ABSTRACT

LUO, MICHELLE LYNN. Understanding and exploiting the properties of Type I CRISPR-Cas systems. (Under the direction of Dr. Chase Beisel).

CRISPR-Cas (clustered regularly interspaced short palindromic repeats-CRISPR-associated) systems offer powerful and versatile biomolecular tools with many applications including genome editing, gene regulation, sequence-specific antimicrobials, and imaging. Because these RNA-directed immune systems are found in most prokaryotes, an opportunity exists to harness endogenous CRISPR-Cas systems as convenient tools in these organisms. Type I systems are particularly promising as this type accounts for the majority of identified CRISPR-Cas systems in both bacteria and archaea. I report that the Type I-E CRISPR-Cas system in the bacterium *Escherichia coli* can be co-opted for programmable transcriptional repression. Deletion of the signature *cas3* gene converted this immune system into a programmable gene regulator capable of reversible and multiplexable silencing of heterologous and endogenous genes.

Utilizing this platform, I explored how the length of the CRISPR RNA impacts the form and function of the Type I-E effector protein complex called Cascade (CRISPR-associated complex for antiviral defense). I discovered that the length of the Type I-E CRISPR RNA molecule is not fixed and can be extended, altering the stoichiometry of the Cascade effector complex while preserving functionality. Interestingly, altering RNA length can elicit significant improvement in transcriptional silencing efficiency, but only for particular target locations. These findings heighten our understanding of CRISPR-Cas systems as they continue to increase in popularity and may offer insight into designing tunable transcriptional regulators.

Understanding and Exploiting the Properties of Type I CRISPR-Cas Systems

by

Michelle Lynn Luo

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Chemical and Biomolecular Engineering

Raleigh, North Carolina

2016

APPROVED BY:

_____	_____
Dr. Chase Beisel Committee Chair	Dr. Rodolphe Barrangou
_____	_____
Dr. Robert Kelly	Dr. Steven Peretti

DEDICATION

To my family for their unwavering support throughout the years.

To Atanas Petkov for his boundless love that keeps me sane when life gets rocky.

BIOGRAPHY

Michelle Lynn Luo was born in Pittsburgh, Pennsylvania in 1990. Shortly after her birth, her family moved to Cincinnati, Ohio where her brother Patrick was born. The family moved again to Columbus, Ohio before finally settling down in Apex, North Carolina when Michelle was 6 years old. Growing up, Michelle always had an aptitude for math and science. Her favorite class in high school was AP Chemistry, so she decided to pursue chemical engineering in college because it seemed like a happy marriage of math and chemistry (little did she know she would end up taking a detour to the world of biology). After graduating high school, Michelle attended Princeton University in New Jersey, where she received her Bachelor of Science in Engineering degree in 2012, majoring in Chemical and Biological Engineering. While at Princeton University, Michelle conducted her senior thesis research under the supervision of Dr. James A. Link. This undergraduate research experience inspired Michelle to pursue additional research opportunities. In the fall of 2012, Michelle returned to her home state and joined the graduate program in the Department of Chemical and Biomolecular Engineering at North Carolina State University. She conducted her graduate studies with the mentorship of Dr. Chase Beisel. During the summer of 2015, Michelle interned at Ventana Medical Systems in Tucson, Arizona, which reaffirmed her desire to seek out research and development opportunities where science and technology interface with healthcare solutions. During the summer of 2016, Michelle taught the CRISPR technologies module of the Synthetic Biology Course at Cold Spring Harbor Laboratory in New York, which ignited an interest in science education and outreach. After receiving her Doctor of Philosophy degree, Michelle will be moving to Florida and marrying her college sweetheart, Atanas Petkov. She plans to pursue a career in seeking innovative healthcare solutions in biotechnology, diagnostics, medical devices, and beyond.

ACKNOWLEDGMENTS

First, I would like to thank my adviser, Chase Beisel, for his support in all my intellectual endeavors. His confidence in my ability to operate independently and his encouragement to pursue outside opportunities have made my graduate experience invaluable.

I would also like to thank all my collaborators, especially past and present members of the Beisel Lab for their technical guidance and expertise. Special thanks go to Adam Mullis, Steven Denny, Ryan Leenay, and Ken Maksimchuk for their contributions to my projects and abundant in-lab witty repartee.

I also thank my team at Ventana Medical Systems. Thank you for hiring me in the first place to do an internship. Thank you for all the wonderful advice and for giving me such an incredible opportunity to explore what industry has to offer. I also give thanks to the 2015 summer intern cohort for all our Arizonan adventures.

Furthermore, I thank all my friends that I have made during my time at NCSU for making Raleigh such a fun place to live. Cheers to all the classes we've endured together, all the meals we've eaten, all the parties we've attended, and all the good times we've enjoyed.

Finally, I would like to thank my family, both old and new. To my parents, thank you for all your support and encouragement. To my fiancé, thank you for listening to absurdly technical talks, for ensuring that I remember there is a life outside of the lab, for entertaining my obsession with all things cute and fluffy, and for being the most wonderful person that I could have ever hoped to spend my life with.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Current and future prospects for CRISPR-based tools in bacteria

Michelle L. Luo, Ryan T. Leenay, Chase L. Beisel

Department of Chemical and Biomolecular Engineering, North Carolina State University, Raleigh, NC 27695

Original publication in: *Biotechnology & Bioengineering* 113(5): 930-943, 2016

ABSTRACT

CRISPR-Cas systems have rapidly transitioned from intriguing prokaryotic defense systems to powerful and versatile biomolecular tools. This article reviews how these systems have been translated into technologies to manipulate bacterial genetics, physiology, and communities. Recent applications in bacteria have centered on multiplexed genome editing, programmable gene regulation, and sequence-specific antimicrobials, while future applications can build on advances in eukaryotes, the rich natural diversity of CRISPR-Cas systems, and the untapped potential of CRISPR-based DNA acquisition. Overall, these systems have formed the basis of an ever-expanding genetic toolbox and hold tremendous potential for our future understanding and engineering of the bacterial world.

1.1 INTRODUCTION

Almost three decades ago, Japanese researchers identified an unusual set of repetitive DNA sequences in the genome of the bacterium *Escherichia coli* (Ishino et al., 1987). These repeats were later found to be part of an expansive family of repetitive DNA sequences termed clustered regularly interspaced short palindromic repeats, or CRISPRs (Jansen et al., 2002). While the repeats attracted initial attention, the intervening "spacer" sequences turned out to be the critical elements in initially defining the function of CRISPR-Cas systems. These spacers were found to be homologous to foreign plasmid and bacteriophage sequences (Bolotin et al., 2005; Mojica et al., 2005), which hinted at a defensive function for CRISPR. The major breakthrough came in 2007, with the report that bacteriophage-resistant strains had acquired spacer sequences that matched the bacteriophage genome (Barrangou et al., 2007). Critically, the acquired spacer and the flanking CRISPR-associated (Cas) genes were essential to confer immunity to the bacteriophage. This seminal work quickly led to our current understanding of these diverse adaptive defense systems in bacteria and archaea now known as CRISPR-Cas systems.

CRISPR-Cas systems consist of two general components: CRISPR RNAs (crRNAs) and Cas proteins. The crRNAs base pair with complementary DNA or RNA sequences associated with an invader, and the Cas proteins clear the recognized genetic material. Because base pairing is straightforward to predict and to design, the biotechnology community was interested in the capacity of these systems to bind and cleave user-defined sequences. The catalyst for the CRISPR-Cas revolution, however, came with the demonstration that a single protein, Cas9, could be harnessed for site-specific DNA binding and cleavage (Gasiunas et al., 2012; Jinek et al., 2012). In the few short years since this demonstration, CRISPR-Cas systems have emerged as powerful and versatile tools in applications ranging from genome editing to molecular imaging. While most of these advances have been reported in eukaryotes, CRISPR-Cas systems also offer promising tools for understanding and engineering bacteria. This article discusses recent applications of CRISPR-Cas systems in bacteria in the realms of genome

editing, gene regulation, and antimicrobials. The review then forecasts upcoming opportunities and challenges associated with further exploiting these versatile prokaryotic immune systems.

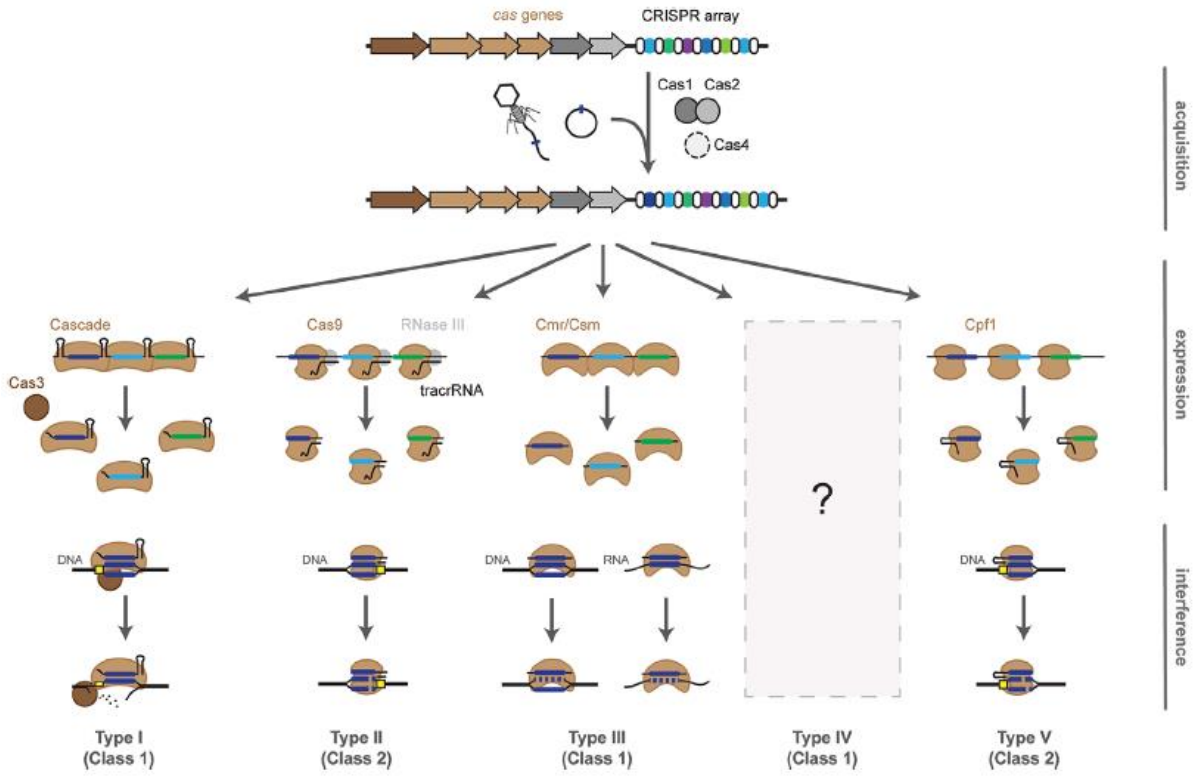

Figure 1.1 Overview of adaptive immunity by CRISPR-Cas systems. Immunity is conferred through three steps: acquisition, expression, and interference. Acquisition: a small piece of the invader DNA is integrated as a new spacer within the CRISPR array. Expression: the CRISPR array is transcribed and undergoes processing by the Cas proteins and accessory factors to form the CRISPR RNA (crRNA). Interference: the spacer portion of the crRNA serves as a recognition element for the Cas proteins to target invading DNA (Type I, II, III, V) or RNA (Type III). Type I, II, and V systems require a protospacer-adjacent motif (PAM, yellow box) for target recognition. The current understanding of Type IV systems is limited to bioinformatics analyses.

1.2 A PRIMER ON CRISPR-CAS SYSTEMS

CRISPR-Cas systems naturally protect bacteria and archaea from foreign genetic elements such as plasmids or bacteriophages. Immunity proceeds in three stages: acquisition, expression, and interference (Figure 1.1). For acquisition, a spacer generated from a short sequence of invading DNA is incorporated at the leading edge of the CRISPR locus. Next, for

expression, the array of alternating repeats and spacers is transcribed and subsequently processed by the Cas proteins and accessory factors into individual crRNAs. Finally, for interference, a ribonucleoprotein complex of the Cas protein(s) and an individual crRNA binds and cleaves nucleic acids that are complementary to the spacer portion of the crRNA. More details on the mechanisms of CRISPR-based immunity can be found in other recent reviews (Barrangou and Marraffini, 2014; Bondy-Denomy and Davidson, 2014; Van der Oost et al., 2014).

CRISPR-Cas systems are remarkably widespread and diverse. To date, the CRISPRdb online database (Grissa et al., 2007) has identified 1302 bacterial and archaeal strains with putative CRISPR arrays out of 2762 genomes analyzed. Each of these arrays is associated with differing families of cas genes that necessitated a standard system for their classification and nomenclature. The latest classification divides CRISPR-Cas systems into two classes according to the configuration of their effector modules (Makarova et al., 2015). Class 1 systems are defined by multisubunit effector complexes while Class 2 systems utilize a single effector protein. Within these two classes, CRISPR-Cas systems can be further divided into five types with sixteen total subtypes, defined based on the distinct proteins that facilitate CRISPR-Cas activities (Chylinski et al., 2014; Jiang and Doudna, 2015; Makarova et al., 2011; Makarova et al., 2015). Class 1 CRISPR-Cas systems include Type I, Type III, and putative Type IV systems; Class 2 systems include Type II systems and putative new Type V systems. These types (and subtypes) vary in the number of Cas proteins, the mechanism of crRNA processing and targeting, and whether the target is DNA or RNA (Table 1.1). These attributes offer distinct capabilities, some of which have been co-opted for biotechnology.

Table 1.1 The five standard types of CRISPR-Cas systems. The naming conventions are based on the most recent nomenclature for CRISPR-Cas systems (Makarova et al., 2015). The types are grouped into two classes based on the use of multiple proteins (Class 1) or a single protein (Class 2) for interference. Cascade: CRISPR associated complex for adaptive defense. PAM: protospacer-adjacent motif.

Class	**Type**	**Subtypes**	**Signature gene**	**# non-acquisition *cas* genes**	**Processing factors**	**Targeting**	**Self/non-self recognition**
1	I	A,B,C,D,E,F,U	*cas3*	4-8	Cascade	Nicks, degrades DNA	PAM 5' of matching target
	III	A,B,C,D	*cas10*	6-8	Cas6	Cleaves DNA/RNA	Base pairing with 5' handle
	IV (putative)	n/a	*csf1*	3-4	Unknown	Unknown	Unknown
2	II	A,B,C	*cas9*	1	tracrRNA, RNaseIII	Cleaves DNA	PAM 3' of matching target
	V (putative)	n/a	*cpf1*	1	Cpf1	Cleaves DNA	PAM 5' of matching target

Type I CRISPR-Cas systems, the most abundant type in both bacteria and archaea, are defined by the presence of the signature Cas3 protein (Jackson et al., 2014b; Sinkunas et al., 2011). Cas3 works together with a large multimeric protein complex termed the CRISPR-associated complex for antiviral defense (Cascade) whose composition is unique to each subtype (I-A through I-F, I-U) (Brendel et al., 2014; Jackson et al., 2014a; Jore et al., 2011; Makarova et al., 2015; Nam et al., 2012; Wiedenheft et al., 2011; Zhao et al., 2014). Cascade is responsible for both processing the crRNAs as well as locating and binding the target sequence (Brouns et al., 2008; Westra et al., 2012). Once Cascade binds to the recognized DNA site, Cas3 is prompted to unwind, cleave, and degrade one strand of the DNA (Sinkunas et al., 2011).

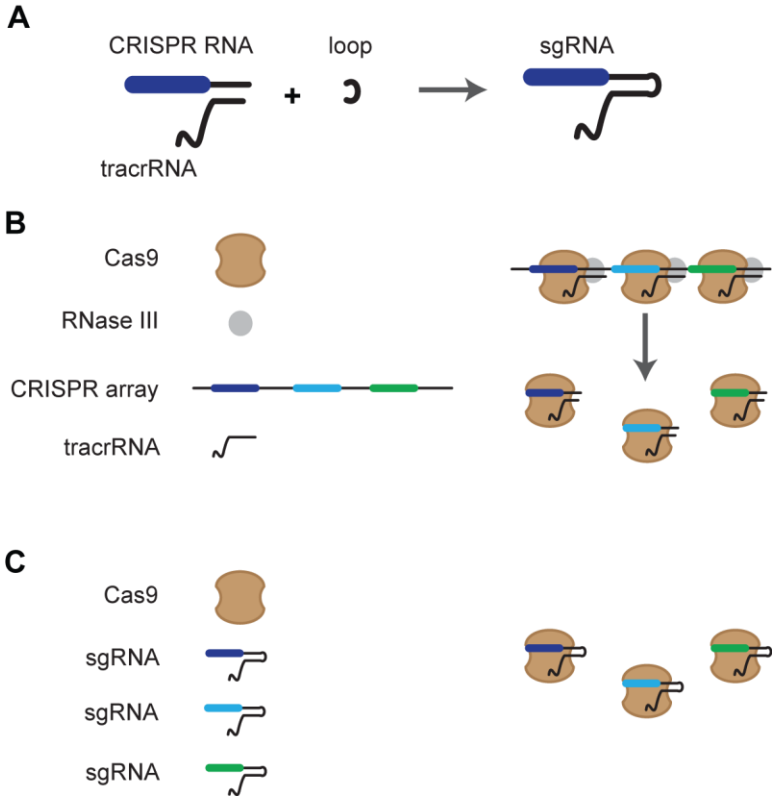

Figure 1.2 Design of the sgRNA (A) The natural crRNA and tracrRNA are connected by a loop to form a single-guide RNA (sgRNA). (B) Processing of the natural Type II CRISPR array requires two additional factors: RNase III and tracrRNA. (C) Use of sgRNAs bypasses the requirement for RNase III and tracrRNA.

Type II systems represent the least abundant systems, including no known examples in archaea. These systems have three subtypes (II-A, II-B, II-C) that are united by the signature Cas9 protein and are differentiated by the Cas proteins involved in acquisition (Chylinski et al., 2014; Makarova et al., 2015). In general, Type II systems are very compact, as they require only Cas9 for interference. Cas9 generates double-stranded DNA breaks via its HNH-nuclease and RuvC-like nuclease domains, which cleave the target strand and non-target strand, respectively (Gasiunas et al., 2012; Jiang et al., 2015a; Jinek et al., 2012; Nishimasu et al., 2014). Along with Cas9, Type II systems require a trans-activating crRNA (tracrRNA) and RNase III as part of crRNA processing (Chylinski et al., 2013; Deltcheva et al., 2011). To simplify this system for practical use, Jinek and coworkers fused the processed version of the tracrRNA and the crRNA from *Streptococcus pyogenes* into a chimeric RNA, termed a single-guide RNA (sgRNA) (Jinek et al., 2012) (Figure 1.2). This sgRNA design eliminated the need for RNase III processing and allowed the expression of a single RNA molecule for DNA targeting. The downside was that the inherent capability of the CRISPR arrays for multiplexing—targeting multiple sequences with a single transcript—was lost. However, various groups have engineered platforms to encode multiple sgRNAs from a single transcript (Nissim et al., 2014; Tsai et al., 2014; Xie et al., 2015).

Type III systems are found in both bacteria and archaea and are typified by the signature *cas10* gene (Makarova et al., 2011; Makarova et al., 2015). These systems drew initial interest because of their natural ability to target RNA (Hale et al., 2009; Staals et al., 2014; Tamulaitis et al., 2014). However, other early accounts indicated that Type III systems instead target DNA (Hatoum-Aslan et al., 2014; Marraffini and Sontheimer, 2008). While these systems were originally thought to target either DNA or RNA, there is emerging evidence that some systems can simultaneously target both nucleic acids. For instance, the Type III-B system in the archaeon *Sulfolobus islandicus* was shown to be capable of targeting DNA and RNA (Peng et al., 2013; Peng et al., 2014) whereas the Type III-A system in the pathogenic bacterium *Staphylococcus epidermidis* cleaves RNA and transcriptionally active DNA via independent active sites within the protein effector complex (Samai et al. 2015). In addition to the III-A and

III-B subtypes discussed here, there are two additional type III subtypes—III-C and III-D—that are less understood (Makarova et al., 2015). Interestingly, Type III systems are phylogenetically related to Type I systems and even share structural similarities between the Csm (III-A) or Cmr (III-B) complex and the Cascade complex from Type I systems (Osawa et al., 2015; Rouillon et al., 2013; Spilman et al., 2013; Staals et al., 2013; Taylor et al., 2015).

Lastly, two new types have been proposed: a Class 1 Type IV system and a Class 2 Type V system (Makarova et al., 2015). The putative Type IV system encodes a predicted minimal multi-subunit effector complex with a unique large subunit, Csf1, which serves as the signature protein for this system. The suggested Type V system utilizes Cpf1 as a single protein for interference. To date, only one Type V system has been characterized. Zetsche et al. (2015) demonstrated that the Type V system of *Francisella novicida* U112 does not require a tracrRNA for crRNA maturation and Cpf1-crRNA complexes are sufficient to cleave DNA target molecules. Furthermore, Cpf1 introduces a staggered double-stranded DNA break with overhangs unlike the blunt-ended break of Type II Cas9.

One of the original conundrums of CRISPR-Cas systems was how crRNAs could differentiate between their own genomic spacer sequence and the identical target sequence present in the invader. The answer was not found within the target sequence (termed the protospacer), but immediately flanking it. For Type I, Type II, and Type V systems, this flanking feature is called a protospacer adjacent motif (PAM) (Mojica et al., 2009; Szczelkun et al., 2014; Zetsche et al., 2015) and is typically 2 – 5 base pairs. For Type I systems, the PAM is located on the 5’ end of the protospacer (herein defined as the strand matching the spacer) and is recognized by Cascade (Sashital et al., 2012; Westra et al., 2013). For Type II systems, the PAM is located on the 3’ end of the protospacer and is recognized by Cas9 (Deveau et al., 2008; Gasiunas et al., 2012; Jinek et al., 2012). For example, the widely used *S. pyogenes* Cas9 recognizes an NGG PAM sequence (where N is any nucleotide) located on the 3’ end of the protospacer. The Type V systems contain a T-rich PAM on the 5’ end of the protospacer (Zetsche et al., 2015). In contrast, Type III systems do not rely on a PAM. Instead, these systems evaluate base pairing between the target and the 5’ handle of the crRNA (Marraffini

and Sontheimer, 2010). Extensive base pairing between the handle and the protospacer results in no targeting while limited base pairing results in targeting. These requirements result in two simple rules when designing crRNAs: (1) identify a PAM (Types I, II, V) or a sequence with poor base-pairing potential to the 5' portion of the crRNA repeat (Type III) and (2) use the flanking 20 – 30 nucleotides as the spacer portion of the crRNA. These remarkably straightforward rules have helped drive the implementation of CRISPR-Cas systems in a wide range of applications.

1.3 CURRENT APPLICATIONS

A number of new applications have been developed in bacteria that take advantage of the programmability and sequence specificity of CRISPR-Cas systems. Cas9 has dominated these applications based on its compactness, although the use of the Type I systems and Type III systems is beginning to gain traction.

1.3.1 Bacterial genome editing

Genome editing has been one of the most visible and celebrated applications of CRISPR to date, with an overwhelming focus on eukaryotes. In 2013, genome editing was demonstrated in human cells (Cho et al., 2013; Cong et al., 2013; Jinek et al., 2013; Mali et al., 2013b) and was quickly expanded to eukaryotic organisms ranging from fungi to monkeys (Bassett et al., 2013; DiCarlo et al., 2013; Dickinson et al., 2013; Friedland et al., 2013; Hwang et al., 2013; Li et al., 2013; Nekrasov et al., 2013; Niu et al., 2014). In these eukaryotic cells and organisms, Cas9 was used to introduce a double-stranded break into a defined location in the genome, followed by repair through the endogenous non-homologous end joining (NHEJ) or homology-directed repair pathways.

Although Cas9-based genome editing in bacteria was also first reported in 2013 (Jiang et al., 2013a), only a handful of publications have ensued. The disparity between these few publications and the volumes published for eukaryotes can be explained by the relatively poor capacity of bacteria to repair double-stranded breaks by CRISPR-Cas systems. Unlike

eukaryotes, bacteria generally cannot repair breaks caused by CRISPR. The lethality of genome targeting is supported by the natural co-occurrence of self-targeting CRISPRs and degenerate *cas* genes (Stern et al., 2010); a putatively large evolutionary change following the natural acquisition of a self-targeting spacer (Aklujkar and Lovley, 2010); the cytotoxicity of expressing a genome-targeting crRNA (Edgar and Qimron, 2010; Vercoe et al., 2013); and the subsequent appearance of disruptive mutations in the *cas* genes, self-targeting crRNA, or the target location to escape self-targeting cell death (Gomaa et al., 2014; Jiang et al., 2013b).

A simple explanation for the lethality of bacterial genome targeting is that, given the defensive function of CRISPR-Cas systems, repairing targeting events would be wholly counterproductive. If bacteriophage or plasmid DNA was attacked by CRISPR and the break was repaired, then the invader would persist. It is worth noting that bacteria often possess multiple pathways for repairing DNA damage (Selle and Barrangou, 2015). We speculate that CRISPR-Cas systems evolved to block many of these repair mechanisms in bacteria, whether by Cas9 staying tightly bound to the DNA end following cleavage (Gasiunas et al., 2012) or by Cas3 degrading a strand of the target DNA through Cas3's exonuclease activity (Westra et al., 2012). See Selle and Barrangou (2015) for a focused review describing mechanisms of genome repair in bacteria and how it relates to genome editing with CRISPR.

These obstacles were originally addressed in the first demonstration of genome editing in bacteria (Jiang et al., 2013a). In this work, Cas9 was used in combination with a CRISPR array and a tracrRNA to generate defined point mutations in *Streptococcus pneumoniae* and in *Escherichia coli*. In both cases, a template oligonucleotide designed for homologous recombination was introduced simultaneously with the CRISPR array. The oligonucleotide mutated the PAM of the target site, preventing Cas9 from recognizing the sequence. As a result, CRISPR-Cas9 served a cleanup role by eliminating cells that did not undergo recombination (Figure 1.3A). By either relying on high recombination rates in *S. pneumoniae* or introducing the λ-red recombination system *E. coli*, the authors reported remarkably high rates of recombination ranging between 65% and 100% of all screened colonies. The arrays could also be multiplexed, which allowed editing at multiple locations at one time.

Following the initial demonstration of genome editing in *S. pneumoniae* and in *E. coli*, Cas9-based genome editing has been reported in ranging bacteria of industrial relevance. These include *Lactobacillus reuteri*, a lactic acid bacterium with probiotic properties (Oh and van Pijkeren, 2014); *Clostridium beijerinckii*, a species widely used as a metabolic host for acetone, butanol, and ethanol production (Wang et al., 2015); multiple *Streptomyces* species, which synthesize natural products with antimicrobial activity (Cobb et al., 2015; Huang et al., 2015; Tong et al., 2015); and additional demonstrations in *E. coli* (Jiang et al., 2015b; Li et al., 2015; Pyne et al., 2015). These examples principally relied on the chimeric sgRNA design (Cobb et al., 2015; Jiang et al., 2015b; Li et al., 2015; Wang et al., 2015), although the dual-expression of tracrRNAs and crRNAs has also been implemented for genome editing (Jiang et al., 2013a; Oh and van Pijkeren, 2014; Pyne et al., 2015). The DNA templates were supplied as single-stranded oligonucleotides or as double-stranded DNA within the sgRNA plasmid. Between these two options, encoding the repair template within the sgRNA plasmid generally yielded greater editing efficiencies.

Many of these examples confirmed the need for high rates of transformation and recombination: transformation determines the number of cells subject to the DNA template and CRISPR-Cas9, whereas recombination determines the number of cells that will survive attack by Cas9. For instance, follow-up demonstrations in *E. coli* were limited to only three simultaneous editing events, even following overexpression of the λ-red recombination machinery (Jiang et al., 2015b; Li et al., 2015). Without high rates of both transformation and recombination, the background mutational rate of the crRNA or Cas9 (typically 10-5 – 10-4) (Gomaa et al., 2014; Jiang et al., 2013b) will overwhelmingly account for surviving transformants. The transformation barrier may be alleviated through inducible control of the crRNA or the Cas9, although tightly controlled inducible systems are not widely available. Furthermore, more universal recombination systems may be needed to drive high rates of homologous recombination in otherwise disparate organisms.

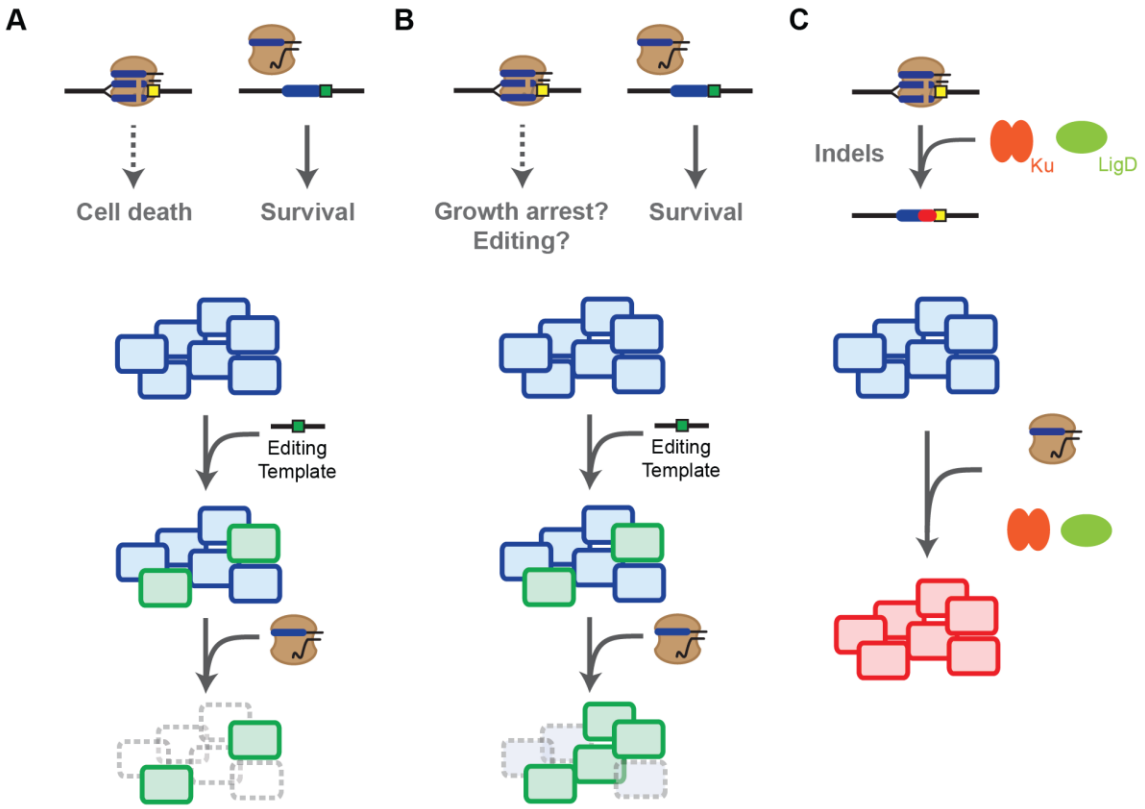

Figure 1.3 Genome editing with CRISPR-Cas9 in bacteria. (A) DNA cleavage by Cas9 is generally lethal, leading to clearance of cells that did not undergo recombination. (B) Employing a nicking Cas9 appears to either temper the lethality of genome targeting or drive genome editing. (C) Utilizing the bacterial non-homologous end-joining pathway composed of Ku and LigD can rescue the lethality of Cas9-based genome targeting and drive indel formation.

One potential strategy to improve the editing efficiency is using a Cas9 that harbors a point mutation in either the RuvC or HNH nuclease domain (D10A or H840A for the S. pyogenes Cas9) (Figure 1.3B). Because each nuclease domain cleaves either strand of the target DNA, the mutated Cas9 nicks the target without introducing a double-stranded break (Gasiunas et al., 2012; Jinek et al., 2012). Xu et al. (2015) hypothesized that a nicking Cas9 could improve editing in bacteria by reducing the lethality of DNA targeting. Supplying *Clostridium cellulolyticum* with a nicking Cas9, an sgRNA, and a template oligonucleotide resulted in small insertions or deletions, with editing efficiencies up to 95% for 200-nucleotide homology arms. In contrast, the regular Cas9 did not yield any colonies. While DNA nicking

did promote greater editing frequencies than double-stranded breaks, the precise mechanism is not fully understood. On one hand, Cas9 and its nicking counterpart could similarly clear unedited members of the population, but the tempered lethality of nicks could allow more time to achieve recombination. On the other hand, the nicking Cas9 could be driving homologous recombination by accelerating strand invasion, similar to nickase activity in eukaryotes (Cong et al., 2013; Mali et al., 2013b). Once the mechanism of editing is fully elucidated, this nickase strategy could be extended to other bacterial organisms that are poorly equipped to repair double-stranded breaks introduced by Cas9.

Another intriguing strategy is recapitulating the NHEJ pathway in bacteria to repair double-stranded breaks by CRISPR-Cas9 (Figure 1.3C). Rather than attempt to import eukaryotic NHEJ pathways and their numerous components, NHEJ pathways found in some bacteria offered simpler opportunities. These bacterial pathways only require two proteins, Ku and LigD, and are found in a small but diverse fraction of bacteria (Aravind and Koonin, 2001; Bowater and Doherty, 2006). To date, only one recent report has coupled NHEJ and Cas9-based genome editing in bacteria (Tong et al., 2015). Here, genome editing was performed in the actinomycetes species *Streptomyces coelicolor*. The *ligD* gene was expressed from a related *Streptomyces* species to reconstitute the incomplete NHEJ pathway in this species. Co-expressing LigD along with Cas9 and a designed sgRNA led to high efficiencies of NHEJ-generated indels—insertions or deletions—within two genes from the biosynthetic pathway of the blue actinorhodin antibiotic. It will be interesting to see if a reconstituted NHEJ pathway along with a homologous template will increase recombination rates or if indel formation will dominate over homology directed repair. If the NHEJ pathway can be imported into other bacteria to alleviate the lethality of genome targeting and boost the overall efficiency of genome editing, Ku and LigD could become synonymous with CRISPR-Cas9 for genome editing in bacteria.

1.3.2 Bacterial gene regulation

Despite the predominant focus on genome editing, CRISPR-Cas systems are also revolutionizing the programmable regulation of gene expression. This feat was first demonstrated with CRISPR-Cas9 in two landmark publications (Bikard et al., 2013; Qi et al., 2013). In both publications, the *S. pyogenes* Cas9 nuclease was converted into a DNA-binding protein by point mutations to the RuvC and HNH domains (D10A and H840A). This catalytically dead Cas9 (dCas9) lost the ability to cleave DNA but retained its DNA binding ability (Jinek et al., 2012). By directing dCas9 to bind to the promoter or open reading frame of a target gene, expression could be repressed by preventing transcription initiation or elongation (Bikard et al., 2013; Qi et al., 2013) (Figure 1.4A). The strongest levels of repression occurred when targeting the promoter, with up to 1,000-fold repression. Targeting the coding region also repressed gene expression, where targeting was much more effective for the non-template strand than the template strand (Bikard et al., 2013; Qi et al., 2013). Expressing multiple sgRNAs was also shown to silence multiple genes at one time or, when targeting the same gene, to further boost silencing as long as the targeted sites did not overlap (Qi et al., 2013). Either an sgRNA (Qi et al., 2013) or the combination of a CRISPR array and a tracrRNA (Bikard et al., 2013) were capable of gene repression; the sgRNA offered a more compact option whereas the CRISPR array:tracrRNA was better suited for multiplexing. Advances in gene regulation with dCas9 were reviewed recently (Bikard and Marraffini, 2013; Fineran and Dy, 2014; Sampson and Weiss, 2014) and a protocol is available based on sgRNAs (Larson et al., 2013).

Beyond *E. coli*, dCas9 has been implemented in a growing collection of other bacterial strains. Bikard et al. (2013) implemented dCas9 in *Streptococcus pneumoniae* to repress β-galactosidase expression and observed up to a 14-fold reduction in activity. Choudhary et al. (2015) utilized dCas9 in mycobacteria to identify essential genes. Tong et al. (2015) applied dCas9 in the actinomycetes species *Streptomyces coelicolor* to reversibly control expression of a gene involved in production of the actinorhodin antibiotic. These examples underscore the broad applicability of dCas9 to gene repression in bacteria. For basic genetic studies and strain

development, dCas9 may even be preferable over Cas9-based editing because multiple genes can be silenced at one time without the need to modify each genetic locus.

The dCas9 protein is also capable of bacterial gene activation (Figure 1.4B). In *E. coli*, this was demonstrated by fusing the ω subunit of RNA polymerase to dCas9 and then expressing this construct in cells lacking the ω subunit gene *rpoZ* (Bikard et al., 2013). This subunit co-purifies with RNA polymerase (RNAP) yet the subunit is dispensable, with no discernible deletion phenotypes (Dove and Hochschild, 1998). To optimize gene activation with dCas9, Bikard et al. (2013) assessed multiple permutations, including N- and C-terminal fusions, targeting the template or the non-template strand, varying the distance between the target site and the transcriptional start site, and testing promoters with varying transcriptional activities. The most successful combination was fusing the ω subunit to the C-terminus of Cas9 (Cas9-ω) and targeting 96 nucleotides upstream of the transcriptional start site of a weak promoter. While this configuration yielded ~23-fold activation (Bikard et al., 2013), most other configurations yielded less than 8-fold activation. Compared to >100-fold repression reported for dCas9 (Qi et al., 2013), gene activation with CRISPR-Cas9 has significant room for improvement.

