrast, whereas the wild-type CRISPR-Cas system tolerated the 4

mismatches and protected cells from ΦNM4γ4 infection, the d*csm6* mutant cells were susceptible to viral attack (**Figure 3-21A** and **B**). Together with **Figure 3-19** and **Figure 3-20**, my results show that Csm6 RNase activity is required to maintain immunity even in the presence of target mutations that decrease the efficiency of DNA targeting, a distinct property of type III systems.

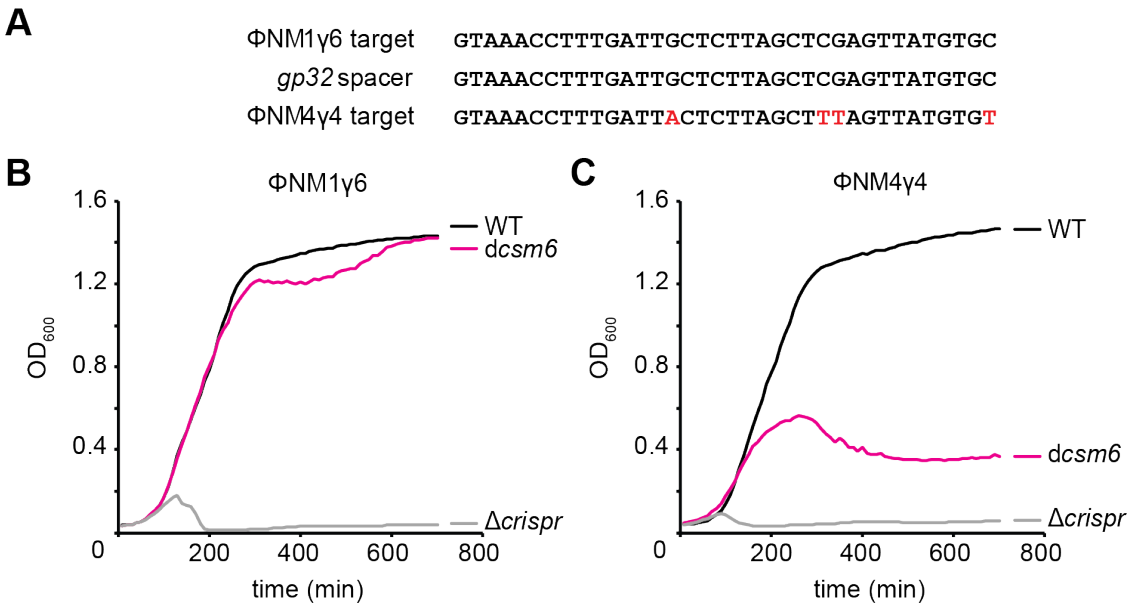


Figure 3-21 Csm6 is required to provide immunity against viruses with target mutations (continued)

(A) The *gp32*-targeting spacer matches phage ΦNM1γ6 genome perfectly but presents four mismatches in the phage ΦNM4γ4 genome. **(B)** *S. aureus* RN4220 cells that harbored a wild-type, d*csm6* or Δ*crispr* type III-A CRISPR-Cas system targeting the *gp32* gene were grown in liquid media and infected with ΦNM1γ6 (at 0 hr) with a multiplicity of infection of five viruses per bacteria (MOI = 5). Optical density at 600nm (OD_{600}) was measured for the following 12 hr to monitor cell survival. Representative growth curves of at least three independent assays are shown. **(C)** Same as **(B)**, but following infection with ΦNM4γ4.

3.9 Discussion

Type III-A CRISPR-Cas systems encode a Cas10-Csm complex that is responsible for crRNA biogenesis and targeting. Previous work showed that the complex can cleave both the genome and the transcripts of invaders (Peng et al. 2015; Samai et al. 2015), and that the DNA- and RNA-targeting activity resides in Cas10 and Csm3 within the complex, respectively. Here I identified a new CRISPR-associated RNase, Csm6, that is not associated with the Cas10-Csm complex. I demonstrated that Csm6 has RNase activity both in vitro and in vivo, and it plays an important role in CRISPR immunity.

Whereas DNA cleavage is fundamental for CRISPR immunity against invaders such as phages, a role for the RNase activity of these systems has not been determined. Previous works showed that the type III CRISPR-Cas immunity requires transcription of the target (Goldberg et al. 2014). This provides a clue, as the transcriptional requirement creates a challenge for the type III CRISPR-Cas system when it targets viral genes that are expressed late in the infection cycle. Since CRISPR immunity cannot mount attacks until the phages have finished replication in this scenario, the immune system can be overwhelmed by the abundance of DNA targets present at this point, and cell viability can be compromised. Indeed, my data showed phage DNA accumulates substantially when CRISPR targets a late- but not early-expressed gene during infection. Under this condition, I showed that the two RNases encoded by the type III-A CRISPR-Cas, Csm3 and Csm6 provide a second line of defense. While mutation of the active site of either protein alone is not sufficient to break immunity, cells

harboring double mutation succumb to phage infection. In contrast, when an early-transcribed gene is targeted, DNA cleavage occurs shortly after genome injection. In this case the endonuclease activity of the Cas10-Csm complex is sufficient to clear the virus; the infectious cycle does not proceed further and I showed that degradation of phage transcripts by the two RNases is not necessary to prevent phage propagation.

Temporal regulation of phage gene expression is universal. I believe that the ability of transcript degradation by type III-A CRISPR-Cas systems would be important to provide immunity against most classes of dsDNA viruses when the crRNA guide targets genes that are expressed late in the infectious cycle. Target selection occurs during the “adaptation” phase of CRISPR-Cas immunity, when new spacer sequences from an invading phage are incorporated into the CRISPR array (Heler et al. 2014). Although little is known about the acquisition of spacers by type III CRISPR-Cas systems, type I and II systems incorporate spacers matching all regions of the viral genome (Datsenko et al. 2012; Paez-Espino et al. 2013; Heler et al. 2015), without any noticeable bias towards early or late genes. If such bias is also absent during spacer acquisition by type III CRISPR-Cas systems, the RNase activity would be necessary to confer immunity to all bacteria that incorporate a spacer specifying a late-expressed gene. At the moment we do not know the genomic position at which type III-A targets are expressed “too late” for viral clearance by the Cas10-Csm DNA cleavage activity; i.e., the targeting region at which the Csm3 or Csm6 RNase activities are required to rescue the host. I speculate that the onset of phage

replication might mark a turning point at which these RNase activities become essential.

My study showed that either Csm3 or Csm6 RNase activity is required for immunity when the target of the Cas10-Csm complex is located in a late-expressed gene. Mutations in the active sites of either of these genes are not sufficient to disrupt immunity. However, in the presence of target mismatches that lead to the accumulation of phage DNA, Csm6, but not Csm3, RNase activity is required for immunity. How can this be explained? Whereas the RNase activity of Csm3 is strictly defined by crRNAs, I showed that Csm6 is a crRNA-independent nuclease that can degrade transcript at least 1 kB away from the target site. Thus, it is conceivable that mismatches between the crRNA and target may disrupt activity of Csm3 but not Csm6, a hypothesis that should be tested by future experiments.

Mismatch tolerance is an important aspect that distinguishes the type III from the type I and II CRISPR immune systems. Whereas type I and II CRISPR immunity is very sensitive to mutations in the target sequence (especially in the “seed” region of the target) (Wiedenheft et al. 2011; Gasiunas et al. 2012; Jinek et al. 2012; Westra et al. 2012), type III immunity is unusually tolerant of such mutations (Manica et al. 2013; Goldberg et al. 2014), allowing the targeting of “escape” or related viruses. In all three systems, target mutations prevent efficient DNA cleavage. Here I demonstrated that in type III systems, transcript degradation by Csm6 results in robust immunity, presumably by stalling the

progression of the phage lytic cycle and allowing for more time for phage DNA clearance.

Whereas Csm3 is part of the Cas10-Csm complex and cleaves sequences that are specified by the crRNA guide (Staals et al. 2014; Tamulaitis et al. 2014; Samai et al. 2015), Csm6 is not part of the complex and has sequence-independent activity. As mentioned above, how Csm6 achieves specificity for phage transcripts is not known. In vitro, this RNase performs multiple cleavages in RNA substrates of different lengths and sequences, in principle lacking such specificity. On the other hand, the Cas10-Csm complex recognizes its RNA targets using crRNA guides, thus displaying a sequence specificity that will result in the cleavage of phage transcripts by Csm3 when the crRNA derives from a phage-matching spacer. Therefore one plausible mechanism to restrict Csm6 activity to the Cas10-Csm target would be the existence of a transient biophysical interaction between them during the targeting stage. This intriguing hypothesis should be examined in the future.

Another intriguing aspect of Csm6 is its requirement to prevent plasmid conjugation (Hatoum-Aslan et al. 2014). Since I found that Csm6 is not involved in DNA degradation, I speculate that it may assist the Cas10-Csm complex in the clearance of the plasmid DNA by degrading plasmid transcripts, some of them possibly important for plasmid replication and maintenance. Additional work focused on Csm6 should address question.

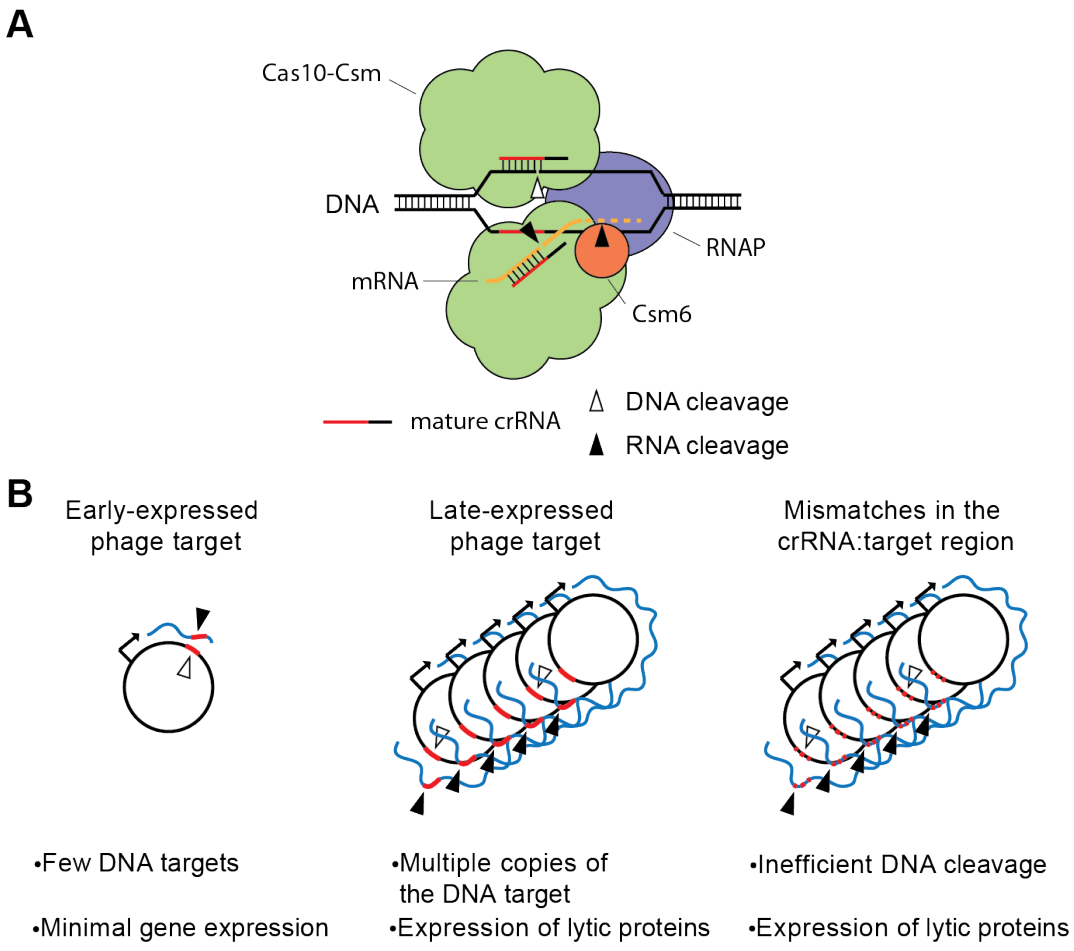


Figure 3-22 A model for co-transcriptional cleavage of target DNA and its transcript during type III-A CRISPR-Cas immunity

(**A**) Type III-A CRISPR-Cas cleaves both DNA (by Cas10) and RNA (by Csm3 and Csm6). Cleavage activity requires target transcription. (**B**) The role of RNA cleavage in type III-A CRISPR immunity. Type III RNases are dispensable when CRISPR targets early-expressed phage genes. However, they are required when CRISPR targets late-expressed phage genes or when crRNA:target region contains mismatches. See text also. Abbreviations: crRNA, CRISPR RNA; RNAP, RNA polymerase.