Despite the intense focus on Type II systems and dCas9, Type I systems are also capable of transcriptional regulation (Figure 1.4C). This capability was recently demonstrated for the Type I-E system in *E. coli* through the expression of Cascade and the elimination of the Cas3 nuclease (Luo et al., 2015; Rath et al., 2015). In the absence of Cas3, the Cascade complex still processes the CRISPR array and binds target DNA sequences. Cascade can be expressed off the endogenous locus in the genome (Luo et al., 2015) or introduced exogenously via a plasmid (Rath et al., 2015). Similar to Type II systems, the Type I-E system was also capable of >100-fold repression when targeting the promoter. Strand bias within the transcribed region was also observed for Cascade-driven repression, although the particular bias did not match between the two reports (Luo et al., 2015; Rath et al., 2015). Because Type I systems comprise a diverse set of subtypes, it will be interesting to see if other subtypes are

capable of gene silencing and whether endogenous Type I systems in other bacteria can be repurposed for programmable gene regulation.

One draw of gene regulation using Type I systems is that multiple mature crRNA molecules can be generated from a single CRISPR array without the need for accessory processing factors. Using a single array, Luo et al. (2015) simultaneously silenced four independent endogenous sugar utilization pathways in *E. coli*, thereby generating a complex growth phenotype. When evaluating the potency of spacers within an array, it was observed that every additional spacer added to the array reduced the efficiency of each individual spacer (Luo et al., 2015). This points to Cascade being a limiting factor in Type I silencing and should be considered when targeting large numbers of genes. Similar analyses remain to be reported for gene regulation with dCas9.

Another emerging opportunity for CRISPR-based gene regulation is mRNA targeting. Hale et al., (2012) first demonstrated this capability by programming the *Pyrococcus furiosus* Type III-B complex to cleave RNAs *in vitro*. This demonstration was followed by Zebec et al. (2014) using the Type III-B system of *Sulfolobus solfataricus* to target mRNA degradation of specific chromosomal genes *in vivo* (Figure 1.4D). The Type III-A complex was also recently discovered to target RNA as demonstrated *in vivo* (Staals et al., 2014; Tamulaitis et al., 2014). Thus, the Type III-A Csm complex and the Type III-B Cmr complex hold promise as genetic tools to for the post-transcriptional regulation of chromosomal genes. Type III systems are also not subject to PAMs, potentially creating greater flexibility when selecting target sequences. A major downside was that gene silencing *in vivo* was relatively modest (~2-fold) in comparison to that achieved by dCas9 and Cascade (Zebec et al., 2014).

There is compounding evidence that some Cas9's naturally target RNA (Figure 1.4E). Sampson et al. (2013) demonstrated that the Type II-B Cas9 from *Francisella novicida* naturally represses the expression of the FTN_1103 bacterial lipoprotein as part of immune avoidance during host infection. Silencing required the tracrRNA as well as a small, CRISPR-Cas-associated RNA (scaRNA) encoded adjacent to the CRISPR array. The scaRNA was predicted to hybridize to the ribosome-binding site and start codon of the FTN_1103 mRNA.

Although the mechanism of silencing remains unclear, this general strategy allowed for the targeted silencing of the Hepatitis C virus in mammalian cells (Price et al., 2015). There is thus tremendous potential for targeted gene silencing with this Cas9 and a need to elucidate design principles.

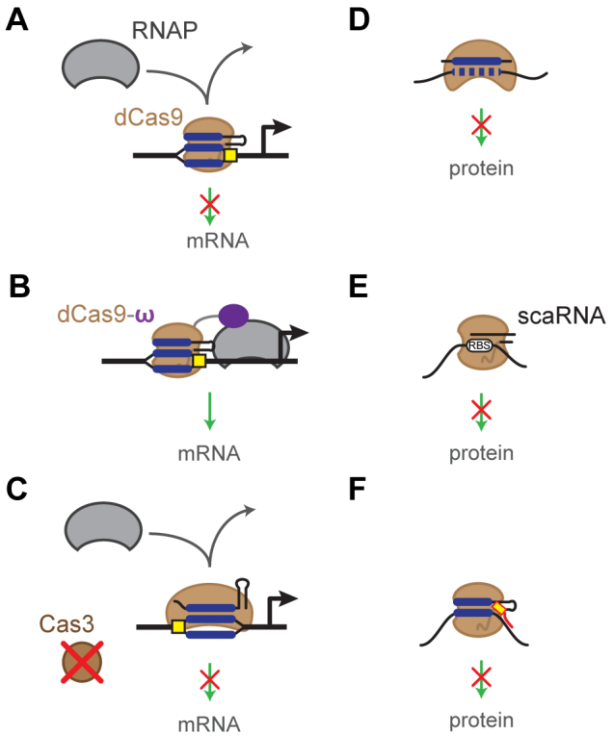

Figure 1.4 CRISPR-based gene regulation. (A) A catalytically dead Cas9 (dCas9) can be targeted to the promoter or coding region of a gene, blocking transcription. (B) A fusion between dCas9 and the ω subunit of RNA polymerase (dCas9-ω) can recruit RNA polymerase to activate transcription. (C) Eliminating the *cas3* gene from Type I systems can allow targeted DNA binding and transcriptional repression with Cascade. (D) Type III systems can be readily co-opted to bind and cleave target mRNAs. (E) The *Francisella novicida* Cas9 utilizes a scaRNA to silence an endogenous gene through putative base-pairing interactions. (F) Introducing a DNA oligonucleotide PAMmer allows Cas9 to bind and cleave RNAs complementary to the guide portion of an sgRNA.

Aside from the natural RNA-targeting ability of some Cas9 proteins, more traditional Cas9's can be coaxed into targeting RNA. This capability was demonstrated *in vitro* by pairing the *S. pyogenes* Cas9 with PAM-presenting oligonucleotides (PAMmers) (O'Connell et al., 2014) (Figure 1.4F). When the PAMmer base pairs with the target RNA, Cas9 recognizes the PAM but proceeds to cleave the RNA base-paired to the sgRNA. While Cas9 also would be expected to recognize and cleave the associated DNA, selecting targets that lacked a PAM allowed Cas9 to only recognize and cleave the PAMmer-bound RNA. Interestingly, RNA cleavage was only observed using a deoxyribonucleotide-based PAMmer but not a ribonucleotide-based PAMmer, suggesting that Cas9 can differentiate the slight structural variations between deoxyribose and ribose moieties within the PAM. Whether the PAMmer can yield efficient gene silencing *in vivo* remains to be investigated.

Beyond the co-regulation of multiple genes, CRISPR-Cas systems are now being implemented for the design of genetic circuits. While the first examples were published in mammalian cells (Kiani et al., 2014; Nissim et al., 2014), there has been one report of circuit design in bacteria (Nielsen and Voigt, 2014). In this work, sgRNAs were designed to target promoters driving expression of other sgRNAs, allowing the construction of complex circuit topologies. The sgRNAs were designed to recognize distinct sequences, preventing any measurable crosstalk. The resulting circuits could interface with endogenous processes to control cellular phenotypes. Specifically, the authors connected an OR logic gate to the expression of the *malT* transcription factor, which regulates maltose utilization and also production of a lambda phage receptor. When at least one input (inducer) was present, the cells exhibited near-normal lambda phage infectivity. However, when both inputs were absent, the cells exhibited a 240-fold reduction in plaque formation. While these circuits all relied on a single Cas9, the availability of orthogonal Cas9's that recognize different sgRNAs (Esvelt et al, 2013) opens the opportunity to create sophisticated gene circuits and functionalities.

Separately from gene regulation, CRISPR-Cas systems can be harnessed for RNA processing. Type I and Type III systems naturally process the transcribed CRISPR array into individual crRNAs through the activity of the Cas6 protein (Carte et al., 2008; Haurwitz et al.,

2010; Li, 2015; Niewoehner et al., 2014). This protein specifically binds the hairpin within the CRISPR repeat, thereby cleaving at the 3' base. Qi et al. (2012) exploited this activity in *E. coli* by encoding the hairpin sequence from the *Pseudomonas aeruginosa* Type I-F CRISPR locus between the 5' untranslated region of a gene and its ribosome-binding site. Co-expressing the system's Cas6 protein Csy4 resulted in efficient cleavage of the hairpin, separating the untranslated region from the coding region of the gene. The result was that the sequence of the untranslated region had little effect on translation—effectively insulating gene expression from the upstream sequence. This same approach yielded predictable gene expression for multi-gene operons when the hairpin was placed in the intervening untranslated regions. The extension of this strategy in the bacterium *Bacillus subtilis*, the eukaryote *Saccharomyces cerevisiae* (Qi et al., 2012), and mammalian cells (Nissim et al., 2014; Tsai et al., 2014) demonstrates the universality of Csy4 for directed RNA processing.

1.3.3 Next-generation antimicrobials

The emergence of multidrug resistance and the trickling pipeline for new small-molecule antibiotics has led to intense interest in the development of novel antimicrobials. While broad-spectrum antimicrobials have been the standard, the increasingly recognized importance of the human microbiota (Lemon et al., 2012; Neish, 2009; Walter and Ley, 2011) underscores the need to selectively eliminate pathogens while maintaining the natural microflora. The specificity, ease of design, and lethality of CRISPR created the possibility of exploiting these defense systems as the basis of programmable antimicrobials (Figure 1.5).

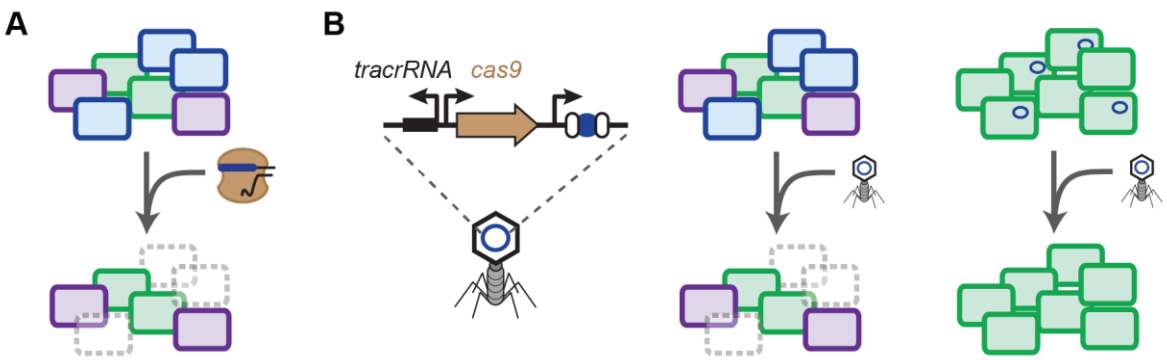

Figure 1.5 CRISPR-based antimicrobials. (A) Targeting the bacterial genome leads to potent and sequence-specific cell killing. (B) Bacteriophages can be employed to deliver plasmids encoding CRISPR-Cas9, leading to targeted killing or plasmid clearance

There was circumstantial evidence that CRISPR could serve as an antimicrobial based on the lethality and specificity of genome targeting (see section on *Bacterial genome editing*). However, the concept of CRISPR-based antimicrobials was first extended and experimentally demonstrated only recently (Gomaa et al., 2014). It was hypothesized that the nuclease of a CRISPR-Cas system could be directed to the bacterial genome, leading to genome cleavage and cell death; specificity would come from selecting target sequences present in some bacteria but not others. To demonstrate this concept, Gomaa et al. (2014) encoded a genome-targeting CRISPR array from the *E. coli* Type I-E system and transformed this plasmid into an *E. coli* strain expressing the associated Cas genes. The transformation efficiency of this plasmid was ~105-fold lower than that of a non-targeting control, indicating potent killing. Remarkably, every location tested—whether in coding or non-coding regions, essential or non-essential genes, top and bottom strands of the genome—resulted in a similar drop in the transformation efficiency. Furthermore, killing could be achieved using exogenous or endogenous CRISPR-Cas systems. The only requirements were the presence of the target sequence and the presence of a flanking PAM. This targeting flexibility was used to differentiate two highly similar K-12 and B lineages of *E. coli* as well as between *E. coli* K-12 and *Salmonella*. By using mixtures of targeting and non-targeting plasmids, the authors were able to quantitatively reduce the

levels of one strain, opening the possibility of finely tuning the composition of a mixed microbial population rather than only removing individual members.

With the framework established for CRISPR-based antimicrobials, the next obstacle was delivery. A promising delivery vehicle involves bacteriophages. Bacteriophages naturally prey on bacteria by injecting their genetic material into cells and hijacking the host machinery to undergo replication. One way to co-opt bacteriophages as delivery vehicles is to incorporate the bacteriophage's packing and replication elements into a plasmid —called a phagemid (Westwater et al., 2002). As part of the lytic cycle, the phagemid undergoes packaging in competition with the bacteriophage genome. The resulting bacteriophage particles then deliver the packaged phagemid to susceptible strains.

Two recent publications exploited this strategy to deliver a phagemid encoding the *S. pyogenes* cas9, tracrRNA, and designed CRISPR array (Bikard et al., 2014; Citorik et al., 2014). This setup delivers the complete CRISPR-Cas9 system, obviating the need for an endogenous, active system. Bikard et al. (2014) employed this strategy to selectively kill an antibiotic-resistant strain of *Staphylococcus aureus*. *S. aureus* was a notable target because methicillin-resistant *S. aureus* (MRSA) is a principal cause of antibiotic-resistant infections. The phagemid was generated by cloning the packaging site and the *rinA*, *terS*, and *terL* genes from the staphylococcal ΦNM1 bacteriophage along with the required elements of CRISPR-Cas9. The CRISPR array was designed to target an antibiotic resistance gene present in the genome of *S. aureus*. Incubating the resulting bacteriophage particles with the *S. aureus* cells resulted in up to a 10^4-fold reduction in the number of viable colonies. This degree of killing was impressive considering that no selective pressure was applied to retain the phagemid. This phagemid system was also tested in a mouse skin colonization model in which the mice were exposed to a mixture of a kanamycin-resistant strain and a kanamycin-sensitive strain of *S. aureus*. As expected, the phagemid targeting the *kanR* gene reduced the proportion of kanamycin-resistant *S. aureus*.

Citorik et al. (2014) similarly used M13-derived phagemids to selectively target antibiotic-resistant strains of *E. coli*. The CRISPR array was designed to target a mutation in

the *gyrA* gene of *E. coli* that confers resistance to quinolone antibiotics. Exposing the resulting bacteriophage particles to *E. coli* cells harboring this mutation resulted in up to a 10^4-fold reduction in the number of viable colonies, paralleling that observed for *S. aureus* (Bikard et al., 2014). This same approach was used to selectively remove the *gyrA* mutants from a mixture of this strain and two other antibiotic-resistant strains of *E. coli*. To simulate treatment of an *in vivo* infection, a strain of enterohaemorrhagic *E. coli* (EHEC) was fed to larvae of the moth *Galleria mellonella*. The phagemid targeting a chromosomally encoded virulence factor was able to moderately improve survival.

Another potential use of CRISPR antimicrobials is the clearance of antibiotic resistance plasmids, as first demonstrated by Garneau et al., (2010). Bikard et al. (2014) designed the phagemid to target plasmid-borne antibiotic resistance genes harbored by a clinical strain of methicillin-resistant *S. aureus* (MRSA). Delivery of the ΦNM1 phagemid eliminated the plasmids without impacting cell viability and prevented any import of the same plasmid via conjugation. Citorik et al. (2014) targeted the pNDM-1 and pSHV-18 plasmids that confer resistance to β-lactam antibiotics. In both cases, delivery of the M13 phagemid resulted in a 10^3-fold reduction in the number of viable colonies. This reduction was associated with addiction systems encoded on the plasmid, which triggered cell death upon plasmid removal. This strategy was quite convenient by simultaneously removing antibiotic resistance and killing the host cells.

In a similar attempt to sensitize bacteria to antibiotics, Yosef et al. (2015) engineered a λ prophage to carry the *E. coli* Type I-E system as well as a CRISPR array encoding spacers that target the β-lactamase genes *ndm-1* and *ctx-M-15*. In this case, the phage would form a stable lysogen in the host, and the encoded CRISPR-Cas system would both clear existing resistance genes and prevent the cells from acquiring resistance in the future. The authors demonstrated that the presence of the lysogen reduced the transformation efficiency of plasmids encoding the target sequences by three orders-of-magnitude. Furthermore, lysogens with the *ndm-1* and *ctx-M-15* spacers resisted T7 phages that housed *ndm-1* and *ctx-M-15* protospacers by over four orders-of-magnitude when compared to lysogens lacking the spacer

array, demonstrating that these antibiotic sensitized bacteria have a selective advantage in their resistance to lytic phages.

Aside from bacteriophages, bacterial conjugation could be used to deliver DNA. Conjugation relies on cell-to-cell contact to transfer plasmids that often have wide host ranges. Citorik et al. (2014) found that conjugative transfer of the plasmid encoding the *S. pyogenes* tracrRNA, *cas9,* and targeting CRISPR array resulted in up to a 60-fold reduction of viable recipients in comparison to non-targeted cells. The limited reduction in the number of viable cells was attributed to the inefficiency of conjugal transfer, a common issue for this delivery method.

Delivery remains the largest hurdle to the implementation of CRISPR antimicrobials. Even in both examples cited above, the simple *in vivo* models exhibited greatly reduced efficacy in comparison to the *in vitro* experiments. Future efforts will need to focus on engineering the bacteriophages for improved delivery and exploring alternative delivery vehicles. The long-term success will also depend on identifying applications uniquely suited to CRISPR antimicrobials—particularly over traditional antibiotics, antimicrobial peptides, and lytic bacteriophages. These applications may extend well beyond human therapeutics and into the realms of diagnostics, agriculture, biomanufacturing, and research tools.

1.4 FUTURE PERSPECTIVES

CRISPR-Cas systems have already proven to be powerful tools for understanding and engineering bacteria. However, engineered CRISPR-Cas systems represent a three-year-old technology and much of their potential remains to be realized. Below, we posit next steps for the development of CRISPR technologies in bacteria and the ensuing opportunities and challenges.

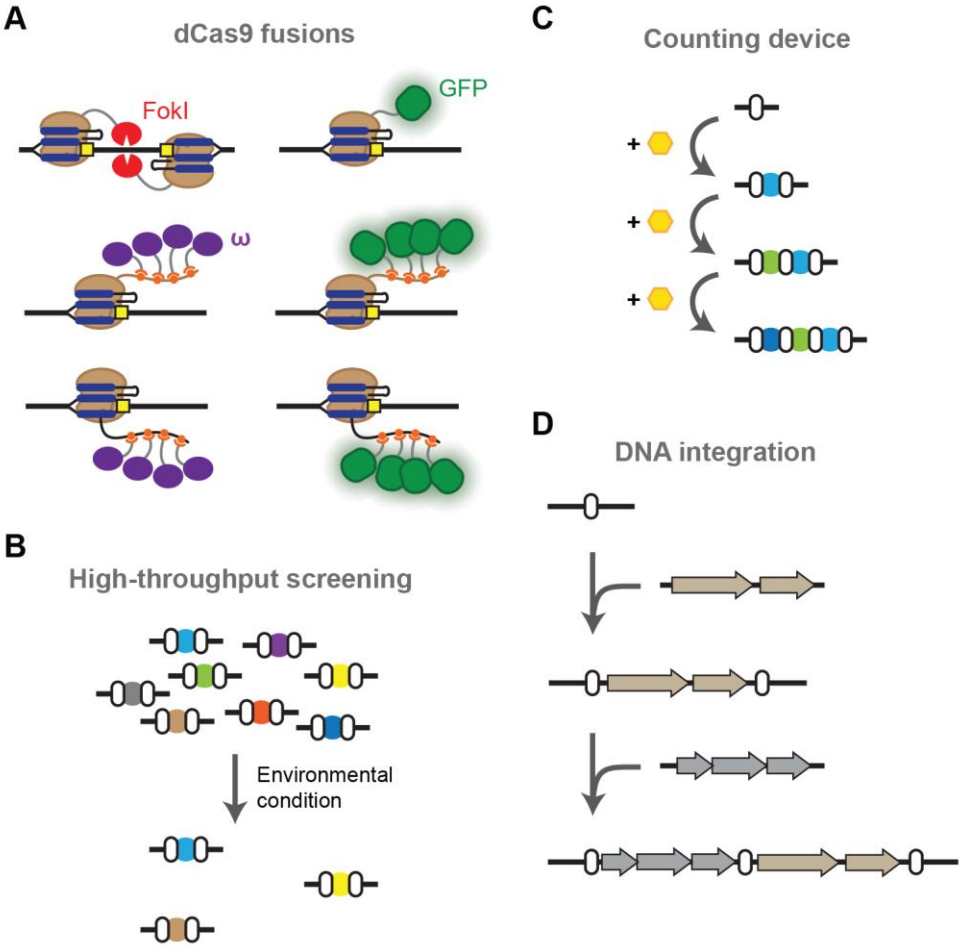

Figure 1.6 Opportunities for CRISPR technologies in bacteria. (A) Building on advances in eukaryotes. The FokI nuclease can be fused to Cas9 as an alternative means of introducing double-stranded breaks as part of genome editing. Fusing binding domains to Cas9 or the 3' end of the sgRNA can recruit other proteins to regulate gene expression (e.g., the ω subunit of RNA polymerase) or to dynamically image genomic loci (e.g., GFP). (B) Libraries of crRNAs or sgRNAs can be compiled and subjected to specific environmental conditions to rapidly screen for genes that either promote or inhibit growth. (C) Acquisition could be used to count transient events such as an environmental stimulus. (D) Acquisition could also be harnessed to integrate large pieces of synthetic DNA into the genome.

1.4.1 Following in the footsteps of CRISPR in eukaryotes

Even though CRISPR-Cas systems are native to prokaryotes, advances in CRISPR technologies have focused almost entirely on eukaryotes. Regardless of the underlying reasons, advances in eukaryotes can serve as a guide for similar developments in prokaryotes—whether to drive existing applications or to institute entirely new ones.

Genome editing remains the dominant and best-developed application of CRISPR-Cas systems in eukaryotes. Most of these developments have been driven by the need to express multiple sgRNAs, initiate homology-directed repair (HDR) over NHEJ, and limit unintended targeting (Hsu et al., 2014). Unintended targeting (or off-target effects) is less likely in bacteria because of their smaller genomes, and natural CRISPR arrays are well suited for multiplexing. However, many of the associated advances in eukaryotes could be useful in bacteria. For instance, fusions of dCas9 and the FokI endonuclease minimize unintended editing events by requiring a FokI dimerization event in order for DNA cleavage to occur (Guilinger et al., 2014; Tsai et al., 2014) (Figure 1.6A). In bacteria, this same fusion protein may generate double-stranded breaks in a manner that can more easily undergo repair by endogenous pathways. Separately, importing pathways for NHEJ or HDR could transition CRISPR from being a selective pressure to actively driving editing, as suggested by genome editing in actinomycetes (Tong et al., 2015). By improving the efficiency of genome editing in bacteria, CRISPR-based editing could be sufficiently effective to generate genome-wide modifications or to conduct high-throughput functional screens that identify genes associated with defined phenotypes (Koike-Yusa et al., 2013; Shalem et al., 2014; Wang et al., 2014; Zhou et al., 2014) (Figure 1.6B).

Gene regulation with CRISPR-Cas9 in eukaryotes has also witnessed major breakthroughs, with a focus on improving the system's overall efficiency and utility. The most notable advance has been the development of protein fusions to drive gene activation and repression (Gilbert et al., 2013; Maeder et al., 2013; Mali et al., 2013a; Perez-Pinera et al., 2013) (Figure 1.6A). One of the key developments was transitioning from single-domain fusions, such as between dCas9 and the VP64 activating domain or the KRAB repressing domain, to extensions of dCas9 or the sgRNA that recruit the regulatory domains (Konermann et al., 2014; Shechner et al., 2015; Tanenbaum et al., 2014; Zalatan et al., 2014). For instance, the 3' end of the sgRNA was extended to encode aptamers that recruited VP64 domains to efficiently activate transcription (Mali et al., 2013a). While gene repression with dCas9 and

Cascade seems to be sufficiently potent in bacteria, there is ample room for improvement for gene activation.

Aside from improving existing applications in bacteria, a notable advancement in eukaryotes has not been reported in bacteria: DNA imaging. Here, dCas9 was fused to GFP (Anton et al., 2014; Chen et al., 2013; Ma et al., 2015) or recruited GFP (Tanenbaum et al., 2014) to dynamically image chromosomal structure and localization in mammalian cells (Figure 1.6A). Bacterial genomes are known to be highly structured and undergo major localization changes based on the environmental conditions (Libby et al., 2012). However, the associated imaging tools rely on previous technologies such as fusing GFP to transcription factors; incorporating dCas9 and GFP offers a more programmable tool that could drive our understanding of chromosomal dynamics in bacteria, shedding new light into the structure of bacterial cells and how this structure changes based on environmental conditions.

1.4.2 Extending beyond imported Cas9

CRISPR technologies have focused on importing Type II systems because of the ease in using Cas9 and an sgRNA. However, Type II systems represent merely one type of CRISPR-Cas systems, where the other systems individually offer distinct attributes and capabilities. For instance, Type I systems naturally process CRISPR arrays without the need for further engineering. Type I systems are also unique in that their signature Cas3 protein does not generate double-stranded breaks but rather is recruited to cleave and degrade DNA through the action of the protein's 3'-to-5' exonuclease activity (Sinkunas et al. 2013). Separately, Type III systems naturally target RNA and may offer targeting flexibility due to the absence of a PAM—whether for mRNA silencing or transcript processing. The challenge will be finding applications that are uniquely suited to a given system and counterbalance the added difficulty of expressing multiple proteins in defined ratios, rather than the single Cas9 protein for Type II systems. The newly categorized Type V systems and their Cpf1 proteins offer additional possibilities. Because Cpf1 processes crRNA arrays without any additional RNA species and the Cpf1-crRNA complex alone can achieve DNA cleavage, these systems may simplify the

design of genome-editing tools (Zetsche et al., 2015). Alternatively, by implementing multiple, truly orthogonal types of CRISPR-Cas systems, multiple functions such as genome editing, plasmid curing, gene activation, gene repression, and invader defense can be performed simultaneously.

One potential opportunity beyond Cas9 is the utilization of endogenous systems already present in a microbe. CRISPR loci have been identified in the genomes of ~50% of bacteria and ~87% of archaea (Makarova et al., 2015), and the vast majority of these are associated with Type I systems and Type III systems. For bacteria harboring an active CRISPR-Cas system, the system may be readily co-opted by merely importing a synthetic CRISPR array. This approach was part of CRISPR antimicrobials (Gomaa et al., 2014), large deletions (Vercoe et al., 2013), and gene repression (Luo et al., 2015; Zebec et al., 2014). The benefit is that all of the protein components are already present—a major advantage for Class 1 systems that require multiple proteins (Table 1.1). In extreme thermophiles and hyperthermophiles, the traditional Cas9 proteins from mesophiles are unlikely to function at typical growth conditions. In contrast, Type I and Type III systems appear to be enriched in these microbes, arguing for their use in CRISPR-based applications. The downside to using endogenous systems is that the system must be confirmed to be active, which could be environment-dependent, and must undergo full characterization—namely, identifying the PAM for Type I, II, and V systems as well as determining the tracrRNA for Type II systems. Furthermore, co-opting the endogenous system may interfere with its natural functions such as in host defense. Accordingly, choosing to import a system (and which type) or to rely on an endogenous system will likely depend on the selected organism and the needs of the project. Finally, even if an imported Cas9 is selected, there is a choice between using CRISPR arrays or sgRNAs. The important takeaway is that importing Cas9 with an sgRNA is not the only option and, in many cases, may be less desirable than the alternatives.

1.4.3 Exploiting CRISPR-Cas acquisition

To date, CRISPR-derived genetic tools have centered on the final step in prokaryotic adaptive immunity—interference. However, an unexplored step is acquisition (Heler et al., 2014). Acquisition represents the first step of acquired immunity wherein the CRISPR-Cas system integrates a small piece of the foreign invader DNA (Figure 1.1). This step is associated with the near-universal proteins Cas1 and Cas2, with some involvement by Cas4 in some subtypes (Chylinski et al., 2014; Makarova et al., 2011; Makarova et al., 2015; Nuñez et al., 2014). Cas1 and Cas2 integrate new repeat-spacers in the CRISPR array, with a preference for stalled replication forks (Levy et al., 2015). However, the Cas proteins involved in interference appear to play a predominant role in spacer selection, whether by selecting spacers with a PAM (Heler et al., 2015; Paez-Espino et al., 2013) or by driving acquisition of sequences in the vicinity of a target sequence (Fineran et al., 2014; Richter et al., 2014). There is still plenty left to learn about the natural mechanisms that drive acquisition, although there may be sufficient information to begin engineering this untapped yet universal aspect of CRISPR-Cas systems.

We envision two avenues in which acquisition can be harnessed for tools in bacteria. In the first avenue, acquisition can be linked to an exogenous or intracellular signal such that transient induction leads to an acquisition event (Figure 1.6C). This could in turn be used as a permanent memory of prior transient events—much like natural CRISPR loci serve as memories of prior infections. This approach could offer a simpler strategy to track these events, particularly in comparison to existing synthetic counters and memory devices (Bonnet et al., 2012; Friedland et al., 2009; Siuti et al., 2013). Second, the system could be harnessed as site-specific recombinases in order to drive the efficient insertion of a synthetic DNA sequence (Figure 1.6D). While the natural systems prefer to integrate ~30-base sequences directly into the leading repeat, it may be possible to reengineer the system to accept larger or shorter sequences that can be inserted at other locations. This capability could greatly expand the existing repertoire of protein integrases and form the next generation of CRISPR technologies.

1.5 CONCLUSIONS

CRISPR-Cas systems have enjoyed tremendous popularity within recent years, owing to the versatility and power of these widespread defense systems. Bacterial applications thus far include streamlined genome engineering, programmable transcriptional regulation, typing and epidemiology of strains, vaccination of bacteria against mobile genetic elements, and smart antibiotics. Most of these applications are in their infancy and there still is ample room for further improvements. Nonetheless, the gains so far have begun reshaping how we pursue basic and applied research in bacteria and offer considerable potential for the treatment of bacterial infections.

ACKNOWLEDGMENTS

We thank K. Selle, R. Barrangou, and G. Williams for critical reading of the manuscript. This work was supported by funding from the National Science Foundation (CBET-1403135 to C.L.B, MCB-1452902 to C.L.B.) and the National Institutes of Health (5T32GM008776-15 to R.T.L.). M.L.L. and C.L.B. are co-inventors on multiple patent applications related to CRISPR-Cas systems and their uses.

REFERENCES

Aklujkar M, Lovley DR. 2010. Interference with histidyl-tRNA synthetase by a CRISPR spacer sequence as a factor in the evolution of *Pelobacter carbinolicus*. *BMC Evol. Biol.* **10**:230.

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.

Aravind L, Koonin EV. 2001. Prokaryotic homologs of the eukaryotic DNA-end-binding protein Ku, novel domains in the Ku protein and prediction of a prokaryotic double-strand break repair system. *Genome Res.* **11**:1365–1374.

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. *Mol. Cell* **54**:234–244.

Bassett AR, Tibbit C, Ponting CP, Liu JL. 2013. Highly efficient targeted mutagenesis of *Drosophila* with the CRISPR/Cas9 system. *Cell Rep.* **4**:220–228.

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. *Nat. Biotechnol.* **32**:1146–1150.

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 Res.* **41**:7429–7437.

Bikard D, Marraffini LA. 2013. Control of gene expression by CRISPR-Cas systems. *F1000Prime Rep.* **5**:47.

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.

Bondy-Denomy J, Davidson AR. 2014. To acquire or resist: The complex biological effects of CRISPR-Cas systems. *Trends Microbiol.* **22**:218–225.

Bonnet J, Subsoontorn P, Endy D. 2012. Rewritable digital data storage in live cells via engineered control of recombination directionality. *Proc. Natl. Acad. Sci. U. S. A.* **109**:8884–8889.

Bowater R, Doherty AJ. 2006. Making ends meet: Repairing breaks in bacterial DNA by non-homologous end-joining. *PLoS Genet.* **2**:e8.

Brendel J, Stoll B, Lange SJ, Sharma K, Lenz C, Stachler AE, Maier LK, Richter H, Nickel L, Schmitz RA, Randau L, Allers T, Urlaub H, Backofen R, Marchfelder A. 2014. A complex of Cas proteins 5, 6, and 7 is required for the biogenesis and stability of clustered regularly interspaced short palindromic repeats (CRISPR)-derived RNAs (crRNAs) in *Haloferax volcanii*. *J. Biol. Chem.* **289**:7164–7177.

Brouns SJJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJH, Snijders APL, Dickman MJ, Makarova KS, Koonin E V, van der Oost J. 2008. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**:960–964.

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 Dev.* **22**:3489–3496.

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

Cho SW, Kim S, Kim JM, Kim JS. 2013. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat. Biotechnol.* **31**:230–232.

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

Chylinski K, Makarova KS, Charpentier E, Koonin EV. 2014. Classification and evolution of type II CRISPR-Cas systems. *Nucleic Acids Res.* **42**:6091–6105.

Chylinski K, Le Rhun A, Charpentier E. 2013. The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems. *RNA Biol.* **10**:726–737.

Citorik RJ, Mimee M, Lu TK. 2014. Sequence-specific antimicrobials using efficiently delivered RNA-guided nucleases. *Nat. Biotechnol.* **32**:1141–1145.

Cobb RE, Wang Y, Zhao H. 2015. High-Efficiency Multiplex genome editing of Streptomyces species using an engineered CRISPR/Cas system. *ACS Synth. Biol.* **4**:723–728.

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

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.

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

DiCarlo JE, Norville JE, Mali P, Rios X, Aach J, Church GM. 2013. Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Res.* **41**:4336–4343.

Dickinson DJ, Ward JD, Reiner DJ, Goldstein B. 2013. Engineering the *Caenorhabditis elegans* genome using Cas9-triggered homologous recombination. *Nat. Methods* **10**:1028–1034.

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

Edgar R, Qimron U. 2010. The *Escherichia coli* CRISPR system protects from λ lysogenization, lysogens, and prophage induction. *J. Bacteriol.* **192**:6291–6294.

Fineran PC, Dy RL. 2014. Gene regulation by engineered CRISPR-Cas systems. *Curr. Opin. Microbiol.* **18**:83–89.

Fineran PC, Gerritzen MJH, Suárez-Diez M, Künne T, Boekhorst J, van Hijum SAFT, Staals RHJ, Brouns SJJ. 2014. Degenerate target sites mediate rapid primed CRISPR adaptation. *Proc. Natl. Acad. Sci. U. S. A.* **111**:E1629–1638.

Friedland AE, Lu TK, Wang X, Shi D, Church G, Collins JJ. 2009. Synthetic gene networks that count. *Science* **324**:1199–1202.

Friedland AE, Tzur YB, Esvelt KM, Colaiácovo MP, Church GM, Calarco JA. 2013. Heritable genome editing in *C. elegans* via a CRISPR-Cas9 system. *Nat. Methods* **10**:741–743.

Garneau JE, Dupuis MÈ, Villion M, Romero DA, Barrangou R, Boyaval P, Fremaux C, Horvath P, Magadán 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. *Proc. Natl. Acad. Sci. U. S. A.* **109**:E2579–2586.

Gilbert LA, Larson MH, Morsut L, Liu Z, Brar GA, Torres SE, Stern-Ginossar N, Brandman O, Whitehead EH, Doudna JA, Lim WA, Weissman JS, Qi LS. 2013. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**:442–451.

Gomaa AA, Klumpe HE, Luo ML, Selle K, Barrangou R, Beisel CL. 2014. Programmable removal of bacterial strains by use of genome-targeting CRISPR-Cas systems. *MBio* **5**:e00928–13.

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.

Guilinger JP, Thompson DB, Liu DR. 2014. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat. Biotechnol.* **32**:577–582.

Hale CR, Majumdar S, Elmore J, Pfister N, Compton M, Olson S, Resch AM, Glover CVC, Graveley BR, Terns RM, Terns MP. 2012. Essential features and rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs. *Mol. Cell* **45**:292–302.

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.

Hatoum-Aslan A, Maniv I, Samai P, Marraffini L a. 2014. Genetic characterization of antiplasmid immunity through a type III-A CRISPR-Cas system. *J. Bacteriol.* **196**:310–7.

Haurwitz RE, Jinek M, Wiedenheft B, Zhou K, Doudna JA. 2010. Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science* **329**:1355–1358.

Heler R, Marraffini LA, Bikard D. 2014. Adapting to new threats: the generation of memory by CRISPR-Cas immune systems. *Mol. Microbiol.* **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.

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 Biochim. Biophys. Sin. (Shanghai).* **47**:231–243.

Hwang WY, Fu Y, Reyon D, Maeder ML, Tsai SQ, Sander JD, Peterson RT, Yeh JR, Joung JK. 2013. Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat. Biotechnol.* **31**:227–229.

Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A. 1987. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J. Bacteriol.* **169**:5429–5433.

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

Jackson RN, Lavin M, Carter J, Wiedenheft B. 2014b. Fitting CRISPR-associated Cas3 into the Helicase family tree. *Curr. Opin. Struct. Biol.* **24**:106–114.

Jansen R, Embden JD, Gaastra W, Schouls LM. 2002. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* **43**:1565–1575.

Jiang F, Doudna JA. 2015. The structural biology of CRISPR-Cas systems. *Curr. Opin. Struct. Biol.* **30**:100–111.

Jiang F, Zhou K, Ma L, Gressel S, Doudna JA. 2015a. A Cas9-guide RNA complex preorganized for target DNA recognition. *Science.* **348**:1477–1481.

Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA. 2013a. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* **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 Genet.* **9**:e1003844.

Jiang Y, Chen B, Duan C, Sun B, Yang J, Yang S. 2015b. Multigene editing in the *Escherichia coli* genome using the CRISPR-Cas9 system. *Appl. Environ. Microbiol.* **81**:2506–2514.

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, East A, Cheng A, Lin S, Ma E, Doudna J. 2013. RNA-programmed genome editing in human cells. *Elife* **2**:e00471.

Jore MM, Lundgren M, van Duijn E, Bultema JB, Westra ER, Waghmare SP, Wiedenheft B, Pul U, Wurm R, Wagner R, Beijer MR, Barendregt A, Zhou K, Snijders APL, Dickman MJ, Doudna JA, Boekema EJ, Heck AJR, van der Oost J, Brouns SJJ. 2011. Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat. Struct. Mol. Biol.* **18**:529–536.

Kiani S, Beal J, Ebrahimkhani MR, Huh J, Hall RN, Xie Z, Li Y, Weiss R. 2014. CRISPR transcriptional repression devices and layered circuits in mammalian cells. *Nat. Methods* **11**:723–726.

Koike-Yusa H, Li Y, Tan E-P, Velasco-Herrera MDC, Yusa K. 2013. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat. Biotechnol.* **32**:267–273.

Konermann S, Brigham MD, Trevino AE, Joung J, Abudayyeh OO, Barcena C, Hsu PD, Habib N, Gootenberg JS, Nishimasu H, Nureki O, Zhang F. 2014. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**:583–588.

Larson MH, Gilbert LA, Wang X, Lim WA, Weissman JS, Qi LS. 2013. CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat. Protoc.* **8**:2180–2196.

Lemon KP, Armitage GC, Relman DA, Fischbach MA. 2012. Microbiota-targeted therapies: an ecological perspective. *Sci. Transl. Med.* **4**:137rv5.

Levy A, Goren MG, Yosef I, Auster O, Manor M, Amitai G, Edgar R, Qimron U, Sorek R. 2015. CRISPR adaptation biases explain preference for acquisition of foreign DNA. *Nature* **520**:505–510.

Li H. 2015. Structural principles of CRISPR RNA processing. *Structure* **23**:13–20.

Li JF, Norville JE, Aach J, McCormack M, Zhang D, Bush J, Church GM, Sheen J. 2013. Multiplex and homologous recombination–mediated genome editing in *Arabidopsis* and *Nicotiana benthamiana* using guide RNA and Cas9. *Nat. Biotechnol.* **31**:688–691.

Li Y, Lin Z, Huang C, Zhang Y, Wang Z, Tang Y, Chen T, Zhao X. 2015. Metabolic engineering of *Escherichia coli* using CRISPR-Cas9 meditated genome editing. *Metab. Eng.* **31**:13–21.

Libby EA, Roggiani M, Goulian M. 2012. Membrane protein expression triggers chromosomal locus repositioning in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **109**:7445–7450.

Luo ML, Mullis AS, Leenay RT, Beisel CL. 2015. Repurposing endogenous Type I CRISPR-Cas systems for programmable gene repression. *Nucleic Acids Res* **43**:674–681.

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

Maeder ML, Linder SJ, Cascio VM, Fu Y, Ho QH, Joung JK. 2013. CRISPR RNA-guided activation of endogenous human genes. *Nat. Methods* **10**:977–979.

Makarova KS, Haft DH, Barrangou R, Brouns SJJ, Charpentier E, Horvath P, Moineau S, Mojica FJM, Wolf YI, Yakunin AF, van der Oost J, Koonin EV. 2011. Evolution and classification of the CRISPR-Cas systems. *Nat. Rev. Microbiol.* **9**:467–477.

Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ, Barrangou R, Brouns SJJ, Charpentier E, Haft DH, Horvath P, Moineau S, Mojica FJM, Terns RM, Terns MP, White MF, Yakunin AF, Garrett RA, van der Oost J, Backofen R, Koonin E V. 2015. An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.* **(in press)**. doi: 10.1038/nrmicro3569

Mali P, Aach J, Stranges PB, Esvelt KM, Moosburner M, Kosuri S, Yang L, Church GM. 2013a. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.* **31**:833–838.

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

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

Marraffini LA, Sontheimer EJ. 2010. Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature* **463**:568–571.

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

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

Nam KH, Haitjema C, Liu X, Ding F, Wang H, DeLisa MP, Ke A. 2012. Cas5d protein processes pre-crRNA and assembles into a Cascade-like interference complex in subtype I-C/Dvulg CRISPR-Cas system. *Structure* **20**:1574–1584.

Neish AS. 2009. Microbes in Gastrointestinal Health and Disease. *Gastroenterology* **136**:65–80.

Nekrasov V, Staskawicz B, Weigel D, Jones JDG, Kamoun S. 2013. Targeted mutagenesis in the model plant *Nicotiana benthamiana* using Cas9 RNA-guided endonuclease. *Nat. Biotechnol.* **31**:691–693.

Nielsen AA, Voigt CA. 2014. Multi-input CRISPR/Cas genetic circuits that interface host regulatory networks. *Mol. Syst. Biol.* **10**:763.

Niewoehner O, Jinek M, Doudna JA. 2014. Evolution of CRISPR RNA recognition and processing by Cas6 endonucleases. *Nucleic Acids Res.* **42**:1341–1353.

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.

Nissim L, Perli SD, Fridkin A, Perez-Pinera P, Lu TK. 2014. Multiplexed and programmable regulation of gene networks with an integrated RNA and CRISPR/Cas toolkit in human cells. *Mol. Cell* **54**:698–710.

Niu Y, Shen B, Cui Y, Chen Y, Wang J, Wang L, Kang Y, Zhao X, Si W, Li W, Xiang AP, Zhou J, Guo X, Bi Y, Si C, Hu B, Dong G, Wang H, Zhou Z, Li T, Tan T, Pu X, Wang F, Ji S, Zhou Q, Huang X, Ji W, Sha J. 2014. Generation of gene-modified cynomolgus monkey via Cas9/RNA-mediated gene targeting in one-cell embryos. *Cell* **156**:836–843.

Nuñez JK, Kranzusch PJ, Noeske J, Wright A V, Davies CW, Doudna J a. 2014. Cas1–Cas2 complex formation mediates spacer acquisition during CRISPR–Cas adaptive immunity. *Nat. Struct. Mol. Biol.* **21**:528–534.

O'Connell MR, Oakes BL, Sternberg SH, East-Seletsky A, Kaplan M, Doudna JA. 2014. Programmable RNA recognition and cleavage by CRISPR/Cas9. *Nature* **516**:263–266.

Oh J-H, van Pijkeren J-P. 2014. CRISPR-Cas9-assisted recombineering in *Lactobacillus reuteri*. *Nucleic Acids Res.* **42**:e131.

Van der Oost J, Westra ER, Jackson RN, Wiedenheft B. 2014. Unravelling the structural and mechanistic basis of CRISPR-Cas systems. *Nat. Rev. Microbiol.* **12**:479–492.

Osawa T, Inanaga H, Sato C, Numata T. 2015. Crystal Structure of the CRISPR-Cas RNA silencing Cmr complex bound to a target analog. *Mol. Cell* **58**:418–430.

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. *Nat. Commun.* **4**:1430.

Peng W, Feng M, Feng X, Liang YX, She Q. 2014. An archaeal CRISPR type III-B system exhibiting distinctive RNA targeting features and mediating dual RNA and DNA interference. *Nucleic Acids Res.* **43**:406–417.

Peng W, Li H, Hallstrøm S, Peng N, Liang YX, She Q. 2013. Genetic determinants of PAM-dependent DNA targeting and pre-crRNA processing in *Sulfolobus islandicus*. *RNA Biol.* **10**:738–748.

Perez-Pinera P, Kocak DD, Vockley CM, Adler AF, Kabadi AM, Polstein LR, Thakore PI, Glass KA, Ousterout DG, Leong KW, Guilak F, Crawford GE, Reddy TE, Gersbach CA. 2013. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat. Methods* **10**:973–976.

Price AA, Sampson TR, Ratner HK, Grakoui A, Weiss DS. 2015. Cas9-mediated targeting of viral RNA in eukaryotic cells. *Proc. Natl. Acad. Sci. U. S. A.* **112**:6164–6169.

Pyne ME, Moo-Young M, Chung DA, Chou CP. 2015. Coupling the CRISPR/Cas9 System with Lambda Red recombineering enables simplified chromosomal gene replacement in *Escherichia coli*. Ed. M. Kivisaar. *Appl. Environ. Microbiol.* **81**:5103–5114.

Qi L, Haurwitz RE, Shao W, Doudna JA, Arkin AP. 2012. RNA processing enables predictable programming of gene expression. *Nat. Biotechnol.* **30**:1002–1006.

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.

Rath D, Amlinger L, Hoekzema M, Devulapally PR, Lundgren M. 2015. Efficient programmable gene silencing by Cascade. *Nucleic Acids Res.* **43**:237–246.

Richter C, Dy RL, McKenzie RE, Watson BNJ, Taylor C, Chang JT, McNeil MB, Staals RHJ, Fineran PC. 2014. Priming in the Type I-F CRISPR-Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer. *Nucleic Acids Res.* **42**:8516–8526.

Rouillon C, Zhou M, Zhang J, Politis A, Beilsten-Edmands V, Cannone G, Graham S, Robinson CV, Spagnolo L, White MF. 2013. Structure of the CRISPR interference complex CSM reveals key similarities with Cascade. *Mol. Cell* **52**:124–134.

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.

Sampson TR, Saroj SD, Llewellyn AC, Tzeng YL, Weiss DS. 2013. A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. *Nature* **497**:254–457.

Sampson TR, Weiss DS. 2014. Exploiting CRISPR/Cas systems for biotechnology. *Bioessays* **36**:34–38.

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

Selle K, Barrangou R. 2015. Harnessing CRISPR-Cas systems for bacterial genome editing. *Trends Microbiol.* **23**:225–232.

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

Shechner DM, Hacisuleyman E, Younger ST, Rinn JL. 2015. Multiplexable, locus-specific targeting of long RNAs with CRISPR-Display. *Nat. Methods* **12**:664–670.

Sinkunas T, Gasiunas G, Fremaux C, Barrangou R, Horvath P, Siksnys V. 2011. Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. *EMBO J.* **30**:1335–1342.

Siuti P, Yazbek J, Lu TK. 2013. Synthetic circuits integrating logic and memory in living cells. *Nat. Biotechnol.* **31**:448–452.

Spilman M, Cocozaki A, Hale C, Shao Y, Ramia N, Terns R, Terns M, Li H, Stagg S. 2013. Structure of an RNA silencing complex of the CRISPR-Cas immune system. *Mol. Cell* **52**:146–152.

Staals RHJ, Agari Y, Maki-Yonekura S, Zhu Y, Taylor DW, van Duijn E, Barendregt A, Vlot M, Koehorst JJ, Sakamoto K, Masuda A, Dohmae N, Schaap PJ, Doudna JA, Heck AJR, Yonekura K, van der Oost J, Shinkai A. 2013. Structure and activity of the RNA-targeting Type III-B CRISPR-Cas complex of *Thermus thermophilus*. *Mol. Cell* **52**:135–145.

Staals RHJ, Zhu Y, Taylor DW, Kornfeld JE, Sharma K, Barendregt A, Koehorst JJ, Vlot M, Neupane N, Varossieau K, Sakamoto K, Suzuki T, Dohmae N, Yokoyama S, Schaap PJ, Urlaub H, Heck AJR, Nogales E, Doudna JA, Shinkai A, van der Oost J. 2014. RNA Targeting by the Type III-A CRISPR-Cas Csm complex of *Thermus thermophilus*. *Mol. Cell* **56**:518–530.

Stern A, Keren L, Wurtzel O, Amitai G, Sorek R. 2010. Self-targeting by CRISPR: gene regulation or autoimmunity? *Trends Genet.* **26**:335–340.

Szczelkun MD, Tikhomirova MS, Sinkunas T, Gasiunas G, Karvelis T, Pschera P, Siksnys V, Seidel R. 2014. Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proc. Natl. Acad. Sci. U. S. A.* **111**:9798–9803.

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

Tanenbaum ME, Gilbert LA, Qi LS, Weissman JS, Vale RD. 2014. A versatile protein tagging system for signal amplification in single molecule imaging and gene regulation. *Cell* **159**:635–646.

Taylor DW, Zhu Y, Staals RHJ, Kornfeld JE, Shinkai A, van der Oost J, Nogales E, Doudna JA. 2015. Structures of the CRISPR-Cmr complex reveal mode of RNA target positioning. *Science.* **348**:581–585.

Tong Y, Charusanti P, Zhang L, Weber T, Lee SY. 2015. CRISPR-Cas9 based engineering of Actinomycetal genomes. *ACS Synth. Biol.* **4**:1020–1029.

Tsai SQ, Wyvekens N, Khayter C, Foden JA, Thapar V, Reyon D, Goodwin MJ, Aryee MJ, Joung JK. 2014. Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat. Biotechnol.* **32**:569–576.

Vercoe RB, Chang JT, Dy RL, Taylor C, Gristwood T, Clulow JS, Richter C, Przybilski R, Pitman AR, Fineran PC. 2013. Cytotoxic Chromosomal Targeting by CRISPR/Cas Systems Can Reshape Bacterial Genomes and Expel or Remodel Pathogenicity Islands. *PLoS Genet.* **9**:e1003454.

Walter J, Ley R. 2011. The Human Gut Microbiome: Ecology and Recent Evolutionary Changes. *Annu. Rev. Microbiol.* **65**:411–429.

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. *J. Biotechnol.* **200**:1–5.

Westra ER, van Erp PBG, Künne T, Wong SP, Staals RHJ, Seegers CLC, Bollen S, Jore MM, Semenova E, Severinov K, de Vos WM, Dame RT, de Vries R, Brouns SJJ, van der Oost J. 2012. CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Mol. Cell* **46**:595–605.

Westra ER, Semenova E, Datsenko KA, Jackson RN, Wiedenheft B, Severinov K, Brouns SJJ. 2013. Type I-E CRISPR-cas systems discriminate target from non-target DNA through base pairing-independent PAM recognition. *PLoS Genet.* **9**:e1003742.

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, Lander GC, Zhou K, Jore MM, Brouns SJJ, van der Oost J, Doudna J a, Nogales E. 2011. Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature* **477**:486–489.

Xie K, Minkenberg B, Yang Y. 2015. Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proc. Natl. Acad. Sci. U. S. A.* **112**:3570–3575.

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. Ed. A. M. Spormann. *Appl. Environ. Microbiol.* **81**:4423–4431.

Yosef I, Manor M, Kiro R, Qimron U. 2015. Temperate and lytic bacteriophages programmed to sensitize and kill antibiotic-resistant bacteria. *Proc. Natl. Acad. Sci. U.S.A.* **112**:7267–7272.

Zalatan JG, Lee ME, Almeida R, Gilbert LA, Whitehead EH, La Russa M, Tsai JC, Weissman JS, Dueber JE, Qi LS, Lim WA. 2014. Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell* **160**:339–350.

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

Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, Volz SE, Joung J, van der Oost J, Regev A, Koonin E V., Zhang F. 2015. Cpf1 Is a Single RNA-guided endonuclease of a Class 2 CRISPR-Cas System. *Cell* **(in press)**. doi: 10.1016/j.cell.2015.09.038

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.

CHAPTER 2

Engineering genes with CRISPR-Cas9

Michelle L. Luo, Chase L. Beisel

Department of Chemical and Biomolecular Engineering, North Carolina State University, Raleigh, NC 27695

Original publication in: *Chemical Engineering Progress,* September 2016: 36-41

2.1 Introduction

CRISPR-Cas9 has emerged as a powerful new gene-editing tool, giving researchers the power to quickly and easily alter DNA. With applications from gene therapy to crop development, genome engineering has the potential to cure disease and solve world hunger (Figure 2.1). These capabilities have also raised major ethical questions that are being actively debated worldwide.

Here, we explain what CRISPR is, how it works, why it's revolutionizing biotechnology, and the multiple debates throughout society that have been catalyzed by the technology. This article is meant to serve as a primer, where many excellent reviews offer greater depth on the natural role of CRISPR (1–3) and its many technological uses (4–7).

2.2 CRISPR-Cas systems are bacterial adaptive immune systems

CRISPR (clustered regularly interspaced short palindromic repeats) technologies have their origins in the prokaryotic world of bacteria and archaea. Like humans, bacteria have their own type of viruses. When these viruses infect a cell, they inject their DNA in an attempt to co-opt the cellular host machinery to produce more copies of the virus. Phage proliferation often leaves behind a damaged or dead bacterial host cell. As a result, bacteria have evolved multilayered antiviral defense systems. CRISPR-Cas systems are a highly adaptive and inheritable system that represents one mechanism of prokaryotic defense. CRISPR-Cas systems are unique in their ability to create memory in reaction to an initial infection, which enhances immune response to subsequent reinfections.

CRISPR-Cas systems provide immunity through a general three-step pathway: acquisition, expression, and interference (Figure 2.2). During the first stage, foreign genetic material infects the cell. Cas proteins recognize this viral DNA (or RNA) and generate fragments of that viral DNA to be copied and pasted into the bacterial genome at the CRISPR locus. This sampling of viral DNA is what drives the adaptive element of the immune response. In this manner, the CRISPR-Cas system allows the bacteria to retain memory of past infections.

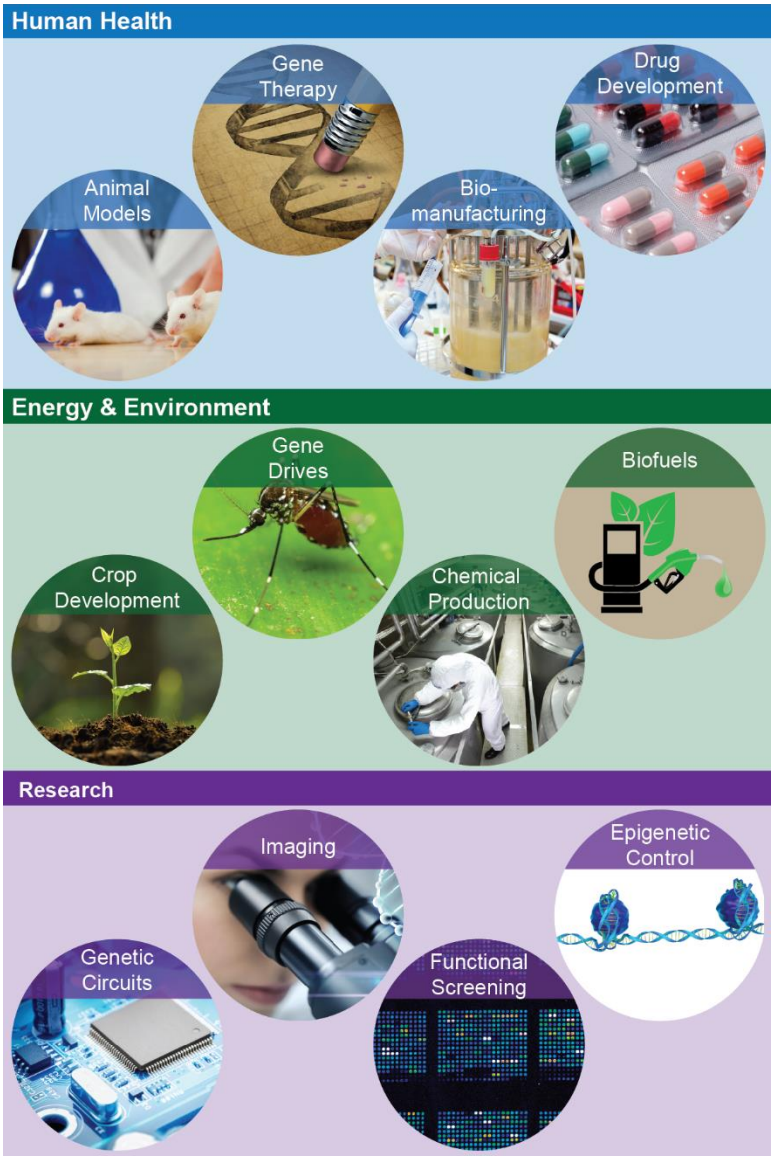

Figure 2.1 CRISPR-Cas technologies have many uses and applications across multiple industries.

During the second stage of immunity, the cell expresses the CRISPR locus and produces CRISPR RNA (crRNA) molecules, which contain an exact replica of the previously seen viral DNA. The crRNAs work together with Cas proteins and accessory factors to protect the cells. Because RNA is a chemical cousin of DNA, it can interact with DNA molecules that have a complementary sequence.

In the final stage, a phage reinfects the bacterium. This time, however, the protein-crRNA effector complex recognizes the viral DNA that matches the crRNA sequence. This recognition triggers the complex to degrade and eliminate the viral DNA, thereby protecting the cell.

While immunity is the natural function of CRISPR-Cas systems, researchers discovered that these effector complexes are programmable, so scientists can engineer these systems to recognize not only viral DNA, but almost any DNA sequence of our choosing. It is this programmable recognition of nucleic acids that makes CRISPR so appealing as a basis for genetic tools.

2.3 Exploiting the power of CRISPR-Cas systems

While there are many different types of CRISPR-Cas systems, the CRISPR craze largely centers around one specific protein, Cas9. To function, Cas9 must be complexed with a crRNA and a trans-activating RNA (tracrRNA). This has been simplified by combining the crRNA and the tracrRNA into an engineered single-guide RNA (sgRNA) (8). As a result, only two components—the Cas9 protein and the sgRNA molecule— are required for this engineered system to function (Figure 2.3).

To recognize a target sequence, Cas9 first searches for a short conserved sequence called the protospacer-associated motif (PAM) (9). Once Cas9 identifies a PAM, it then checks for complementarity between the target and the sgRNA. If the sequence is confirmed as a match, Cas9 cuts the target DNA, generating a blunt-ended double-stranded break.

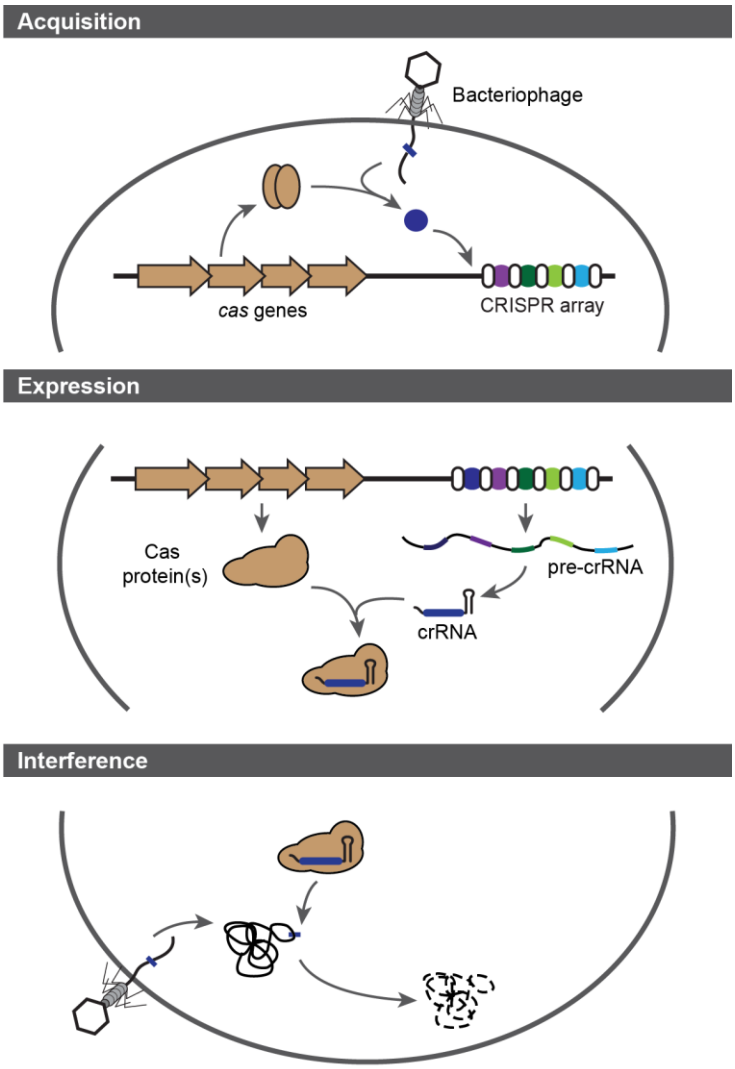

Figure 2.2 CRISPR-Cas systems provide bacteria with adaptive immunity. Immunity is conferred through three steps: acquisition, expression, and interference. Acquisition: Cas proteins incorporate a small piece of foreign DNA as a new spacer within the CRISPR array. Expression: The CRISPR array is transcribed and undergoes processing by the Cas protein(s) and accessory factors to form the CRISPR RNA (crRNA). Interference: The effector complex, comprised of the crRNA and Cas protein(s), targets and degrades invading DNA.

The double-stranded break alone does not result in editing. Editing events occur when the cell tries to repair the damage via either non-homologous end joining or homology-directed repair (10).

The non-homologous end-joining repair process is often error-prone. In an attempt to mend the gap in the DNA, the repair machinery will often generate small insertions or deletions in the affected area. These random mutations can inactivate genes, which can allow researchers to understand gene function.

Although knocking out genes in this way enables researchers to identify gene function, sometimes changes need to be better defined. Homology-directed repair can be used to generate more systematic changes. In this technique, a piece of donor DNA that carries the desired template is supplied. Once Cas9 makes a cut, this donor DNA acts as a bandage, pairing up with the cut ends and replacing the original sequence with the new version. The ability to introduce very controlled mutations would allow scientists to treat genetic disease by fixing harmful mutations or by replacing a defective gene with a healthy copy.

While there have been several attempts over the years to modify genomes and control gene function, Cas9 in particular holds great promise for genome engineering for several reasons.

First, the system is very compact. Introducing components into organisms is non-trivial. As the possibilities of genome engineering expand to include virtually any organism, the fewer parts that have to made, the simpler the technology. With the engineered sgRNA, the components simply include one protein and one RNA molecule.

Second, the system is highly programmable. CRISPR-Cas systems naturally rely on variable crRNA sequences to confer immunity. Researchers can design their own sgRNAs to target almost any desired sequence. The one limitation is that a target site must be adjacent to a PAM sequence. Luckily, these sequences are short and can often be found widely dispersed throughout the genome.

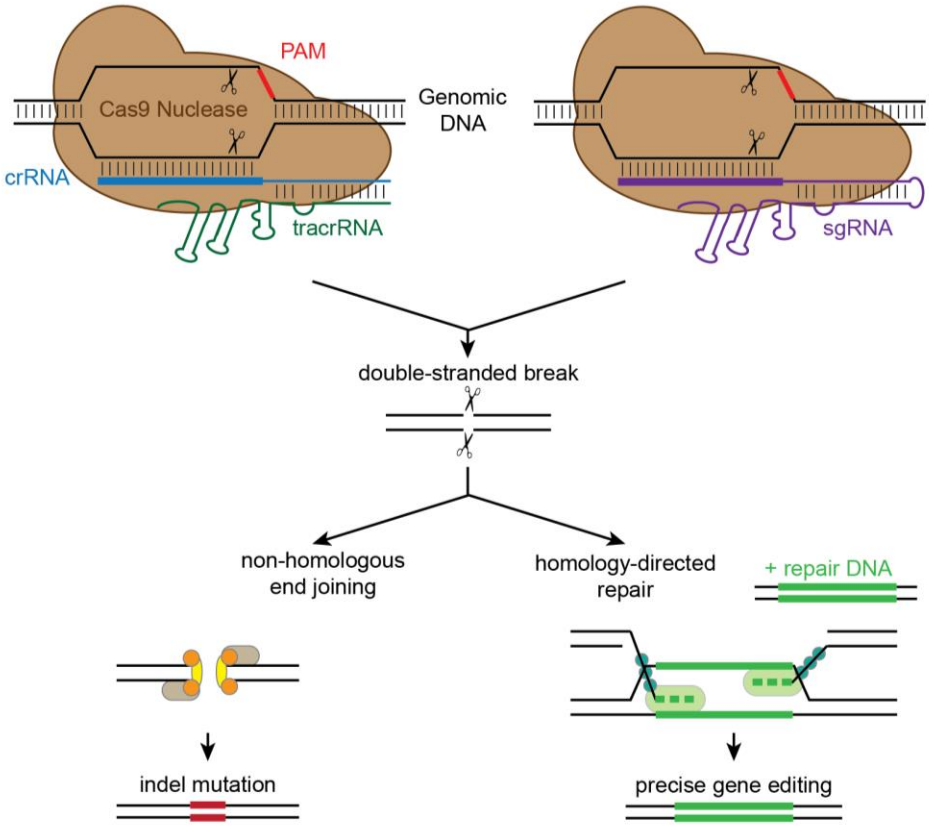

Figure 2.3 The CRISPR-Cas9 mechanism of action. Cas9 naturally relies on two pieces of short RNA: crRNA and tracrRNA. The engineered sgRNA combines these two elements into a single RNA molecule. In order to find the target, Cas9 scans the DNA for a sequence termed the protospacer adjacent motif (PAM). The Cas9 then unzips the DNA and determines whether the sgRNA matches the target DNA. If they do match, the Cas9 generates a double-strated DNA break. When this occurs, the cell tries to repair the cut.

Third, the system can be multiplexed. Because CRISPR-Cas systems naturally utilize arrays composed of multiple crRNAs, these systems are well-equipped to target numerous sequences simultaneously. By supplying multiple sgRNAs, multiple sequences can be targeted in order to study complex interactions or to cure human diseases associated with multiple mutations.

Lastly, the system is incredibly easy to design. Previous methods to edit genomes—transcription activator-like effector nucleases (TALENs), zinc-finger nucleases (ZFNs) and meganucleases—have relied on complex and sometimes proprietary protein design (11,12).

Because CRISPR-Cas systems rely on RNA molecules to direct activity, nucleic acid sequences can be readily and cheaply designed.

Because of the many potential benefits of CRISPR, researchers have been doing a lot to further their development as genetic tools. We can now make specific deletions, insertions, and modifications to genes. Moreover, we can also control gene expression by regulating transcription and modifying epigenetic markers.

Controlling gene expression requires inactivating the cleavage activity of the system. With Cas9, this is accomplished by mutating two catalytic sites that each cut one strand of the DNA target (13). The resulting dead Cas9 (dCas9) is able to bind to DNA but cannot cleave it. Directing dCas9 to a promoter or open reading frame represses transcription, providing a simple way to turn off expression of any gene. Similar results have also been achieved using other Cas proteins (14,15).

dCas9 can also be fused to effector domains to improve gene repression of even achieve gene activation (16–18). Additionally, epigenetic modifiers fused to dCas9 can result in DNA acetylation or methylation for targeted epigenetic regulation (19). Although the CRISPR-Cas9 nuclease offers a powerful means to permanently alter the genomic sequence, the deactivated dCas9 allows for transient and more nuanced control of gene expression.

2.4 CRISPR-Cas9 applications

All of these techniques have a number of applications, with genome engineering being the most prominently featured. Although the principle of genome editing has been around for several decades, CRISPR-Cas9 represents the latest iteration of increasingly sophisticated technologies. After the first CRISPR-Cas9 genome editing demonstrations in 2013 (20–22), it has since been established in a number of other organisms: plants, worms, flies, fish, mice, pigs, dogs, and primates, among others (23). We have even edited human cells.

For such a young technology, CRISPR-Cas9 has already had a number of successful demonstrations for correcting genetic disease. Scientists have reversed cataract formation in mice (24) and excised latent HIV from human cells (25,26). In addition to actively treating

disease, genome editing has enabled researchers to better study disease mechanisms. Scientists can use CRISPR-Cas9 to develop animal models with specific genetic mutations that mimic human disease. This has been particularly useful in modeling diseases such as cancer that often have complex genomic aberrations (27). Finally, CRISPR is proving useful for high-throughput screening in order to identify new drug targets (28–30).

Outside of medicine, CRISPR-Cas9 gene editing has been a game-changer for agriculture. The first plants featured in the slew of CRISPR-Cas9 genome engineering papers were *Arabidopsis* and tobacco – two common laboratory plants (31,32). CRISPR-Cas9 has now been extended to include many crop varieties such as wheat, rice, soybeans, potatoes, oranges and tomatoes (33). These crops incorporate changes that boost resistance to pests, fungal disease, and drought. Excitingly, an anti-browning mushroom and a waxy corn engineered with CRISPR-Cas9 have also recently received U.S. Food and Drug Administration (FDA) approval (34), paving the way for CRISPR-edited foods to soon appear on grocery shelves (35).

While CRISPR-Cas9 has undoubtedly simplified gene editing, there are still many technical challenges. First, a pervading safety concern is the likelihood of Cas9 accidentally cutting an unintended site in the genome—what is called on off-target effect. Since the early reports of off-target effects (36), rules have been developed for sgRNA design that effectively minimize these events from occurring (37).

Second, researchers are still working on efficiently delivering the CRISPR machinery to the target of interest. Cas9 is immensely powerful at editing once it has access to DNA, but getting inside cells is not trivial, especially when we move up in complexity from cell culture to whole organisms. For gene therapy, both viral and non-viral delivery vehicles have been investigated (38). These are being explored not only for their ability to deliver their CRISPR cargo to specific locations and cells, but also for their ability to limit off-target effects and an immune response.

2.5 Bioethics of genome engineering

The enormous opportunity to transform society also presents important ethical questions that warrant discussion. In January 2015, a group of almost 20 academic scientists and bioethicists met in Napa, California to discuss the ethical implications of genome engineering. This conference bore striking resemblance to the famous Asilomar conference in 1975 on recombinant DNA, which gave us one of the most famous examples of scientific self-regulation.

Following the 2015 meeting, *Science* magazine published a perspective that made four recommendations on how the scientific community should proceed with CRISPR-Cas9 and genome editing technology (39). One recommendation asked that steps be taken to "strongly discourage…any attempts at germline genome modification for clinical application in humans, while societal, environmental, and ethical implications of such activity are discussed among scientific and governmental organizations."

The human germline consists of the cells that make up sperm and eggs. Editing the human germline is particularly contentious because these changes would be passed down from generation to generation. This stands in contrast to somatic (non-reproductive) cells.

As an example, let us consider cystic fibrosis, a fatal single-gene disorder that causes the body to produce thick, sticky mucus that primarily clogs the lungs and leads to infection. There is no cure. If an individual with cystic fibrosis had the genomic DNA in their lungs corrected to produce thin secretions, the alteration—a somatic mutation—would be limited to only that individual's DNA while their offspring would still inherit the cystic fibrosis variant gene. However, if the same mutation were made to the germline, all subsequent generations of offspring would have the healthy copy of the gene—as well as any unintended side-effects. Given the novelty of this transformative technology, it is critical to fully understand all consequences before making changes that would affect future generations.

Another controversial issue is using CRISPR-Cas9 to alter wild populations by promoting inheritance of particular genes. This is accomplished by generating what is known as a gene drive (40). During normal sexual reproduction, each copy of a gene has a 50% change

of being inherited by a particular offspring. Gene drives bias this inheritance rate (Figure 2.4). If an organism contains one copy of a CRISPR-Cas9 gene drive, the drive will duplicate itself onto the other chromosome, ensuring 100% of the organism's offspring will inherit the gene drive. This has potential to eradicate vector-borne diseases or control invasive species, but involves intentionally releasing a genetically modified organism to modify a population's genetic makeup.

Before we take action, we must consider the ecological implications of altering populations, the back-up mechanisms needed to reverse a gene drive, and the path to regulatory and societal approval.

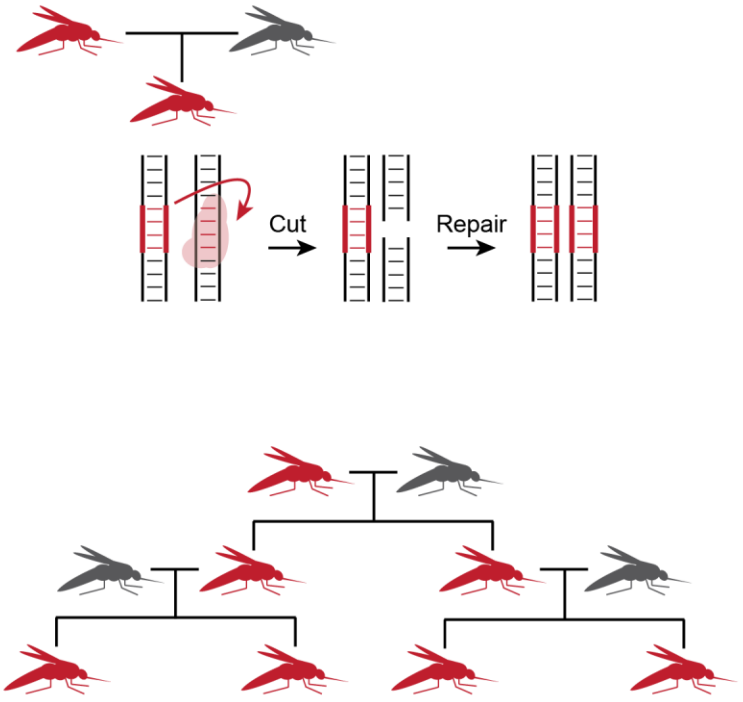

Figure 2.4 CRISPR-driven gene drives. During normal sexual reproduction, each copy of a gene has a 50% chance of being inherited by a particular offspring. Gene drives bias this inheritance rate. One copy of the drive converts the wild-type gene into the gene drive, such that the altered gene is almost always inherited. Over time and over multiple generations, this amplifies the copies of the altered gene.