In summary, my results allow the formulation of a model for the molecular mechanisms underlying type III-A CRISPR-Cas immunity (**Figure 3-22**). The type

III-A Cas10-Csm complex performs co-transcriptional cleavage of the target DNA and its transcripts. Within this complex, Cas10 contains the DNase activity and Csm3 is an RNase. Csm6 is another type III-A RNase that degrades target transcripts. This molecular mechanism of immunity allows for the rapid attack of the viral genome when early-expressed targets are specified by the crRNA guide, which leads to fast and efficient degradation of the invader's genetic material and the clearance of the infection without the need of RNase activity. In contrast, the targeting of late-expressed genes allows viral replication and transcription before DNA cleavage can occur. The accumulated genomes are not cleared efficiently by the endonuclease activity of the Cas10-Csm complex, and the degradation of phage transcripts by Csm3 or Csm6 is required to prevent the completion of the infectious cycle and the lysis of the host cell. Similarly, the presence of crRNA:target mismatches within the phage population prevents efficient DNA cleavage that also leads to the accumulation of phage genomes in the cell. In this scenario, the Csm6 RNase is required for transcript degradation and survival.

CHAPTER 4 PERSPECTIVES

CRISPR-Cas is a sequence-specific adaptive immune system that defends prokaryotic cells against infectious viruses. This immune system also acts as a barrier to horizontal gene transfer of plasmids and other mobile genetic elements, which is a major route through which bacteria exchange beneficial genes such as those that encode antibiotic resistance. From a basic science point of view, studying CRISPR-Cas can allow scientists to better understand evolution, such as the arms race between bacteria and phages, and perhaps discover novel methods to prevent the spread of antibiotic resistance. CRISPR-Cas may even impact on human health. It is well known that the bacteria living on and inside us, known as the human microbiota, play an indispensable role in maintaining normal physiological function and providing many benefits for human beings. Since CRISPR-Cas is widely present in prokaryotes and protect them from viral predation, this immune system may play a role in shaping the bacterial population and diversity in this microenvironment.

CRISPR-Cas encodes a suite of fascinating molecular machineries such as Cas9 that can cleave DNA with incredible precision. The protein has not only been developed for genetic manipulation in prokaryotic cells, but also in eukaryotic cells (Cong et al. 2013; Mali et al. 2013). Compared to conventional targeted genome editing tools such as zinc finger nucleases (ZFNs) (Carroll 2011) and transcription activator-like effector nucleases (TALENs) (Joung and Sander 2013), one of the biggest advantages of the CRISPR-Cas is that sequence-specificity is determined by small crRNAs or sgRNAs (see Chapter 2.3). Not only

are the design and generation of the targeting reagents made much easier, it also opens up opportunities to design multiple sgRNAs in order to interrogate gene function on a genome-wide scale. Indeed, this is an obvious future direction seen by many scientists, as a few independent groups demonstrated that by harnessing the cellular NHEJ pathways, Cas9 with a library of sgRNAs can be used to generate knockout libraries of human cells. This approach was successfully applied to identify genes responsible for drug resistance (Shalem et al. 2014; Wang et al. 2014), genes important for bacterial toxicity (Zhou et al. 2014), and genes conferring resistance to bacterial toxins (Koike-Yusa et al. 2014). In addition to genome-wide screening, Cas9 has seen scores of applications in basic biology, synthetic biology, gene therapy, cancer therapeutics, drug development and beyond (Hsu et al. 2014; Jiang and Marraffini 2015).

We and others also showed that the two nuclease domains of Cas9 can be inactivated to make the protein a sequence-specific DNA-binding protein (Bikard et al. 2013; Qi et al. 2013). This catalytically dead Cas9, dCas9, can mediate transcription repression and activation. As mentioned in Chapter 2.4.3, there are similarities and differences between dCas9- and sRNA-mediated gene regulation, and future endeavors should be made to compare and contrast the two methods in order to discover rules governing the efficacy of repression. The activation of gene expression by dCas9 when fused to ω was moderate. Future experiments should consider using different activator domains and varying the linker distance between the activator and dCas9 to enhance the level of activation. Since dCas9

is a sequence-specific DNA-binding protein, this general property can be principally exploited to direct any protein of interest to desired locations in the genome for many purposes. Indeed, dCas9 was fused to machineries involved in epigenetic modification to modulate gene expression (Mendenhall et al. 2013; Hilton et al. 2015), or enhanced green fluorescent protein (eGFP) to visualize the spatiotemporal dynamics of DNA sequences in living cells (Chen et al. 2013; Anton et al. 2014).

CRISPR-Cas has three types and three stages (**Figure 1-4**). Current technological development primarily focuses on the targeting stage of the type II systems, and this frontier is being extended at a breathtaking speed. What is not yet developed? Adaptation is the first stage of the immune system in which cells use Cas machineries to integrate foreign DNA or RNA fragments into the host chromosome. Given that integrases have been successfully developed as tools for site-specific recombination historically, it is highly plausible that the CRISPR adaptation machineries can be engineered for technological uses as well. Secondly, works performed by us and others revealed that the type III CRISPR-Cas systems cleave both DNA and RNA, suggesting that these systems may be exploited to manipulate RNA molecules.

The type III CRISPR-Cas encodes elaborate molecular scissors that cleave both DNA and RNA. My work identified Csm6, a new CRISPR-associated ribonuclease and elucidated at least one biological function of the RNA cleavage – when the activity of DNA cleavage is weakened (i.e. when targeting late-expressed phage genes or regions bearing mismatches), type III RNases

degrade phage transcripts and prevent the production of toxic enzymes, thus acting as a second line of defense in addition to the DNA-cleaving module. As mentioned in Chapter 3.9 , there are many questions remained to be explored. Csm6 is not associated with the Cas10-Csm complex, what is the mechanism through which the protein gets recruited to the target transcript, and what is the processivity of it? When the crRNA and the target contain mismatches, the RNase activity of Csm6 but not Csm3 is required for immunity against phage infection. Given that the activity of Csm3 is strictly defined by crRNAs while that of Csm6 can be extended to at least 1 kB away from the target site, could this mean that mismatches between the crRNA and target disrupt activity of Csm3 but not Csm6? Csm6 is not involved in DNA degradation, and yet it is required to prevent plasmid conjugation. Could this be attributed to its RNase activity towards plasmid transcripts essential for replication and maintenance? Understanding these fundamental problems can strengthen the current model in which the type III RNases function as a second line of defense.

Forty years ago, the discovery of restriction enzymes unleashed a revolution in modern molecular biology. Today, CRISPR-Cas is driving numerous innovative applications from basic biology to biotechnology and medicine. Despite being forty years apart, both restriction enzymes and CRISPR-Cas systems share remarkably similar pathways. They both originated from basic studies of bacterial immune systems that cleave viral DNA, and have unequivocally revolutionized biological sciences by facilitating manipulation of DNA sequences. Restriction enzymes and CRISPR-Cas systems are testaments

to the practical value of studying the microbial world, which harbors a reservoir of amazing molecular machineries yet to be discovered.

CHAPTER 5 MATERIALS AND METHODS

5.1 Bacterial strains and growth conditions

Cultivation of *E.coli* strains DH5α, MG1655, HME63 (derived from MG1655, Δ(argF-lac) U169 λ cI857 Δcro-bioA galK tyr 145 UAG mutS<>amp) (Costantino and Court 2003) and their derivatives was done in LB medium (BD) or Terrific Broth medium (Fisher Scientific) at 37 °C. Whenever applicable, media were supplemented with ampicillin (100 µg/ml), chloramphenicol (25 µg/ml) or streptomycin (50 µg/ml) to ensure plasmid maintenance or selection of *rpsL* mutant cells.

Cultivation of *S. aureus* strain RN4220 (Nair et al. 2011) and derivatives was done in TSB medium (BD) at 37 °C unless otherwise indicated. Whenever applicable, media were supplemented with chloramphenicol (10 µg/ml) or erythromycin (10 µg/ml) to ensure plasmid maintenance. When appropriate, anhydrotetracycline (aTc) was used at a concentration of 0.25 µg/ml (unless otherwise indicated) to initiate transcription from the P_{tet} promoter.

Cultivation of *S. epidermidis* strain RP62a (Gill et al. 2005) and derivatives was done in BHI medium (BD) at 37 °C. Whenever applicable, media were supplemented with 10 µg/ml chloramphenicol or 5µg/ml mupirocin to ensure plasmid maintenance.

Liquid culture of *S. pneumoniae* strain R6 (Hoskins et al. 2001) and derivatives were grown in THYE medium (30 g/l Todd-Hewitt agar, 5 g/l yeast

extract). Plating was done on tryptic soy agar (TSA) medium (BD) supplemented with 5% defibrinated sheep blood. When appropriate, media were supplemented with kanamycin (400 µg/ml), chloramphenicol (5 µg/ml), erythromycin (1 µg/ml), streptomycin (100 µg/ml) or spectinomycin (100 µg/ml) for selection.

5.2 Bacterial strain construction

Construction of *E.coli* JEN202 was described in (Bikard et al. 2013).

Construction of *S. pneumoniae* strains crR6, $R6^{8232.5}$ and $R6^{370.1}$ were described in (Bikard et al. 2012). crR6M, crR6Rk, crR6Rc, JEN38 and JEN62 were described in (Jiang et al. 2013a).

Construction of *S. epidermidis* Δ*crispr*, known as LAM104, was described in (Marraffini and Sontheimer 2008).

5.3 Plasmid Cloning

5.3.1 Cloning in *E.coli*

Cloning used *E.coli* DH5α electrocompetent cells. To clone WT Csm6 for purification, PCR was performed using pWJ30β (Hatoum-Aslan et al. 2013) as template and primers PS11 and PS12. The PCR product was digested with restriction enzymes NdeI and XhoI and ligated to the vector pET23a-His6 (C-terminal) digested with the same enzymes, making plasmid pPS10. The Csm6 mutants R364A (plasmid pPS42), H369A (plasmid pPS43) and R364A-H369A (plasmid pPS44) were constructed using the plasmid pPS10 as a backbone with

three sets of primers PS245/PS246, PS243/PS244 and PS247/PS248, respectively.

Construction of plasmids pCas9, pCRISPR::Ø and pCRISPR::rpsL were described in (Jiang et al. 2013a). pdCas9, pDB127, pWJ66, pWJ68, pWJ89, pWJ96 and pWJ97 were described in (Bikard et al. 2013).

5.3.2 Cloning in *S. pneumoniae*

Construction of plasmid pDB97 was described in (Jiang et al. 2013a).

5.3.3 Cloning in *S. aureus*

Cloning used S. aureus RN4220 electrocompetent cells. For type III-A pCRISPR-Cas plasmids, new spacers were cloned by ligation of annealed oligonucleotide pairs and BsaI-digested parent vector, pGG-BsaI-R (Goldberg et al. 2014). The sequences of the spacers cloned are provided in **Table 1**. To construct Δcsm6 plasmids, PCR was performed using WT plasmid as template and primers L342/L343. PCR product was restriction digested with PspOMI and EagI (NEB), followed by ligation by T4 DNA Ligase (NEB). To construct dcsm6 plasmids, PCR was performed using WT plasmid as template and primers W852/PS248 and primers PS247/W614. The two PCR products were then ligated using Gibson assembly. To construct dcsm3 plasmids, PCR was performed using WT plasmid as template and primers W852/PS466 and primers PS465/W614. The two PCR products were then ligated using Gibson assembly. To construct pCsm6 overexpression plasmid, one PCR was performed using

pWJ153 (Goldberg et al. 2014) as template and primers W1129/W1113. Another PCR was performed using pWJ30β as template and primers W1127/W1128. The two PCR products were then ligated using Gibson assembly.

Construction plasmid pG0(mut) was described in (Marraffini and Sontheimer 2008). Plasmid pSpc1was known as pCRISPR in (Marraffini and Sontheimer 2008). pWJ30β was described as p*crispr* in (Hatoum-Aslan et al. 2013). pWJ40 was described in (Goldberg et al. 2014). pTarget (pWJ153) was described in (Goldberg et al. 2014); pTarget (pWJ267) was constructed by PCR using pWJ153 as a template and primers W1105/W1106.

5.4 Plasmid DNA preparation

Plasmid DNA was purified from 2 to 6 ml of *E. coli* DH5α or *S. aureus* RN4220 overnight cultures. For preparation from *S. aureus* cultures, cells were pelleted, re-suspended in 100 µl TSM buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 0.5 M sucrose) then treated with 5 µl lysostaphin (2 mg ml^{-1}) at 37 °C for 1 h before treatment with plasmid miniprep reagents from Qiagen. Purification used Qiagen or EconoSpin columns.

5.5 Conjugation

Conjugation was carried out by filter mating as described elsewhere (Morton et al. 1995). Briefly, donor (*S. aureus* RN4220 carrying pG0400 or pG0(mut)) and recipient cells were cultured overnight with appropriate antibiotics. The following day, equal amount (10^9 CFU) of donors and recipients were mixed in 5ml of fresh

BHI medium and vacuum-filtered through 0.45 mM filters (Millipore). Filters were incubated on BHI agar plates at 37 °C for 18 hours and grown bacteria were resuspended in 3 ml of fresh BHI. Serial dilutions were then plated on BHI agar containing the appropriate antibiotics for the enumeration of recipients and transconjugants.

5.6 Preparation of electrocompetent *S. aureus* cells

S. aureus RN4220 or derivative strains were grown overnight in TSB medium, diluted 1:100 in fresh medium without antibiotics, then allowed to grow to an OD_{600} reading of 0.8–1.0. At this point, cells were pelleted at 4 °C, and two or three washes were performed using chilled, sterile dH_2O. Cells were ultimately re-suspended in 1/100th (of the original culture) volume of chilled, sterile 10% glyercol and 50 µl aliquots were distributed for storage at -80 °C.

5.7 Preparation of competent *E.coli* cells

Chemically competent and electrocompetent cells were made according to (Renzette 2011).

5.8 *E. coli* λ-Red recombination

E.coli strain HME63 was used for all λ-Red recombination experiments. Recombineering cells were prepared and handled according to a previously published protocol (Sharan et al. 2009). Briefly, a 2 ml overnight culture (LB medium) inoculated from a single colony obtained from a plate was grown at

30 °C. The overnight culture was diluted 100-fold and grown at 30 °C with shaking (200 r.p.m.) until the OD_{600} was 0.4–0.5 (~3 h). For λ-Red induction, the culture was transferred to a 42 °C water bath to shake at 200 r.p.m. for 15 min. Immediately after induction, the culture was swirled in an ice-water slurry and chilled on ice for 5–10 min. Cells were then washed and aliquoted according to the protocol. For electro-transformation, 50 µl of cells were mixed with 1 nmol of salt-free oligos (IDT) or 100–150 ng of plasmid DNA (prepared by QIAprep Spin Miniprep Kit, Qiagen). Cells were electroporated using 1 mm Gene Pulser cuvette (Bio-rad) at 1.8 kV and were immediately re-suspended in 1 ml of room temperature LB medium. Cells were recovered at 30 °C for 1–2 h before being plated on LB agar with appropriate antibiotic resistance and incubated at 32 °C overnight.

5.9 *S. aureus* transformation

Electrocompetent *S. aureus* RN4220 or derivative strains were thawed, mixed with appropriate amount (~100 ng) of dialysed plasmid DNA and incubated at room temperature for 5 min. Electroporation was performed using a GenePulser Xcell (BioRad) with the following parameters: 2900 V, 25 mF, 100V, 2mm. After electroporation, cells were immediately re-suspended in 200 µl – 1 ml of TSB and recovered at 37 °C for 2 h with shaking. Serial dilutions were then prepared before plating with appropriate antibiotics. Serial dilutions were then plated on tryptic soy agar containing appropriate antibiotics for transformants-counting.

Transformation efficiency was expressed either as colony-forming unit (CFU)/µg of DNA or CFU/ml of recovered cells.

5.10 *S. pneumoniae* transformation

Competent cells were prepared as described previously (Bikard et al. 2012). Briefly, an overnight THYE culture was diluted to OD_{600} of 0.05 in M1 medium (THYE with 0.2% BSA, 1 mM $CaCl_2$, pH 7.1) and grown until OD_{600} of 0.15. Cells were pelleted and re-suspended in $1/10^{th}$ (of the original volume) in M2 medium (THYE with 0.2% BSA, 1 mM $CaCl_2$, 10% glycerol, pH 7.9), and frozen at −80°C for later use. For all genome editing transformations, cells were gently thawed on ice, and re-suspended in 10 volumes of M3 medium (THYE with 0.2% BSA, 1 mM $CaCl_2$, 100 ng/ml of competence-stimulating peptide CSP1 (Havarstein et al. 1995), pH 7.9). Editing constructs were added to cells at a final concentration of 0.7 ng/µl to 2.5 µg/µl, followed by incubation for 20 min at 37 °C before the addition of 2 µl of targeting constructs (similar concentration as the editing templates) and then incubated 40 min at 37 °C. Serial dilutions were then plated on agar containing appropriate antibiotics for transformants counting.

5.11 Preparation of *S. pneumoniae* genomic DNA

For transformation purposes, *S. pneumoniae* genomic DNA was extracted using the Wizard Genomic DNA Purification Kit, following instructions provided by the manufacturer (Promega). For genotyping purposes, 700 µl of overnight *S. pneumoniae* cultures were pelleted, re-suspended in 60 µl of lysozyme solution

(2 mg/ml) and incubated 30 min at 37 °C. The genomic DNA was extracted using QIAprep Spin Miniprep Kit (Qiagen).

5.12 Analysis of targeting requirement (PAM and seed sequences) by Cas9 complex in *S. pneumoniae*

The five nucleotides following the canonical *spc1* target were randomized through amplification of $R6^{8232.5}$ genomic DNA with primers W377/L426. This PCR product was then assembled with the *cat* gene and the *srtA* upstream region that were amplified from the same template with primers L422/W376. 80 ng of the assembled DNA was used to transform strains R6 and crR6. Samples for the randomized targets were prepared using the following primers: B280-B290/L426 to randomize bases 1–10 of the target and B269-B278/L426 to randomize bases 10–20. Primers L422/B268 and L422/B279 were used to amplify the *cat* gene and *srtA* upstream region to be assembled with the first and last ten PCR products, respectively. The assembled constructs were pooled and 30 ng was used to transform R6 and crR6. After transformation, cells were plated on chloramphenicol selection. For each sample more than 2×10^5 cells were pooled together in 1 ml of THYE and genomic DNA was extracted with the Promega Wizard kit. Primers B250/B251 were used to amplify the target region. PCR products were tagged and run on one Illumina MiSeq paired-end lane using 300 cycles.

5.13 Generation of targeting constructs and editing templates for genome editing in *S. pneumoniae*

Targeting constructs used for genome editing were made by Gibson assembly (Gibson et al. 2009) of Left PCRs and Right PCRs (**Table 2**). Editing constructs were made by SOEing PCR (Horton et al. 1993) fusing PCR products A (PCR A), PCR products B (PCR B) and PCR products C (PCR C) when applicable (**Table 2**). The CRISPR::Ø and CRISPR::ermAM(stop) targeting constructs were generated by PCR amplification of JEN62 and crR6 genomic DNA respectively, with oligos L409 and L481.

5.14 β-galactosidase (Miller) assay

β-galactosidase (Miller) assays were performed as previously described (Zahner and Hakenbeck 2000).

5.15 Fluorescence measurements for dCas9-mediated transcription repression and activation

Fluorescence was measured in a Tecan microplate reader. In all experiments, background fluorescence, or auto-fluorescence, was measured using a control strain lacking the GFP reporter. Auto-fluorescence was subtracted from the fluorescence readings and relative fluorescence was normalized to cells expressing a non-targeting crRNA (encoded by the BsaI sequences designed for spacer cloning). All transcription repression studies were performed in *E.coli* MG1655 cells.

Transcription activation of the *lacZ* gene (encoding β-galactosidase) was studied in *E. coli* KS1ΔZ, a strain carrying the *lacZ* gene under the control of a weak promoter that can be induced by a cI-ω fusion (Dove and Hochschild 1998). KS1ΔZ also carried a deletion of *rpoZ*, the gene encoding the ω subunit of RNAP. Green fluorescence protein (GFP) assays were performed in an *E. coli* MG1655 mutant, JEN202 (Bikard et al. 2013), in which *rpoZ* was replaced by a spectinomycin resistance gene.

All spacers used for dCas9-mediated transcription repression and activation are provided in **Table 3**.

5.16 Northern blot analysis

RNA was extracted from overnight cultures using TRIzol (Invitrogen) following the manufacturer's protocol. For each sample, 10 mg of RNA were separated on a 5% polyacrylamide gel. The RNA was electro-transferred to a charged membrane and hybridized either to a probe upstream (P511) or downstream (P510) of the T10 and B10 target sites. A probe annealing to the 5S rRNA (B507) was used as a control.

5.17 Plasmid-curing assay

S. aureus RN4220 cells harboring both the wild-type CRISPR-Cas plasmid or its variant and the target plasmid (pWJ153 or pWJ267 contained the *spc1* or *gp43* targets, respectively) were cultured in TSB supplemented with chloramphenicol (10 μg/ml). ATc was added to a final concentration of 0.25 μg/ml

during log phase (at OD_{600} of 0.6-0.8). Plasmid DNA was prepared at designated time points, linearized with the common single cutter BamHI and subjected to agarose gel electrophoresis.

5.18 Purification of Csm6

The pPS10, pPS42, pPS43 and pPS44 plasmids were transformed into *E. coli* BL21 (DE3) Rosetta 2 cells (Merck Millipore). Cultures (1 liter) were grown at 37 °C in Terrific Broth medium (Fisher Scientific) containing 100 µg/ml ampicilin and 34 µg/ml chloramphenicol until the OD600 reached 0.6. The cultures were adjusted to 0.3 mM isopropyl-1-thio-β-d-galactopyranoside and incubation was continued for 16 h at 16 °C with constant shaking. The cells were harvested by centrifugation and the pellets stored at −80 °C. All subsequent steps were performed at 4 °C. Thawed bacteria were resuspended in 35 ml of buffer A (50 mM Tris–HCl, pH 7.5, 500 mM NaCl, 200 mM Li_2SO_4, 20% sucrose, 15 mM Imidazole) containing one complete EDTA free protease inhibitor tablet (Roche). Triton X-100 and lysozyme were added to final concentrations of 0.1 % and 0.1 mg/ml, respectively. After 1 hour, the lysate was sonicated to reduce viscosity. Insoluble material was removed by centrifugation for 1 hour at 15,000 rpm in a Beckman JA-3050 rotor. The soluble extract was mixed for 1 hour with 5 ml of Ni2+ -Nitrilotriacetic acid-agarose resin (Thermo) that had been pre-equilibrated with buffer A. The resin was recovered by centrifugation, then first washed with 50 ml of buffer A, followed by washing with 50 ml of IMAC buffer (50 mM Tris-HCl pH 7.5, 250 mM NaCl, 10% glycerol) containing 50 mM imidazole. The resin was

subsequently resuspended in 10 ml of IMAC buffer containing 100 mM imidazole, and then poured into a column. The column was then eluted step-wise with 10 ml aliquots of IMAC buffer (50 mM Tris-HCl pH 7.5, 250 mM NaCl, 10% glycerol) containing 200, and 500 mM imidazole. The 500 mM imidazole elutes containing the protein was pooled together and dialyzed against 50 mM Tris-HCl pH 7.5, 50 mM NaCl, 10% glycerol. Subsequently Csm6 was purified using a 5 ml HiTrap Q Sepharose Fast Flow (GE Life Sciences), eluting with a linear gradient of 50 mM - 2 M NaCl. The peak fraction from the Q sepharose column was further purified by hydrophobic interaction chromatography using butyl sepharose 4 FF (GE Healthcare Life Sciences), eluting with stepwise lowering of ammonium sulphate concentration from 1 M to 50 mM. The final purification step was performed using size exclusion chromatography with Superdex 200 10/300 GL (GE Healthcare) column, using buffer B (50 mM Tris-HCl pH 7.5, 5% glycerol, 150 mM NaCl, 1 mM TCEP).

5.19 Csm6 RNA cleavage assay

RNA cleavage reactions were performed at 37°C with 1 μM of 5’-radiolabeled (R55 and R24) and 3’-radiolabeled ssRNA (R55) substrates and 10 μM of wild type or mutant Csm6. The reaction was carried out in reaction buffer containing 50 mM Tris-HCl [pH7.5], 30 mM NaCl, 2 mM DTT and 1% glycerol. Reaction mixtures were withdrawn at specified time intervals and subsequently quenched with 90% formamide and 50 mM EDTA. Reaction products were separated by

denaturing PAGE and the gel was visualized by phosphorimaging. The 5'-radiolabeled decade RNA ladder (Life Technologies) was used as a size marker.