2.6 Closing thoughts

Presently, the invention of CRISPR-Cas9 technology is the key question at the heart of a bitter patent battle (41). One side is led by Jennifer Doudna at the University of California, Berkeley, and Emmanuelle Charpentier, now at the Max Planck Institute for Infection Biology in Berlin and Umeå University in Sweden. The other side is headed by Feng Zhang of the Broad Institute and the Massachusetts Institute of Technology. While the Broad Institute currently holds the patent for the technology, the US Patent and Trademark Office agreed to hear patent interference proceedings in January 2016. The outcomes of these proceedings are undoubtedly worth millions to the research institutions involved. Furthermore, the commercialization of CRISPR-Cas9 has birthed a number of biotechnology startups, who all have a vested interest in the final outcome of this case.

While parties are warring over CRISPR-Cas9 patents, other CRISPR discoveries and innovations have continued to surface. Recently, a new Cas protein called Cpf1 has risen to prominence that may have improved performance over Cas9 (42). In contrast to Cas9, Cpf1 can readily process multiple crRNAs from the CRISPR array and creates cleavage products with ‘sticky ends,’ which may allow for easier modification of targeted DNA.

Regardless of who comes out on top of the patent battle or whether other proteins supplant Cas9 as the go-to gene editing device, these discoveries are a win for science, providing researchers with an ever-expanding toolkit. Keep in mind that engineered CRISPR-Cas systems are a fledgling technology that is only a few years old. There is still much more ground to be covered and more opportunities to be exploited. Stay tuned.

REFERENCES

1. Makarova, K. S. *et al.* An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.* **13,** 722–736 (2015).

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

3. Doudna, J. A. & Charpentier, E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346,** 1258096 (2014).

4. Dominguez, A. A., Lim, W. A. & Qi, L. S. Beyond editing: repurposing CRISPR-Cas9 for precision genome regulation and interrogation. *Nat. Rev. Mol. Cell Biol.* **17,** 5–15 (2016).

5. Wright, A. V., Nuñez, J. K. & Doudna, J. A. Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. *Cell* **164,** 29–44 (2016).

6. Wang, H., La Russa, M. & Qi, L. S. CRISPR/Cas9 in Genome Editing and Beyond. *Annu. Rev. Biochem.* **85,** 227–64 (2016).

7. Hsu, P. D., Lander, E. S. & Zhang, F. Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell* **157,** 1262–1278 (2014).

8. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (80-.).* **337,** 816–21 (2012).

9. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155,** 733–40 (2009).

10. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8,** 2281–308 (2013).

11. Gaj, T., Gersbach, C. a & Barbas, C. F. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.* **31,** 397–405 (2013).

12. Kim, H. & Kim, J.-S. A guide to genome engineering with programmable nucleases. *Nat. Rev. Genet.* **15,** 321–34 (2014).

13. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152,** 1173–1183 (2013).

14. Luo, M. L., Mullis, a. S., Leenay, R. T. & Beisel, C. L. Repurposing endogenous type I CRISPR-Cas systems for programmable gene repression. *Nucleic Acids Res.* **43,** 674–681 (2014).

15. Rath, D., Amlinger, L., Hoekzema, M., Devulapally, P. R. & Lundgren, M. Efficient programmable gene silencing by Cascade. *Nucleic Acids Res.* **43,** 237–246 (2015).

16. Maeder, M. L. *et al.* CRISPR RNA-guided activation of endogenous human genes. *Nat. Methods* **10,** 977–979 (2013).

17. Gilbert, L. A. *et al.* CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. *Cell* **154,** 442–451 (2013).

18. Bikard, D. *et al.* Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res.* **41,** 7429–7437 (2013).

19. Thakore, P. I., Black, J. B., Hilton, I. B. & Gersbach, C. A. Editing the epigenome: technologies for programmable transcription and epigenetic modulation. *Nat. Methods* **13,** 127–37 (2016).

20. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339,** 819–823 (2013).

21. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339,** 823–826 (2013).

22. Jinek, M. *et al.* RNA-programmed genome editing in human cells. *Elife* **2,** e00471 (2013).

23. Sander, J. D. & Joung, J. K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* (2014). doi:10.1038/nbt.2842

24. Wu, Y. *et al.* Correction of a genetic disease in mouse via use of CRISPR-Cas9. *Cell Stem Cell* **13,** 659–62 (2013).

25. Ebina, H., Misawa, N., Kanemura, Y. & Koyanagi, Y. Harnessing the CRISPR/Cas9 system to disrupt latent HIV-1 provirus. *Sci. Rep.* **3,** 2510 (2013).

26. Hu, W. *et al.* RNA-directed gene editing specifically eradicates latent and prevents new HIV-1 infection. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 11461–11466 (2014).

27. Sánchez-Rivera, F. J. & Jacks, T. Applications of the CRISPR-Cas9 system in cancer biology. *Nat. Rev. Cancer* **15,** 387–95 (2015).

28. Shalem, O., Sanjana, N. E. & Zhang, F. High-throughput functional genomics using CRISPR-Cas9. *Nat. Rev. Genet.* **16,** 299–311 (2015).

29. Zhou, Y. *et al.* High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature* **509,** 487–91 (2014).

30. Shalem, O. *et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343,** 84–87 (2014).

31. Li, J.-F. *et al.* Multiplex and homologous recombination-mediated genome editing in Arabidopsis and Nicotiana benthamiana using guide RNA and Cas9. *Nat. Biotechnol.* **31,** 688–91 (2013).

32. Nekrasov, V., Staskawicz, B., Weigel, D., Jones, J. D. G. & Kamoun, S. Targeted mutagenesis in the model plant Nicotiana benthamiana using Cas9 RNA-guided endonuclease. *Nat. Biotechnol.* **31,** 691–3 (2013).

33. Bortesi, L. & Fischer, R. The CRISPR/Cas9 system for plant genome editing and beyond. *Biotechnol. Adv.* **33,** 41–52 (2015).

34. Waltz, E. CRISPR-edited crops free to enter market, skip regulation. *Nature biotechnology* **34,** 582 (2016).

35. National Academies of Sciences, Engineering, and Medicine**. "**Genetically Engineered Crops: Experiences and Prospects." Washington, DC, The National Academies Press (2016) doi:10.17226/23395

36. Fu, Y. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* **31,** 822–6 (2013).

37. Graham, D. B. & Root, D. E. Resources for the design of CRISPR gene editing experiments. *Genome Biol.* **16,** 260 (2015).

38. Nelson, C. E. & Gersbach, C. A. Engineering Delivery Vehicles for Genome Editing. *Annu. Rev. Chem. Biomol. Eng.* **7,** 637–62 (2016).

39. Baltimore, D. *et al.* A prudent path forward for genomic engineering and germline gene modification. *Science* **348,** 36 (2015).

40. Esvelt, K. M., Smidler, a. L., Catteruccia, F. & Church, G. M. Concerning RNA-guided gene drives for the alteration of wild populations. *Elife* **3,** e03401 (2014).

41. Ledford, H. Bitter fight over CRISPR patent heats up. *Nature* **529,** 265 (2016).

42. Zetsche, B. *et al.* Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163,** 759–71 (2015).

CHAPTER 3

Repurposing endogenous Type I CRISPR-Cas systems for programmable gene repression

Michelle L. Luo, Adam S. Mullis, Ryan T. Leenay, and Chase L. Beisel

Department of Chemical and Biomolecular Engineering, North Carolina State University, Raleigh, NC 27695

Original publication in: *Nucleic Acids Research* 43(1): 674-681, 2015.

ABSTRACT

CRISPR-Cas systems have shown tremendous promise as heterologous tools for genome editing and transcriptional regulation. Because these RNA-directed immune systems are found in most prokaryotes, an opportunity exists to harness the endogenous systems as convenient tools in these organisms. Here, we report that the Type I-E CRISPR-Cas system in *Escherichia coli* can be co-opted for programmable transcriptional repression. We found that deletion of the signature *cas3* gene converted this immune system into a programmable gene regulator capable of reversible gene silencing of heterologous and endogenous genes. Targeting promoter regions yielded the strongest repression, whereas targeting coding regions showed consistent strand bias. Furthermore, multi-targeting CRISPR arrays could generate complex phenotypes. This strategy offers a simple approach to convert many endogenous Type I systems into transcriptional regulators, thereby expanding the available toolkit for CRISPR-mediated genetic control while creating new opportunities for genome-wide screens and pathway engineering.

3.1 INTRODUCTION

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)-Cas (CRISPR-associated) systems provide prokaryotes with adaptive immunity against foreign invaders (1, 2). Recognition of these invaders is conducted by CRISPR RNAs (crRNAs) that direct Cas proteins to cleave complementary nucleic acid sequences. The crRNAs are processed from transcribed arrays of identical repeats and intervening target-specific spacers, conferring immunity against multiple unique sequences. To manage the diversity of proteins associated with CRISPR-Cas systems, three general types have been defined each with a collection of subtypes (3). Apart from the particular suite of Cas proteins, these types can be distinguished by the molecular target, the mode of target recognition, and the mechanism of crRNA processing. For instance, Type I systems target DNA with a protospacer-adjacent motif (PAM) flanking the 3' end of the target sequence and rely on the Cas proteins for crRNA processing (4, 5), whereas Type II systems target DNA with a PAM flanking the 5' end of the target sequence and rely on a tracrRNA and RNase III for crRNA processing (6, 7). Of the three types, Type I systems are the most prevalent in both bacteria and archaea (3).

A defining feature of CRISPR-Cas systems is that crRNAs can be readily designed to guide the specific targeting of virtually any sequence. Aside from genome editing and DNA imaging, this capability has opened widespread opportunities in programmable gene regulation (8–11). Towards this goal, Type II systems were recently engineered to bind but not cleave target DNA through point mutations in the endonucleolytic domains of the signature *cas9* gene (12). Designed crRNAs directed the resulting catalytically dead Cas9 (dCas9) protein to bind specific promoters and coding regions, thereby modulating the recruitment of or extension by RNA polymerase (13–17). Separately, Type III-B systems, the only CRISPR-Cas systems known to naturally target RNA (18), can be directed to cleave chromosomal mRNAs (19, 20). While these efforts demonstrated the capacity of Type II and III systems for gene regulation, what remains unexplored is the capacity of Type I systems to exhibit this same phenomenon. If confirmed, this prevalent type could be harnessed to advance genetic control with CRISPR-Cas systems and provide insights into their potential roles as natural gene regulators.

Type I CRISPR-Cas systems generally involve two protein elements for DNA targeting: Cascade and Cas3. Cascade, a multimeric complex of three to six different Cas proteins, is responsible for processing CRISPR arrays (4) and for binding target DNA sequences through PAM and protospacer recognition (4, 21–23). Cas3, the signature protein of Type I systems, is responsible for cleaving and degrading target DNA (21, 24–26). Recent biochemical studies of different Type I subtypes revealed that Cascade is a stable complex that recruits Cas3 only after DNA binding (21–23, 25, 27–29). Based on these insights, we hypothesized that removal of Cas3 from an endogenous CRISPR-Cas system would allow Cascade to tightly bind target DNA sequences without subsequent degradation. As a result, designed CRISPR arrays would be sufficient to direct targeted DNA binding, thereby blocking RNA polymerase recruitment or extension (13, 17). Using the Type I-E CRISPR-Cas system in *Escherichia coli* K-12 as a model, we found that deleting the *cas3* gene from the *E. coli* genome allowed the targeted and multiplexed regulation of gene expression using the endogenous CRISPR-Cas system. CRISPR arrays could also be generated for the coordinated silencing of multiple endogenous genes and the generation of complex phenotypes.

3.2 MATERIALS AND METHODS

3.2.1 Strains and plasmid construction

See Supplementary Table 3.S1 for a list of all *E. coli* K-12 strains used in this work. To generate BW25113 Δ*cas3*::*cat* and MG1655 Δ*cas3*::*cat*, the *cat* resistance cassette was PCR-amplified from the pKD3 plasmid (30) using oligonucleotides that append the synthetic constitutive promoter J23119 (BBa_J23119 in the registry for standard biological parts, www.partsregistry.org) (J23119-pKD3.for, J23119-pKD3.rev). Following a second PCR amplification to introduce homology arms (HR-cas3.for, HR-cas3.rev), the resulting PCR product was recombineered into NM500 by mini-λ-mediated recombination (31). The insertion replaced the native *cas3* gene and the native promoter for the Cascade operon with the *cat* cassette and the J23119 promoter. Successful recombination was verified by sequencing. P1 transduction was then used to transfer the *cat* cassette and the synthetic promoter into

BW25113 and into MG1655. Successful transduction was verified by PCR. To generate BW25113 Δ*cas3*, the *cat* cassette from BW25113 Δ*cas3*::*cat* was excised using the pCP20 plasmid as described previously (32). To generate NM500 *cas3*$^{+}$, the *cat* resistance cassette was PCR-amplified from the pKD3 plasmid using oligonucleotides that append the constitutive promoter J23119 (J23119-pKD3.for, HR-casA.rev). Following a second PCR amplification to introduce homology arms (HR-cas3.for, HR-casA.rev), the resulting PCR product was recombined into NM500. This NM500 *cas3*$^{+}$ strain replaces the native promoter for the Cascade operon with a constitutive promoter while retaining the native *cas3* gene. To generate BW25113 ΔCRISPR-Cas::*cat*, the *cat* resistance cassette was PCR-amplified from the pKD3 plasmid (HR-CRISPR.for, HR-cas3.rev), and recombineered into NM500, followed by P1 transduction into BW25113. This BW25113 ΔCRISPR-Cas eliminates the entire CRISPR locus as well as *cas3*, the Cascade operon, and the CRISPR1 locus.

See Supplementary Table 3.S2 for a list of all plasmids used in this work. The GFP reporter plasmids were based on the pUA66 plasmid (low-copy sc101 origin-of-replication) (33) and are reported in previous work (34). To construct the arabinose-inducible pcrRNA.ind plasmid (medium-copy pBR322 origin-of-replication), oligonucleotides were designed to encode a single repeat and a synthetic rho-independent terminator (BBa_B1006 in the registry for standard biological parts) (pcrRNA.ind.for, pcrRNA.ind.rev). These oligonucleotides were annealed, 5' phosphorylated using polynucleotide kinase (PNK), and ligated into the pBAD18 plasmid digested with KpnI-HF and HindIII-HF. To construct the constitutively expressed pcrRNA.con plasmid, oligonucleotides encoding the synthetic constitutive promoter J23119 (pcrRNA.con.for, pcrRNA.con.rev) were annealed, 5' phosphorylated with PNK, and ligated into the pcrRNA.ind plasmid digested with NsiI and NheI. The insertion replaced the *araC* gene and P_{araB} promoter with the synthetic constitutive promoter. To insert new repeat-spacer pairs into pcrRNA.con or pcrRNA.ind, oligonucleotides encoding the palindromic repeat and crRNA spacers were annealed, 5' phosphorylated with PNK, and ligated into either plasmid digested with KpnI and XhoI. See Supplementary Figure 3.S2 for an illustration of the cloning scheme.

All plasmid cloning was verified by sequencing. See Supplementary Table 3.S3 for a list of all oligonucleotides used in this work. All oligonucleotides were chemically synthesized by IDT. All enzymes were purchased from NEB.

3.2.2 Growth conditions

All strains were cultured in 14 ml round-bottom polypropylene tubes at 37°C and 250 RPM in up to 5 ml of LB medium (10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl) or M9 minimal medium (1X M9 salts, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 10 μg/ml thiamine) containing the indicated combination of 0.4% glycerol, 0.2% indicated sugar, and 0.2% casamino acids. All strains were plated on LB agar (LB medium with 1.2% agar) in 100x15mm polystyrene petri dishes. To maintain any plasmids, cells were cultured in liquid medium or on agar plates containing appropriate antibiotics at the following concentration: 50 μg/ml of ampicillin, 34 μg/ml of chloramphenicol, 50 μg/ml of kanamycin.

3.2.3 Spacer design

See Supplementary Table 3.S4 for a list of all of protospacers targeted in this work. Protospacers were selected by identifying a PAM (CTT, CCT, CAT, CTC located at the 3' end of the target sequence) for the Type I-E system in *E. coli* (23). Note that only CTT and CCT were used in this work based on our previous experience with these PAM sequences (35). The 32 nucleotides immediately downstream of the PAM were then used as the spacer. The cloning scheme required changing the final two nucleotides of the spacer to TC (Supplementary Figure 3.S2), which is not expected to impact crRNA activity (5).

3.2.4 Transformation assays

The transformation assay as shown in Supplementary Figure 3.S4B was conducted similar to previous work (35). Briefly, *E. coli* BW25113 Δ*cas3*::*cat* or NM500 *cas3*$^+$ cells harboring pUA66-lacZ were cultured overnight in LB medium. Cultures were back-diluted 1:25 into 25 ml of LB medium in 125 ml Erlenmeyer flasks and grown to an ABS_{600} of 0.6 –

0.8, which was quantified using a Nanodrop 2000c spectrophotometer (Thermo Scientific). The cells were then washed in ice-cold 10% glycerol and concentrated by a factor of ~100. A total of 50 µl of the concentrated cells were transformed with 50 ng of plasmid DNA using a MicroPulser electroporator (Bio-Rad). Transformed cells were recovered in 500 µl SOC medium for 1 hr at 37ºC. After the recovery period, the cells were diluted by factors of 10^4 – 10^6 and 250 µl of the dilution were plated on LB agar with appropriate antibiotics and inducers.

3.2.5 Flow cytometry analysis

Cells grown overnight in M9 minimal medium containing 0.2% casamino acids and 0.4% glycerol were back-diluted to an ABS_{600} of 0.01 into M9 minimal medium with the specified combination of 0.1 mM IPTG and 0.2% of the indicated inducing sugar. Upon reaching an ABS_{600} of ~0.2 after ~3-4 hours of growth, the cultures were diluted 1:100 in 1X PBS and run on an Accuri C6 Flow Cytometer (Becton Dickinson) equipped with CFlow plate sampler, a 488 nm laser, and a 530 ± 15 nm bandpass filter. Events reflecting cells were gated based on forward scatter (FSC-H) and side scatter (SSC-H) with respective lower cutoffs of 11,500 and 600 to reduce the measurement of particulates. The gate was set using *E. coli* cells stained with the DRAQ5 dye (Thermo Scientific). The fluorescence of the gated cells was then measured in FL1-H. At least 20,000 events were analyzed for each sample.

For the reversibility experiments, cells were grown overnight in M9 minimal medium containing 0.2% casamino acids, 0.4% glycerol, and 0.1 mM IPTG, with or without 0.2% L-arabinose. Overnight cultures were pelleted and resuspended twice in M9 minimal media with 0.2% casamino acids, 0.4% glycerol, and 0.1 mM IPTG to remove any residual L-arabinose. The washed cultures were then back-diluted to an ABS_{600} of ~0.001 in 30 ml of the same medium without or with 0.2% L-arabinose, respectively. Every hour, 800 µl of culture was withdrawn for flow cytometry analysis and measurement of the ABS_{600}.

3.2.6 Doubling-time measurements

Cells were grown overnight in M9 minimal medium with 0.4% glycerol. The overnight cultures were pelleted and resuspended twice in M9 minimal medium with no carbon source. The washed cultures were then back-diluted to an ABS_{600} of ~0.001 into 25 ml of M9 minimal medium containing 0.2% of the indicated sugar in 125 ml Erlenmeyer flasks. Every 30 minutes, 800 µl of culture was withdrawn for measurement of the ABS_{600}.

3.2.7 Quantitative real-time PCR

Cells were grown overnight in M9 minimal medium containing 0.2% casamino acids and 0.4% glycerol. Overnight cultures were back-diluted 1:250 in M9 minimal medium containing 0.2% casamino acids, 0.4% glycerol, and 0.2% of the indicated sugar. Once cultures reached an ABS_{600} of ~0.4, total RNA was isolated as reported previously (36) followed by treatment with DNase I. cDNAs were generated from 2 µg of the resulting RNA using random primers and SuperScript III reverse-transcriptase (Invitrogen) followed by treatment with RNase H. Quantitative PCR was conducted on cDNA samples using SYBR Green (Bio-Rad) and the gene-specific primers (X-qPCR.fwd/rev, where X is the target gene) listed in Supplementary Table 3.S3. cDNAs were run on a Mastercycler ep realplex2 real-time PCR system (Eppendorf) according to the manufacturer's instructions. For the PCR runs, each cDNA was heated to 95ºC for 2 minutes followed by 50 cycles of a 15-second denaturing step at 95ºC, a 15-second annealing step at 55ºC, and a 30-second extension step at 72ºC. At the end of the run a melt curve was generated to ensure the absence of non-specific products. Relative quantitation of gene expression was calculated using the ΔC_t method.

3.2.8 Growth assays

Cells were inoculated into M9 minimal medium containing 0.4% glycerol and grown overnight. After 24 hours, cells were pelleted and resuspended in 2 ml of M9 minimal medium with no carbon source two times to remove glycerol as a possible source of growth. The washed cultures were then back-diluted to an ABS_{600} of 0.001 into 2 ml of M9 minimal medium

containing 0.2% of the indicated sugar(s). Finally, the cultures were grown for 24 hours until the ABS_{600} was measured.

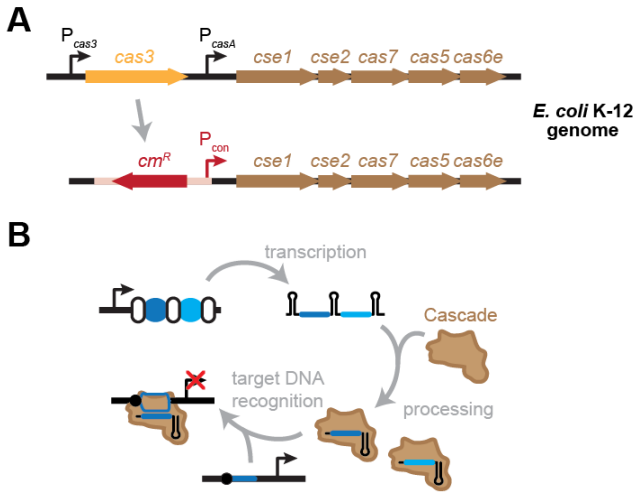

Figure 3.1 Repurposing the Type I-E CRISPR-Cas in *E. coli* K-12 for programmable gene repression. Conversion of the Type I-E CRISPR-Cas system into a programmable repressor. The deletion of *cas3* and insertion of a constitutive promoter upstream of the Cascade operon allows crRNA-directed DNA binding without cleavage. (B) Putative mechanism of crRNA-directed gene repression. Cascade processes the transcribed CRISPR array into individual crRNAs. The Cascade–crRNA complex then binds target DNA sequences (blue line) flanked by a PAM (black circle), leading to transcriptional repression.

3.3 RESULTS

3.3.1 Targeted gene repression following deletion of cas3

To explore the capacity of Type I systems for gene regulation, we employed the Type I-E CRISPR-Cas system in *Escherichia coli* K-12 (Supplementary Figure 3.S1), the best characterized Type I system in a genetically tractable bacterium (4, 5, 21, 23). Because the operon encoding Cascade (*cse1-cse2-cas7-cas5-cas6e*) is strongly repressed under normal growth conditions (37, 38), we replaced *cas3* and the native *cse1* promoter with a constitutive promoter in one round of homologous recombination (Figure 3.1). The resulting strain (BW25113 Δ*cas3*::*cat*) was transformed with a medium-copy plasmid encoding L-arabinose-inducible single-spacer arrays (Supplementary Figure 3.S2) and a low-copy reporter plasmid

encoding the green fluorescent protein (*gfp*) gene downstream of the *lacZ* promoter (pUA66-lacZ, Supplementary Table 3.S5). The spacers were designed to target ten locations in the promoter and *gfp* coding region as well as two locations far upstream of the promoter (Figure 3.2A, Supplementary Table 3.S4). Using flow cytometry analysis, the fluorescence of individual cells was then measured following induction of GFP and crRNA expression.

In comparison to the spacer-free plasmid (pCRISPR.ind, Supplementary Figure 3.S2), we observed ranging extents of repression that depended on which region of pUA66-lacZ was targeted (Figure 3.2B). Targeting either strand of the promoter region strongly reduced GFP fluorescence (~200-fold). Targeting the transcribed region moderately reduced GFP fluorescence, but only when targeting anywhere along the non-template strand or in the vicinity of the RNA polymerase footprint on the template strand (39). Interestingly, the strand bias observed when targeting the template versus non-template strand mirrored that observed for dCas9 in bacteria (13, 17). As expected, targeting upstream of the promoter region negligibly reduced fluorescence. In all cases, the extent of gene silencing was uniform across the entire bacterial population (Supplementary Figure 3.S3). Importantly, GFP levels were similar for the no-spacer plasmid and a plasmid encoding a spacer targeting the *mviM* gene in Salmonella enterica (Figure 3.2B), ruling out potential differences due to the assembly of Cascade. We also found that GFP silencing was reversible based on the change in fluorescence following addition or removal of L-arabinose (Figure 3.2C). The associated dynamics can be attributed to the stability of GFP similar to previous work (13).

We next performed a series of control experiments to assess the impact of deleting *cas3* and constitutively expressing the Cascade operon. We first measured GFP fluorescence in the original wild-type strain in which the Cascade operon was tightly repressed and *cas3* was still present (BW25113) and in a strain in which *cas3* and the Cascade operon were both deleted (BW25113 ΔCRISPR-Cas::*cat*). The fluorescence levels were similar regardless of whether a targeting or non-targeting spacer was used (Supplementary Figure 3.S4A), indicating that Cascade must be present for gene silencing. Next, to assess the impact on DNA integrity, we

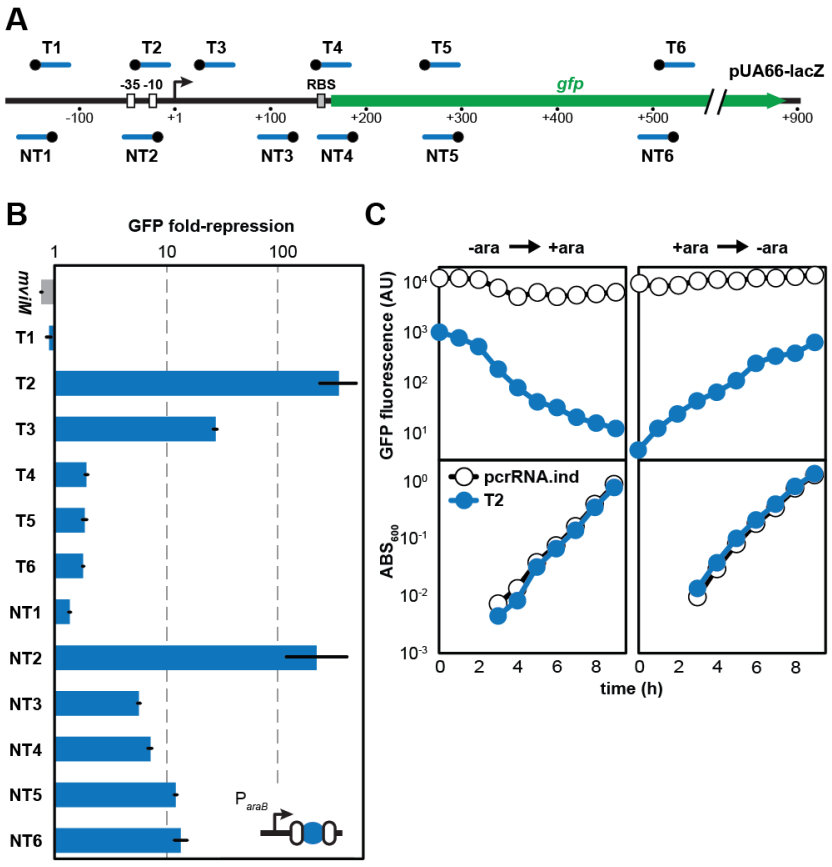

Figure 3.2 RNA-mediated transcriptional repression with the repurposed Type I-E CRISPR-Cas system in *E. coli* K-12. (A) Targeted silencing of plasmid-based GFP expression. The *gfp* gene is under the control of the *lacZ* promoter in the low-copy plasmid pUA66-lacZ. Each spacer sequence (blue line) and PAM (black circle) match the closest strand of the protospacer. RBS, ribosome-binding site. (B) Location-dependent and strand-dependent repression of GFP expression. BW25113 Δ*cas3*::*cat* harboring the medium-copy pUA66-lacZ and the indicated single-spacer plasmid were subjected to flow cytometry analysis following induction with IPTG and L-arabinose. The non-targeting spacer serves as a negative control. Repression is calculated as the ratio of the autofluorescence-subtracted fluorescence for the inducible no-spacer plasmid (pcrRNA.ind) and each single-spacer plasmid. See Supplementary Figure 3.S3A for representative histograms from the flow cytometry analysis. (C) Reversibility of gene silencing. BW25113 Δ*cas3*::*cat* cells harboring pUA66-lacZ and either the no-spacer plasmid (pcrRNA.ind, white circles) or the T2 single-spacer plasmid (T2, blue circles) either were pre-induced with only IPTG and switched to both IPTG and L-arabinose (left) or were pre-induced with both IPTG and L-arabinose and switched to only IPTG (right). Following the addition or removal of L-arabinose at t = 0, the autofluorescence-subtracted fluorescence for individual cells and the turbidity of the culture were followed over time. GFP fluorescence was ~10-fold lower for cells with the targeting plasmid versus the spacer-free plasmid, which we attribute to leaky expression from the P_{araB} promoter under these growth conditions.

measured the transformation efficiencies for targeting and non-targeting plasmids in strains with Cascade constitutively expressed and *cas3* present (NM500 *cas3*$^{+}$) or absent (BW25113 Δ*cas3*::*cat*). Surprisingly, we observed similar transformation efficiencies for the targeting and non-targeting plasmids even when *cas3* was present (Supplementary Figure 3.S4B), suggesting that Cas3 is poorly expressed or inactive in this particular strain. As further support, the strain with *cas3* present could still strongly silence GFP (Supplementary Figure 3.S4C). Finally, to gauge the impact of the resistance cassette, we excised the cassette used to delete *cas3* and measured gene silencing. The resulting strain (BW25113 Δ*cas3*) and the original strain (BW25113 Δ*cas3*::*cat*) exhibited similar silencing efficiencies (Supplementary Figure 3.S4C), indicating a negligible impact of the resistance cassette.

3.3.2 Impact of array length and spacer position

One beneficial feature of Cascade is that it can process multiple crRNAs from a single spacer array. However, little is known about how the composition of natural or synthetic multi-spacer arrays quantitatively impacts individual targets. To evaluate the impact of array length, we generated arrays with one promoter-targeting spacer (T2) followed by zero to three non-targeting spacers (*mviM*) (Figure 3.3). Flow cytometry analysis revealed a gradual decrease in silencing efficiency with each additional spacer. We speculate that this decrease may be due to non-targeting spacers diluting available Cascade complexes for targeting crRNAs, as observed with other RNA-based systems (40). In support of this assertion, the single-spacer array and an array of four targeting spacers exhibited statistically indistinguishable extents of silencing (two-tailed t-test, $t(4) = 1.05$, $p = 0.35$) (Figure 3.3). To evaluate the impact of spacer position, we generated arrays with different permutations of one targeting and three non-targeting spacers (Figure 3.3). With the exception of a targeting spacer in the first position of the four-spacer array, the extent of *gfp* silencing was similar regardless of spacer position (one-way ANOVA, $F(2,6) = 0.15$, $p = 0.86$). These results suggest that longer arrays can reduce the potency of individual spacers, whereas the exact location of a spacer within an array has a lesser contribution to the potency of silencing.

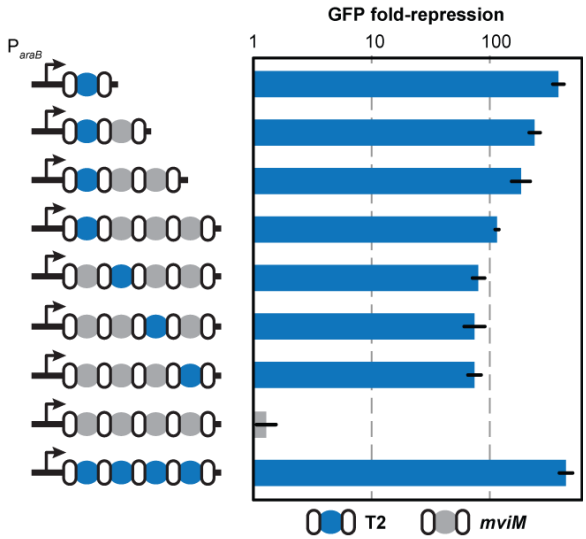

Figure 3.3 Impact of array length and spacer location on silencing efficiency. BW25113 Δ*cas3* cells harboring pUA66-lacZ and the indicated inducible CRISPR array plasmid were subjected to flow cytometry analysis following induction with IPTG and L-arabinose. Repression is calculated as the ratio of the autofluorescence-subtracted fluorescence for the inducible no-spacer plasmid (pcrRNA.ind) and each multi-spacer plasmid. Repeats, white ovals; T2 spacers, blue circles; non-targeting spacers matching the *S. enterica mviM* gene, gray circles. See Supplementary Figure 3.S3B for representative histograms from the flow cytometry analysis. Values represent geometric mean and SEM from independent experiments staring with three separate colonies.

3.3.3 Multiplexed repression of endogenous genes

As a complement to targeting heterologous genes such as *gfp*, we explored the ability of spacers to regulate endogenous targets. We focused on operons involved in the catabolism of the sugars L-arabinose (*araBAD*), L-rhamnose (*rhaBAD*), D-xylose (*xylAB*), and D-lactose (*lacZYA*) (Figure 3.4A) because these operons are well characterized and are required for growth on their cognate sugar (41–44). As the *araBAD*, *rhaBAD*, and *lacZYA* operons are disrupted in BW25113, we imported the *cas3* deletion and synthetic promoter into another strain of *E. coli* K-12 (MG1655 Δ*cas3*::*cat*). We also placed each single-spacer array under the control of the strong, constitutive promoter J23119 to circumvent the need for L-arabinose as an inducer (Supplementary Figure 3.S2).

To assess silencing of promoter activity, we cloned the promoter of each operon upstream of *gfp* in the pUA66 plasmid (Supplementary Table 3.S5) and measured the ability of each spacer to repress its target promoter by flow cytometry analysis (Figure 3.4B,

Supplementary Figure 3.S3). In comparison to the spacer-free plasmid (pCRISPR.con, Supplementary Figure 3.S2), each targeting plasmid greatly reduced fluorescence (~80-fold to 900-fold). As expected, combining the spacers into one array strongly reduced fluorescence for each promoter (Figure 3.4B), although the degree of silencing was generally less than that observed for the individual spacers (Figure 3.3).

To evaluate silencing of the endogenous genes, we measured mRNA levels of each operon for cells with each single-spacer plasmid. In comparison to the no-spacer plasmid, the single-spacer plasmids greatly reduced mRNA levels (~11-fold to 2,200-fold) of the target operons (Figure 3.4C), paralleling that observed for the GFP reporters (Figure 3.4B). This wide range in repression matches the variability in gene silencing observed with dCas9 (13, 17).

Finally, we explored whether targeting endogenous genes could generate defined phenotypes. Because each operon is required for the catabolism of its cognate sugar, we measured growth on each sugar as well as on two non-targeted sugars D-glucose and N-acetyl-D-glucosamine (GlcNAc). We cultured MG1655 Δ*cas3*::*cat* expressing a single-spacer or four-spacer array with the different sugars as sole carbon sources and measured the turbidity of the culture after 24 hours of growth (Figure 3.4D, Supplementary Figure 3.S5A). We found that targeting each operon limited growth on the cognate sugar, whether using a single-spacer array or the four-spacer array. The four-spacer array silenced all target operons in individual cells, as this array limited growth in medium containing all four targeted sugars (Supplementary Figure 3.S5A). Critically, growth was unhampered for all non-targeted sugars, supporting the specificity of targeting. The final turbidity was generally lower for all cultures grown in L-rhamnose (Figure 3.4E, Supplementary Figure 3.S5A), which we attribute to L-rhamnose being a poor carbon source (Supplementary Figure 3.S5B). We thus conclude that the Type I-E system in *E. coli* can be programmed to silence multiple endogenous genes and generate complex phenotypes.

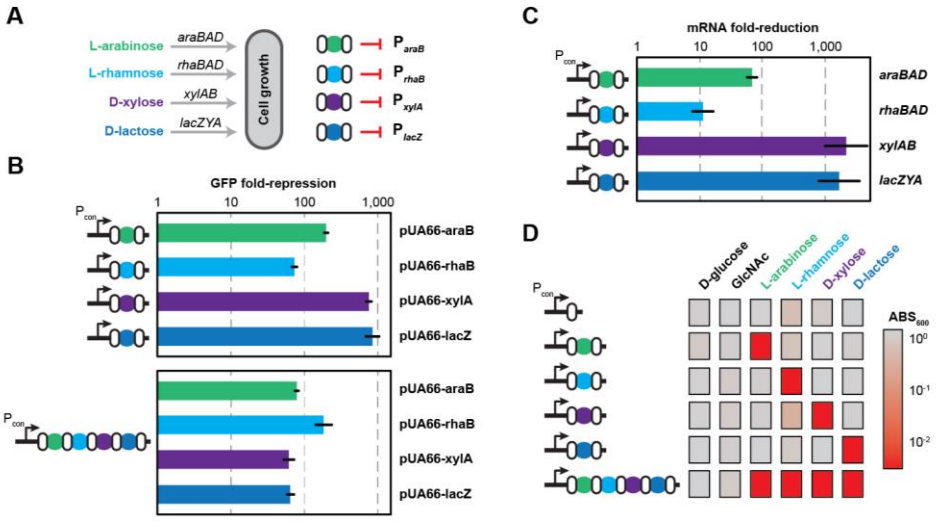

Figure 3.4 Targeted repression of endogenous genes and pathways. Spacers were designed to target the promoter of each catabolic operon required for growth on its cognate sugar. (B) Repression of promoter activity. Each promoter was cloned upstream of the *gfp* gene in pUA66. The resulting plasmids were then tested in MG1655 Δ*cas3*::*cat* cells harboring the corresponding single-spacer plasmid (top) or multi-spacer plasmid (bottom) by flow cytometry analysis following promoter induction with the cognate sugar. Repression is calculated as the ratio of the autofluorescence-subtracted fluorescence for the constitutive no-spacer plasmid (pcrRNA.con) and each single-spacer or multi-spacer plasmid. Values represent geometric mean and SEM from independent experiments with three colonies. See Supplementary Figure 3.S3C for representative histograms from the flow cytometry analysis. (C) Repression of endogenous genes. MG1655 Δ*cas3*::*cat* cells harboring the indicated single-spacer plasmid were harvested for total RNA following induction with the cognate sugar and subjected to qRT-PCR analysis. Repression is calculated as the ratio of the relative mRNA levels from the no-spacer plasmid (pcrRNA.con) and the indicated single-spacer plasmid. Values represent the geometric mean and SEM for quadruple technical replicates. (D) Targeted suppression of growth. MG1655 Δ*cas3*::*cat* cells harboring the indicated single-spacer or multi-spacer plasmid were grown on each sugar as the sole carbon source and turbidity was measured after 24 h of growth. Values represent the geometric mean of the measured ABS_{600} values from independent experiments starting with three separate colonies.

3.4 DISCUSSION

We found that the Type I-E CRISPR-Cas system in *E. coli* can be repurposed for programmable gene repression through the deletion of *cas3* and constitutive expression of the Cascade operon. An ensuing question is the extent to which this phenomenon applies to the other five Type I subtypes. Structural and phylogenetic data suggest that this same phenomenon would apply to Type I-B, I-C, and I-F systems based on the stability of Cascade in the absence of Cas3 and the ability of this complex to process transcribed CRISPR arrays (3, 22, 27, 29). Type I-A and I-B systems appear to be exceptions, as two distinct Cas3 proteins (Cas3' and Cas3") are required for stabilization of the Type I-A Cascade and the uncharacterized Type I-B Cascade is most closely related to Type I-A systems (1, 45). However, these *cas3* genes could be catalytically inactivated (24) as performed with Cas9 (12), albeit via point mutations that are harder to introduce with rudimentary genetic tools.

With this demonstration, another question is whether Type I systems or Type II systems should be employed for transcriptional regulation. Type II systems in the form of dCas9 are highly attractive because they offer a compact heterologous system that can be imported into diverse organisms. However, exploiting endogenous Type I systems does offer some potential advantages. For instance, once *cas3* is deleted, only the CRISPR array totaling at most a few hundred bases must be introduced. Another potential advantage is that the native Type I system would be well suited for thermophilic and hyperthermophilic microorganisms that thrive in environmental conditions that would prevent proper folding of common Cas9 proteins. Type I systems also offer PAMs that are distinct from those associated with known Type II systems, including a different orientation and a bias toward T/C-rich sequences (1, 23, 46). Finally, Type I systems are naturally found in diverse industrially and medically relevant strains, including *Escherichia coli*, *Streptococcus thermophilus*, *Clostridium autoethanogenum*, and *Acinetobacter baumannii* (47). A major drawback to this strategy is that the strains would lose immunity against some invading pathogens. Overexpression of Cascade in the absence of Cas3 may also inadvertently impact the transcriptional landscape, although this remains to be explored even for dCas9.

One interesting parallel observed for transcriptional regulation with Type I and Type II systems is the strand bias when targeting transcribed regions (Figure 3.2B) (13, 17). Previous work with dCas9 demonstrated that targeting the non-template strand but not the template strand strongly interfered with RNA polymerase extension. We observed the same trend with the Type I-E Cascade (Figure 3.2B) despite structural differences and opposing PAM locations in comparison to dCas9 (21, 23, 48, 49). Based on this parallel, we speculate that RNA polymerase extension is more sensitive to protein binding on the non-template strand rather than the particular orientation of the interfering protein or titration by the encoded mRNA. However, further investigation of the mechanisms of transcriptional repression is warranted.

An emerging concern with CRISPR technologies is the degree of off-target effects (50–53). These concerns stem from the Cas proteins accommodating mismatches between the crRNA spacer and the DNA target (5, 35, 48, 54), potentially recognizing similar sites elsewhere in the genome. While recent genome-wide screens in human embryonic stem cells failed to detect any unintended editing events with Cas9 (55, 56), off-targets would be expected to vary with the selected spacer sequence. Fortunately, using CRISPR-Cas systems for transcriptional regulation in prokaryotes would be far less likely to produce off-target effects: prokaryotes possess much smaller genomes than eukaryotes, limiting the probability of similar sequences appearing at other sites; and transcriptional repression can only occur within defined regions and strands of the genome (13, 17). Accordingly, only one off-target has been reported to-date for dCas9 in multiple studies in bacteria and in mammalian cells (13–15), wherein the off-target contained a recognized PAM and strong homology to the target sequence (17).

The regulatory capacity of the Type I-E system in the absence of *cas3* hints at the possibility of Type I systems naturally controlling gene expression. A previous bioinformatics search for genome-targeting spacers—a potential indicator of gene regulation—identified numerous instances in natural arrays (57). The authors concluded that accidental self-targeting forced deactivation of the endogenous CRISPR-Cas system because many of the *cas* genes were missing. However, accidental self-targeting could also drive the loss or disruption of Cas3, thereby converting the system into a gene regulator. The identification of such systems

would complement the single CRISPR-Cas system known to regulate cellular processes (58, 59).

In summary, our findings offer a novel strategy for exploiting an organism's native Type I CRISPR-Cas system for transcriptional regulation. In the future, we intend to explore additional Type I subtypes, broadening the scope of this method. Utilizing native Type I CRISPR-Cas systems would require elucidating the PAM, deleting *cas3*, and validating the functionality of Cascade (48). However, once achieved, these systems would further augment the genetic toolbox available for programmable gene regulation and offer novel approaches for genome-wide screens and strain engineering. Moreover, our findings provide a framework to identify natural Type I systems that naturally regulate gene expression, potentially expanding the list of known systems that orchestrate cellular processes (58, 59).

ACKNOWLEDGMENTS

We thank A. May and R. Barrangou for helpful discussions and critical reading of the manuscript. We thank S. Carrell and A. Jermusyk for technical assistance with qRT-PCR analysis and M. Shah for assistance with cloning and recombineering.

FUNDING

The work was made possible by the Kenan Institute of Engineering, Science & Technology and the National Science Foundation [CBET-1403135] (to C.L.B.); and the National Institutes of Health [5T32GM008776-15] (to R.T.L.). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

AUTHOR CONTRIBUTIONS

M.L.L. and C.L.B. designed the study and wrote the manuscript. M.L.L., A.S.M., and R.T.L. performed the experiments.

CONFLICT OF INTEREST STATEMENT

The authors declare a competing financial interest in the form of a filed patent application.

REFERENCES

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

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

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

4. Brouns, S.J.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J.H., Snijders, A.P.L., Dickman, M.J., Makarova, K.S., Koonin, E.V. and van der Oost, J. (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, **321**, 960–964.

5. Semenova, E., Jore, M.M., Datsenko, K.A., Semenova, A., Westra, E.R., Wanner, B., van der Oost, J., Brouns, S.J.J. and Severinov, K. (2011) Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl. Acad. Sci. U. S. A.*, **108**, 10098–10103.

6. Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J. and Charpentier, E. (2011) CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, **471**, 602–607.

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

8. Hsu, P.D., Lander, E.S. and Zhang, F. (2014) Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell*, **157**, 1262–1278.

9. Sander, J.D. and Joung, J.K. (2014) CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.*, **32**, 347–355.

10. Fineran, P.C. and Dy, R.L. (2014) Gene regulation by engineered CRISPR-Cas systems. *Curr. Opin. Microbiol.*, **18**, 83–89.

11. Bikard, D. and Marraffini, L.A. (2013) Control of gene expression by CRISPR-Cas systems. *F1000prime Rep.*, **5**, 47.

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

13. Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P. and Lim, W.A. (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, **152**, 1173–1183.

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

15. Perez-Pinera, P., Kocak, D.D., Vockley, C.M., Adler, A.F., Kabadi, A.M., Polstein, L.R., Thakore, P.I., Glass, K.A., Ousterout, D.G., Leong, K.W., et al. (2013) RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat. Methods*, **10**, 973–976.

16. Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S., Yang, L. and Church, G.M. (2013) CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.*, **31**, 833–8.

17. Bikard, D., Jiang, W., Samai, P., Hochschild, A., Zhang, F. and Marraffini, L.A. (2013) Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res.*, **41**, 7429–37.

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

19. Zebec, Z., Manica, A., Zhang, J., White, M.F. and Schleper, C. (2014) CRISPR-mediated targeted mRNA degradation in the archaeon *Sulfolobus solfataricus*. *Nucleic Acids Res.*, **42**, 5280–5288.

20. Hale, C.R., Majumdar, S., Elmore, J., Pfister, N., Compton, M., Olson, S., Resch, A.M., Glover, C.V.C., 3rd, Graveley, B.R., Terns, R.M., et al. (2012) Essential features and rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs. *Mol. Cell*, **45**, 292–302.

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

22. Nam, K.H., Haitjema, C., Liu, X., Ding, F., Wang, H., DeLisa, M.P. and Ke, A. (2012) Cas5d protein processes pre-crRNA and assembles into a Cascade-like interference complex in subtype I-C/Dvulg CRISPR-Cas system. *Struct. Lond. Engl. 1993*, **20**, 1574–1584.

23. Westra, E.R., van Erp, P.B.G., Künne, T., Wong, S.P., Staals, R.H.J., Seegers, C.L.C., Bollen, S., Jore, M.M., Semenova, E., Severinov, K., et al. (2012) CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Mol. Cell*, **46**, 595–605.

24. Sinkunas, T., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P. and Siksnys, V. (2011) Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. *EMBO J.*, **30**, 1335–1342.

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

26. Jackson, R.N., Lavin, M., Carter, J. and Wiedenheft, B. (2014) Fitting CRISPR-associated Cas3 into the helicase family tree. *Curr. Opin. Struct. Biol.*, **24**, 106–114.

27. Brendel, J., Stoll, B., Lange, S.J., Sharma, K., Lenz, C., Stachler, A.-E., Maier, L.-K., Richter, H., Nickel, L., Schmitz, R.A., et al. (2014) A Complex of Cas Proteins 5, 6, and 7 Is Required for the Biogenesis and Stability of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-derived RNAs (crRNAs) in *Haloferax volcanii*. *J. Biol. Chem.*, **289**, 7164–7177.

28. Richter, C., Gristwood, T., Clulow, J.S. and Fineran, P.C. (2012) *In vivo* protein interactions and complex formation in the *Pectobacterium atrosepticum* subtype I-F CRISPR/Cas System. *PloS One*, **7**, e49549.

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

30. Datsenko, K.A. and Wanner, B.L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.*, **97**, 6640–6645.

31. Court, D.L., Swaminathan, S., Yu, D., Wilson, H., Baker, T., Bubunenko, M., Sawitzke, J. and Sharan, S.K. (2003) Mini-lambda: a tractable system for chromosome and BAC engineering. *Gene*, **315**, 63–69.

32. Cherepanov, P.P. and Wackernagel, W. (1995) Gene disruption in *Escherichia coli*: Tc^R and Km^R cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. *Gene*, **158**, 9–14.

33. Zaslaver, A., Bren, A., Ronen, M., Itzkovitz, S., Kikoin, I., Shavit, S., Liebermeister, W., Surette, M.G. and Alon, U. (2006) A comprehensive library of fluorescent transcriptional reporters for *Escherichia coli*. *Nat. Methods*, **3**, 623–628.

34. Afroz, T., Biliouris, K., Kaznessis, Y. and Beisel, C.L. (2014) Bacterial sugar utilization gives rise to distinct single-cell behaviours. *Mol. Microbiol.*, **93**, 1093–1103.

35. Gomaa, A.A., Klumpe, H.E., Luo, M.L., Selle, K., Barrangou, R. and Beisel, C.L. (2014) Programmable removal of bacterial strains by use of genome-targeting CRISPR-Cas systems. *mBio*, **5**, e00928–00913.

36. Stead, M.B., Agrawal, A., Bowden, K.E., Nasir, R., Mohanty, B.K., Meagher, R.B. and Kushner, S.R. (2012) RNAsnap^{TM}: a rapid, quantitative and inexpensive, method for isolating total RNA from bacteria. *Nucleic Acids Res.*, **40**, e156.

37. Pul, U., Wurm, R., Arslan, Z., Geissen, R., Hofmann, N. and Wagner, R. (2010) Identification and characterization of *E. coli* CRISPR-*cas* promoters and their silencing by H-NS. *Mol. Microbiol.*, **75**, 1495–1512.

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

39. Kovacic, R.T. (1987) The 0 °C closed complexes between *Escherichia coli* RNA polymerase and two promoters, T7-A3 and *lac*UV5. *J. Biol. Chem.*, **262**, 13654–13661.

40. Hussein, R. and Lim, H.N. (2011) Disruption of small RNA signaling caused by competition for Hfq. *Proc. Natl. Acad. Sci. U. S. A.*, **108**, 1110–1115.

41. Gross, J. and Englesberg, E. (1959) Determination of the order of mutational sites governing L-arabinose utilization in *Escherichia coli* B/r bv transduction with phage Plbt. *Virology*, **9**, 314–331.

42. Power, J. (1967) The L-rhamnose genetic system in *Escherichia coli* K-12. *Genetics*, **55**, 557–568.

43. Lawlis, V.B., Dennis, M.S., Chen, E.Y., Smith, D.H. and Henner, D.J. (1984) Cloning and sequencing of the xylose isomerase and xylulose kinase genes of *Escherichia coli*. *Appl. Environ. Microbiol.*, **47**, 15–21.

44. Beckwith, J.R. (1967) Regulation of the *lac* operon. Recent studies on the regulation of lactose metabolism in *Escherichia coli* support the operon model. *Science*, **156**, 597–604.

45. Plagens, A., Tripp, V., Daume, M., Sharma, K., Klingl, A., Hrle, A., Conti, E., Urlaub, H. and Randau, L. (2014) *In vitro* assembly and activity of an archaeal CRISPR-Cas type I-A Cascade interference complex. *Nucleic Acids Res.*, **42**, 5125–5138.

46. Esvelt, K.M., Mali, P., Braff, J.L., Moosburner, M., Yaung, S.J. and Church, G.M. (2013) Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat. Methods*, **10**, 1116–21.

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

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

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

50. Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K. and Sander, J.D. (2013) High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.*, **31**, 822–6.

51. Pattanayak, V., Lin, S., Guilinger, J.P., Ma, E., Doudna, J.A. and Liu, D.R. (2013) High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat. Biotechnol.*, **31**, 839–43.

52. Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., et al. (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.*, **31**, 827–832.

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

54. Fineran, P.C., Gerritzen, M.J.H., Suárez-Diez, M., Künne, T., Boekhorst, J., van Hijum, S.A.F.T., Staals, R.H.J. and Brouns, S.J.J. (2014) Degenerate target sites mediate rapid primed CRISPR adaptation. *Proc. Natl. Acad. Sci. U. S. A.*, **111**, E1629–1638.

55. Smith, C., Gore, A., Yan, W., Abalde-Atristain, L., Li, Z., He, C., Wang, Y., Brodsky, R.A., Zhang, K., Cheng, L., et al. (2014) Whole-Genome Sequencing Analysis Reveals High Specificity of CRISPR/Cas9 and TALEN-Based Genome Editing in Human iPSCs. *Cell Stem Cell*, **15**, 12–13.

56. Veres, A., Gosis, B.S., Ding, Q., Collins, R., Ragavendran, A., Brand, H., Erdin, S., Talkowski, M.E. and Musunuru, K. (2014) Low Incidence of Off-Target Mutations in Individual CRISPR-Cas9 and TALEN Targeted Human Stem Cell Clones Detected by Whole-Genome Sequencing. *Cell Stem Cell*, **15**, 27–30.

57. Stern, A., Keren, L., Wurtzel, O., Amitai, G. and Sorek, R. (2010) Self-targeting by CRISPR: gene regulation or autoimmunity? *Trends Genet. TIG*, **26**, 335–340.

58. Sampson, T.R., Napier, B.A., Schroeder, M.R., Louwen, R., Zhao, J., Chin, C.-Y., Ratner, H.K., Llewellyn, A.C., Jones, C.L., Laroui, H., et al. (2014) A CRISPR-Cas system enhances envelope integrity mediating antibiotic resistance and inflammasome evasion. *Proc. Natl. Acad. Sci. U. S. A.*, **111**, 11163–8.

59. Sampson, T.R., Saroj, S.D., Llewellyn, A.C., Tzeng, Y.-L. and Weiss, D.S. (2013) A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. *Nature*, **497**, 254–257.

SUPPLEMENTARY INFORMATION

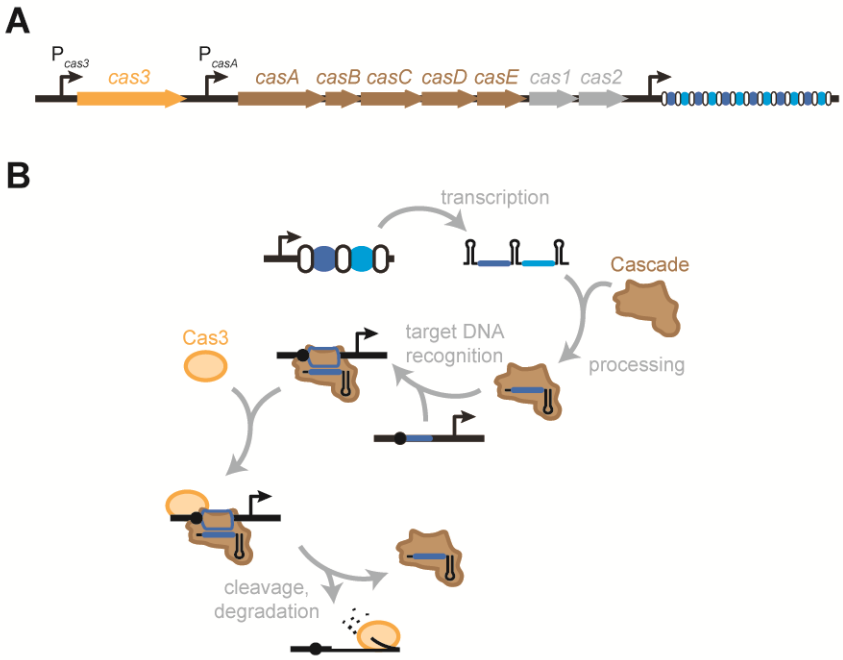

Figure 3.S1 The Type I-E CRISPR-Cas system in *E. coli* K-12. (A) Genetic locus of the Type I-E system. The *cas3* gene is located upstream of the *cse1-cse2-cas7-cas5-cas6e* operon encoding the Cascade protein complex. The two downstream genes *cas1* and *cas2* are involved in spacer acquisition. The native spacer array is composed of identical repeats (white ovals) and intervening spacers (blue circles). (B) Mechanism of DNA destruction based on previous work (1, 2). The transcribed array is processed into individual crRNAs by Cascade. The spacer portion of the array is then used to identify complementary DNA sequences flanked by a PAM (black circle). DNA binding leads to recruitment of Cas3, which cleaves and degrades the target DNA.

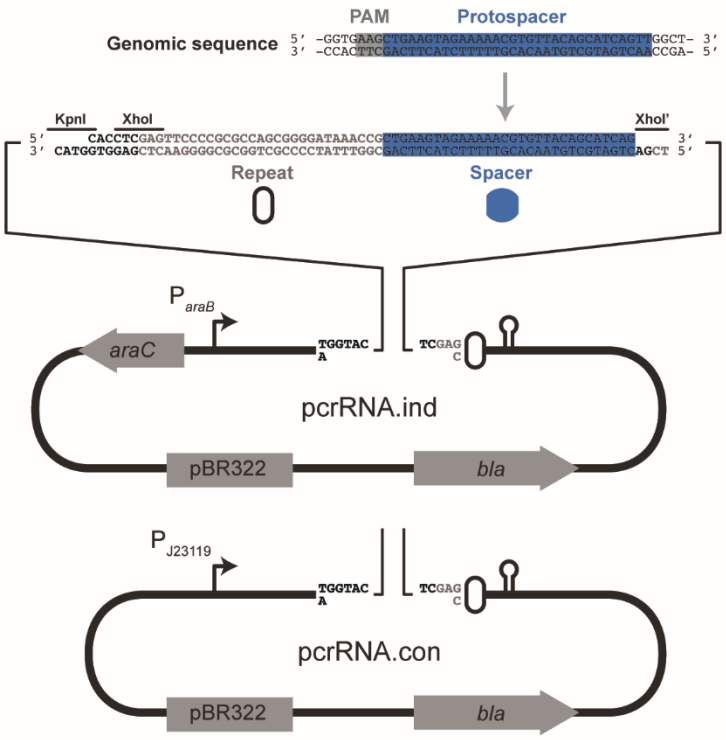

Figure 3.S2 Cloning scheme for the synthetic Type I-E CRISPR arrays. Following identification of a PAM in the target sequence, the downstream 32 base pairs serve as the protospacer. All but the last two base pairs are copied into two annealed oligonucleotides that, when annealed, form a repeat-spacer pair. The spacer-repeat pair contains the overhangs for a cleaved KpnI restriction site (left) and cleaved XhoI restriction site (right) along with an internal XhoI restriction site. As long as the 30th base in the protospacer is not a C, ligation of the annealed oligonucleotides into either plasmid digested with KpnI/XhoI disrupts the original XhoI restriction site. Consequently, additional repeat-spacer pairs can be sequentially inserted into the KpnI/XhoI restriction sites.

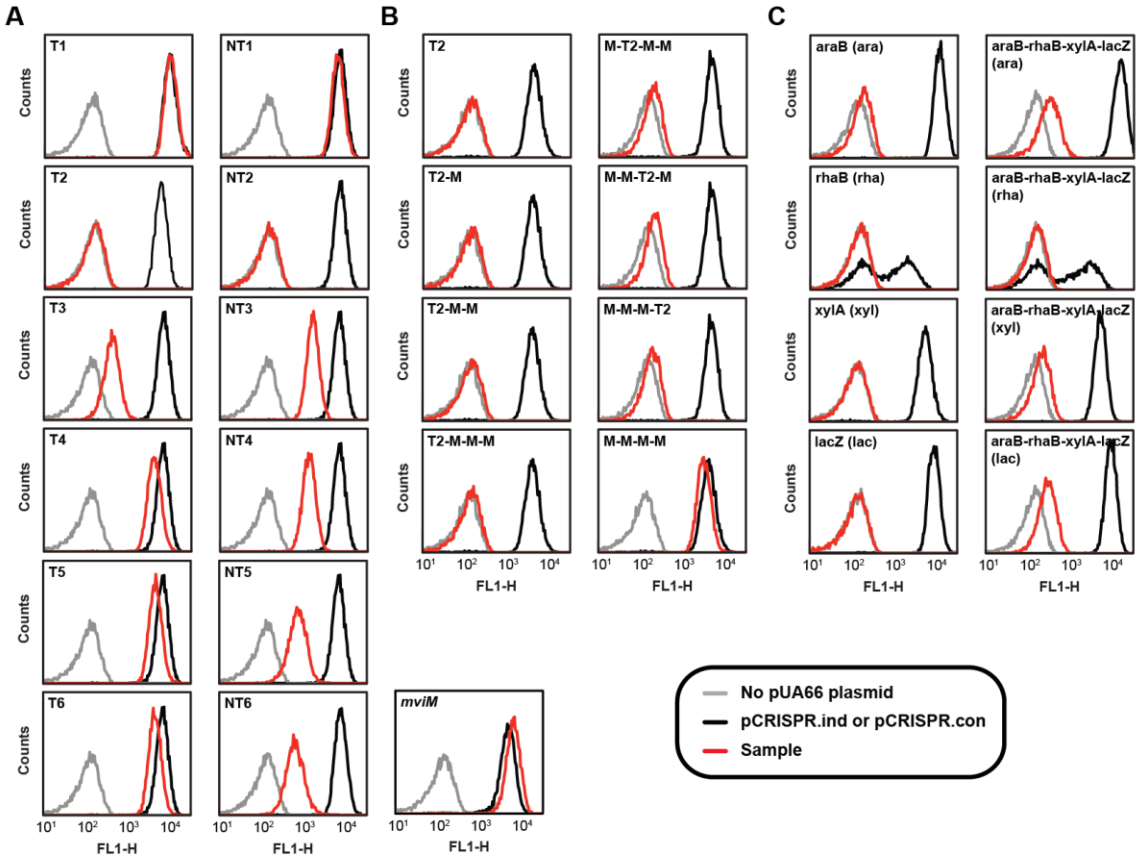

Figure 3.S3 Representative histograms from the flow cytometry analysis. (A) Histograms from Figure 3.2B. (B) Histograms from Figure 3.3. (C) Histograms for Figure 3.4B. The gray histograms are the no-GFP control for autofluorescence, the black histograms are the no spacer plasmids (either pCRISPR.ind or pCRISPR.con), and the red histograms are the indicated single-spacer or multi-spacer plasmids. See the corresponding figure legend in the main text for more information. Sugars in parentheses were included in the growth medium. Note that the bimodal response to L-rhamnose was reported previously (3). Histograms are representative of independent experiments starting with three separate colonies.

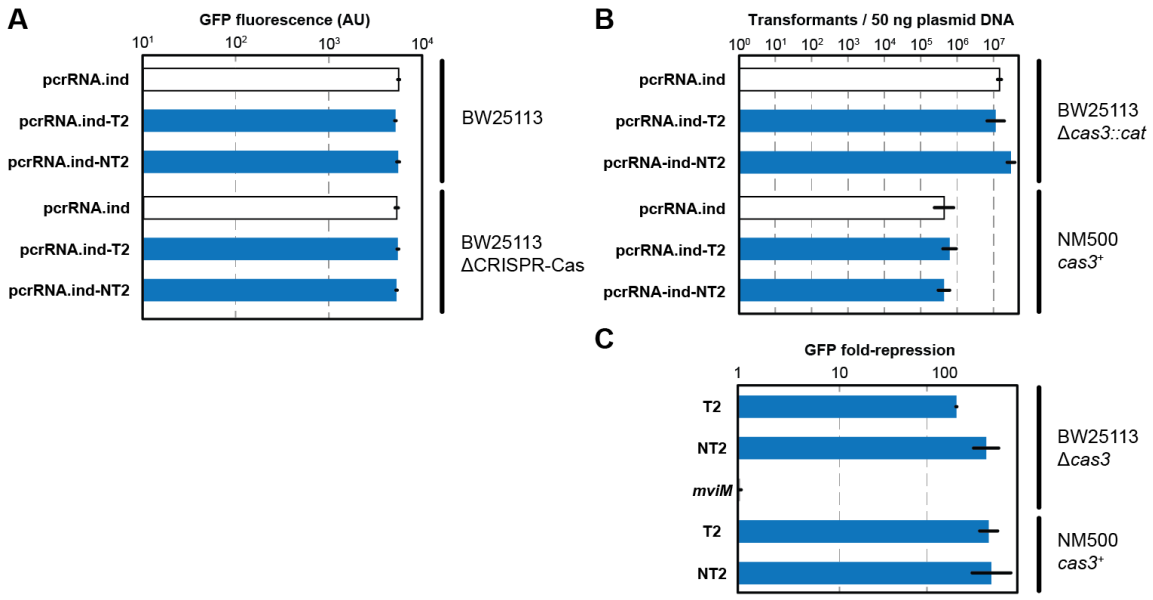

Figure 3.S4 GFP expression and DNA transformation in variants of the parent strain BW25113 or NM500. (A) GFP fluorescence of BW25113 (top) or BW25113 ΔCRISPR-Cas (bottom) harboring pUA66-lacZ and the indicated plasmid. Cells were grown for ~3-4 hours in M9 minimal medium containing 0.2% casamino acids, 0.4% glycerol, 0.2% L-arabinose, and 0.1 mM IPTG to ABS600 ~0.2 prior to flow cytometry analysis. The reported values are the absolute fluorescence minus autofluorescence from cells lacking GFP. (B) Transformation efficiencies in the absence or presence of *cas3*. BW25113 Δ*cas3*::*cat* or NM500 *cas3*$^{+}$ cells were transformed with 50 ng of the indicated plasmid and plated on LB agar with ampicillin and kanamycin, and the number of colonies was counted. The differences in transformation efficiencies between strains may be attributed to switching cuvette manufacturers. (C) GFP repression following excision of the resistance cassette or in the presence of cas3. BW25113 Δ*cas3* or NM500 *cas3*$^{+}$ cells harboring pUA66-lacZ and pCRISPR.ind, pCRISPR.ind-T2 (T2), pCRISPR.ind-NT2 (NT2), or pCRISPR.ind-mviM (*mviM*) were grown as indicated in A. Repression is calculated as the ratio of the autofluorescence-subtracted fluorescence for pcrRNA.ind and each single-spacer plasmid. See Figure 3.2 for more information. Values represent the geometric mean and S.E.M. from independent experiments starting with three separate colonies.

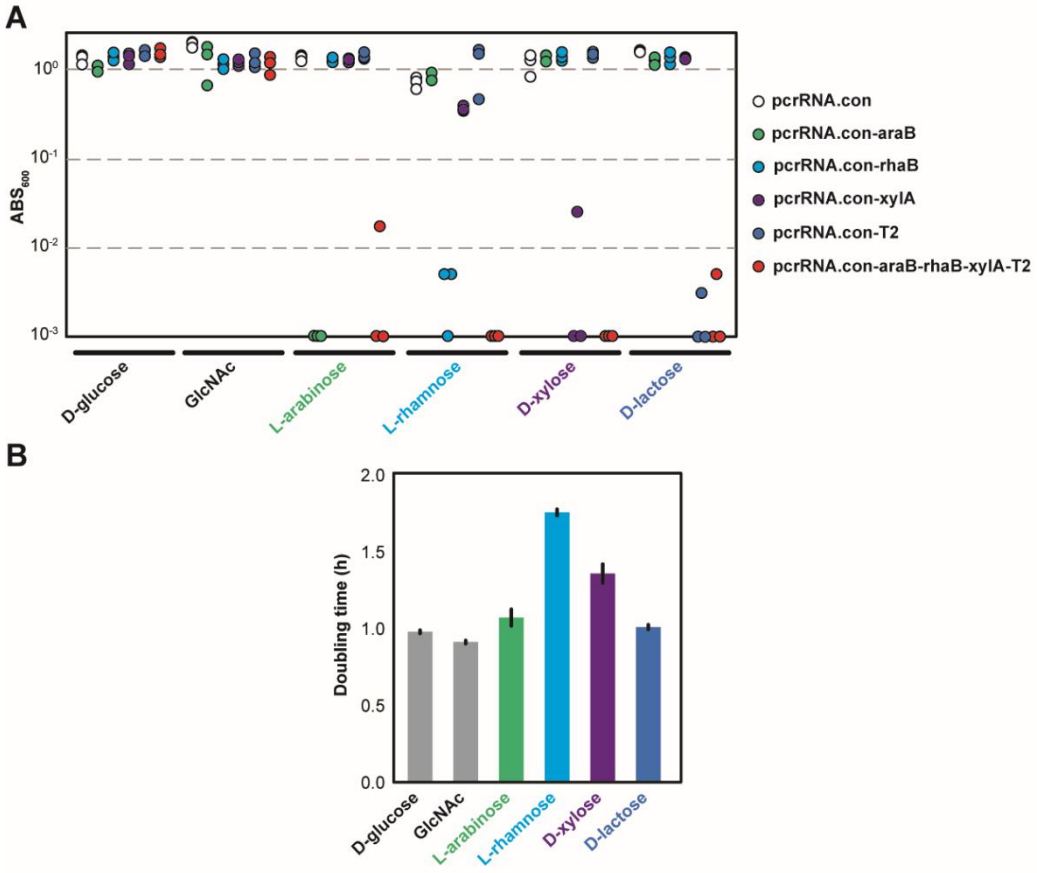

Figure 3.S5 Extended information for the growth assays. (A) Individual ABS_{600} values for the growth assays. Raw data values are reported. The data represent those shown in Figure 3.4D with the exception of the last column showing growth in media containing four sugars. Dots represent individual measurements from independent cultures. (B) Doubling times of MG1655 Δ*cas3*::*cat* cells harboring the constitutive pcrRNA.con plasmid grown in minimal medium with the indicated sugar as the sole carbon source. Values represent the geometric mean and S.E.M. from independent experiments starting with three separate colonies.

Table 3.S1 Strains used in this work

Strains	Genotype	Source	Stock #
BW25113	*Escherichia coli* K-12 F^- DE(*araD-araB*)567 lacZ4787(del)(::rrnB-3) λ^- rph-1 DE(*rhaD-rhaB*)568 *hsd*R514	CGSC [a] #7636	pCB294
BW25113 Δ*cas3::cat*	BW25113 [Δ*cas3* P_{cse1}]*::*[*cat* P_{J23119}]	This study	pCB385
BW25113 Δ*cas3*	BW25113 [Δ*cas3* P_{cse1}]*::*[P_{J23119}]	This study	pCB400
BW25113 ΔCRISPR-Cas	BW25113 [Δ*cas3-cse1-cse2-cas7-cas5-cas6e*-CRISPR1]::*cat*	This study	pCB401
BW25113 *cas3*$^+$	BW25113 [ΔP_{cse1}]*::*[*cat* P_{J23119}]	This study	pCB402
MG1655	*Escherichia coli* K-12 F^- λ^- *ilv*G- *rfb*-50 *rph*-1	Storz lab (NIH)	pCB1
MG1655 Δ*cas3::cat*	MG1655 [Δ*cas3* P_{cse1}]*::*[*cat* P_{J23119}]	This study	pCB386

[a] CGSC: Coli genetic stock center (http://cgsc.biology.yale.edu).