5.20 Csm6 DNA cleavage assay

DNA cleavage reactions were performed at 37°C for up to 2 hours with 1 μM of 5'-radiolabeled ssDNA (PS362) and dsDNA (PS362/PS363) substrates and 10 μM of wild type Csm6. The reaction was carried out in reaction buffer containing 50 mM Tris-HCl [pH7.5], 10 mM $MgCl_2$, 30 mM NaCl, 2 mM DTT and 1% glycerol. Reaction mixtures were withdrawn at specified time intervals and subsequently quenched with 90% formamide and 50 mM EDTA. Reaction products were separated by denaturing PAGE and the gel was visualized by phosphorimaging. The 5'-radiolabeled 10 bp DNA ladder (Promega) was used as a size marker.

5.21 Co-transcriptional DNA cleavage assay

Elongation complexes (ECs) were reconstituted essentially as described in (Samai et al. 2015). Typically, 2 μl 1 pmol/μl of template strand (TS) and 1 μl of 4 pmol/μl RNA oligos were mixed in 1 × transcription buffer and incubated at 65°C for 5 min, followed by gradual cooling to room temperature. After addition of 1.5 μl E. coli RNAP core enzyme (NEB), the reaction was incubated at 25°C for 25–30 min and at 37°C for 1 min. Then, 4 μl 1.25 pmol/μl nontemplate strand (NTS) (pretreated by heating to 65°C for 5 min, then on ice for 2 min, and finally at 37°C for 2 min) was added and incubated for 10–15 min at 37°C. The final concentration of TS was 0.10 pmol/μl after adding supplement buffer to obtain

transcription conditions. Assembled ECs were kept on ice until use. In a transcription coupled DNA cleavage assay, 5 nM of EC was used and, Cas10-Csm complex was added to a final concentration of 5 nM and 50 nM. Transcription was initiated with the addition of 2.5 mM of RNTPs. All the reactions were performed at 37°C. For all the DNA cleavage time course experiments, RNTPs were added to the elongation complex (EC), prior to the addition of Cas10-Csm complex. After addition of Cas10-Csm, the samples were collected at 30 min, and quenched by mixing with Proteinase K (NEB) and 20 mM EDTA. The DNA/RNA samples were then extracted using phenolchloroform-isoamyl alcohol (25:24:1), ethanol precipitated and resuspended into loading buffer (90% formamide). The DNA products were heater at 95°C for 5 min before loading onto the gel. Cleavage products were resolved on a 12% denaturing polyacrylamide gels containing 7 M urea and visualized by phosphorimaging (Typhoon, GE Life Sciences).

5.22 Phage infections and plate reader growth curves

S. aureus RN4220 cells were infected with phages ΦNM1γ6 or ΦNM4γ4 during early log phase (at OD_{600} of 0.3-0.4). Plate reader growth curves were measured as previously described (Goldberg et al. 2014) with slight modifications. Briefly, overnight cultures were launched from single colonies and diluted 1:250 in TSB broth supplemented with 5mM $CaCl_2$ and appropriate antibiotics. After 1 h of growth, phage was added at a multiplicity of infection (MOI) of 5. Measurements were taken every 10 minutes.

5.23 Measurement of average burst size

S. aureus RN4220 cells with appropriate CRISPR-Cas plasmids were grown in TSB supplemented with 5mM $CaCl_2$ and appropriate antibiotics to an OD_{600} of 0.3-0.5. Cells were infected by ΦNM1γ6 at MOI = 0.1 for 5 min. Cells were immediately washed in TSB twice at 4 °C and re-suspended in equal initial volume. An aliquot of cells were spotted on heart infusion (BD) soft agar plates with a sensitive lawn (i.e. RN4220). The rest of the cells were incubated at 37 °C for another 75 min before an aliquot of cells were spotted on a sensitive lawn. Agar plates were incubated at 37 °C for 16-20 hours before plaques were enumerated. Average burst size was calculated as the ratio of plaques formed at 80 min to plaques formed at 5 min for each strain of interest.

5.24 Extraction of total RNA in *S. aureus*

10-25 ml of *S. aureus* culture were pelleted and immediately frozen at -20 °C. Pellets were gently thawed at 4 °C and washed with 1 ml ice-cold TE pH 6.8. Pellets were re-suspended in 100 µl of ice-cold TE pH 6.8 and mixed with 750 µl of ice-cold TRIzol®. The mix was transferred into a 2 ml microtubes pre-filled with 0.25 cm^3 of 0.1 mm glass beads on ice. Cells were disrupted using MiniBeadbeater-1 (BioSpec Products) at an intensity setting of 42 for 30 seconds twice at 4 °C. 200 µl of chloroform was added to the disrupted mix was and the rest of RNA extraction protocol was followed according to TRIzol®.

5.25 Extraction of total DNA in *S. aureus*

10-25 ml of *S. aureus* culture were pelleted and immediately frozen at -20 °C. Pellets were gently thawed at 4 °C and washed with 1 ml ice-cold TE pH 8.0. Pellets were re-suspended in 400 µl of ice-cold TE pH 8.0 and mixed with 500 µl of ice-cold Phenol/Chloroform/Isoamyl alcohol (25:24:1) (Fisher Scientific). The mix was transferred into a 2 ml microtubes pre-filled with 0.25 cm^3 of 0.1 mm glass beads on ice. Cells were disrupted using Mini-Beadbeater-1 (BioSpec Products) at an intensity setting of 42 for 30 seconds twice at 4 °C. The disrupted mix was centrifuged at 16,000 rcf for 10 minutes at room temperature. The aqueous phase was collected and mixed with 500 µl of chloroform and centrifuged as above. The aqueous phase was collected again and mixed with 1 ml of isopropanol. Precipitated DNA was washed with 1 ml of 75% ethanol, air dried and dissolved in 50-300 µl of water.

5.26 qPCR

S. aureus RN4220 were infected by ΦNM1γ6 (MOI=5) during early log phase (at OD_{600} of 0.3-0.4). qPCR was performed using Fast SYBR® Green Master Mix (Life technologies) and 7900HT Fast Real-Time PCR System (Applied Biosystems). For RNA samples, total RNA was treated with DNase I (Sigma-Aldrich). 1 µg of DNase I treated RNA samples were subjected to reverse transcription using M-MuLV Reverse Transcriptase (NEB) and 100 ng of random hexamer (Invitrogen) according to the NEB protocol. The resulting cDNA was diluted 5 times as stocks. 500 nM of primers were used and 0.2 µl of the cDNA

stock was used as template for a 10 µl reaction according to the Fast SYBR® Green Master Mix protocol. For DNA samples, 25 ng of total DNA were used as template. The house-keeping *rho* gene was used as endogenous control for normalization (Theis et al. 2007). Primers used for amplification are shown in **Table 5**.

5.27 RNA sequencing

Total RNA was treated with DNase I (Sigma-Aldrich) and subjected to TruSeq Stranded mRNA Library Prep Kit (Illumina) without rRNA depletion and Illumina NextSeq. Reads were aligned to reference genomes using Bowtie and sorted using Samtools. Using a custom script, sorted reads were accessed via Pysam, normalized as reads per million values, and plotted as the average over consecutive windows of 500 base pairs using matplotlib tools for IPython.

5.28 Southern blot

20 µg of total DNA prepared from infected cells were digested with restriction enzymes EcoRI and PsiI for 5 hours and resolved on a 1% argarose gel. DNA fragments were transferred from the gel via capillary action to a Hybond membrane (GE Healthcare) using alkaline transfer (Sambrook 1989). Probes for the upstream and downstream fragments were produced via PCR of ΦNM1γ6 DNA using primers W863/W864 and W865/W866, respectively, and α-^{32}P-dATP in addition to regular dNTPs. Hybridization was performed at 65°C overnight in Church buffer (Sambrook 1989).

Table 1 Spacers used in Chapter 3

Target	5-nt upstream	Spacer sequence (5'-3')	5-nt downstream
gp5	TTTCG	CCATTCATCTAATTTCAAGGCTATGTTTGATGTAG	-
gp14	TTCTA	CTACGTCCGTAATGCTAGGATTTGCAAATTTCTTA	-
gp19	TTCTA	CACCCATATCATCTAGTACAAGTAAATCAATATCA	-
gp32	CAACT	GTAAACCTTTGATTGCTCTTAGCTCGAGTTATGTGC	-
gp43	TTCTA	ATTCGTCATCTTCAAGTAATGCCTCTAAATCAATA	-
gp43 (3 mm)	TTCTA	TTTCGTCTTCTTCATGTAATGCCTCTAAATCAATA	-
gp43 (4 mm)	TTCTA	TTTCGTCTTCATCATGTAATGCCTCTAAATCAATA	-
gp43 (5 mm)	TTCTA	TTTGGTCTTCATCATGTAATGCCTCTAAATCAATA	-
gp43 (type II)	-	ACTTCACACAAGATAACATTATTGATTTAG	AGGCA
gp50	TTCTA	GTCCAATATTTTCTGCGATTTCATCTAGTGCTTCA	-
gp59	TTCTA	ATCGCGTTAAACGCCAATCTTGTTCGTGTCGTTTG	-
nes	AGAGA	ACGTATGCCGAAGTATATAAATCATCAGTACAAAG	

Table 2 Targeting constructs and editing templates used for genome editing in *S. pneumonaie*

	Targeting Constructs					Editing Templates						
Edition	Template DNA	Left PCR	Right PCR	Spacer sequence	PAM	Template DNA	PCR A	PCR B	PCR C	SOEing PCR	Name of resulting strains	Primers used to verify edited genotype
bgaA R>A	crR6Rk	W256/W391	W392/L403	GCTCACTAGCAGGGTTGTGGGTTGTACGGA	TGG	R6	W403/W397	W398/W404	N/A	W403/W404	JEN56	W403/W404
bgaA NE>AA	crR6Rk	W256/W391	W392/L403	GCTCACTAGCAGGGTTGTGGGTTGTACGGA	TGG	R6	W403/W431	W432/W433	W434/W404	W403/W404	JEN60	W403/W404
ΔbgaA	crR6Rk	W256/W391	W392/L403	GCTCACTAGCAGGGTTGTGGGTTGTACGGA	TGG	R6	B255/B256	B257/B258	N/A	B255/B258	JEN52	W393/W405
ΔsrtA	crR6Rc	W256/B218	B217/L403	TCCTAGCAGGATTTCTGATATTACTGTCAC	TGG	R6	B230/W463	W464/B229	N/A	B230/B229	JEN51	W422/W426
ermB Stop	crR6Rk	W256/W356	W357/L403	TTTAAAAGAAACCGATACCGTTTACGAAAT	TGG	JEN38	L422/W370	W371/L426	N/A	L422/L426	JEN43	L457/L458
ΔsrtA ΔbgaA	JEN51 (for Left PCR) and JEN52 (for Right PCR)	W256/W465	W466/W403	same as the ones used for ΔsrtA and ΔbgaA	TGG	same as the ones used for ΔsrtA and ΔbgaA					JEN64	same as the ones used for ΔsrtA and ΔbgaA

Table 3 Spacers used for dCas9-mediated transcription repression and activation

Name	Spacer sequence (5'-3')	Target
T0	CTACGGAACTCTTGTGCGTAAGGAAAAGTA	pDB127
T1	CTGAGCAATCACCTATGAACTGTCGACTCG	pDB127
T3	TGTCGACTCGAGGCCTTGACAGGTACCTCA	pDB127
T4	CCTTGACAGGTACCTCATGGATACCTATAA	pDB127
T5	CAGGTACCTCATGGATACCTATAATGGTTC	pDB127
T6	ACCTATAATGGTTCCGGAATTCCTTTAAAG	pDB127
T7	CCTTTAAAGAGGAGAAATCTAGATGAGTAA	pDB127
T8	CTAGATGAGTAAAGGAGAAGAACTTTTCAC	pDB127
T9	TGGAGTTGTCCCAATTCTTGTTGAATTAGA	pDB127
T10	TCAAGAGTGCCATGCCCGAAGGTTATGTAC	pDB127
T11	GATACCCTTGTTAATAGAATCGAGTTAAAA	pDB127
B0	TCTTCAGAAATGAGCTTTTGCTCCTCTGCT	pDB127
B1	CCTTTACTCATCTAGATTTCTCCTCTTTAA	pDB127
B2	TAGGTGATTGCTCAGGACATTTCTGTTAGA	pDB127
B3	CTCGAGTCGACAGTTCATAGGTGATTGCTC	pDB127
B4	ACCTGTCAAGGCCTCGAGTCGACAGTTCAT	pDB127
B5	CCATTATAGGTATCCATGAGGTACCTGTCA	pDB127
B7	TCCTCTTTAAAGGAATTCCGGAACCATTAT	pDB127
B8	CATCTAGATTTCTCCTCTTTAAAGGAATTC	pDB127
B9	ATTAACATCACCATCTAATTCAACAAGAAT	pDB127
B10	CTTGAAAAAGTCATGCTGTTTCATATGATC	pDB127
B11	CAATACCTTTTAACTCGATTCTATTAACAA	pDB127
Z1	ACAACACGCACGGTGTTACATTAGGCACCC	KS1ΔZ lacZ promoter
Z2	GATCTTCGACAACACGCACGGTGTTACATT	KS1ΔZ lacZ promoter
Z3	GGCTGCAGGTCGGATCTTCGACAACACGCA	KS1ΔZ lacZ promoter
Z4	ATTCTGTCGAAGATCAGCTTGGCTGCAGGT	KS1ΔZ lacZ promoter
W101	GATCTTCCACAACACGCACGGTGTTACATT	pWJ89
W102	GGCTGCAGGCCGGATCTTCCACAACACGCA	pWJ89
W103	CCTCTATGGATTATCACCTTGGCTGCAGGC	pWJ89
W104	GTCGACTCGAGCCTCTATGGATTATCACCT	pWJ89
W105	TCACCTATGAACTGTCGACTCGAGCCTCTA	pWJ89
W106	GGTATGCCTAATGTAACACCGTGCGTGTTG	pWJ89
W107	AATGTAACACCGTGCGTGTTGTGGAAGATC	pWJ89
W108	GCGTGTTGTGGAAGATCCGGCCTGCAGCCA	pWJ89
W109	CCGGCCTGCAGCCAAGGTGATAATCCATAG	pWJ89
W110	AATCCATAGAGGCTCGAGTCGACAGTTCAT	pWJ89

Table 4 DNA oligonucleotides used in Chapter 2

Name	Sequence (5'-3')
B217	TCAGGCTTTCAAACAGATACAAGGGCGACCCGCTTTGTGATCTAGGCACCAATAACTGCC
B218	GTGACAGTAATATCAGAAATCCTGCTAGGAGTTTTGGGACCATTCAAAACAGC
B229	GGGTTTCAAGTCTTTGTAGCAAGAG
B230	GCCAATGAACGGGAACCCTTGGTC
B250	NNNNGACGAGGCAATGGCTGAAATC
B251	NNNNTTATTTGGCTCATATTTGCTG
B255	CTTTACACCAATCGCTGCAACAGAC
B256	CAAAATTTCTAGTCTTCTTTGCCTTTCCCCATAAAACCCTCCTTA
B257	AGGGTTTTATGGGGAAAGGCAAAGAAGACTAGAAATTTTGATACC
B258	CTTACGGTGCATAAAGTCAATTTCC
B269	TGGCTCGATTTCAGCCATTGC
B270	CTTTGACGAGGCAATGGCTGAAATCGAGCCAANAAAGCGCAAG
B271	CTTTGACGAGGCAATGGCTGAAATCGAGCCAAANAAGCGCAAG
B272	CTTTGACGAGGCAATGGCTGAAATCGAGCCAAAANAGCGCAAG
B273	CTTTGACGAGGCAATGGCTGAAATCGAGCCAAAAANGCGCAAG
B274	CTTTGACGAGGCAATGGCTGAAATCGAGCCAAAAAANCGCAAG
B275	CTTTGACGAGGCAATGGCTGAAATCGAGCCAAAAAAGNGCAAG
B276	CTTTGACGAGGCAATGGCTGAAATCGAGCCAAAAAAGCNCAAGAAG
B277	CTTTGACGAGGCAATGGCTGAAATCGAGCCAAAAAAGCGNAAGAAG
B278	CTTTGACGAGGCAATGGCTGAAATCGAGCCAAAAAAGCGCNAGAAG
B279	GCGCTTTTTTGGCTCGATTTCAG
B280	CAATGGCTGAAATCGAGCCAAAAAAGCGCANGAAGAAATC
B281	CAATGGCTGAAATCGAGCCAAAAAAGCGCAANAAGAAATC
B282	CAATGGCTGAAATCGAGCCAAAAAAGCGCAAGNAGAAATC
B283	CAATGGCTGAAATCGAGCCAAAAAAGCGCAAGANGAAATC
B284	CAATGGCTGAAATCGAGCCAAAAAAGCGCAAGAANAAATC
B285	CAATGGCTGAAATCGAGCCAAAAAAGCGCAAGAAGNAATCAACC
B286	CAATGGCTGAAATCGAGCCAAAAAAGCGCAAGAAGANATCAACC
B287	CAATGGCTGAAATCGAGCCAAAAAAGCGCAAGAAGAANTCAACC
B288	CAATGGCTGAAATCGAGCCAAAAAAGCGCAAGAAGAAANCAACC
B289	CAATGGCTGAAATCGAGCCAAAAAAGCGCAAGAAGAAATNAACCAGC
B290	CAATGGCTGAAATCGAGCCAAAAAAGCGCAAGAAGAAATCNACCAGC
B296	GATCCTCCATCCGTACAACCCACAACCCTGG
B297	AATTCCAGGGTTGTGGGTTGTACGGATGGAG
B298	CATGGATCCTATTTCTTAATAACTAAAAATATGG
B299	CATGAATTCAACTCAACAAGTCTCAGTGTGCTG
B300	AAACATTTTTTCTCCATTTAGGAAAAAGGATGCTG
B301	AAAACAGCATCCTTTTTCCTAAATGGAGAAAAAAT
B302	AAACCTTAAATCAGTCACAAATAGCAGCAAAATTG
B303	AAAACAATTTTGCTGCTATTTGTGACTGATTTAAG
B304	AAACTTTTCATCATACGACCAATCTGCTTTATTTG
B305	AAAACAAATAAAGCAGATTGGTCGTATGATGAAAA
B306	AAACTCGTCCAGAAGTTATCGTAAAAGAAATCGAG
B307	AAAACTCGATTTCTTTTACGATAACTTCTGGACGA
B308	AAACAATCTCTCCAAGGTTTCCTTAAAAATCTCTG
B309	AAAACAGAGATTTTTAAGGAAACCTTGGAGAGATT
B310	AAACGCCATCGTCAGGAAGAAGCTATGCTTGAGTG
B311	AAAACACTCAAGCATAGCTTCTTCCTGACGATGGC
B312	AAACATCTCTATACTTATTGAAATTTCTTTGTATG

B313	AAAACATACAAAGAAATTTCAATAAGTATAGAGAT
B314	AAACTAGCTGTGATAGTCCGCAAAACCAGCCTTCG
B315	AAAACGAAGGCTGGTTTTGCGGACTATCACAGCTA
B316	AAACATCGGAAGGTCGAGCAAGTAATTATCTTTTG
B317	AAAACAAAAGATAATTACTTGCTCGACCTTCCGAT
B318	AAACAAGATGGTATCGCAAAGTAAGTGACAATAAG
B319	AAAACTTATTGTCACTTACTTTGCGATACCATCTT
B320	GAGACCTTTGAGCTTCCGAGACTGGTCTCAGTTTTGGGACCATTCAAAACAG
B321	TGAGACCAGTCTCGGAAGCTCAAAGGTCTCGTTTTAGAGCTATGCTGTTTTG
B337	GACGCTATTTGTGCCGATAGCTAAGCCTATTGAGTATTTC
B338	GAAATACTCAATAGGCTTAGCTATCGGCACAAATAGCGTC
B339	GGAAACTTTGTGGAACAATGGCATCGACATCATAATCACT
B340	AGTGATTATGATGTCGATGCCATTGTTCCACAAAGTTTCC
B352	AAACTACTTTACGCAGCGCGGAGTTCGGTTTTTTG
B353	AAAACAAAAAACCGAACTCCGCGCTGCGTAAAGTA
B368	TTTGGATCCTTATTTGTATAGTTCATCCATGCC
B369	TCGAGGCCTTGACAGGTACCTCATGGATACCTATAATGGTTCCGG
B370	AATTCCGGAACCATTATAGGTATCCATGAGGTACCTGTCAAGGCC
B371	ACAGAATTCCTTTAAAGAGGAGAAATCTAGATGAGTAAAGGAGAAGAACTTTTC
B441	TCCTGAACAGTTACGCGTGCAGCTGCGTCACCTCCTAGCTGACTCAAATC
B442	TGAGTCAGCTAGGAGGTGACGCAGCTGCACGCGTAACTGTTCAGGACGCTG
B446	CTATTGCTGAAGGTCGTCGTGCTGCAGATAAGAAATACTCAATAGGCTTAG
B448	CCTATTGAGTATTTCTTATCTGCAGCACGACGACCTTCAGCAATAGCGG
B507	TCGGCGCTACGGCGTTTCACTTCTGAGTTCGG
H001	GGGCACTTTTTCACTCATTTTAGCTTCCTTAGCTCCTGAAAATC
H002	GGTGCCAGCCAATGATTTTTTTAAGGCAGTTATTGG
H003	GCTAAGGAAGCTAAAATGAGTGAAAAAGTGCCCGCC
H004	ACTGCCTTAAAAAAATCATTGGCTGGCACCAAGCAG
L402	TTTCCCTTGAACTAGTCGAAGG
L403	AGTCATCCCAGCAACAAATGG
L403	AGTCATCCCAGCAACAAATGG
L409	CGTGGTAAATCGGATAACGTTCCAAGTGAAG
L422	TGCTCTTCTTCACAAACAAGGG
L426	AAGCCAAAGTTTGGCACCACC
L430	GTAGCTTATTCAGTCCTAGTGG
L444	CGTTTGTTGAACTAATGGGTGCAAATTACGAATCTTCTCCTGACG
L445	CGTCAGGAGAAGATTCGTAATTTGCACCCATTAGTTCAACAAACG
L446	GATATTATGGAGCCTATTTTTGTGGGTTTTTAGGCATAAAACTATATG
L447	CATATAGTTTTATGCCTAAAAACCCACAAAAATAGGCTCCATAATATC
L448	ATTATTTCTTAATAACTAAAAATATGG
L457	CGTGTACAATTGCTAGCGTACGGC
L458	GCACCGGTGATCACTAGTCCTAGG
L459	CCTAGGACTAGTGATCACCGGTGCAAATATGAGCCAAATAAATATAT
L461	GCCGTACGCTAGCAATTGTACACGTTTGTTGAACTAATGGGTGC
L481	TTCAAATTTTCCCATTTGATTCTCC
L488	CCATATTTTTAGTTATTAAGAAATAATACCAGCCATCAGTCACCTCC
P510	AAACTTGACTTCAGCACGTGTCTTGTAGTTCC
P511	CAACAAGAATTGGGACAACTCCAGTGAAAAGTTC
W256	AGACGATTCAATAGACAATAAGG
W286	GTTTTGGGACCATTCAAAACAGCATAGCTCTAAAACCTCGTAGAC