Table 3.S2 Plasmids used in this work

Plasmid	Description	Resistance marker	Source	Stock #
pUA66-lacZ	*lacZ* promoter upstream of GFP	Kanamycin	OpenBiosystems	pCB338
pUA66-araB	*araB* promoter upstream of GFP	Kanamycin	OpenBiosystems	pCB208
pUA66-xylA	*xylA* promoter upstream of GFP	Kanamycin	Ref. (3)	pCB289
pUA66-rhaB	*rhaB* promoter upstream of GFP	Kanamycin	Ref. (3)	pCB292
pBAD18	L-arabinose-inducible plasmid with *araC* regulator	Ampicillin	Ref. (3)	pCB284
pcrRNA.ind	pBAD18 with single repeat	Ampicillin	This study	pCB359
pcrRNA.ind-T1	pcrRNA.ind with spacer T1	Ampicillin	This study	pCB360
pcrRNA.ind-T2	pcrRNA.ind with spacer T2	Ampicillin	This study	pCB361
pcrRNA.ind-T3	pcrRNA.ind with spacer T3	Ampicillin	This study	pCB362
pcrRNA.ind-T4	pcrRNA.ind with spacer T4	Ampicillin	This study	pCB363
pcrRNA.ind-T5	pcrRNA.ind with spacer T5	Ampicillin	This study	pCB364
pcrRNA.ind-T6	pcrRNA.ind with spacer T6	Ampicillin	This study	pCB365
pcrRNa.ind-NT1	pcrRNA.ind with spacer NT1	Ampicillin	This study	pCB366
pcrRNA.ind-NT2	pcrRNA.ind with spacer NT2	Ampicillin	This study	pCB367

Table 3.S2 Continued

Plasmid	Description	Resistance marker	Source	Stock #
pcrRNA.ind-NT3	pcrRNA.ind with spacer NT3	Ampicillin	This study	pCB368
pcrRNA.ind-NT4	pcrRNA.ind with spacer NT4	Ampicillin	This study	pCB369
pcrRNA.ind-NT5	pcrRNA.ind with spacer NT5	Ampicillin	This study	pCB370
pcrRNA.ind-NT6	pcrRNA.ind with spacer NT6	Ampicillin	This study	pCB371
pcrRNA.ind-LM	pcrRNA.ind with spacers T2-*mviM*	Ampicillin	This study	pCB372
pcrRNA.ind-LMM	pcrRNA.ind with spacers T2-*mviM-mviM*	Ampicillin	This study	pCB373
pcrRNA.ind-LMMM	pcrRNA.ind with spacers T2-*mviM-mviM-mviM*	Ampicillin	This study	pCB374
pcrRNA.ind-MLMM	pcrRNA.ind with spacers *mviM*-T2-*mviM-mviM*	Ampicillin	This study	pCB375
pcrRNA.ind-MMLM	pcrRNA.ind with spacers *mviM-mviM*-T2-*mviM*	Ampicillin	This study	pCB376
pcrRNA.ind-MMML	pcrRNA.ind with spacers *mviM-mviM-mviM*-T2	Ampicillin	This study	pCB377
pcrRNA.ind-MMMM	pcrRNA.ind with spacers *mviM-mviM-mviM-mviM*	Ampicillin	This study	pCB378
pcrRNA.con	pcrRNA.con with synthetic constitutive promoter	Ampicillin	This study	pCB379
pcrRNA.con-lacZ	pcrRNA.con with spacer T2	Ampicillin	This study	pCB380
pcrRNA.con-araB	pcrRNA.con with spacer araB	Ampicillin	This study	pCB381

Table 3.S2 Continued

Plasmid	Description	Resistance marker	Source	Stock #
pcrRNA.con-xylA	pcrRNA.con with spacer xylA	Ampicillin	This study	pCB382
pcrRNA.con-rhaB	pcrRNA.con with spacer rhaB	Ampicillin	This study	pCB383
pcrRNA.con-araB/rhaB/xylA/T2	pcrRNA.con with spacers araB-rhaB-xylA-T2	Ampicillin	This study	pCB384

Table 3.S3 Oligonucleotides used in this work

Name	Sequence
J23119-pKD3.for	GCTAGCATTATACCTAGGACTGAGCTAGCTGTCAATCCATATGAATATCCTCCTTAG
J23119-pKD3.rev	TGTAGGCTGGAGCTGCTT
HR-cas3.for	TACAATTAACCTATACATATATTAAGATGTGTTGAATTGTGCTAGCATTATACCTAGGAC
HR-cas3.rev	TGATATCATCGATAATACTAAAAAAACAGGGAGGCTATTATGTAGGCTGGAGCTGCTT
HR-CRISPR.rev	ACCGCAGAGGCGGGGGAACTCCAAGTGATATCCATCATTCCATATGAATATCCTCCTTAG
HR-casA.rev	CTTTTAATTTCCCGGTATGAGATTTTATATTCACAGTATGtgtaggctggagctgctt
pcrRNA.ind.for	CCACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGAAAAAAAAACCCCGCCCCTGACAGGGCGGGGTTTTTTTTA
pcrRNA.ind.rev	AAGCTTAAAAAAAACCCCGCCCTGTCAGGGGCGGGGTTTTTTTTTCGGTTTATCCCCGCTGGCGCGGGGAACTCGAGGTGGTACC
pcrRNA.con.for	TTTGACAGCTAGCTCAGTCCTAGGTATAATGCTAGCG
pcrRNA.con.rev	CTAGCGCTAGCATTATACCTAGGACTGAGCTAGCTGTCAAATGCA
T2.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGCTTTACACTTTATGCTTCCGGCTCGTATGT
T2.rev	TCGAACATACGAGCCGGAAGCATAAAGTGTAAAGCGGTTTATCCCCGCTGGCGCGGGGAACTCGAGGTGGTAC
NT2.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGCATAAAGTGTAAAGCCTGGGGTGCCTAATG
NT2.rev	TCGACATTAGGCACCCCAGGCTTTACACTTTATGCGGTTTATCCCCGCTGGCGCGGGGAACTCGAGGTGGTAC
T3.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGAAACAGCTATGACCATGATTACGGATTCAC

Table 3.S3 Continued

Name	Sequence
T3.rev	TCGAGTGAATCCGTAATCATGGTCATAGCTGTTTCGGTTTATCCCCGCTGGC GCGGGGAACTCGAGGTGGTAC
NT3.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGGCGATTAAGTTGGGTAA CGCCAGGGTTTTC
NT3.rev	TCGAGAAAACCCTGGCGTTACCCAACTTAATCGCCGGTTTATCCCCGCTGGC GCGGGGAACTCGAGGTGGTAC
T1.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGCCCTTTCGTCTTCACAC TCGAGCACGACAG
T1.rev	TCGACTGTCGTGCTCGAGTGTGAAGACGAAAGGGCGGTTTATCCCCGCTGGC GCGGGGAACTCGAGGTGGTAC
T4.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGAGATATACATATGAGTA AAGGAGAAGAACT
T4.rev	TCGAAGTTCTTCTCCTTTACTCATATGTATATCTCGGTTTATCCCCGCTGGC GCGGGGAACTCGAGGTGGTAC
T5.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGTGATGCAACATACGGAA AACTTACCCTTAA
T5.rev	TCGATTAAGGGTAAGTTTTCCGTATGTTGCATCACGGTTTATCCCCGCTGGC GCGGGGAACTCGAGGTGGTAC
T6.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGTGATACCCTTGTTAATA GAATCGAGTTAAA
T6.rev	TCGATTTAACTCGATTCTATTAACAAGGGTATCACGGTTTATCCCCGCTGGC GCGGGGAACTCGAGGTGGTAC
NT1.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGACGAAAGGGCCTCGTGA TACGCCTATTTTT
NT1.rev	TCGAAAAAATAGGCGTATCACGAGGCCCTTTCGTCGGTTTATCCCCGCTGGC GCGGGGAACTCGAGGTGGTAC
NT4.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGTTCTTCTCCTTTACTCA TATGTATATCTCC

Table 3.S3 Continued

Name	Sequence
NT4.rev	TCGAGGAGATATACATATGAGTAAAGGAGAAGAACGGTTTATCCCCGCTGGC GCGGGGAACTCGAGGTGGTAC
NT5.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGGTAAGTTTTCCGTATGT TGCATCACCTTCA
NT5.rev	TCGATGAAGGTGATGCAACATACGGAAAACTTACCGGTTTATCCCCGCTGGC GCGGGGAACTCGAGGTGGTAC
NT6.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGGTATCACCTTCAAACTT GACTTCAGCACGT
NT6.rev	TCGAACGTGCTGAAGTCAAGTTTGAAGGTGATACCGGTTTATCCCCGCTGGC GCGGGGAACTCGAGGTGGTAC
araB.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGATTAGCGGATCCTACCT GACGCTTTTTATC
araB.rev	TCGAGATAAAAAGCGTCAGGTAGGATCCGCTAATCGGTTTATCCCCGCTGGC GCGGGGAACTCGAGGTGGTAC
xylA.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGGAGTGCCCAATATTACG ACATCATCCATCA
xylA.rev	TCGATGATGGATGATGTCGTAATATTGGGCACTCCGGTTTATCCCCGCTGGC GCGGGGAACTCGAGGTGGTAC
rhaB.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGGTCGCGAATTCAGGCGC TTTTTAGACTGGT
rhaB.rev	TCGAACCAGTCTAAAAAGCGCCTGAATTCGCGACCGGTTTATCCCCGCTGGC GCGGGGAACTCGAGGTGGTAC
mviM.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGAGCGCGGGCAGGGTATT CTCATCAAACCCA
mviM.rev	TCGATGGGTTTGATGAGAATACCCTGCCCGCGCTCGGTTTATCCCCGCTGGC GCGGGGAACTCGAGGTGGTAC
lacZ-qPCR.fwd	CGGCGTATCGCCAAAATCAC
lacZ-qPCR.rev	ATGGGTAACAGTCTTGGCGG

Table 3.S3 Continued

Name	Sequence
araB-qPCR.fwd	TACCAGTGCGTTAGGCTGTG
araB-qPCR.rev	CTGGACCCGATCCTCAATCG
xylA-qPCR.fwd	CGCCCCACAGGACATAGTTT
xylA-qPCR.rev	GGAACGGCCAACTGCTTTAC
rhaB-qPCR.fwd	TCACTTTCCGGGATCGGTTG
rhaB-qPCR.rev	TTCAGCGAGTGCTTCAGGAG

Table 3.S4 Protospacers in this work

Spacer name	Target strand[a]	Distance from TSS[b]	Protospacer sequence[c]
T1	T	-141	**AGG**CCCTTTCGTCTTCACaCTCGAGCACGACAG
T2/lacZ	T	-37	**AGG**CTTTACACTTTATGCTTCCGGCTCGTATGT
T3	T	+27	**AGG**AAACAGCTATGACCATGATTACGGATTCAC
T4	T	+149	**AGG**AGATATACATATGAGTAAAGGAGAAGAACT
T5	T	+263	**AGG**TGATGCAACATACGGAAAACTTACCCTTAA
T6	T	+506	**AGG**TGATACCCTTGTTAATAGAATCGAGTTAAA
NT1	N	-129	**AAG**ACGAAAGGGCCTCGTGATACGCCTATTTTT
NT2	N	-20	**AAG**CATAAAGTGTAAAGCCTGGGGTGCCTAATG
NT3	N	+119	**AAG**GCGATTAAGTTGGGTAACGCCAGGGTTTTC
NT4	N	+180	**AAG**TTCTTCTCCTTTACTCATATGTATATCTCC
NT5	N	+290	**AGG**GTAAGTTTTCCGTATGTTGCATCACCTTCA
NT6	N	+515	**AGG**GTATCACCTTCAAACTTGACTTCAGCACGT
araB	T	-53	**AAG**ATTAGCGGATCCTACCTGACGCTTTTTATC
xylA	T	-19	**AGG**GAGTGCCCAATATTACGACATCATCCATCA
rhaB	T	-33	**AAG**GTCGCGAATTCAGGCGCTTTTTAGACTGGT
mviM[d]	N/A	N/A	**AAG**AGCGCGGGCAGGGTATTCTCATCAAACCCA

a) Characteristics of the target strand, which is complementary to the spacer: T, template strand of gene; N, non-template strand of the gene.

b) Distance from the transcriptional start site (TSS) to the closest end of the PAM. Negative and positive values are upstream and downstream of the TSS, respectively.

c) PAMs are in bold red lettering. CRISPR spacers were designed to match the protospacer sequence.

d) Targets a protospacer in *Salmonella typhimurium* LT2 and has been shown to be non-targeting in *E. coli*(4).

Table 3.S5 Promoter sequences

Promoter	Sequence[a]
lacZ	CTTTCGTCTTCACACTCGAGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGC GCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTA TGCTTCCGGCTCGTATGTTGTGTGG**A**ATTGTGAGCGGATAACAATTTCACACAGGAA ACAGCTATGACCATGATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTGACTGG GAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACAGGATCCTCTAGATTTA AGAA
araB	CCTGTCTCTTGATCAGATCTGGCCTCAATCGGCGTTAAACCCGCCACCAGATGGGCG TTAAACGAGTATCCCGGCAGCAGGGGATCATTTTGCGCTTCAGCCATACTTTTCATA CTCCCACCATTCAGAGAAGAAACCAATTGTCCATATTGCATCAGACATTGCCGTCAC TGCGTCTTTTACTGGCTCTTCTCGCTAACCCAACCGGTAACCCCGCTTATTAAAAGC ATTCTGTAACAAAGCGGGACCAAAGCCATGACAAAAACGCGTAACAAAAGTGTCTAT AATCACGGCAGAAAAGTCCACATTGATTATTTGCACGGCGTCACACTTTGCTATGCC ATAGCATTTTTATCCATAAGATTAGCGGATCCTACCTGACGCTTTTTATCGCAACTC TCTACTGTTTCTCCAT**A**CCCGTTTTTTTGGATGGAGTGAAACGATGGCGATTGCAAT TGGCCTCGATTTTGGCAGTGATTCTGTGCGAGCTTTGGCGGTGGACTGCGCTACCGG TGAAGCTCGAGGGGATCCTCTAGA
xylA	CGAGGCCCTTTCGTCTTCACGGTGTAGGGCCTTCTGTAGTTAGAGGACAGTTTTAAT AAGTAACAATCACCGCGATAAACGTAACCAATTTTTAGCAACTAAACAGGGGAAAAC AATTACAGATTTTTATCTTTCGATTACGATTTTTGGTTTATTTCTTGATTTATGACC GAGATCTTACTTTTGTTGCGCAATTGTACTTATTGCATTTTTCTCTTCGAGGAATTA CCCAGTTTCATCATTCCATTTTATTTTGCGAGCGAGCGCACACTTGTGAATTATCTC AATAGCAGTGTGAAATAACATAATTGAGCAACTGAAAGGGAGTGCCCAATATTACGA C**A**TCATCCATCACCCGCGGCATTACCTGATTATGGAGTTCAATATGCAAGCCTATTT TGACCAGCTCGATCGCGTTCGTTATGAAGGCTCAAAATCCTCAAACCCGTTAGCATT CCGTCACTACAATCCCGACGAACTGGTGTTGGGTAAGCGTATGTAATCTAGATTTAA GAAGGAGAT
rhaB	CCTGTCTCTTGATCAGATCTGTTCTATCGCCACGGACGCGTTACCAGACGGAAAAAA ATCCACACTATGTAATACGGTCATACTGGCCTCCTGATGTCGTCAACACGGCGAAAT AGTAATCACGAGGTCAGGTTCTTACCTTAAATTTTCGACGGAAAACCACGTAAAAAA CGTCGATTTTTCAAGATACAGCGTGAATTTTCAGGAAATGCGGTGAGCATCACATCA CCACAATTCAGCAAATTGTGAACATCATCACGTTCATCTTTCCCTGGTTGCCAATGG CCCATTTTCCTGTCAGTAACGAGAAGGTCGCGAATTCAGGCGCTTTTTAGACTGGTC GT**A**ATGAAATTCAGCAGGATCACATTATGACCTTTCGCAATTGTGTCGCCGTCGATC TCGGCGCATCCAGTGGGCGCGTGATGCTGGCGCGTTACGAGCGTGAATGGGATCCTC TAGATTTAAGAA

a) Sequences highlighted in gray are from pUA66, indicating where each promoter was inserted into the plasmid. The underlined and bolded base is the previously mapped transcriptional start site.

SUPPLEMENTARY REFERENCES

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

2. Brouns, S.J.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J.H., Snijders, A.P.L., Dickman, M.J., Makarova, K.S., Koonin, E.V. and van der Oost, J. (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, **321**, 960–964.

3. Afroz, T., Biliouris, K., Kaznessis, Y. and Beisel, C.L. (In press) Bacterial sugar utilization gives rise to distinct single-cell behaviors. *Mol. Microbiol.* doi: 10.1111/mmi.12695

4. Gomaa, A.A., Klumpe, H.E., Luo, M.L., Selle, K., Barrangou, R. and Beisel, C.L. (2014) Programmable removal of bacterial strains by use of genome-targeting CRISPR-Cas systems. *mBio*, **5**, e00928–00913.

CHAPTER 4

The CRISPR RNA-guided surveillance complex in *Escherichia coli* accommodates extended RNA spacers

Michelle L. Luo[1], Ryan N. Jackson[2], Steven R. Denny[1], Monika Tokmina-Lukaszewska[3], Kenneth R. Maksimchuk[1], Wayne Lin[2], Brian Bothner[3], Blake Wiedenheft[2], Chase L. Beisel[1]

[1]Department of Chemical and Biomolecular Engineering,
North Carolina State University, Raleigh, NC 27695

[2]Department of Microbiology and Immunology,
Montana State University, Bozeman, MT 59717

[3]Department of Chemistry and Biochemistry,
Montana State University, Bozeman, MT 59717

Original publication in: *Nucleic Acids Research* 44(15): 7385-7394, 2016

ABSTRACT

Bacteria and archaea acquire resistance to foreign genetic elements by integrating fragments of foreign DNA into CRISPR (clustered regularly interspaced short palindromic repeats) loci. In *Escherichia coli*, CRISPR-derived RNAs (crRNAs) assemble with Cas proteins into a multi-subunit surveillance complex called Cascade (CRISPR-associated complex for antiviral defense). Cascade recognizes DNA targets via protein-mediated recognition of a PAM (protospacer adjacent motif) and complementary base pairing between the crRNA spacer and the DNA target. Previously determined structures of Cascade showed that the crRNA is stretched along an oligomeric protein assembly, leading us to ask how crRNA length impacts the assembly and function of this complex. We found that extending the spacer portion of the crRNA resulted in larger Cascade complexes with altered stoichiometry and preserved *in vitro* binding affinity for target DNA. Longer spacers also preserved the *in vivo* ability of Cascade to repress target gene expression and to recruit the Cas3 endonuclease for target degradation. Finally, longer spacers exhibited enhanced silencing at particular target locations and were sensitive to mismatches within the extended region. These findings demonstrate the flexibility of the Type I-E CRISPR machinery and suggest that spacer length can be modified to fine-tune Cascade activity.

4.1 INTRODUCTION

CRISPR-Cas (clustered regularly interspaced short palindromic repeats and CRISPR-associated proteins) systems are widespread and diverse adaptive immune systems in bacteria and archaea (1, 2). These systems identify and degrade foreign genetic material associated with plasmids and bacteriophages through three distinct stages termed acquisition, expression, and interference (3–5). During acquisition, small pieces of foreign DNA are inserted as new spacers into the CRISPR array composed of alternating repeats and spacers. During expression, the CRISPR array is processed into mature CRISPR RNAs (crRNAs) composed of a spacer and flanking portions of the repeat, yielding an RNA-protein complex composed of the crRNA and Cas effector proteins. During interference, the complex binds foreign genetic material complementary to the spacer portion of the crRNA, thereby activating *cis* or *trans*-acting nucleases that cleave the bound target.

Despite their common function in adaptive immunity, CRISPR-Cas systems are phylogenetically and functionally diverse, where the latest classifications define two classes, six types, and nineteen subtypes (2, 6). Class 1 systems rely on multi-subunit surveillance complexes while Class 2 systems rely on a single effector protein. The two main classes are subdivided into types and subtypes that further delineate the particular form and function of each system. The phylogenetic distribution of these systems vary widely, with the Class 1 Type I system as the most abundant and widespread in both bacteria and archaea (2). This type is defined by the signature *cas3* gene encoding a protein with helicase and endonuclease activities (2, 7–10).

Our current knowledge of Type I systems has largely stemmed from studies of the Type I-E system in *Escherichia coli* K-12. In *E. coli*, the Cas3 helicase-nuclease is recruited to DNA targets bound by a multi-subunit complex termed the CRISPR-associated complex for antiviral defense (Cascade) (9, 11). Cascade is a 405 kDa complex composed of an uneven stoichiometry of five different Cas proteins ($Cse1_1Cse2_2Cas7_6Cas5_1Cas6e_1$) and a 61-nt crRNA (12). This complex forms a seahorse-shaped architecture with subunits that represent the head (Cas6e), backbone (Cas7), belly (Cse2), and tail (Cse1, Cas5). Cascade engages

foreign DNA by searching for a sequence called a protospacer-adjacent motif (PAM) (13, 14). PAM recognition is thought to distort the DNA in a way that facilitates crRNA-guided interrogation of the flanking DNA sequence for complementarity with the crRNA spacer (9, 12, 14–18). Target binding triggers a conformational change that recruits Cas3, which degrades the non-target strand in the 3'-to-5' direction (7, 11, 19).

The Type I-E system from *E. coli* has proven to be a versatile tool for programmable gene silencing and DNA destruction (20). Gene silencing relies on the capacity of Cascade to stably bind without cleaving target DNA in the absence of Cas3 (21, 22). Directing Cascade to bind a promoter repressed expression of the downstream gene up to ~1,000-fold, whereas directing Cascade to the coding region led to more modest, strand-dependent silencing (21, 22). Cascade has also been used in the presence of Cas3 to target and degrade the bacterial genome in a sequence-specific manner, spurring applications in sequence-specific antimicrobials, genome editing, and biocontainment (23–25).

Studies to understand and exploit the Type I-E system from *E. coli* have relied on a 32-nt spacer. This length aligns with the extremely narrow range observed in naturally occurring CRISPR arrays across *E. coli* strains (26). Crystal structures of Cascade suggest that the RNA serves as a scaffold for the oligomerization of the six Cas7 proteins forming the backbone of the complex (15, 16, 27). Based on this insight, we speculated that we could control Cascade assembly by varying the length of the crRNA spacer and that these engineered complexes may represent a novel method for controlling target recognition. We found that extending the spacer resulted in larger Cascade complexes with altered protein subunit stoichiometry. Using a combination of *in vitro* binding studies with *in vivo* gene silencing and targeted killing assays, we demonstrate that these enlarged complexes preserve function in target binding, transcriptional repression, and DNA degradation. We also found that longer spacers can enhance gene silencing depending on the targeted site. Finally, we showed that mismatches in the extended region disrupted DNA binding, gene silencing, and targeted killing. These findings reveal that Cascade assembly is dependent on spacer length and that this length may be used as method for fine-tuning the activity and specificity of Cascade.

4.2 MATERIALS AND METHODS

4.2.1 Strains and plasmid construction

See Supplementary Table 4.S1 for a list of all *E. coli* K-12 strains used in this work. To generate BW25113 Δ*cas3*::*cat lacZ*$^+$, the *lacZ* gene from MG1655 was transferred by P1 transduction into BW25113 Δ*cas3*::*cat*. Cells were selected on M9 agar plates containing chloramphenicol and 0.2% D-lactose as the sole carbon source. Next, the *cat* cassette from BW25113 Δ*cas3*::*cat lacZ*$^+$ was excised using the pCP20 plasmid as described previously (28) to generate BW25113 *Δcas3 lacZ*$^+$.

See Supplementary Table 4.S2 for a list of all plasmids used in this work. To generate cBAD33, the pBAD33 plasmid was digested with NsiI and SacI to remove *araC* through the *araB* promoter. Oligonucleotides encoding the J23108 promoter (pBAD33.J23108.fwd2/ pBAD33.J23108.rev2) were 5' phosphorylated with T4 polynucleotide kinase and annealed by slowly cooling from 95°C to 25°C. The resulting dsDNA was ligated with T4 DNA ligase into the digested pBAD33 and transformed into NovaBlue electrocompetent cells. The pCas3 plasmid was generated by PCR-amplifying the *cas3* gene from *E. coli* MG1655 genomic DNA using primers to introduce a strong ribosome binding site, an upstream KpnI restriction site, and a downstream XbaI restriction site (pCas3.for/pCas3.rev). The *cas3* amplicon and cBad33 vector were digested with KpnI and XbaI, ligated together with T4 DNA ligase, and transformed into NovaBlue electrocompetent cells.

The GFP reporter plasmids were based on the pUA66 plasmid (29). To generate pUA66lacZ-NT3PAM-mutant from pUA66lacZ, mutagenic primers (NT3-PAM.Q5.for/NT3-PAM.Q5.rev) were used with the Q5® Site-Directed Mutagenesis kit from New England Biolabs (NEB).

The crRNA expression plasmids, pcrRNA.ind and pcrRNA.con, were based on previous work (21). To insert new repeat-spacer pairs into pcrRNA.con or pcrRNA.ind, either plasmid first digested with KpnI and XhoI. Then, oligonucleotides encoding the palindromic repeat and crRNA spacers were annealed, 5' phosphorylated with T4 polynucleotide kinase, ligated into the digested vector with T4 DNA ligase, and transformed into TOP10

electrocompetent cells. All plasmid cloning was verified by Sanger sequencing. See Supplementary Table 4.S3 for a list of all oligonucleotides used in this work. All oligonucleotides were chemically synthesized by Integrated DNA Technologies (IDT) or Eurofins Genomics. All enzymes were purchased from New England Biolabs (NEB)

4.2.2 Cascade expression and purification

Cascade and Cascade variants were expressed and purified using previously described methods (12, 17). Briefly, crRNAs and *E. coli* K-12 Cas proteins were co-expressed in *E. coli* BL21 (DE3) cells on three different expression vectors; a pCDF vector containing Cse2 fused to an N-terminal Strep tag, Cas7, Cas5 and Cas6; a pRSF vector containing Cse1; and either pcrRNA.con or pcrRNA.ind vectors (described above) containing the CRISPR with a normal (32-nt) or extended spacer sequence (e.g. 32 +6, 32 +12). Cells were grown at 37°C in LB-media under antibiotic selection to an OD_{600nm} of 0.5, then were induced with a final concentration of 0.2 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG) and 0.02 % L-arabinose. After induction, cells were cultured overnight at 16°C, pelleted by centrifugation (5,000g for 10 min), suspended in lysis buffer (100 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA, 1 mM TCEP and 5% glycerol), and frozen at -80°C. Cells were lysed by sonication and lysates were clarified by centrifugation (22,000g for 30 min). Cascade and Cascade variants self-assemble *in vivo* and were affinity purified on StrepTrap HP resin (GE) with the N-terminal Strep-II tag on Cse2. Cascade was eluted off StrepTrap HP resin with lysis buffer supplemented with 2.5 mM desthiobiotin. Cascade was concentrated and then further purified by gel filtration chromatography using a 26/60 Superdex 200 (GE Healthcare) equilibrated with 50 mM Tris-HCl pH 7.5, 100 mM NaCl, 1 mM TCEP and 5% glycerol. The elution profile was visualized at 280 and 260 nm wavelengths to identify changes in complex size.

To identify size differences between Cascade variants, 600 µl of each complex was purified by gel filtration chromatography using a 16/60 Superdex 200 (GE Healthcare) equilibrated with 100 mM HEPES pH 7.0. Eluted complexes were concentrated, mixed with 5X loading buffer (100 mM HEPES pH 7.5, 375 mM NaCl, 50% glycerol) and loaded onto a

5% Blue Native Polyacrylamide gel and run at 85 V for 2 hours (30). To analyze protein subunit abundance, 20 µg of each Cascade variant was loaded onto a 10% denaturing SDS Polyacrylamide Gel and run at 150 V for 45 minutes. To determine crRNA length, RNA was isolated by phenol/chloroform extraction and loaded onto a TBE (Tris Borate EDTA), 7M-Urea, 14% polyacrylamide gel and ran at 190 V for 30 minutes. Proteins were stained with Coomassie dye and RNA with SYBR Gold (Thermo Fisher Scientific).

4.2.3 SDS PAGE gel densitometry and subunit stoichiometry estimation

Coomassie stained SDS-PAGE gels were analyzed with GelQuant.Net (biochemlabsolutions.com). Bands intensities were quantified with local background correction. Band intensity fractions were determined with the equation:

$$\text{band intensity fraction} = \frac{\text{individual band intensity}}{\text{sum of all lane intensities}}$$

Stoichiometry was estimated using the equation:

$$\#\text{ of subunits} = \text{band intensity fraction} \times \frac{\text{Cascade complex molecular weight}}{\text{subunit molecular weight}}$$

Average stoichiometry and standard deviations were calculated from five different SDS PAGE gels stained with Coomassie Blue.

4.2.4 Protein and native mass spectrometry

Cascade was buffer exchanged into 100 mM ammonium acetate, pH 7 (Sigma) by repeated washing over 3 kDa molecular weight cutoff spin filters (Pall Corporation). A total of 2 µM Cascade was injected using gold-coated borosilicate glass capillaries and analyzed on a SYNAPT G2-Si electrospray time-of-flight instrument (Waters) in positive mode. The system was calibrated with sodium iodide dissolved in solution of 2-propanol and water (Waters). Gold-coated borosilicate glass capillaries were prepared as previously reported (31) with the following changes: Borosilicate glass capillaries, 1.2/0.68 OD/ID mm (World Precision Instruments) were pulled on model *P*-97 (Sutter Instrument Company) in one step, with heat 540, pull 90, velocity 50, time 150. Solutions were sprayed at a rate of 90 nl/min.

Capillaries were coated with two layers (inner: Cr, 3 nm to increase coat resistance and prolong spray time; outer: Au, 60 nm to provide conductivity) using an AMod Evaporator System.

To assure optimal instrument performance in high mass-to-charge range, several adjustments to previously published protocol (32) were made. To minimize complex dissociation and maximize ion transmission, source temperature was set to 30ºC, capillary voltage to 1.7 kV, cone gas to 40 L/h. Sampling cone and source offset were set to 40 V and 50 V, respectively. Inlet pressure was fixed at 3 mbar. To improve desolvation/declustering, trap bias voltage was adjusted to 16 V and argon pressure in the collision cell (trap) was 7 ml/min. Transfer collision energy was kept at constant level of 5 V while trap energy varied between 10-200 V. To determine accurate protein subunit masses, Cascade complexes were denatured by dilution in a 50:50 solution of 1% formic acid (Sigma) and acetonitrile. Collected spectra of Cascade complexes in native and denatured conditions were processed and analyzed in MassLynx software version 4.1 (Waters).

4.2.5 Electrophoretic Mobility Shift Assays (EMSA)

Oligonucleotides (Operon) were 5'-end labeled with γ^{32}P-ATP (PerkinElmer) using T4 polynucleotide kinase (NEB), and purified by phenol/chloroform extraction followed by MicroSpin G-25 column (GE Healthcare) filtration. Labeled oligonucleotides were hybridized with >5 fold molar excess of complementary strand in hybridization buffer (20mM HEPES pH 7.5, 75mM NaCl, 2mM EDTA, 10% glycerol, and 0.01% bromophenol blue) by incubating at 95°C for 5 minutes, and gradually cooling to 25°C in a thermocycler. DNA duplexes were gel purified, ethanol precipitated, and recovered in hybridization buffer.

Varying concentrations of Cascade or Cascade variants were incubated with ^{32}P labeled oligonucleotides in hybridization buffer with 1mM TCEP for 15 minutes at 37 °C. Reactions were loaded onto a 6% native polyacrylamide gel, and run for 3 hours at 150 V at 4 °C. After electrophoresis, gels were dried, exposed to phosphor storage screens, and scanned with a Typhoon (GE Healthcare) phosphorimager. Bound and unbound DNA fractions were quantified using ImageQuant software (GE Healthcare), and the fractions of bound

oligonucleotides were plotted against total Cascade concentration. The data were fit by least-squares analysis using a standard binding isotherm

$$y = \frac{x}{K_d + x}$$

where *y* is the fraction of bound DNA, *x* is the concentration of the Cascade complex, and K_d is the apparent dissociation constant.

4.2.6 Growth conditions

All strains were cultured at 37°C and 250 RPM in up to 5 ml of LB medium (10 g/l tryptone, 5 g/l yeast extract, 5 g/l NaCl) or M9 minimal medium (1X M9 salts, 2 mM MgSO4, 0.1 mM CaCl2, 10 μg/ml thiamine) containing 0.4% glycerol and 0.2% casamino acids. All strains were plated on LB agar (LB medium with 1.2% agar) or on M9 agar with lactose (M9 medium with 1.2% agar and 0.2% D-lactose) in 100x15 mm polystyrene petri dishes. To maintain any plasmids, cells were cultured in liquid medium or on agar plates containing appropriate antibiotics at the following concentrations: 50 μg/ml of ampicillin, 34 μg/ml of chloramphenicol, 50 μg/ml of kanamycin.

4.2.7 Spacer design

See Supplementary Table 4.S4 for a list of all of protospacers targeted in this work and Supplementary Table 4.S5 for pUA66 promoter sequences. Protospacers were selected by identifying a PAM located at the 5' end of the protospacer strand matching the spacer for the Type I-E system in *E. coli* (33). The AAG, AGG, and AAC PAMs were used in this work. A non-functional ACG PAM, shown previously to not support Cascade-based repression, was used as a negative control (34). The 26-56 nucleotides immediately downstream of the PAM were then used as the spacer. The cloning scheme required fixing the final two nucleotides of the spacer to TC, as described previously (23).

4.2.8 Transformation assays

The transformation assay was conducted similar to previous work (23). Briefly, *E. coli* BW25113 Δ*cas3 lacZ*$^{+}$ harboring pCas3 were cultured overnight in LB medium. Cultures were back-diluted 1:250 into 25 ml of LB medium and grown to an ABS_{600} of 0.6 – 0.8 as quantified with a Nanodrop 2000c spectrophotometer (Thermo Fisher Scientific). The cells were then washed in ice-cold 10% glycerol and concentrated by a factor of ~100. A total of 50 µl of the concentrated cells were transformed with 50 ng of plasmid DNA using a MicroPulser electroporator (Bio-Rad). Plasmids encoding the constitutively expressed spacers (pcrRNA.con-XX, where XX is replaced with the spacer name) were used for this assay. Transformed cells were recovered in 300 µl SOC medium for 1 hr at 37ºC. After the recovery period, the cells were diluted up to factors of 10^5 and 200 µl of the dilution were plated on LB agar with appropriate antibiotics.

4.2.9 Flow cytometry analysis

Flow cytometry was conducted similar to previous work (21). Briefly, cells grown overnight in M9 minimal medium with 0.2% casamino acids and 0.4% glycerol were back-diluted to an ABS_{600} of 0.01 into M9 minimal medium containing 0.2% casamino acids, 0.4% glycerol, and appropriate inducers. For the BW25113 Δc*as3*::*cat* cells containing arabinose-inducible plasmids (pcrRNA.ind-XX, where XX is replaced with the spacer name) and pUA66-lacZ, 0.2% L-arabinose and 0.1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) were added as inducers. For the MG1655 Δ*cas3*::*cat* cells harboring constitutively-expressed pcrRNA.con-xylA-s1 with pUA66-xylA, the only inducer required was 0.2% D-xylose. Upon reaching an ABS_{600} of ~0.2 after 3-4 hours of growth, the cultures were diluted 1:100 in phosphate buffered saline (PBS) and run on an Accuri C6 Flow Cytometer (Becton Dickinson) equipped with CFlow plate sampler, a 488 nm laser, and a 530 ± 15 nm bandpass filter. Events reflecting cells were gated based on forward scatter (FSC-H) and side scatter (SSC-H) with respective lower cutoffs of 14,000 and 600 to reduce the measurement of particulates. The gate was set using *E. coli* cells stained with the DRAQ5 dye (Thermo Fisher Scientific). The

fluorescence of the gated cells was then measured in FL1-H. At least 30,000 events were analyzed for each sample.

4.2.10 Statistical analyses

All statistical analyses were conducted using a one-tailed Student's *t*-test with unequal variance. Statistical analyses for the apparent dissociation constants (K_d) assumed that the variance is geometrically distributed, resulting in standard errors of measurement reported as fractions of the geometric mean.

4.3 RESULTS

4.3.1 Spacer length determines Cascade size and composition

Previously determined structures of Cascade revealed a seahorse-shaped complex composed of a Cas6e head, a backbone composed of six Cas7 subunits, a belly consisting of two Cse2 subunits, and a tail composed of one Cse1 and one Cas5 (Supplementary Figure 4.S1A) (15–17, 27). The crRNA is an integral component of this complex that is stretched from head to tail, where each of the Cas7 backbone subunits interacts with six nucleotides of the crRNA spacer sequence. Furthermore, each Cse2 subunit makes direct contacts with two Cas7 subunits (35). Based on this structural information, we hypothesized that the composition and stoichiometry of Cascade might be affected by spacer length. We speculated that extending the spacer would result in an additional Cas7 subunit for every six nucleotides of spacer extension and an additional Cse2 subunit for an extension of twelve nucleotides (Figure 4.1A and Supplementary Figure 4.S1A). To test this hypothesis, we purified Cascade expressed with a CRISPR array containing a naturally occurring spacer of 32 nts (+0) or a spacer extended to 38 (+6) or 44 (+12) nts. Analytical gel filtration and Blue native PAGE of Cascade complexes assembled using the extended crRNAs revealed increases in size that directly correlated with spacer length (Supplementary Figure 4.S1B). Protein and RNA subunits of each complex were resolved using denaturing PAGE (Figure 4.1B). As expected, the larger complexes contained crRNAs that were 6 or 12 nts longer than the crRNA associated with wildtype Cascade. The

extended complexes contained all five of the Cas protein subunits, where densitometry analysis of the denaturing PAGE gels suggested changes in protein subunit stoichiometry in line with our structural predictions (Supplementary Figure 4.S1C).

To determine the subunit stoichiometry of the wildtype and engineered Cascade complexes, we performed native mass spectrometry to measure intact masses of the +0, +6, and +12 complexes (Figure 4.1A and Supplementary Table 4.S6). The mass of the wildtype (+0) complex was 405,238.2 ± 45.8 Da, while the +6 and +12 complexes had masses of 446,555.1 ± 209.9 Da and 510,127.2 ± 48.9 Da, respectively. The +6 complex was ~41 kDa larger than the wildtype (+0) complex, which is consistent with the addition of a Cas7 subunit (~40 kDa) plus the mass of an additional 6 nts to the crRNA (~1.6 kDa). The addition of 12 nts resulted in a complex ~105 kDa larger than the wildtype complex, which matches the predicted addition of two Cas7 subunits (~80 kDa), one Cse2 subunit (~21 kDa), and 12 nts to the crRNA (~3.5 kDa). Collectively, these results provide evidence that additional Cas7 subunits are incorporated into Cascade when the spacer is extended in 6-nt increments, and additional Cse2 subunits are incorporated in 12-nt increments.

To determine how spacer length impacts target binding, we measured the binding affinity of Cascade with a wildtype spacer (+0) or with a spacer extended by 12 nucleotides (+12) to a dsDNA target. For two different spacers (s1 and p1), the +0 and +12 complexes bound target DNA with high affinity (Figure 4.1C and Supplementary Figure 4.S2) (12, 35). The apparent dissociation constant was marginally lower for the +12 complex around the level of significance (P = 0.062 for s1, P = 0.041 for p1, n = 3) (Supplementary Figure 4.S2), suggesting that the spacer extension at least maintained affinity for the target.