W287	GCTATGCTGTTTTGAATGGTCCCAAAACCATTATTTTAACACACGAGGTG
W288	GCTATGCTGTTTTGAATGGTCCCAAAACGCACCCATTAGTTCAACAAACG
W326	AATTCTTTTCTTCATCATCGGTC
W327	AAGAAAGAATGAAGATTGTTCATG
W341	GGTACTAATCAAAATAGTGAGGAGG
W354	GTTTTTCAAAATCTGCGGTTGCG
W355	AAAAATTGAAAAAATGGTGGAAACAC
W356	ATTTCGTAAACGGTATCGGTTTCTTTTAAAGTTTTGGGACCATTCAAAACAGC
W357	TTTAAAAGAAACCGATACCGTTTACGAAATGTTTTAGAGCTATGCTGTTTTGA
W365	AAACGGTATCGGTTTCTTTTAAATTCAATTGTTTTGGGACCATTCAAAACAGC
W366	AATTGAATTTAAAAGAAACCGATACCGTTTGTTTTAGAGCTATGCTGTTTTGA
W370	GTTCCTTAAACCAAAACGGTATCGGTTTCTTTTAAATTC
W371	GAAACCGATACCGTTTTGGTTTAAGGAACAGGTAAAGGGCATTTAAC
W376	CGATTTCAGCCATTGCCTCGTC
W377	GCCTTTGACGAGGCAATGGCTGAAATCGNNNNNAAAAAGCGCAAGAAGAAATCAAC
W391	TCCGTACAACCCACAACCCTGCTAGTGAGCGTTTTGGGACCATTCAAAACAGC
W392	GCTCACTAGCAGGGTTGTGGGTTGTACGGAGTTTTAGAGCTATGCTGTTTTGA
W393	TTGTTGCCACTCTTCCTTCTTTC
W397	CAGGGTTGTGGGTTGTTGCGATGGAGTTAACTCCCATCTCC
W398	GGGAGTTAACTCCATCGCAACAACCCACAACCCTGCTAGTG
W403	GTGGTATCTATCGTGATGTGAC
W404	TTACCGAAACGGAATTTATCTGC
W405	AAAGCTAGAGTTCCGCAATTGG
W431	GTGGGTTGTACGGATTGAGTTAACTCCCATCTCCTTC
W432	GATGGGAGTTAACTCAATCCGTACAACCCACAACCCTG
W433	GCTTCACCTATTGCAGCACCAATTGACCACATGAAGATAG
W434	GTGGTCAATTGGTGCTGCAATAGGTGAAGCTAATGGTGATG
W463	CTGATTTGTATTAATTTTGAGACATTATGCTTCACCTTC
W464	GCATAATGTCTCAAAATTAATACAAATCAGTGAAATCATG
W465	GTTTTGGGACCATTCAAAACAGCATAGCTCTAAAACGTGACAGTAATATCAG
W466	GTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACGCTCACTAGCAGGGTTG
W542	ATACTTTACGCAGCGCGGAGTTCGGTTTTGTAGGAGTGGTAGTATATACACGAGTACAT
W550	CTGTATTACTGCATTTATTAAGAGTATTAACGACGACCTTCAGCAATAG
W551	TTGCTGAAGGTCGTCGTTAATACTCTTAATAAATGCAGTAATACAGGG
W552	CAGGTTCAGTCTGCCAACAATTTGTTCAATAATAGTTTTAATGACCTCCG
W553	CGGAGGTCATTAAAACTATTATTGAACAAATTGTTGGCAGACTGAACCTG
W573	TCATGCCCAGTCATTTCTTCACCTGTGGAGCTTTTTAAGTCCTGTTGATACCGGGAAGCC
W574	TCAGGCTTTCAAACAGATACAAGGGCGACCCGCTTTGTGATCTAGGCACCAATAACTGCC

Table 5 DNA oligonucleotides used in Chapter 3

Name	Sequence (5'-3')	Purpose
L342	aaaGGGCCCAAATAATATTTTCATTATAGCACCTC	Cloning ofΔcsm6
L343	aaaCGGCCGGAAAAAAATAAGGAATTTAAAGAGC	Cloning ofΔcsm6
PS11	CGCCATATGAAAATATTATTTAGTCCAATAGG	Cloning of Csm6
PS12	CGCCTCGAGTAATAGCTCTTTAAATTCC	Cloning of Csm6
PS243	GGTTTAAGAAATTCCATAGCCGCTAATTTAGATAC	Cloning of Csm6 (R364A)
PS244	GTATCTAAATTAGCGGCTATGGAATTTCTTAAACC	Cloning of Csm6 (R364A)
PS245	CGATATAAATGGTTTAGCAAATTCCATAGCCC	Cloning of Csm6 (H3694A)
PS246	GGGCTATGGAATTTGCTAAACCATTTATATCG	Cloning of Csm6 (H3694A)
PS247	GATATAAATGGTTTAGCAAATTCCATAGCCGCTAATTTAGATAC	Cloning of Csm6(R364A, H369A)
PS248	GTATCTAAATTAGCGGCTATGGAATTTGCTAAACCATTTATATC	Cloning of Csm6(R364A, H369A)
PS362	GTATAGGCACAGCGGGAATAAGGCTATCACTGATGTGCTCGAGTAACTTAACAGC	DNA cleavage assay
PS363	GCTGTTAAGTTACTCGAGCACATCAGTGATAGCCTTATTCCCGCTGTGCCTATAC	DNA cleavage assay
PS465	GAATCTAGTATGATTGGAGCAATTGcTTCTCCTGTAGTTAGAGATTTGCAAACC	Cloning of Csm3(D32A)
PS466	GGTTTGCAAATCTCTAACTACAGGAGAAgCAATTGCTCCAATCATACTAGATTC	Cloning of Csm3(D32A)
W614	GGTTATACTAAAAGTCGTTTGTTGG	Cloning
W852	CCAACAAACGACTTTTAGTATAACC	Cloning
W863	TATGTGGCCGAAAAAACCAAGC	Probe for southern blot
W864	TTGGATATCCATAGTTTTTACACC	Probe for southern blot
W865	ATGACATCAGAAGCGGTTGACG	Probe for southern blot
W866	TGGTTTAACAGTGCGTCTAATCC	Probe for southern blot
W893	TCCATTCGGTAAATCAATTGCAC	gp43 qPCR
W894	TGTTTTTGAGATAAACGCATTTGC	gp43 qPCR
W897	GAAGAATCAGATGGAGATAATGG	gp42 qPCR
W898	AAGACGCTTGTTATATTCTTCTTG	gp42 qPCR
W901	TGCAGTTAAACGCTACAACAGG	gp44 qPCR
W902	CTTCATACTCCTTGAAATCGTTC	gp44 qPCR
W905	TTATAGTAAGAAAACAGCAGAGTC	gp37 qPCR
W906	AAACGCTCTTCTTGTATCTGTTC	gp37 qPCR
W909	TGAATGCATTCAGCGGATCATC	gp52 qPCR
W910	GATTGTCCAACTTGTTCAGACC	gp52 qPCR
W915	GTCAATGACCATAACGCAGAAG	rho qPCR as endogenous control
W916	CAATCGGTGTTACTAAATCCATG	rho qPCR as endogenous control
W1085	GATTAGACATTCACCTTCAATAAC	gp14 qPCR
W1086	TTGCGCTTGTCCTGTGATTTTC	gp14 qPCR
W1105	CATCTTCAAGTAATGCCTCTAAATCAATAATGTTATCGATCTCCTAGGTCATTTGATATG	Construction of pWJ267
W1106	GATTTAGAGGCATTACTTGAAGATGACGAATTAGAAGCATATTTATCAGAGCTCGTGCTA	Construction of pWJ267
W1113	TATGAGATAATGCCGACTGTACTTTTTACAGTCGGTATCAGAGCTCGTGCTATAATT	Cloning of pCsm6
W1127	GCTCTCTATCATTGATAGAGTGAGTTAAACAATGAGGTGCTATA*A* TG	Cloning of pCsm6
W1128	GTAAAAAGTACAGTCGGCATTATCTCATATTTATCATAATAGCTCTTTAAATTCC	Cloning of pCsm6
W1129	TCACTCTATCAATGATAGAGAGC	Cloning of pCsm6

Table 6 RNA oligonucleotides used in Chapter 3

Name	Sequence (5'-3')
R24	CGUGUCGCCCUUAUUCCGAUAGUG
R55	GCUGUUAAGUUACUCGAGCACAUCAGUGAUAGCCUUAUUCCCGCUGUGCCUAUAC

REFERENCES

Anantharaman V, Makarova KS, Burroughs AM, Koonin EV, Aravind L. 2013. Comprehensive analysis of the HEPN superfamily: identification of novel roles in intra-genomic conflicts, defense, pathogenesis and RNA processing. *Biology direct* **8**: 15.

Anton T, Bultmann S, Leonhardt H, Markaki Y. 2014. Visualization of specific DNA sequences in living mouse embryonic stem cells with a programmable fluorescent CRISPR/Cas system. *Nucleus* **5**: 163-172.

Bae T, Baba T, Hiramatsu K, Schneewind O. 2006. Prophages of Staphylococcus aureus Newman and their contribution to virulence. *Molecular microbiology* **62**: 1035-1047.

Bae T, Schneewind O. 2006. Allelic replacement in Staphylococcus aureus with inducible counter-selection. *Plasmid* **55**: 58-63.

Barberis A, Pearlberg J, Simkovich N, Farrell S, Reinagel P, Bamdad C, Sigal G, Ptashne M. 1995. Contact with a component of the polymerase II holoenzyme suffices for gene activation. *Cell* **81**: 359-368.

Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**: 1709-1712.

Barrangou R, Marraffini LA. 2014. CRISPR-Cas systems: Prokaryotes upgrade to adaptive immunity. *Molecular cell* **54**: 234-244.

Bikard D, Euler CW, Jiang W, Nussenzweig PM, Goldberg GW, Duportet X, Fischetti VA, Marraffini LA. 2014. Exploiting CRISPR-Cas nucleases to produce sequence-specific antimicrobials. *Nature biotechnology* **32**: 1146-1150.

Bikard D, Hatoum-Aslan A, Mucida D, Marraffini LA. 2012. CRISPR interference can prevent natural transformation and virulence acquisition during in vivo bacterial infection. *Cell host & microbe* **12**: 177-186.

Bikard D, Jiang W, Samai P, Hochschild A, Zhang F, Marraffini LA. 2013. Programmable repression and activation of bacterial gene expression

using an engineered CRISPR-Cas system. *Nucleic acids research* **41**: 7429-7437.

Blomfield IC, Vaughn V, Rest RF, Eisenstein BI. 1991. Allelic exchange in Escherichia coli using the Bacillus subtilis sacB gene and a temperature-sensitive pSC101 replicon. *Molecular microbiology* **5**: 1447-1457.

Bolotin A, Quinquis B, Sorokin A, Ehrlich SD. 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151**: 2551-2561.

Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJ, Snijders AP, Dickman MJ, Makarova KS, Koonin EV, van der Oost J. 2008. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**: 960-964.

Burgess RR. 1969. Separation and characterization of the subunits of ribonucleic acid polymerase. *The Journal of biological chemistry* **244**: 6168-6176.

Carroll D. 2011. Zinc-finger nucleases: a panoramic view. *Current gene therapy* **11**: 2-10.

Carte J, Wang R, Li H, Terns RM, Terns MP. 2008. Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes & development* **22**: 3489-3496.

Carter DM, Radding CM. 1971. The role of exonuclease and beta protein of phage lambda in genetic recombination. II. Substrate specificity and the mode of action of lambda exonuclease. *The Journal of biological chemistry* **246**: 2502-2512.

Chen B, Gilbert LA, Cimini BA, Schnitzbauer J, Zhang W, Li GW, Park J, Blackburn EH, Weissman JS, Qi LS et al. 2013. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* **155**: 1479-1491.

Chopin MC, Chopin A, Bidnenko E. 2005. Phage abortive infection in lactococci: variations on a theme. *Current opinion in microbiology* **8**: 473-479.

Choudhary E, Thakur P, Pareek M, Agarwal N. 2015. Gene silencing by CRISPR interference in mycobacteria. *Nature communications* **6**: 6267.

Claverys JP, Martin B, Polard P. 2009. The genetic transformation machinery: composition, localization, and mechanism. *FEMS microbiology reviews* **33**: 643-656.

Cobb RE, Wang Y, Zhao H. 2015. High-efficiency multiplex genome editing of Streptomyces species using an engineered CRISPR/Cas system. *ACS synthetic biology* **4**: 723-728.

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA et al. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**: 819-823.

Cormack BP, Valdivia RH, Falkow S. 1996. FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* **173**: 33-38.

Costantino N, Court DL. 2003. Enhanced levels of lambda Red-mediated recombinants in mismatch repair mutants. *Proceedings of the National Academy of Sciences of the United States of America* **100**: 15748-15753.

Court DL, Sawitzke JA, Thomason LC. 2002. Genetic engineering using homologous recombination. *Annual review of genetics* **36**: 361-388.

d'Herelle F. 1917. An invisible antagonist microbe of dysentery bacillus. *Cr Hebd Acad Sci* **165**: 373-375.

Datsenko KA, Pougach K, Tikhonov A, Wanner BL, Severinov K, Semenova E. 2012. Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nature communications* **3**: 945.

Datsenko KA, Wanner BL. 2000. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proceedings of the National Academy of Sciences of the United States of America* **97**: 6640-6645.

Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, Eckert MR, Vogel J, Charpentier E. 2011. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**: 602-607.

Deng L, Garrett RA, Shah SA, Peng X, She Q. 2013. A novel interference mechanism by a type IIIB CRISPR-Cmr module in Sulfolobus. *Molecular microbiology* **87**: 1088-1099.

Deveau H, Barrangou R, Garneau JE, Labonte J, Fremaux C, Boyaval P, Romero DA, Horvath P, Moineau S. 2008. Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. *Journal of bacteriology* **190**: 1390-1400.

Deveau H, Garneau JE, Moineau S. 2010. CRISPR/Cas system and its role in phage-bacteria interactions. *Annual review of microbiology* **64**: 475-493.

Dove SL, Hochschild A. 1998. Conversion of the omega subunit of Escherichia coli RNA polymerase into a transcriptional activator or an activation target. *Genes & development* **12**: 745-754.

Dove SL, Joung JK, Hochschild A. 1997. Activation of prokaryotic transcription through arbitrary protein-protein contacts. *Nature* **386**: 627-630.

Ellis HM, Yu D, DiTizio T, Court DL. 2001. High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *Proceedings of the National Academy of Sciences of the United States of America* **98**: 6742-6746.

Elmore JR, Sheppard NF, Ramia N, Deighan T, Li H, Terns RM, Terns MP. 2016. Bipartite recognition of target RNAs activates DNA cleavage by the Type III-B CRISPR-Cas system. *Genes & development* **30**: 447-459.

Estrella MA, Kuo FT, Bailey S. 2016. RNA-activated DNA cleavage by the Type III-B CRISPR-Cas effector complex. *Genes & development* **30**: 460-470.

Esvelt KM, Mali P, Braff JL, Moosburner M, Yaung SJ, Church GM. 2013. Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nature methods* **10**: 1116-1121.

Farrell S, Simkovich N, Wu Y, Barberis A, Ptashne M. 1996. Gene activation by recruitment of the RNA polymerase II holoenzyme. *Genes & development* **10**: 2359-2367.

Garneau JE, Dupuis ME, Villion M, Romero DA, Barrangou R, Boyaval P, Fremaux C, Horvath P, Magadan AH, Moineau S. 2010. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**: 67-71.

Gasiunas G, Barrangou R, Horvath P, Siksnys V. 2012. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences of the United States of America* **109**: E2579-2586.

Gaudreau L, Schmid A, Blaschke D, Ptashne M, Horz W. 1997. RNA polymerase II holoenzyme recruitment is sufficient to remodel chromatin at the yeast PHO5 promoter. *Cell* **89**: 55-62.

Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, 3rd, Smith HO. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature methods* **6**: 343-345.

Gill SR, Fouts DE, Archer GL, Mongodin EF, Deboy RT, Ravel J, Paulsen IT, Kolonay JF, Brinkac L, Beanan M et al. 2005. Insights on evolution of virulence and resistance from the complete genome analysis of an early methicillin-resistant Staphylococcus aureus strain and a biofilm-producing methicillin-resistant Staphylococcus epidermidis strain. *Journal of bacteriology* **187**: 2426-2438.

Goldberg GW, Jiang W, Bikard D, Marraffini LA. 2014. Conditional tolerance of temperate phages via transcription-dependent CRISPR-Cas targeting. *Nature* **514**: 633-637.

Gong B, Shin M, Sun J, Jung CH, Bolt EL, van der Oost J, Kim JS. 2014. Molecular insights into DNA interference by CRISPR-associated nuclease-helicase Cas3. *Proceedings of the National Academy of Sciences of the United States of America* **111**: 16359-16364.

Grissa I, Vergnaud G, Pourcel C. 2007. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC bioinformatics* **8**: 172.

Gust B, Chandra G, Jakimowicz D, Yuqing T, Bruton CJ, Chater KF. 2004. Lambda red-mediated genetic manipulation of antibiotic-producing Streptomyces. *Advances in applied microbiology* **54**: 107-128.

Hale C, Kleppe K, Terns RM, Terns MP. 2008. Prokaryotic silencing (psi)RNAs in Pyrococcus furiosus. *Rna* **14**: 2572-2579.

Hale CR, Zhao P, Olson S, Duff MO, Graveley BR, Wells L, Terns RM, Terns MP. 2009. RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* **139**: 945-956.

Hamilton CM, Aldea M, Washburn BK, Babitzke P, Kushner SR. 1989. New method for generating deletions and gene replacements in Escherichia coli. *Journal of bacteriology* **171**: 4617-4622.

Hannon GJ, Rossi JJ. 2004. Unlocking the potential of the human genome with RNA interference. *Nature* **431**: 371-378.

Hatfull GF. 2008. Bacteriophage genomics. *Current opinion in microbiology* **11**: 447-453.

Hatoum-Aslan A, Maniv I, Marraffini LA. 2011. Mature clustered, regularly interspaced, short palindromic repeats RNA (crRNA) length is measured by a ruler mechanism anchored at the precursor processing site. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 21218-21222.

Hatoum-Aslan A, Maniv I, Samai P, Marraffini LA. 2014. Genetic characterization of antiplasmid immunity through a type III-A CRISPR-Cas system. *Journal of bacteriology* **196**: 310-317.

Hatoum-Aslan A, Samai P, Maniv I, Jiang W, Marraffini LA. 2013. A Ruler Protein in a Complex for Antiviral Defense Determines the Length of Small Interfering CRISPR RNAs. *The Journal of biological chemistry* **288**: 27888-27897.

Havarstein LS, Coomaraswamy G, Morrison DA. 1995. An unmodified heptadecapeptide pheromone induces competence for genetic transformation in Streptococcus pneumoniae. *Proceedings of the National Academy of Sciences of the United States of America* **92**: 11140-11144.

Heler R, Marraffini LA, Bikard D. 2014. Adapting to new threats: the generation of memory by CRISPR-Cas immune systems. *Molecular microbiology* **93**: 1-9.

Heler R, Samai P, Modell JW, Weiner C, Goldberg GW, Bikard D, Marraffini LA. 2015. Cas9 specifies functional viral targets during CRISPR-Cas adaptation. *Nature* **519**: 199-202.

Helmann JD, Chamberlin MJ. 1988. Structure and function of bacterial sigma factors. *Annual review of biochemistry* **57**: 839-872.

Herskowitz I, Hagen D. 1980. The lysis-lysogeny decision of phage lambda: explicit programming and responsiveness. *Annual review of genetics* **14**: 399-445.

Hilton IB, D'Ippolito AM, Vockley CM, Thakore PI, Crawford GE, Reddy TE, Gersbach CA. 2015. Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nature biotechnology* **33**: 510-517.

Horinouchi S, Weisblum B. 1982. Nucleotide sequence and functional map of pC194, a plasmid that specifies inducible chloramphenicol resistance. *Journal of bacteriology* **150**: 815-825.

Horton RM, Ho SN, Pullen JK, Hunt HD, Cai Z, Pease LR. 1993. Gene splicing by overlap extension. *Methods in enzymology* **217**: 270-279.

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

Hosaka T, Tamehiro N, Chumpolkulwong N, Hori-Takemoto C, Shirouzu M, Yokoyama S, Ochi K. 2004. The novel mutation K87E in ribosomal protein S12 enhances protein synthesis activity during the late growth phase in Escherichia coli. *Molecular genetics and genomics : MGG* **271**: 317-324.

Hoskins J, Alborn WE, Jr., Arnold J, Blaszczak LC, Burgett S, DeHoff BS, Estrem ST, Fritz L, Fu DJ, Fuller W et al. 2001. Genome of the bacterium Streptococcus pneumoniae strain R6. *Journal of bacteriology* **183**: 5709-5717.

Hsu PD, Lander ES, Zhang F. 2014. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**: 1262-1278.

Huang H, Zheng G, Jiang W, Hu H, Lu Y. 2015. One-step high-efficiency CRISPR/Cas9-mediated genome editing in Streptomyces. *Acta biochimica et biophysica Sinica* **47**: 231-243.

Huo Y, Nam KH, Ding F, Lee H, Wu L, Xiao Y, Farchione MD, Jr., Zhou S, Rajashankar K, Kurinov I et al. 2014. Structures of CRISPR Cas3 offer mechanistic insights into Cascade-activated DNA unwinding and degradation. *Nature structural & molecular biology* **21**: 771-777.

Jackson RN, Golden SM, van Erp PB, Carter J, Westra ER, Brouns SJ, van der Oost J, Terwilliger TC, Read RJ, Wiedenheft B. 2014. Structural biology. Crystal structure of the CRISPR RNA-guided surveillance complex from Escherichia coli. *Science* **345**: 1473-1479.

Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA. 2013a. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature biotechnology* **31**: 233-239.

Jiang W, Maniv I, Arain F, Wang Y, Levin BR, Marraffini LA. 2013b. Dealing with the evolutionary downside of CRISPR immunity: bacteria and beneficial plasmids. *PLoS genetics* **9**: e1003844.

Jiang W, Marraffini LA. 2015. CRISPR-Cas: New Tools for Genetic Manipulations from Bacterial Immunity Systems. *Annual review of microbiology* **69**: 209-228.

Jiang WY, Samai P, Marraffini LA. 2016. Degradation of Phage Transcripts by CRISPR-Associated RNases Enables Type III CRISPR-Cas Immunity. *Cell* **164**: 710-721.

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

Jinek M, Jiang F, Taylor DW, Sternberg SH, Kaya E, Ma E, Anders C, Hauer M, Zhou K, Lin S et al. 2014. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **343**: 1247997.

Joung JK, Sander JD. 2013. TALENs: a widely applicable technology for targeted genome editing. *Nature reviews Molecular cell biology* **14**: 49-55.

Karakousis G, Ye N, Li Z, Chiu SK, Reddy G, Radding CM. 1998. The beta protein of phage lambda binds preferentially to an intermediate in DNA renaturation. *Journal of molecular biology* **276**: 721-731.

Karam JD, Drake, J.W. 1994. *Molecular Biology of Bacteriophage T4*. American Society for Microbiology.

Karlinsey JE. 2007. lambda-Red genetic engineering in Salmonella enterica serovar Typhimurium. *Methods in enzymology* **421**: 199-209.

Karu AE, Sakaki Y, Echols H, Linn S. 1975. The gamma protein specified by bacteriophage gamma. Structure and inhibitory activity for the recBC enzyme of Escherichia coli. *The Journal of biological chemistry* **250**: 7377-7387.

Kehoe JW, Kay BK. 2005. Filamentous phage display in the new millennium. *Chemical reviews* **105**: 4056-4072.

Kiro R, Shitrit D, Qimron U. 2014. Efficient engineering of a bacteriophage genome using the type I-E CRISPR-Cas system. *RNA biology* **11**: 42-44.

Kleinstiver BP, Prew MS, Tsai SQ, Topkar VV, Nguyen NT, Zheng Z, Gonzales AP, Li Z, Peterson RT, Yeh JR et al. 2015. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523**: 481-485.

Kmiec E, Holloman WK. 1981. Beta protein of bacteriophage lambda promotes renaturation of DNA. *The Journal of biological chemistry* **256**: 12636-12639.

Koike-Yusa H, Li Y, Tan EP, Velasco-Herrera Mdel C, Yusa K. 2014. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nature biotechnology* **32**: 267-273.

Koonin EV. 2012. The logic of chance: the nature and origin of biological evolution. *Choice: Current Reviews for Academic Libraries* **49**: 1286-1287.

Koonin EV, Krupovic M, Yutin N. 2015. Evolution of double-stranded DNA viruses of eukaryotes: from bacteriophages to transposons to giant viruses. *Annals of the New York Academy of Sciences* **1341**: 10-24.

Krebs JE, Goldstein, E.S., Lewin, B., Kilpatrick, S.T. 2010. Antitermination may be a regulated event. *Lewin's Essential Genes*: 287-291.

Li M, Moyle H, Susskind MM. 1994. Target of the transcriptional activation function of phage lambda cI protein. *Science* **263**: 75-77.

Li Y, Pan S, Zhang Y, Ren M, Feng M, Peng N, Chen L, Liang YX, She Q. 2016. Harnessing Type I and Type III CRISPR-Cas systems for genome editing. *Nucleic acids research* **44**: e34.

Little JW. 1967. An exonuclease induced by bacteriophage lambda. II. Nature of the enzymatic reaction. *The Journal of biological chemistry* **242**: 679-686.

Makarova KS, Grishin NV, Shabalina SA, Wolf YI, Koonin EV. 2006. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biology direct* **1**: 7.

Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, Moineau S, Mojica FJ, Wolf YI, Yakunin AF et al. 2011. Evolution and classification of the CRISPR-Cas systems. *Nature reviews Microbiology* **9**: 467-477.

Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM. 2013. RNA-guided human genome engineering via Cas9. *Science* **339**: 823-826.

Man S, Cheng R, Miao C, Gong Q, Gu Y, Lu X, Han F, Yu W. 2011. Artificial trans-encoded small non-coding RNAs specifically silence the selected gene expression in bacteria. *Nucleic acids research* **39**: e50.

Manica A, Zebec Z, Steinkellner J, Schleper C. 2013. Unexpectedly broad target recognition of the CRISPR-mediated virus defence system in the archaeon Sulfolobus solfataricus. *Nucleic acids research* **41**: 10509-10517.

Manica A, Zebec Z, Teichmann D, Schleper C. 2011. In vivo activity of CRISPR-mediated virus defence in a hyperthermophilic archaeon. *Molecular microbiology* **80**: 481-491.