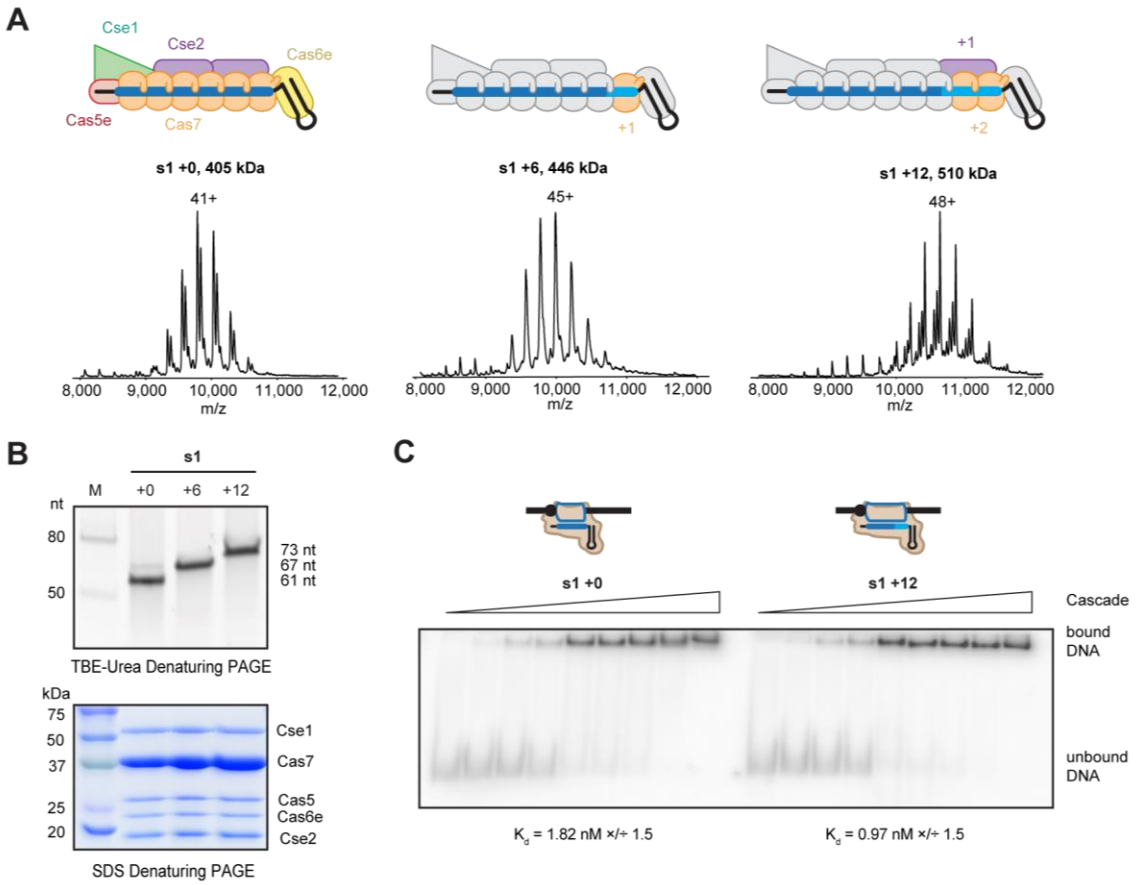

Figure 4.1 *Escherichia coli* Type I-E Cascades with extended spacers form stable complexes with altered stoichiometry. Native mass spectra of Cascade complexes containing crRNA with spacers that are either 32 (+0), 38 (+6) or 44 nts (+12) in length and schematic representations of the corresponding Cascade complexes. Extending the crRNA spacer by 6 or 12 nts (light blue) results in complexes that are ~41 kDa or ~105 kDa larger than wild-type. The additional mass is consistent with the mass of the extended crRNA (light blue), Cas7 (orange) and in the case of the +12 complex an additional Cse2 subunit (purple). (B) SYBR-stained denaturing PAGE of crRNAs purified from Cascade complexes containing a 61-nt crRNA with a 32-nt spacer (+0), a 67-nt crRNA with a 38-nt spacer (+6), or a 73-nt crRNA with a 44-nt spacer (+12) (top) and SDS-PAGE stained with Coomassie blue (bottom). (C) Cascade complexes with a 32-nt spacer (+0, left) or a 44-nt spacer (+12, right) bind DNA targets with high affinity. Shown below each gel is the apparent dissociation constant (K_d), reported as the geometric mean and S.E.M. for three independent experiments. Experimental data and curve fits for the individual experiments are shown in Supplementary Figure 4.S2.

4.3.2 Extended spacers permit transcriptional silencing and DNA interference

Our data demonstrated that Cascade accommodates extended spacers *in vitro*, although it remained unclear how longer spacers would impact Cascade activity *in vivo*. To determine the *in vivo* effect of elongated spacers on Cascade function, we used a previously developed gene repression assay (21, 22). This system relies on an *E. coli* strain (BW25113 Δ*cas3*::*cat*) in which *cas3* was deleted and a constitutive promoter was introduced upstream of the *cse1-cse2-cas7-cas5-cas6e* operon (21). We transformed this strain with an L-arabinose-inducible plasmid encoding a designed CRISPR array and the pUA66-lacZ reporter plasmid encoding the green fluorescent protein (*gfp*) gene downstream of the *lacZ* promoter (Figure 4.2A). The spacers were designed to be 32 nts (+0) or 44 nts (+12) and target locations within the promoter and either strand of the transcribed region. For all sites targeted, the +12 spacers exhibited similar or improved silencing of GFP compared to the +0 spacers, indicating that extended Cascade complexes are capable of repressing expression of a target gene (Figure 4.2B). We observed the strongest silencing when targeting the promoter, intermediate silencing when targeting the sense strand of the transcribed region, and weak silencing when targeting the antisense strand of the transcribed region (Figure 4.2C), which is in line with previous work (21).

We next asked how extended crRNAs impact the ability of Cascade to recruit Cas3 and elicit targeted DNA degradation (Figure 4.2D). Hybridization of the crRNA spacer sequence to the complementary strand of a dsDNA target forms a displaced R-loop that is critical for Cas3 recruitment and DNA target degradation. (12, 36). To measure Cas3 activity, we constitutively expressed *cas3* from a plasmid and measured the transformation efficiency of plasmids encoding genome-targeting crRNAs with 32-nt (+0) or 44-nt (+12) spacers. Previous work has shown that CRISPR-mediated genome targeting is lethal, resulting in greatly reduced transformation efficiencies in comparison to a non-targeting control (23, 37). We used the *lacZ*-targeting spacers from the transcriptional silencing assay (p1, p2, as1, and s1), which required restoring the genomic *lacZ* locus by P1 transduction to generate BW25113 Δ*cas3* *lacZ*$^{+}$. The transformation assays showed that all spacers resulted in ~10^{3}-10^{4} reduction in

transformation efficiency when compared to a spacer-free control (Figure 4.2E). Therefore, extended Cascade complexes can recruit and activate Cas3 for target destruction, demonstrating that extended Cascade complexes can direct CRISPR-mediated interference.

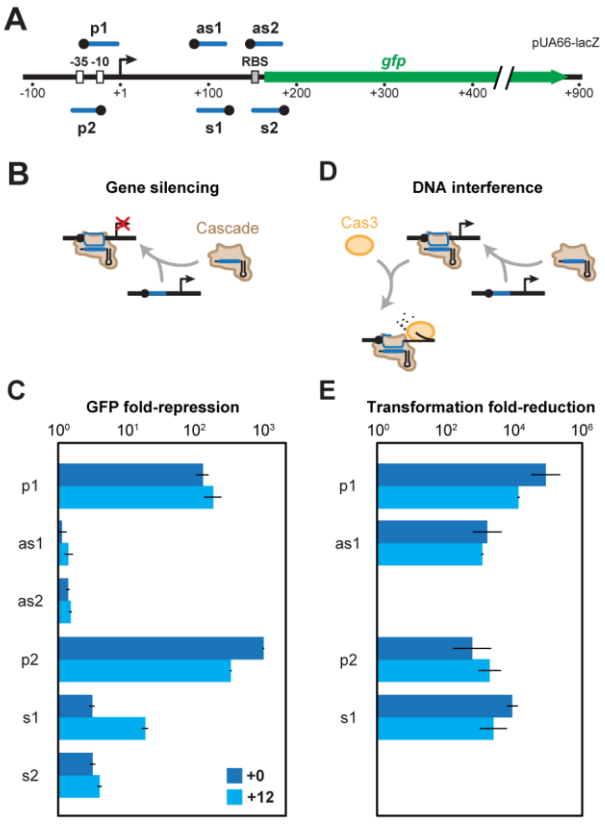

Figure 4.2 Extended Cascade assemblies repress gene expression and impair transformation efficiency *in vivo*. (A) Cascade complexes were designed to target sequences in the pUA66-lacZ reporter. Blue bars above or below the reporter represent spacers that match the top and bottom strands, respectively. Locations include the promoter (p) as well as the sense (s) and antisense (as) strand of *lacZ*. Black dots represent PAMs. Protospacer and PAM sequences can be found in Supplementary Table 4.S4. (B) A schematic illustrating how Cascade binding may block gene transcription. (C) Impact of extending the spacer by 0 (dark blue) or 12 (light blue) nts on GFP silencing by Cascade. Fold-repression is calculated in comparison to a non-targeting control. (D) A schematic illustrating how the addition of Cas3 results in autoimmunity and reduced colonies after transformation of the plasmid containing Cas3. (E) Impact of extending the spacer by 0 (dark blue) or 12 (light blue) nts on DNA interference by Cascade and Cas3 is indicated (bottom). The transformation fold-reduction is calculated in comparison to a spacer-free control. The as2 and s2 spacers were not tested because the associated protospacers were not present in the genomic *lacZ* locus. Values represent the mean and S.E.M. of at least three measurements starting from independent colonies.

4.3.3 Extended spacers can improve Cascade-mediated silencing

Extending the spacer markedly improved Cascade-mediated silencing for the s1 spacer complementary to the sense strand upstream of the ribosomal binding site (from 3-fold to 20-fold) (Figure 4.2C). To determine the relationship between the length of s1 and silencing activity, we designed spacers modified by multiples of six based on the interval of Cas7 binding (Figure 4.3A and Supplementary Figure 4.S1A) (15). Silencing generally improved with each additional six nucleotides at the s1 site (Figure 4.3B), while no trend was observed when extending a promoter-targeting spacer (p1) or an antisense-targeting spacer (as2) (Supplementary Figure 4.S3). Extended spacers that deviated from multiples of six, (+8 and +10) also showed improvement in silencing efficiency compared to wildtype, whereas spacers shortened by six nucleotides lost silencing (Figure 4.3B).

Interestingly, enhanced silencing appears to be related to target location, as two spacers (s3 and s4) with targets near s1 improved silencing when extended (Figure 4.3B). This phenomenon was not limited to the *lacZ* promoter, as we observed enhanced silencing when targeting a similar location within the *xylA* promoter (Figure 4.3C and 4.D), demonstrating that elongated Cascade complexes can be utilized to control silencing in different genetic contexts. To our knowledge, this serves as the first instance of quantitatively enhancing CRISPR-mediated silencing by increasing spacer length (38).

Enhanced silencing was only associated with spacers targeting the sense strand upstream of the coding region. This orientation could allow the crRNA to base pair with the transcribed *gfp* mRNA, potentially resulting in mRNA destabilization and translational inhibition; accordingly, longer crRNAs may be more efficient at base pairing for antisense regulation. To test whether the CRISPR array acts through antisense regulation (i.e. crRNA-guided binding of complementary mRNA), we assessed GFP silencing in the absence of the Cascade proteins. However, GFP silencing by the +0 and the +12 spacers was negligible (Figure 4.3F). To test whether enhanced silencing could be attributed to Cascade binding the *gfp* mRNA, we mutated the PAM to disrupt dsDNA binding but preserve PAM-independent RNA binding (Figure 4.3E) (17, 39). Again, the +0 and +12 spacers yielded negligible GFP

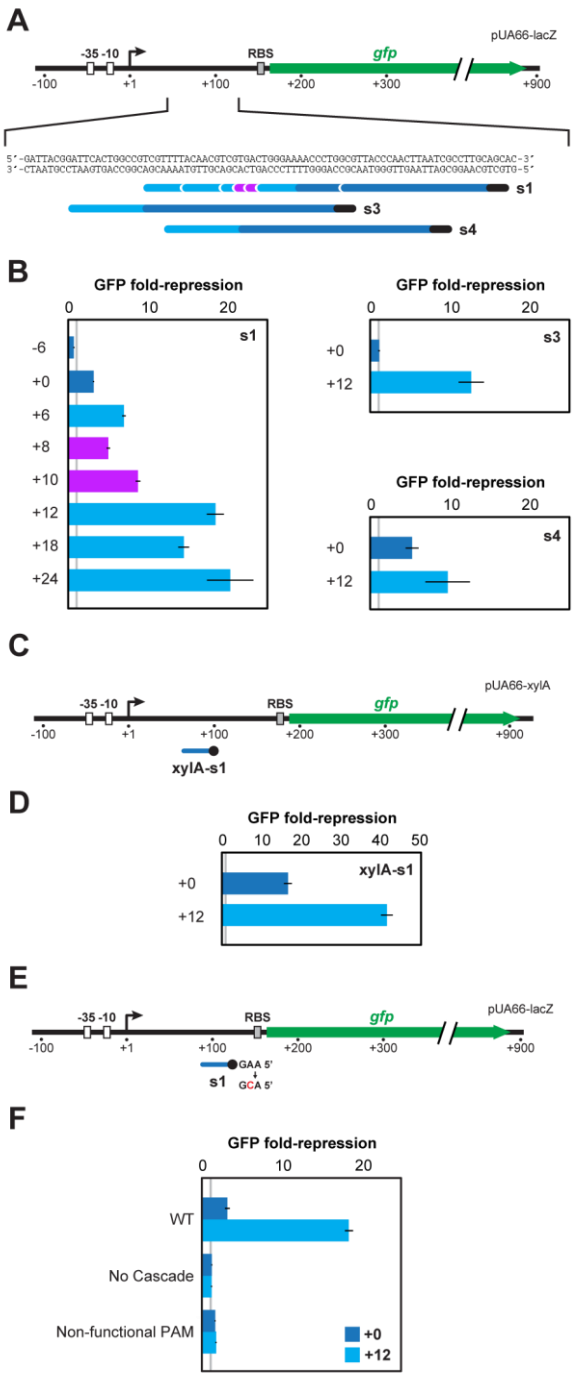

Figure 4.3 Spacer sequence extension can improve Cascade-mediated gene silencing. (A) Target locations within close proximity to the s1 spacer. The 32-nt spacer (dark blue) was extended by lengths divisible by six (light blue) or not divisible by six (purple). The 32-nt spacer for s1 was also shortened by 6 nts (dark blue). Black bars represent the PAM. Sequences match the bottom strand of the reporter

construct. (B) Gene silencing resulting from Cascade complexes with spacers s1, s3 and s4 of varying sizes. The gray line indicates a fold-repression value of 1 reflecting no change in GFP fluorescence. (C) Targeting the sense strand of the untranslated region of the *xylA* promoter. The black dot indicates the PAM. To evaluate GFP silencing, the pUA66-xylA reporter construct was transformed into MG1655Δ*cas3*::*cat* along with a plasmid encoding a constitutively expressed CRISPR array containing the xylA-s1 spacer. (D) Impact of extending the *xylA*-targeting spacer by +0 nts (dark blue) or +12 (light blue) nts on GFP silencing by Cascade. Fold-repression is calculated in comparison to a non-targeting control. (E) Evaluating the basis of spacer length-dependent silencing with Cascade. Testing a non-functional PAM required mutating the PAM of the s1 protospacer in the reporter construct. (F) Potential RNA-based antisense regulation was evaluated in the absence (No Cascade) or presence (Non-functional PAM) of Cascade. Values represent the mean and S.E.M. of at least three measurements starting from independent colonies.

silencing (Figure 4.3F). Therefore, antisense regulation by the free crRNA or Cascade-bound crRNA cannot explain enhanced silencing with longer spacers.

4.3.4 Mismatches between the DNA target and the extended spacer impair Cascade activity

We finally asked whether base pairing through the extended region of the longer crRNAs is important for Cascade binding and activity. One possibility is that a longer crRNA with a mutated extended region would still be capable of base pairing through the first 32 nucleotides of the natural spacer. However, recent work has shown that base pairing at the 3' end of a normal spacer contributes to Cascade binding (36, 40–42). To initially explore the importance of the extended spacer region, we measured the *in vitro* binding affinity of purified Cascade complexes in which the extended region of the p1+12 spacer was mutated to disrupt base-pairing potential with the target (Supplementary Figure 4.S4). Mutations that perturb base pairing in the extended region of the crRNA significantly lowered the apparent binding affinity ($P = 0.004$, $n = 3$) (Figure 4.4A and Supplementary Figure 4.S2), indicating that the extended region is important for Cascade to bind DNA *in vitro*. These mutations also required higher concentrations of Cascade to achieve saturated binding of the DNA target (Figure 4.4A and Supplementary Figure 4.S2), suggesting that higher Cascade concentrations interfered with target binding.

To determine how mutations in the extended region impact transcriptional repression by Cascade, we tested different numbers of mismatches introduced at the 3' end of the +12

spacer (Supplementary Figure 4.S4). As a basis of comparison, we made similar mutations to the +0 spacer. The p1 spacer targeting the *lacZ* promoter was selected because the +0 and +12 versions of the spacer yielded similar levels of gene repression (Figure 4.2C). We found that gene silencing for either spacer was reduced with each additional mutation (Figure 4.4B). While the +0 spacer was more susceptible to the number of mutations, the susceptibilities were similar when accounting for mutations as a fraction of the total spacer length (Figure 4.4C). We also observed similar susceptibilities for the +12 variant of the s1 spacer (Supplementary Figure 4.S5), suggesting that the mutational susceptibilities applied to different protospacer sequences. These results demonstrate that the extended portion of the spacer is critical for Cascade-mediated gene repression *in vivo*.

To assess how mutations in the extended spacer impact DNA interference, we tested the +0 and +12 variants of the p1 spacer with the same set of mutations at the 3' end in the presence of Cas3 expression (Figure 4.4B). We found that more mismatches at the 3' end led to more transformants, paralleling the trend observed with the transcriptional silencing assay. However, the +12 spacer maintained the lower transformation efficiency for up to six mismatches while the +0 spacer showed gradually improved transformation efficiencies with additional mismatches. Thus, extended and regular spacers exhibit different propensities for DNA interference and escape based on the number of mismatches at the 3' end.

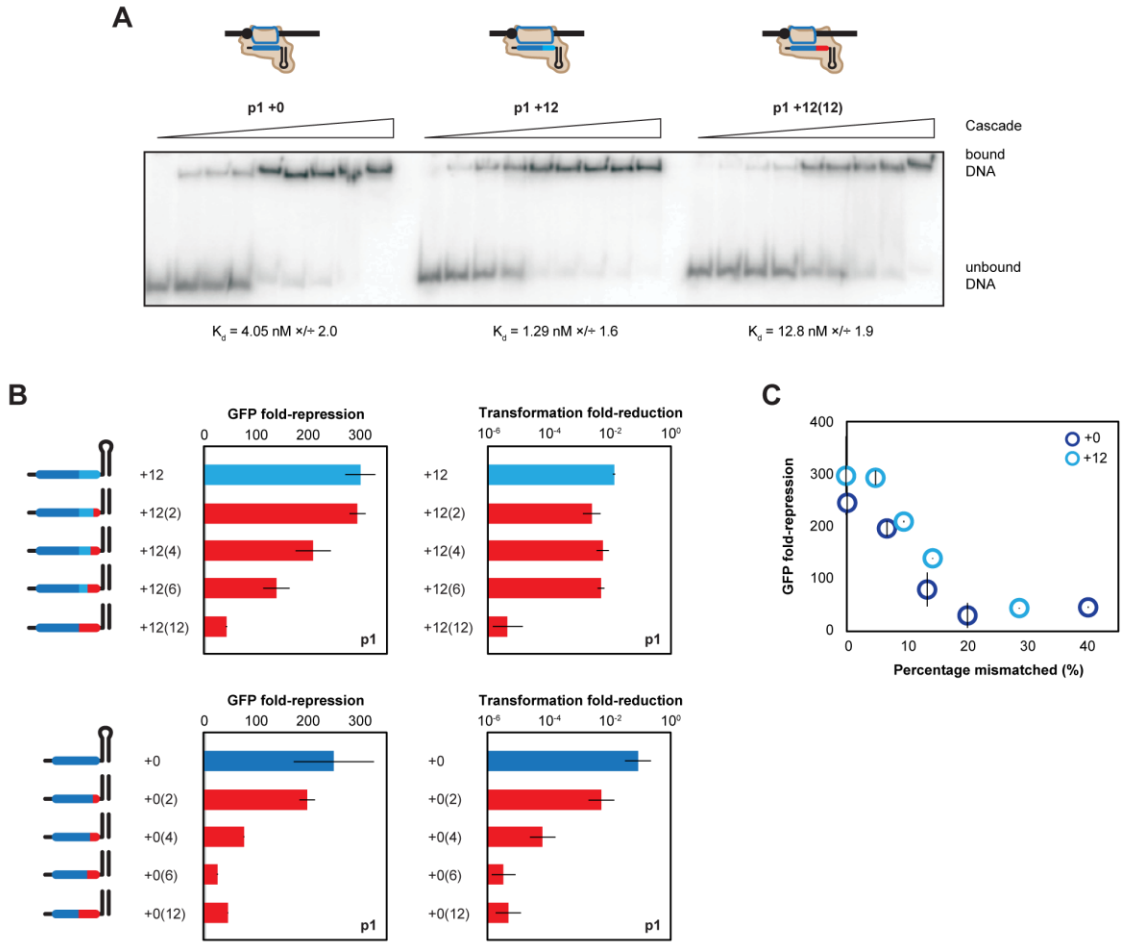

Figure 4.4 Mismatches between spacer extensions and DNA targets impair Cascade activity.(A) Schematics of Cascade complexes are depicted (top) with regular spacer length colored dark blue, complementary extensions colored in light blue and mismatching extensions in red. Representative gel images from electrophoretic mobility shift assays for each of the depicted complexes are shown. Shown below each gel is the apparent dissociation constant (K_d), reported as the geometric mean and S.E.M. for three independent experiments. Experimental data and curve fits for the individual experiments are shown in Supplementary Figure 4.S2. (B) Impact of mutating the 3' end of the spacer on gene silencing (left) and DNA interference (right) for the extended spacer variant (top) and the wild type length (bottom). The number in parentheses indicates the number of mutations made to the 3' end of the s1 +0 or s1 +12 spacer. See Supplementary Figure 4.S5 for the sequence of each mutated spacer. (C) Silencing efficiency plotted by the mismatch percentage of the total spacer length for the p1 spacer. Dark blue circles refer to the +0 spacer. Light blue circles refer to the +12 spacer. Values represent the mean and S.E.M. of at least three measurements starting from independent colonies.

4.4 DISCUSSION

We found that extending the crRNA spacer altered the size and composition of the I-E Cascade complex from *E. coli*. Our *in vitro* data indicated that Cascade incorporates an extra Cas7 subunit with every 6 nts added to the crRNA spacer and an extra Cse2 subunit with every 12 nts added to the spacer. Furthermore, our *in vivo* repression data demonstrated that the altered complex maintained stable DNA binding and Cas3-mediated DNA degradation, where both functions required base pairing through the extended region of the spacer. Given that these functions require the wildtype complex to undergo large conformational rearrangements (16, 27, 35), our results suggest that the additional subunits still permit the necessary conformational changes to stably bind DNA and recruit Cas3 for target degradation.

We also found that longer spacers exhibited enhanced silencing when targeting the DNA sense strand between the promoter and the coding regions of *lacZ* and *xylA*. This trend suggests that improved silencing with extended spacers may be a general phenomenon. We were able to eliminate antisense interactions as possible explanations, though pinpointing the responsible mechanism will require further studies. Nevertheless, our results offer the possibility of tuning the extent of silencing simply by adjusting the number of nucleotides in the spacer. While we extended the spacer by up to 80% of its wildtype length (+24 nts), the Type I-E Cascade may be able to accommodate even longer spacers.

Despite the ability of the Type I-E Cascade to tolerate extended spacers, the naturally occurring spacer lengths for Type I-E systems are largely fixed (26). The natural selection of 32-nt spacers may reflect the mechanism of protospacer acquisition, which consistently integrates spacers of defined length. Accordingly, overexpression of the universal Cas1 and Cas2 acquisition proteins in *E. coli* leads to the incorporation of ~32-nt spacers into the CRISPR locus (43–46).

The unique relationship between spacer length and complex size may be applicable across Class 1 systems based on the general assembly of their effector complexes (47). Evidence for this comes from Type III systems, which naturally generate different crRNA lengths and complex sizes. For example, the Type III-A system of *Streptococcus thermophilus*

processes crRNAs to two distinct lengths where the longer crRNA results in a larger Csm effector complex (48). Similarly, the Type III-B systems from *Thermus thermophilus* and *Pyrococcus furiosus* have been shown to bind crRNAs of multiple lengths that differ by 6 nts (49–51). These insights from multiple species and CRISPR-Cas subtypes suggest that the assemblies of multi-subunit Class 1 complexes are templated by the length of the crRNA. In contrast, the single protein effectors of Class II systems appear to maintain a crRNA spacer that is not amenable to extension (52–57).

In summary, our work provides further insight into the flexibility and extent to which crRNAs can be modified. The length of Type I-E crRNAs is not fixed and can be substantially extended, altering Cascade stoichiometry while maintaining *in vivo* functionality. These observations present additional opportunities to explore how other Class 1 systems respond to alterations to the crRNAs. Lastly, we uncovered a unique situation that offers the possibility of designing tunable transcriptional regulators whose design is dependent solely on the length of the crRNA. These insights are expected to inform our understanding of CRISPR biology and how these adaptive immune systems can be altered and improved toward applications in medicine and biotechnology.

ACKNOWLEDGMENTS

We thank Rodolphe Barrangou for critical discussions and Charlotte Uetrecht technical insights that were critical for measuring the intact masses of Cascade. Authors' contributions: M.L.L and C.L.B. conceived the project. R.N.J. and W.L. performed the *in vitro* experiments. M.T-L and B.B. performed the mass spectrometry. M.L.L., S.R.D., and K.R.M. performed the *in vivo* work. M.L.L., R.N.J., B.W., and C.L.B. wrote the manuscript.

FUNDING

The work was supported by the National Science Foundation [CBET-1403135 to C.L.B., IIP-1549648 to C.L.B., EPS-110134 to B.W.]; the National Institutes of Health [F32GM108436 to R.N.J., P20GM103500 to B.W., R01GM108888 to B.W., 5P20RR02437 to B.B.]; the

Department of Energy, [DE-SC0012518 to B.B.], the Kenan Institute of Engineering, Technology, and Science to C.L.B.; the M.J. Murdock Charitable Trust to B.W. and B.B.; a young investigator award from Amgen to B.W.; and the Montana State University Agricultural Experimental Station to B.W.

REFERENCES

1. Van der Oost,J., Westra,E.R., Jackson,R.N. and Wiedenheft,B. (2014) Unravelling the structural and mechanistic basis of CRISPR-Cas systems. *Nat. Rev. Microbiol.*, **12**, 479–492.

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

3. van der Oost,J., Jore,M.M., Westra,E.R., Lundgren,M. and Brouns,S.J.J. (2009) CRISPR-based adaptive and heritable immunity in prokaryotes. *Trends Biochem. Sci.*, **34**, 401–407.

4. Bhaya,D., Davison,M. and Barrangou,R. (2011) CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annu. Rev. Genet.*, **45**, 273–297.

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

6. Shmakov,S., Abudayyeh,O.O., Makarova,K.S., Wolf,Y.I., Gootenberg,J.S., Semenova,E., Minakhin,L., Joung,J., Konermann,S., Severinov,K., *et al.* (2015) Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. *Mol. Cell*, **60**, 385–397.

7. Sinkunas,T., Gasiunas,G., Fremaux,C., Barrangou,R., Horvath,P. and Siksnys,V. (2011) Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. *EMBO J.*, **30**, 1335–1342.

8. Huo,Y., Nam,K.H., Ding,F., Lee,H., Wu,L., Xiao,Y., Farchione,M.D., Zhou,S., Rajashankar,K., Kurinov,I., *et al.* (2014) Structures of CRISPR Cas3 offer mechanistic insights into Cascade-activated DNA unwinding and degradation. *Nat. Struct. Mol. Biol.*, **21**, 771–777.

9. Westra,E.R., van Erp,P.B.G., Künne,T., Wong,S.P., Staals,R.H.J., Seegers,C.L.C.,

Bollen,S., Jore,M.M., Semenova,E., Severinov,K., *et al.* (2012) CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Mol. Cell*, **46**, 595–605.

10. Jackson,R.N., Lavin,M., Carter,J. and Wiedenheft,B. (2014) Fitting CRISPR-associated Cas3 into the Helicase Family Tree. *Curr. Opin. Struct. Biol.*, **24**, 106–114.

11. Mulepati,S. and Bailey,S. (2013) *In vitro* reconstitution of an *Escherichia coli* RNA-guided immune system reveals unidirectional, ATP-dependent degradation of DNA target. *J. Biol. Chem.*, **288**, 22184–92.

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

13. Redding,S., Sternberg,S.H., Marshall,M., Gibb,B., Bhat,P., Guegler,C.K., Wiedenheft,B., Doudna,J.A. and Greene,E.C. (2015) Surveillance and Processing of Foreign DNA by the *Escherichia coli* CRISPR-Cas System. *Cell*, **163**, 854–865.

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

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

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

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

18. Brouns,S.J.J., Jore,M.M., Lundgren,M., Westra,E.R., Slijkhuis,R.J.H., Snijders,A.P.L., Dickman,M.J., Makarova,K.S., Koonin,E.V. and van der Oost,J. (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, **321**, 960–964.

19. Sinkunas,T., Gasiunas,G., Waghmare,S.P., Dickman,M.J., Barrangou,R., Horvath,P. and Siksnys,V. (2013) *In vitro* reconstitution of Cascade-mediated CRISPR immunity in Streptococcus thermophilus. *EMBO J.*, **32**, 385–394.

20. Luo,M.L., Leenay,R.T. and Beisel,C.L. (2016) Current and future prospects for CRISPR-based tools in bacteria. *Biotechnol. Bioeng.*, **113**, 930–43.

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

22. Rath,D., Amlinger,L., Hoekzema,M., Devulapally,P.R. and Lundgren,M. (2015) Efficient programmable gene silencing by Cascade. *Nucleic Acids Res.*, **43**, 237–246.

23. Gomaa,A.A., Klumpe,H.E., Luo,M.L., Selle,K., Barrangou,R. and Beisel,C.L. (2014) Programmable removal of bacterial strains by use of genome-targeting CRISPR-Cas systems. *MBio*, **5**, e00928–13.

24. Caliando,B.J. and Voigt,C.A. (2015) Targeted DNA degradation using a CRISPR device stably carried in the host genome. *Nat. Commun.*, **6**, 6989.

25. Li,Y., Pan,S., Zhang,Y., Ren,M., Feng,M., Peng,N., Chen,L., Liang,Y.X. and She,Q. (2016) Harnessing Type I and Type III CRISPR-Cas systems for genome editing. *Nucleic Acids Res.*, **44**, e34.

26. Díez-Villaseñor,C., Almendros,C., García-Martínez,J. and Mojica,F.J.M. (2010) Diversity of CRISPR loci in *Escherichia coli*. *Microbiology*, **156**, 1351–1361.

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

28. Cherepanov,P.P. and Wackernagel,W. (1995) Gene disruption in *Escherichia coli*: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. *Gene*, **158**, 9–14.

29. Zaslaver,A., Bren,A., Ronen,M., Itzkovitz,S., Kikoin,I., Shavit,S., Liebermeister,W., Surette,M.G. and Alon,U. (2006) A comprehensive library of fluorescent transcriptional reporters for *Escherichia coli*. *Nat. Methods*, **3**, 623–628.

30. Wittig,I., Braun,H.-P. and Schägger,H. (2006) Blue native PAGE. *Nat. Protoc.*, **1**, 418–428.

31. Kirshenbaum,N., Michaelevski,I. and Sharon,M. (2010) Analyzing large protein complexes by structural mass spectrometry. *J. Vis. Exp.*, **40**, e1954.

32. Michaelevski,I., Kirshenbaum,N. and Sharon,M. (2010) T-wave ion mobility-mass spectrometry: basic experimental procedures for protein complex analysis. *J. Vis. Exp.*, **41**, e1985.

33. Westra,E.R., Semenova,E., Datsenko,K.A., Jackson,R.N., Wiedenheft,B., Severinov,K. and Brouns,S.J.J. (2013) Type I-E CRISPR-cas systems discriminate target from non-target DNA through base pairing-independent PAM recognition. *PLoS Genet.*, **9**, e1003742.

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

35. van Erp,P.B.G., Jackson,R.N., Carter,J., Golden,S.M., Bailey,S. and Wiedenheft,B. (2015) Mechanism of CRISPR-RNA guided recognition of DNA targets in *Escherichia coli*. *Nucleic Acids Res.*, **43**, 8381–8391.

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

37. Vercoe,R.B., Chang,J.T., Dy,R.L., Taylor,C., Gristwood,T., Clulow,J.S., Richter,C., Przybilski,R., Pitman,A.R. and Fineran,P.C. (2013) Cytotoxic Chromosomal Targeting by CRISPR/Cas Systems Can Reshape Bacterial Genomes and Expel or Remodel Pathogenicity Islands. *PLoS Genet.*, **9**, e1003454.

38. Mougiakos,I., Bosma,E.F., de Vos,W.M., van Kranenburg,R. and van der Oost,J. (2016) Next Generation Prokaryotic Engineering: The CRISPR-Cas Toolkit. *Trends Biotechnol.*, doi:10.1016/j.tibtech.2016.02.004.

39. Sashital,D.G., Wiedenheft,B. and Doudna,J.A. (2012) Mechanism of foreign DNA selection in a bacterial adaptive immune system. *Mol. Cell*, **46**, 606–615.

40. Blosser,T.R., Loeff,L., Westra,E.R., Vlot,M., Künne,T., Sobota,M., Dekker,C., Brouns,S.J.J. and Joo,C. (2015) Two Distinct DNA Binding Modes Guide Dual Roles of a CRISPR-Cas Protein Complex. *Mol. Cell*, **58**, 60–70.

41. Szczelkun,M.D., Tikhomirova,M.S., Sinkunas,T., Gasiunas,G., Karvelis,T., Pschera,P., Siksnys,V. and Seidel,R. (2014) Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proc. Natl. Acad. Sci. U. S. A.*, **111**, 9798–9803.

42. Fineran,P.C., Gerritzen,M.J.H., Suárez-Diez,M., Künne,T., Boekhorst,J., van Hijum,S.A.F.T., Staals,R.H.J. and Brouns,S.J.J. (2014) Degenerate target sites mediate rapid primed CRISPR adaptation. *Proc. Natl. Acad. Sci. U. S. A.*, **111**, E1629–1638.

43. Yosef,I., Goren,M.G. and Qimron,U. (2012) Proteins and DNA elements essential for the

CRISPR adaptation process in *Escherichia coli*. *Nucleic Acids Res.*, **40**, 5569–5576.

44. Arslan,Z., Hermanns,V., Wurm,R., Wagner,R. and Pul,Ü. (2014) Detection and characterization of spacer integration intermediates in type I-E CRISPR-Cas system. *Nucleic Acids Res.*, **42**, 7884–7893.

45. Nuñez,J.K., Lee,A.S.Y., Engelman,A. and Doudna,J.A. (2015) Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. *Nature*, **519**, 193–198.

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

47. Jackson,R.N. and Wiedenheft,B. (2015) A Conserved Structural Chassis for Mounting Versatile CRISPR RNA-Guided Immune Responses. *Mol. Cell*, **58**, 722–728.

48. Tamulaitis,G., Kazlauskiene,M., Manakova,E., Venclovas,C., Nwokeoji,A.O., Dickman,M.J., Horvath,P. and Siksnys,V. (2014) Programmable RNA Shredding by the Type III-A CRISPR-Cas System of Streptococcus thermophilus. *Mol. Cell*, **56**, 506–517.

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

50. Hale,C., Kleppe,K., Terns,R.M. and Terns,M.P. (2008) Prokaryotic silencing (psi)RNAs in Pyrococcus furiosus. *RNA*, **14**, 2572–2579.

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

52. Ran,F.A., Hsu,P.D., Lin,C.-Y., Gootenberg,J.S., Konermann,S., Trevino,A.E., Scott,D.A., Inoue,A., Matoba,S., Zhang,Y., *et al.* (2013) Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. *Cell*, **154**, 1380–1389.

53. Shechner,D.M., Hacisuleyman,E., Younger,S.T. and Rinn,J.L. (2015) Multiplexable, locus-specific targeting of long RNAs with CRISPR-Display. *Nat. Methods*, **12**, 664–670.

54. Qi,L.S., Larson,M.H., Gilbert,L.A., Doudna,J.A., Weissman,J.S., Arkin,A.P. and Lim,W.A. (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, **152**, 1173–1183.

55. Mali,P., Aach,J., Stranges,P.B., Esvelt,K.M., Moosburner,M., Kosuri,S., Yang,L. and Church,G.M. (2013) CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.*, **31**, 833–838.

56. Dahlman,J.E., Abudayyeh,O.O., Joung,J., Gootenberg,J.S., Zhang,F. and Konermann,S. (2015) Orthogonal gene knockout and activation with a catalytically active Cas9 nuclease. *Nat. Biotechnol.*, **33**, 1159–1161.

57. Zalatan,J.G., Lee,M.E., Almeida,R., Gilbert,L.A., Whitehead,E.H., La Russa,M., Tsai,J.C., Weissman,J.S., Dueber,J.E., Qi,L.S., *et al.* (2015) Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell*, **160**, 339–350.