Manting EH, Driessen AJ. 2000. Escherichia coli translocase: the unravelling of a molecular machine. *Molecular microbiology* **37**: 226-238.

Marraffini LA. 2010. Impact of CRIPSR immunity on the emergence of bacterial pathogens. *Future microbiology* **5**: 693-695.

Marraffini LA, Dedent AC, Schneewind O. 2006. Sortases and the art of anchoring proteins to the envelopes of gram-positive bacteria. *Microbiology and molecular biology reviews : MMBR* **70**: 192-221.

Marraffini LA, Sontheimer EJ. 2008. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322**: 1843-1845.

Martel B, Moineau S. 2014. CRISPR-Cas: an efficient tool for genome engineering of virulent bacteriophages. *Nucleic acids research* **42**: 9504-9513.

Matsuura S, Komatsu J, Hirano K, Yasuda H, Takashima K, Katsura S, Mizuno A. 2001. Real-time observation of a single DNA digestion by lambda exonuclease under a fluorescence microscope field. *Nucleic acids research* **29**: E79.

Mendenhall EM, Williamson KE, Reyon D, Zou JY, Ram O, Joung JK, Bernstein BE. 2013. Locus-specific editing of histone modifications at endogenous enhancers. *Nature biotechnology* **31**: 1133-1136.

Miller ES, Kutter E, Mosig G, Arisaka F, Kunisawa T, Ruger W. 2003. Bacteriophage T4 genome. *Microbiology and molecular biology reviews : MMBR* **67**: 86-156, table of contents.

Mimee M, Tucker AC, Voigt CA, Lu TK. 2015. Programming a Human Commensal Bacterium, , to Sense and Respond to Stimuli in the Murine Gut Microbiota. *Cell systems* **1**: 62-71.

Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Almendros C. 2009. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**: 733-740.

Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Soria E. 2005. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of molecular evolution* **60**: 174-182.

Monk IR, Shah IM, Xu M, Tan MW, Foster TJ. 2012. Transforming the untransformable: application of direct transformation to manipulate

genetically Staphylococcus aureus and Staphylococcus epidermidis. *mBio* **3**.

Morton TM, Johnston JL, Patterson J, Archer GL. 1995. Characterization of a conjugative staphylococcal mupirocin resistance plasmid. *Antimicrobial agents and chemotherapy* **39**: 1272-1280.

Mosberg JA, Lajoie MJ, Church GM. 2010. Lambda red recombineering in Escherichia coli occurs through a fully single-stranded intermediate. *Genetics* **186**: 791-799.

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

Muniyappa K, Radding CM. 1986. The homologous recombination system of phage lambda. Pairing activities of beta protein. *The Journal of biological chemistry* **261**: 7472-7478.

Murphy KC. 1991. Lambda Gam protein inhibits the helicase and chi-stimulated recombination activities of Escherichia coli RecBCD enzyme. *Journal of bacteriology* **173**: 5808-5821.

-. 1998. Use of bacteriophage lambda recombination functions to promote gene replacement in Escherichia coli. *Journal of bacteriology* **180**: 2063-2071.

Nair D, Memmi G, Hernandez D, Bard J, Beaume M, Gill S, Francois P, Cheung AL. 2011. Whole-genome sequencing of Staphylococcus aureus strain RN4220, a key laboratory strain used in virulence research, identifies mutations that affect not only virulence factors but also the fitness of the strain. *Journal of bacteriology* **193**: 2332-2335.

Nishimasu H, Ran FA, Hsu PD, Konermann S, Shehata SI, Dohmae N, Ishitani R, Zhang F, Nureki O. 2014. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**: 935-949.

Nobrega FL, Costa AR, Kluskens LD, Azeredo J. 2015. Revisiting phage therapy: new applications for old resources. *Trends in microbiology* **23**: 185-191.

Oh JH, van Pijkeren JP. 2014. CRISPR-Cas9-assisted recombineering in Lactobacillus reuteri. *Nucleic acids research* **42**: e131.

Oppenheim AB, Kobiler O, Stavans J, Court DL, Adhya S. 2005. Switches in bacteriophage lambda development. *Annual review of genetics* **39**: 409-429.

Paez-Espino D, Morovic W, Sun CL, Thomas BC, Ueda K, Stahl B, Barrangou R, Banfield JF. 2013. Strong bias in the bacterial CRISPR elements that confer immunity to phage. *Nature communications* **4**: 1430.

Peng W, Feng M, Feng X, Liang YX, She Q. 2015. An archaeal CRISPR type III-B system exhibiting distinctive RNA targeting features and mediating dual RNA and DNA interference. *Nucleic acids research* **43**: 406-417.

Pines G, Freed EF, Winkler JD, Gill RT. 2015. Bacterial Recombineering: Genome Engineering via Phage-Based Homologous Recombination. *ACS synthetic biology* **4**: 1176-1185.

Pitcher RS, Brissett NC, Doherty AJ. 2007. Nonhomologous end-joining in bacteria: a microbial perspective. *Annual review of microbiology* **61**: 259-282.

Pourcel C, Salvignol G, Vergnaud G. 2005. CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* **151**: 653-663.

Ptashne M, Gann A. 1997. Transcriptional activation by recruitment. *Nature* **386**: 569-577.

Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA. 2013. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**: 1173-1183.

Renzette N. 2011. Generation of Transformation Competent E. coli. *Current Protocols in Microbiology* **22:3L:A.3L.1–A.3L.5.**

Reyes A, Semenkovich NP, Whiteson K, Rohwer F, Gordon JI. 2012. Going viral: next-generation sequencing applied to phage populations in the human gut. *Nature reviews Microbiology* **10**: 607-617.

Reyrat JM, Pelicic V, Gicquel B, Rappuoli R. 1998. Counterselectable markers: untapped tools for bacterial genetics and pathogenesis. *Infection and immunity* **66**: 4011-4017.

Samai P, Pyenson N, Jiang W, Goldberg GW, Hatoum-Aslan A, Marraffini LA. 2015. Co-transcriptional DNA and RNA Cleavage during Type III CRISPR-Cas Immunity. *Cell* **161**: 1164-1174.

Sambrook J, Fritsch, E.F., and Maniatis, T. . 1989. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press.

Sapranauskas R, Gasiunas G, Fremaux C, Barrangou R, Horvath P, Siksnys V. 2011. The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. *Nucleic acids research* **39**: 9275-9282.

Sashital DG, Wiedenheft B, Doudna JA. 2012. Mechanism of foreign DNA selection in a bacterial adaptive immune system. *Molecular cell* **46**: 606-615.

Semenova E, Jore MM, Datsenko KA, Semenova A, Westra ER, Wanner B, van der Oost J, Brouns SJ, Severinov K. 2011. Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 10098-10103.

Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelsen TS, Heckl D, Ebert BL, Root DE, Doench JG et al. 2014. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**: 84-87.

Sharan SK, Thomason LC, Kuznetsov SG, Court DL. 2009. Recombineering: a homologous recombination-based method of genetic engineering. *Nature protocols* **4**: 206-223.

Sharma V, Yamamura A, Yokobayashi Y. 2012. Engineering artificial small RNAs for conditional gene silencing in Escherichia coli. *ACS synthetic biology* **1**: 6-13.

Smith GP. 1985. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* **228**: 1315-1317.

Smith GP, Petrenko VA. 1997. Phage Display. *Chemical reviews* **97**: 391-410.

Sokolowski RD, Graham S, White MF. 2014. Cas6 specificity and CRISPR RNA loading in a complex CRISPR-Cas system. *Nucleic acids research* **42**: 6532-6541.

Staals RH, Zhu Y, Taylor DW, Kornfeld JE, Sharma K, Barendregt A, Koehorst JJ, Vlot M, Neupane N, Varossieau K et al. 2014. RNA Targeting by the Type III-A CRISPR-Cas Csm Complex of Thermus thermophilus. *Molecular cell* **56**: 518-530.

Sternberg SH, Redding S, Jinek M, Greene EC, Doudna JA. 2014. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**: 62-67.

Storz G, Vogel J, Wassarman KM. 2011. Regulation by small RNAs in bacteria: expanding frontiers. *Molecular cell* **43**: 880-891.

Suttle CA. 2007. Marine viruses--major players in the global ecosystem. *Nature reviews Microbiology* **5**: 801-812.

Tamulaitis G, Kazlauskiene M, Manakova E, Venclovas C, Nwokeoji AO, Dickman MJ, Horvath P, Siksnys V. 2014. Programmable RNA Shredding by the Type III-A CRISPR-Cas System of Streptococcus thermophilus. *Molecular cell* **56**: 506-517.

Terns MP, Terns RM. 2011. CRISPR-based adaptive immune systems. *Current opinion in microbiology* **14**: 321-327.

Theis T, Skurray RA, Brown MH. 2007. Identification of suitable internal controls to study expression of a Staphylococcus aureus multidrug resistance system by quantitative real-time PCR. *Journal of microbiological methods* **70**: 355-362.

Tock MR, Dryden DT. 2005. The biology of restriction and anti-restriction. *Current opinion in microbiology* **8**: 466-472.

Tong Y, Charusanti P, Zhang L, Weber T, Lee SY. 2015. CRISPR-Cas9 Based Engineering of Actinomycetal Genomes. *ACS synthetic biology* **4**: 1020-1029.

Twort FW. 1915. An investigation on the nature of ultra-microscopic viruses. *Lancet* **2**: 1241-1243.

van Kessel JC, Hatfull GF. 2008. Mycobacterial recombineering. *Methods in molecular biology* **435**: 203-215.

Vogel J, Luisi BF. 2011. Hfq and its constellation of RNA. *Nature reviews Microbiology* **9**: 578-589.

Wang T, Wei JJ, Sabatini DM, Lander ES. 2014. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343**: 80-84.

Wang Y, Zhang ZT, Seo SO, Choi K, Lu T, Jin YS, Blaschek HP. 2015. Markerless chromosomal gene deletion in Clostridium beijerinckii using CRISPR/Cas9 system. *Journal of biotechnology* **200**: 1-5.

Waters LS, Storz G. 2009. Regulatory RNAs in bacteria. *Cell* **136**: 615-628.

Westra ER, van Erp PB, Kunne T, Wong SP, Staals RH, Seegers CL, Bollen S, Jore MM, Semenova E, Severinov K et al. 2012. CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Molecular cell* **46**: 595-605.

Westwater C, Schofield DA, Schmidt MG, Norris JS, Dolan JW. 2002. Development of a P1 phagemid system for the delivery of DNA into Gram-negative bacteria. *Microbiology* **148**: 943-950.

Wiedenheft B, Sternberg SH, Doudna JA. 2012. RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **482**: 331-338.

Wiedenheft B, van Duijn E, Bultema JB, Waghmare SP, Zhou K, Barendregt A, Westphal W, Heck AJ, Boekema EJ, Dickman MJ et al. 2011. RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 10092-10097.

Wilson GG. 1991. Organization of restriction-modification systems. *Nucleic acids research* **19**: 2539-2566.

Wommack KE, Colwell RR. 2000. Virioplankton: viruses in aquatic ecosystems. *Microbiology and molecular biology reviews : MMBR* **64**: 69-114.

Xu T, Li Y, Shi Z, Hemme CL, Li Y, Zhu Y, Van Nostrand JD, He Z, Zhou J. 2015. Efficient Genome Editing in Clostridium cellulolyticum via CRISPR-Cas9 Nickase. *Applied and environmental microbiology* **81**: 4423-4431.

Yu D, Ellis HM, Lee EC, Jenkins NA, Copeland NG, Court DL. 2000. An efficient recombination system for chromosome engineering in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* **97**: 5978-5983.

Zahner D, Hakenbeck R. 2000. The Streptococcus pneumoniae beta-galactosidase is a surface protein. *Journal of bacteriology* **182**: 5919-5921.

Zebec Z, Manica A, Zhang J, White MF, Schleper C. 2014. CRISPR-mediated targeted mRNA degradation in the archaeon Sulfolobus solfataricus. *Nucleic acids research* **42**: 5280-5288.

Zhang J, Rouillon C, Kerou M, Reeks J, Brugger K, Graham S, Reimann J, Cannone G, Liu H, Albers SV et al. 2012. Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity. *Molecular cell* **45**: 303-313.

Zhang Y, Buchholz F, Muyrers JP, Stewart AF. 1998. A new logic for DNA engineering using recombination in Escherichia coli. *Nature genetics* **20**: 123-128.

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

Zhou Y, Zhu S, Cai C, Yuan P, Li C, Huang Y, Wei W. 2014. High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature* **509**: 487-491.