SUPPLEMENTARY INFORMATION

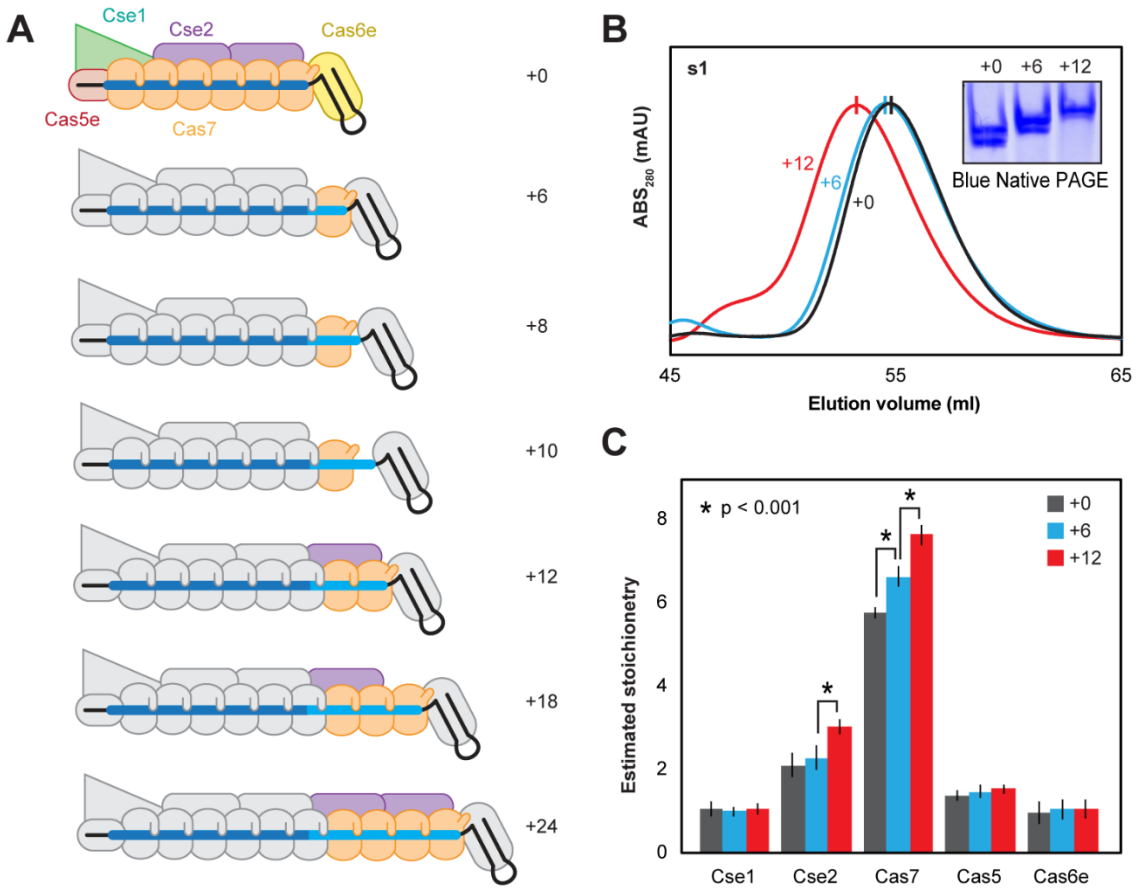

Figure 4.S1 Spacer extension changes protein subunit stoichiometry of *E. coli* Type I-E Cascade. (A) Schematics of putative subunit stoichiometry for Cascade complexes bound to a CRISPR RNA with spacer lengths of 32 nts (+0), 38 nts (+6), 40 nts (+8), 42 nts (+10), 44 nts (+12), 50 nts (+18), and +56 nts (+24) depicting how changes in CRISPR RNA spacer length may allow for the addition of Cas7 and Cse2 subunits. (B) Gel filtration and Blue Native PAGE (inset) of Cascade complexes (+0), (+6), and (+12) show distinct differences in complex size. (C) Band quantification and densitometry analysis of SDS-PAGE assays indicates (+0), (+6) and (+12) complexes consist of the respective subunit stoichiometry depicted in A. The star indicates $p < 0.001$ for $n = 5$. Native Mass Spectrometry also confirmed the predicted stoichiometry for (+0), (+6), and (+12) (see Figure 4.1).

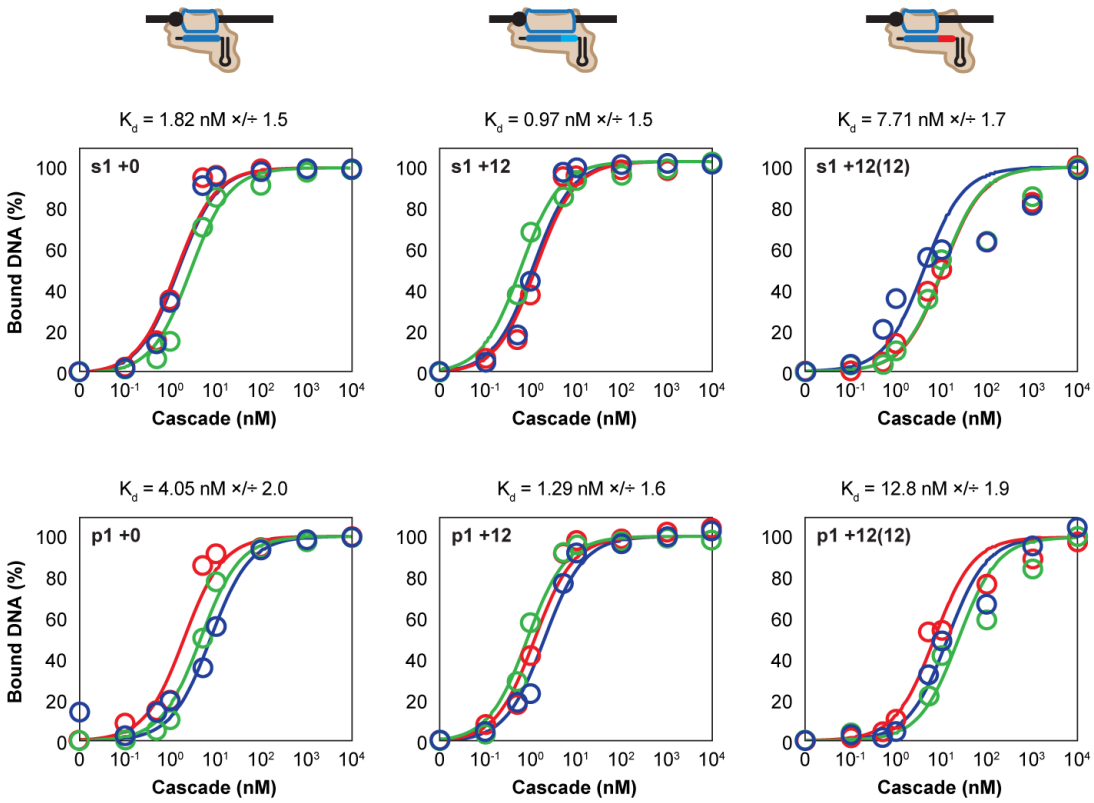

Figure 4.S2 Curve fits for electrophoretic mobility shift assays. The spacer associated with each binding curve is indicated in the upper-left of each panel. Each color represents an independent assay, where circles represent experimental measurements and the curves represent fit Hill equations. The fit values for the apparent dissociation constant (K_d) are shown above each plot.

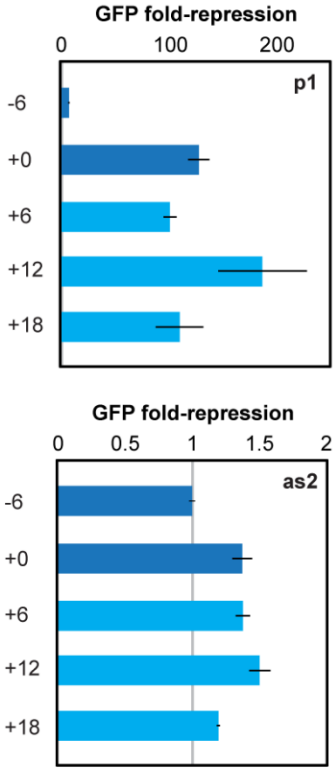

Figure 4.S3 Impact of shortening or extending the spacer on gene silencing by Cascade for the p1 and as2 spacers. Fold-repression is calculated from comparison to a non-targeting control. The gray line indicates a fold-repression value of 1 reflecting no change in GFP fluorescence from the non-targeting control. The depicted values represent the mean and S.E.M. of at least three measurements starting from independent colonies.

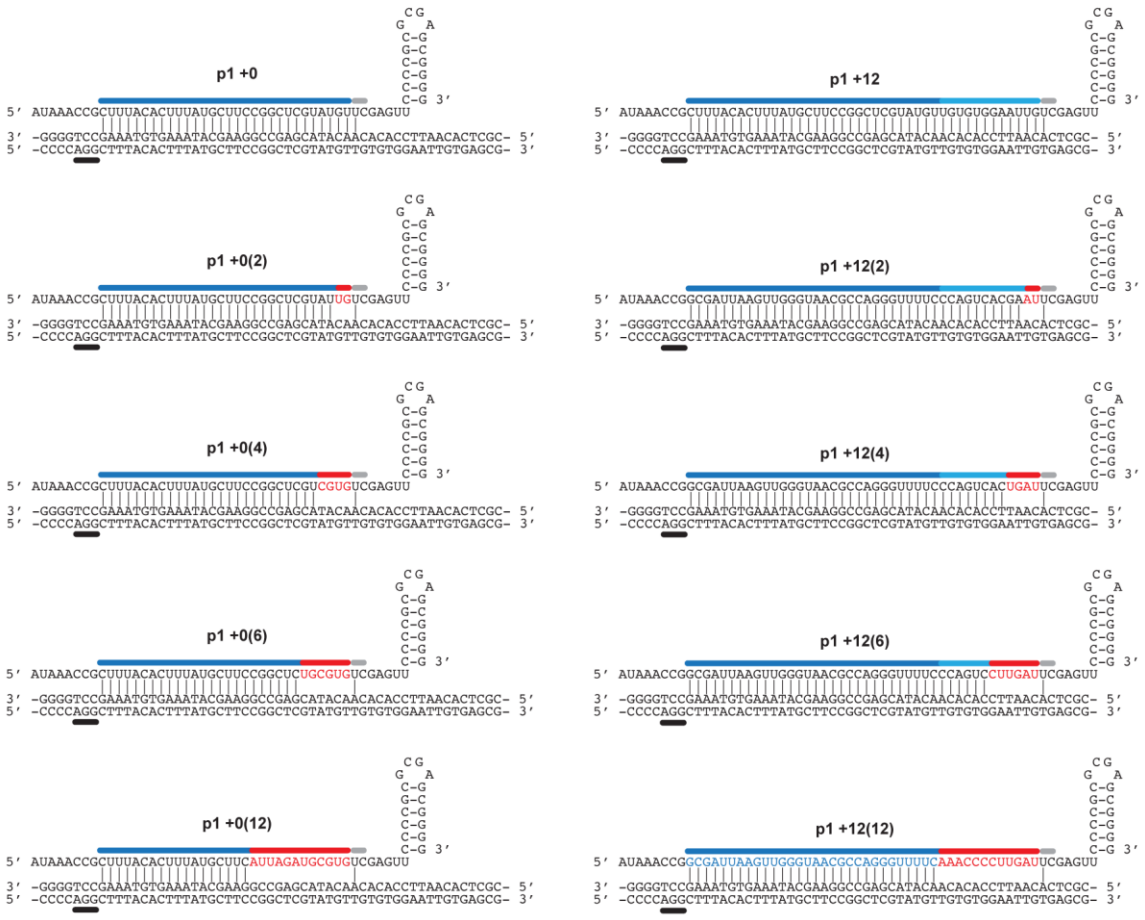

Figure 4.S4 Sequences of p1 spacer variants. The number in parentheses captures the number of mutations made to the 3' end of the +0 or +12 p1 spacer. Black bars indicate the PAM sequence. Dark blue bars indicate the +0 spacer sequence, light blue bars indicate the extended spacer sequence, and red bars indicate mismatched nucleotides.

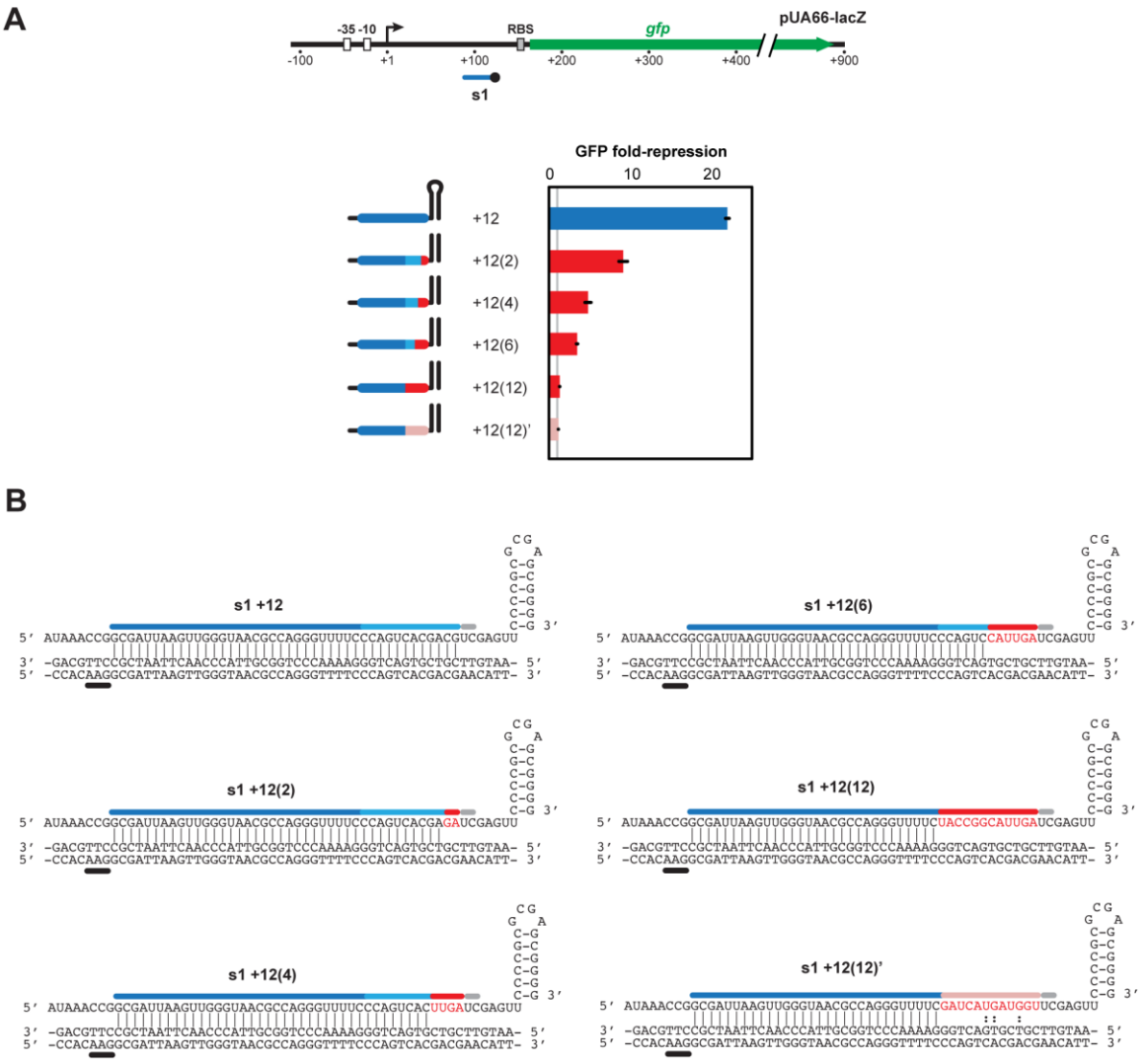

Figure 4.S5 Mismatches on the 3' end of the crRNA reduce gene silencing efficacy with the s1 +12 spacer. (A) A schematic showing the location of the s1 target is depicted (top), and a histogram depicting GFP fold-repression for each spacer length (bottom). The gray line indicates a fold-repression value of 1 reflecting no change in GFP fluorescence from the non-targeting control. The depicted values represent the mean and S.E.M. of at least three measurements starting from independent colonies. (B) Shown are the crRNA and protospacer sequences for the variants of the s1 +12 spacer. The number in parentheses captures the number of mutations made to the 3' end of the s1+12 spacer. Black bars indicate the PAM sequence. Dark blue bars indicate the +0 spacer sequence, light blue bars indicate the extended spacer sequence, and red bars indicate mismatched nucleotides.

Table 4.S1 Strains used in this work

Strains	Genotype	Source	Stock #
BW25113 Δ*cas3::cat*	BW25113 [Δ*cas3* P_{cse1}]*::*[*cat* P_{J23119}]	Supp. Ref. (1)	pCB385
BW25113 Δ*cas3::cat lacZ*$^+$	BW25113 [Δ*cas3* P_{cse1}]*::*[*cat* P_{J23119}] *lacZ*$^+$	This study	pCB489
BW25113 Δ*cas3 lacZ*$^+$	BW25113[Δ*cas3* P_{cse1}]*::*[P_{J23119}] *lacZ*$^+$	This study	pCB490
BW25113 ΔCRISPR-Cas	BW25113 [Δ*cas3-cse1-cse2-cas7-cas5-cas6e*-CRISPR1]::*cat*	Supp. Ref. (1)	pCB401
MG1655 Δ*cas3::cat*	MG1655 [Δ*cas3* P_{cse1}]*::*[*cat* P_{J23119}]	Supp. Ref. (1)	pCB386

Table 4.S2 Plasmids used in this work.

Plasmid	Description	Resistance marker	Source	Stock #
pWUR408	*cse1* in pRSF-1b, no tags	Kanamycin	Supp. Ref. (2)	pWUR408
pWUR656	*cse2* (N-terminal Strep II tag) –*cas7-cas5-cas6e* in pCDF-1b	Streptomycin	Supp. Ref. (3)	pWUR656
pUA66-lacZ	*lacZ* promoter upstream of GFP	Kanamycin	OpenBiosystems	pCB338
pUA66-xylA	*xylA* promoter upstream of GFP	Kanamycin	Supp. Ref. (4)	pCB289
pBAD33	L-arabinose-inducible plasmid with *araC* regulator	Chloramphenicol	Supp. Ref. (5)	pCB442
cBAD33	pBAD33 with constitutive J23108 promoter	Chloramphenicol	This study	pCB491
pCas3	Constitutively expressed *cas3*	Chloramphenicol	This study	pCB492
pUA66lacZ-NT3PAM-mutant	pUA66lacZ with the PAM of the NT3 spacer mutated	Kanamycin	This study	pCB493
pcrRNA.ind	L-arabinose-inducible CRISPR array with single repeat	Ampicillin	Supp. Ref. (1)	pCB359
pcrRNA.con	Constitutive CRISPR array with single repeat	Ampicillin	Supp. Ref. (1)	pCB379
pcrRNA.ind-p1	pcrRNA.ind with spacer p1	Ampicillin	Supp. Ref. (1) pcrRNA.ind-T2	pCB361
pcrRNA.ind-p1+12	pcrRNA.ind with spacer p1+12	Ampicillin	This study	pCB494
pcrRNA.ind-p2	pcrRNA.ind with spacer p2	Ampicillin	Supp. Ref. (1) pcrRNA.ind-NT2	pCB367
pcrRNA.ind-p2+12	pcrRNA.ind with spacer p2+12	Ampicillin	This study	pCB495

Table 4.S2 Continued

Plasmid	Description	Resistance marker	Source	Stock #
pcrRNA.ind-s1	pcrRNA.ind with spacer s1	Ampicillin	Supp. Ref. (1) pcrRNA.ind-NT3	pCB368
pcrRNA.ind-s1+6	pcrRNA.ind with spacer s1+6	Ampicillin	This study	pCB496
pcrRNA.con-p1+12	pcrRNA.con with spacer p1+12	Ampicillin	This study	pCB510
pcrRNA.con-p2	pcrRNA.con with spacer p2	Ampicillin	This study	pCB511
pcrRNA.con-p2+12	pcrRNA.con with spacer p2+12	Ampicillin	This study	pCB512
pcrRNA.con-s1	pcrRNA.con with spacer s1	Ampicillin	This study	pCB513
pcrRNA.con-s1+12	pcrRNA.con with spacer s1+12	Ampicillin	This study	pCB514
pcrRNA.con-s2	pcrRNA.con with spacer s2	Ampicillin	This study	pCB515
pcrRNA.con-s2+12	pcrRNA.con with spacer s2+12	Ampicillin	This study	pCB516
pcrRNA.con-as1	pcrRNA.con with spacer as1	Ampicillin	This study	pCB517
pcrRNA.con-as1+12	pcrRNA.con with spacer as1+12	Ampicillin	This study	pCB518
pcrRNA.con-as2	pcrRNA.con with spacer as2	Ampicillin	This study	pCB519
pcrRNA.con-as2+12	pcrRNA.con with spacer as2+12	Ampicillin	This study	pCB520
pcrRNA.ind-p1+12(2)	pcrRNA.ind with spacer p1+12 containing 2 mismatches	Ampicillin	This study	pCB521

Table 4.S2 Continued

Plasmid	Description	Resistance marker	Source	Stock #
pcrRNA.ind-p1+12(4)	pcrRNA.ind with spacer p1+12 containing 4 mismatches	Ampicillin	This study	pCB522
pcrRNA.ind-p1+12(6)	pcrRNA.ind with spacer p1+12 containing 6 mismatches	Ampicillin	This study	pCB523
pcrRNA.ind-p1+12(12)	pcrRNA.ind with spacer p1+12 containing 12 mismatches	Ampicillin	This study	pCB524
pcrRNA.ind-p1(2)	pcrRNA.ind with spacer p1 containing 2 mismatches	Ampicillin	This study	pCB525
pcrRNA.ind-p1(4)	pcrRNA.ind with spacer p1 containing 4 mismatches	Ampicillin	This study	pCB526
pcrRNA.ind-p1(6)	pcrRNA.ind with spacer p1 containing 6 mismatches	Ampicillin	This study	pCB527
pcrRNA.ind-p1(12)	pcrRNA.ind with spacer p1 containing 12 mismatches	Ampicillin	This study	pCB528
pcrRNA.con-p1+12(2)	pcrRNA.con with spacer p1+12 containing 2 mismatches	Ampicillin	This study	pCB529
pcrRNA.con-p1+12(4)	pcrRNA.con with spacer p1+12 containing 4 mismatches	Ampicillin	This study	pCB530
pcrRNA.con-p1+12(6)	pcrRNA.con with spacer p1+12 containing 6 mismatches	Ampicillin	This study	pCB531
pcrRNA.con-p1+12(12)	pcrRNA.con with spacer p1+12 containing 12 mismatches	Ampicillin	This study	pCB532
pcrRNA.con-p1(2)	pcrRNA.con with spacer p1 containing 2 mismatches	Ampicillin	This study	pCB533
pcrRNA.con-p1(4)	pcrRNA.con with spacer p1 containing 4 mismatches	Ampicillin	This study	pCB534
pcrRNA.con-p1(6)	pcrRNA.con with spacer p1 containing 6 mismatches	Ampicillin	This study	pCB535

Table 4.S2 Continued

Plasmid	Description	Resistance marker	Source	Stock #
pcrRNA.con-p1(12)	pcrRNA.con with spacer p1 containing 12 mismatches	Ampicillin	This study	pCB536
pcrRNA.ind-s1+12(2)	pcrRNA.ind with spacer s1+12 containing 2 mismatches	Ampicillin	This study	pCB537
pcrRNA.ind-s1+12(4)	pcrRNA.ind with spacer s1+12 containing 4 mismatches	Ampicillin	This study	pCB538
pcrRNA.ind-s1+12(6)	pcrRNA.ind with spacer s1+12 containing 6 mismatches	Ampicillin	This study	pCB539
pcrRNA.ind-s1+12(12)	pcrRNA.ind with spacer s1+12 containing 12 mismatches	Ampicillin	This study	pCB540
pcrRNA.ind -1+12(12)'	pcrRNA.ind with spacer s1+12 containing 12 mismatches	Ampicillin	This study	pCB541
pcrRNA.ind-s1-6	pcrRNA.con with spacer s1-6	Ampicillin	This study	pCB542
pcrRNA.ind-p1-6	pcrRNA.con with spacer p1-6	Ampicillin	This study	pCB543
pcrRNA.ind-as2-6	pcrRNA.con with spacer as2-6	Ampicillin	This study	pCB544

Table 4.S3 Oligonucleotides used in this work

Name	Sequence
pBad33.J23108. fwd2	tctgacagctagctcagtcctaggtataatgctagcgagct
pBad33.J23108. rev2	cgctagcattatacctaggactgagctagctgtcagatgca
pCas3.for	AAATGGTACCAGGAGGATCAGATGGAACCTTTTAAATATATATGCCATTACTGG
pCas3.rev	TATTTCTAGATTATTTGGGATTTGCAGGGATGAC
NT3-PAM.Q5.for	ACTTAATCGCcgtGCAGCACAGG
NT3-PAM.Q5.rev	TGGGTAACGCCAGGGTTTTC
p1.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGCTTTACACTTTATGCTTCCGGCTCGTATGT
p1.rev	TCGAACATACGAGCCGGAAGCATAAAGTGTAAAGCGGTTTATCCCCGCTGGCGCGGGGAACTCGAGGTGGTAC
p1+12.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTG
p1+12.rev	TCGACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCGGTTTATCCCCGCTGGCGCGGGGAACTCGAGGTGGTAC
p2.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGCATAAAGTGTAAAGCCTGGGGTGCCTAATG
p2.rev	TCGACATTAGGCACCCCAGGCTTTACACTTTATGCGGTTTATCCCCGCTGGCGCGGGGAACTCGAGGTGGTAC
p2+12.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACT
p2+12.rev	TCGAAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCGGTTTATCCCCGCTGGCGCGGGGAACTCGAGGTGGTAC
s1.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGGCGATTAAGTTGGGTAACGCCAGGGTTTTC
s1.rev	TCGAGAAAACCCTGGCGTTACCCAACTTAATCGCCGGTTTATCCCCGCTGGCGCGGGGAACTCGAGGTGGTAC

Table 4.S3 Continued

Name	Sequence
s1+6.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGGCGATTAAGTTGGG TAACGCCAGGGTTTTCCCAGTC
s1+6.rev	TCGAGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCGGTTTATC CCCGCTGGCGCGGGGAACTCGAGGTGGTAC
s1+8.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGGCGATTAAGTTGGG TAACGCCAGGGTTTTCCCAGTCAC
s1+8.rev	TCGAGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCGGTTTA TCCCCGCTGGCGCGGGGAACTCGAGGTGGTAC
s1+10.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGGCGATTAAGTTGGG TAACGCCAGGGTTTTCCCAGTCACGA
s1+10.rev	TCGATCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCGGTT TATCCCCGCTGGCGCGGGGAACTCGAGGTGGTAC
s1+12.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGGCGATTAAGTTGGG TAACGCCAGGGTTTTCCCAGTCACGACG
s1+12.rev	TCGACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCGG TTTATCCCCGCTGGCGCGGGGAACTCGAGGTGGTAC
s1+18.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGGCGATTAAGTTGGG TAACGCCAGGGTTTTCCCAGTCACGACGTTGTAA
s1+18.rev	TCGATTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAAT CGCCGGTTTATCCCCGCTGGCGCGGGGAACTCGAGGTGGTAC
s1+24.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGGCGATTAAGTTGGG TAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGAC
s1+24.rev	TCGAGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAA CTTAATCGCCGGTTTATCCCCGCTGGCGCGGGGAACTCGAGGTGGTAC
s2.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGTTCTTCTCCTTTAC TCATATGTATATCTCC
s2.rev	TCGAGGAGATATACATATGAGTAAAGGAGAAGAACGGTTTATCCCCGCT GGCGCGGGGAACTCGAGGTGGTAC

Table 4.S3 Continued

Name	Sequence
s2+12.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGTTCTTCTCCTTTAC TCATATGTATATCTCCTTCTTAAATCTA
s2+12.rev	TCGATAGATTTAAGAAGGAGATATACATATGAGTAAAGGAGAAGAACGG TTTATCCCCGCTGGCGCGGGGAACTCGAGGTGGTAC
s3.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGGTTTTCCCAGTCAC GACGTTGTAAAACGAC
s3.rev	TCGAGTCGTTTTACAACGTCGTGACTGGGAAAACCGGTTTATCCCCGCT GGCGCGGGGAACTCGAGGTGGTAC
s3+12.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGGTTTTCCCAGTCAC GACGTTGTAAAACGACGGCCAGTGAATC
s3+12.rev	TCGAGATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCGG TTTATCCCCGCTGGCGCGGGGAACTCGAGGTGGTAC
s4.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGTTGGGTAACGCCAG GGTTTTCCCAGTCACG
s4.rev	TCGACGTGACTGGGAAAACCCTGGCGTTACCCAACGGTTTATCCCCGCT GGCGCGGGGAACTCGAGGTGGTAC
s4+12.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGTTGGGTAACGCCAG GGTTTTCCCAGTCACGACGTTGTAAAAC
s4+12.rev	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGTTGGGTAACGCCAG GGTTTTCCCAGTCACGACGTTGTAAAAC
as1.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGGTCGTGACTGGGAA AACCCTGGCGTTACCC
as1.rev	TCGAGGGTAACGCCAGGGTTTTCCCAGTCACGACCGGTTTATCCCCGCT GGCGCGGGGAACTCGAGGTGGTAC
as1+12.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGGTCGTGACTGGGAA AACCCTGGCGTTACCCAACTTAATCGCC
as1+12.rev	TCGAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACCGG TTTATCCCCGCTGGCGCGGGGAACTCGAGGTGGTAC

Table 4.S3 Continued

Name	Sequence
as2.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGAGATATACATATGA GTAAAGGAGAAGAACT
as2.rev	TCGAAGTTCTTCTCCTTTACTCATATGTATATCTCGGTTTATCCCCGCT GGCGCGGGGAACTCGAGGTGGTAC
as2+12.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGAGATATACATATGA GTAAAGGAGAAGAACTTTTCACTGGAGT
as2+12.rev	TCGAACTCCAGTGAAAAGTTCTTCTCCTTTACTCATATGTATATCTCGG TTTATCCCCGCTGGCGCGGGGAACTCGAGGTGGTAC
xylA-s1.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGATTTTGAGCCTTCA TAACGAACGCGATCGA
xylA-s1.rev	TCGATCGATCGCGTTCGTTATGAAGGCTCAAAATCGGTTTATCCCCGCT GGCGCGGGGAACTCGAGGTGGTAC
xylA-s1+12.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGATTTTGAGCCTTCA TAACGAACGCGATCGAGCTGGTCAAAAT
xylA-s1+12.rev	TCGAATTTTGACCAGCTCGATCGCGTTCGTTATGAAGGCTCAAAATCGG TTTATCCCCGCTGGCGCGGGGAACTCGAGGTGGTAC
p1+12.2mut.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGCTTTACACTTTATG CTTCCGGCTCGTATGTTGTGTGGAATAT
p1+12.2mut.rev	TCGAATATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCGG TTTATCCCCGCTGGCGCGGGGAACTCGAGGTGGTAC
p1+12.4mut.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGCTTTACACTTTATG CTTCCGGCTCGTATGTTGTGTGGATGAT
p1+12.4mut.rev	TCGAATCATCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCGG TTTATCCCCGCTGGCGCGGGGAACTCGAGGTGGTAC
p1+12.6mut.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGCTTTACACTTTATG CTTCCGGCTCGTATGTTGTGTGCTTGAT
p1+12.6mut.rev	TCGAATCAAGCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCGG TTTATCCCCGCTGGCGCGGGGAACTCGAGGTGGTAC

Table 4.S3 Continued

Name	Sequence
p1+12.12mut.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGCTTTACACTTTATG CTTCCGGCTCGTATGTAAACCCCTTGAT
p1+12.12mut.rev	TCGAATCAAGGGGTTTACATACGAGCCGGAAGCATAAAGTGTAAAGCGG TTTATCCCCGCTGGCGCGGGGAACTCGAGGTGGTAC
p1.2mut.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGCTTTACACTTTATG CTTCCGGCTCGTATTG
p1.2mut.rev	TCGACAATACGAGCCGGAAGCATAAAGTGTAAAGCGGTTTATCCCCGCT GGCGCGGGGAACTCGAGGTGGTAC
p1.4mut.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGCTTTACACTTTATG CTTCCGGCTCGTCGTG
p1.4mut.rev	TCGACACGACGAGCCGGAAGCATAAAGTGTAAAGCGGTTTATCCCCGCT GGCGCGGGGAACTCGAGGTGGTAC
p1.6mut.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGCTTTACACTTTATG CTTCCGGCTCTGCGTG
p1.6mut.rev	TCGACACGCAGAGCCGGAAGCATAAAGTGTAAAGCGGTTTATCCCCGCT GGCGCGGGGAACTCGAGGTGGTAC
p1.12mut.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGCTTTACACTTTATG CTTCATTAGATGCGTG
p1.12mut.rev	TCGACACGCATCTAATGAAGCATAAAGTGTAAAGCGGTTTATCCCCGCT GGCGCGGGGAACTCGAGGTGGTAC
s1+12.2mut.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGGCGATTAAGTTGGG TAACGCCAGGGTTTTCCCAGTCACGAGA
s1+12.2mut.rev	TCGATCTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCGG TTTATCCCCGCTGGCGCGGGGAACTCGAGGTGGTAC
s1+12.4mut.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGGCGATTAAGTTGGG TAACGCCAGGGTTTTCCCAGTCACTTGA
s1+12.4mut.rev	TCGATCAAGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCGG TTTATCCCCGCTGGCGCGGGGAACTCGAGGTGGTAC

Table 4.S3 Continued

Name	Sequence
s1+12.6mut.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGGCGATTAAGTTGGG TAACGCCAGGGTTTTCCCAGTCCATTGA
s1+12.6mut.rev	TCGATCAATGGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCGG TTTATCCCCGCTGGCGCGGGGAACTCGAGGTGGTAC
s1+12.12mut.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGGCGATTAAGTTGGG TAACGCCAGGGTTTTCTACCGGCATTCA
s1+12.12mut.rev	TCGATGAATGCCGGTAGAAAACCCTGGCGTTACCCAACTTAATCGCCGG TTTATCCCCGCTGGCGCGGGGAACTCGAGGTGGTAC
s1+12.12mut'.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGGCGATTAAGTTGGG TAACGCCAGGGTTTTCGATCATGATGGT
s1+12.12mut'.rev	TCGAACCATCATGATCGAAAACCCTGGCGTTACCCAACTTAATCGCCGG TTTATCCCCGCTGGCGCGGGGAACTCGAGGTGGTAC
s1-6.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGGCGATTAAGTTGGG TAACGCCAGG
s1-6.rev	TCGACCTGGCGTTACCCAACTTAATCGCCGGTTTATCCCCGCTGGCGCG GGGAACTCGAGGTGGTAC
p1-6.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGCTTTACACTTTATG CTTCCGGCTC
p1-6.rev	TCGAGAGCCGGAAGCATAAAGTGTAAAGCGGTTTATCCCCGCTGGCGCG GGGAACTCGAGGTGGTAC
as2-6.for	CACCTCGAGTTCCCCGCGCCAGCGGGGATAAACCGAGATATACATATGA GTAAAGGAGA
as2-6.rev	TCGATCTCCTTTACTCATATGTATATCTCGGTTTATCCCCGCTGGCGCG GGGAACTCGAGGTGGTAC

Table 4.S4 Protospacers targeted in this work

Spacer name	Target strand	Protospacer sequence
p1-6	AS	**AGG**CTTTACACTTTATGCTTCCGGCTC
p1	AS	**AGG**CTTTACACTTTATGCTTCCGGCTCGTATGT
p1+12	AS	**AGG**CTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTG
p2	S	**AAG**CATAAAGTGTAAAGCCTGGGGTGCCTAATG
p2+12	S	**AAG**CATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACT
s1-6	S	**AAG**GCGATTAAGTTGGGTAACGCCAGG
s1	S	**AAG**GCGATTAAGTTGGGTAACGCCAGGGTTTTC
s1+6	S	**AAG**GCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTC
s1+8	S	**AAG**GCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCAC
s1+10	S	**AAG**GCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGA
s1+12	S	**AAG**GCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACG
s1+18	S	**AAG**GCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAA
s1+24	S	**AAG**GCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAA ACGAC
s2	S	**AAG**TTCTTCTCCTTTACTCATATGTATATCTCC
s2+12	S	**AAG**TTCTTCTCCTTTACTCATATGTATATCTCCTTCTTAAATCTA
s3	S	**AGG**GTTTTCCCAGTCACGACGTTGTAAAACGAC
s3+12	S	**AGG**GTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAATC
s4	S	**AAG**TTGGGTAACGCCAGGGTTTTCCCAGTCACG
s4+12	S	**AAG**TTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAAC
as1	AS	**AAC**GTCGTGACTGGGAAAACCCTGGCGTTACCC
as1+12	AS	**AAC**GTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCC
as2-6	AS	**AGG**AGATATACATATGAGTAAAGGAGA
as2	AS	**AGG**AGATATACATATGAGTAAAGGAGAAGAACT
as2+12	AS	**AGG**AGATATACATATGAGTAAAGGAGAAGAACTTTTCACTGGAGT
xylA-s1	S	**AGG**ATTTTGAGCCTTCATAACGAACGCGATCGA
xylA-s1+12	S	**AGG**ATTTTGAGCCTTCATAACGAACGCGATCGAGCTGGTCAAAAT

a) Characteristics of the target strand, which is complementary to the spacer: S, sense strand of gene; AS, antisense strand of the gene.

b) PAMs are in bold red lettering. CRISPR spacers were designed to match the protospacer sequence.

Table 4.S5 pUA66 promoter sequences

Promoter	Sequence
lacZ	CTTTCGTCTTCACACTCGAGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGC GCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTA TGCTTCCGGCTCGTATGTTGTGTGG**A**ATTGTGAGCGGATAACAATTTCACACAGGAA ACAGCTATGACCATGATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTGACTGG GAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACAGGATCCTCTAGATTTA AGAAGGAGATATACAT
xylA	CGAGGCCCTTTCGTCTTCACGGTGTAGGGCCTTCTGTAGTTAGAGGACAGTTTTAAT AAGTAACAATCACCGCGATAAACGTAACCAATTTTTAGCAACTAAACAGGGGAAAAC AATTACAGATTTTTATCTTTCGATTACGATTTTTGGTTTATTTCTTGATTTATGACC GAGATCTTACTTTTGTTGCGCAATTGTACTTATTGCATTTTTCTCTTCGAGGAATTA CCCAGTTTCATCATTCCATTTTATTTTGCGAGCGAGCGCACACTTGTGAATTATCTC AATAGCAGTGTGAAATAACATAATTGAGCAACTGAAAGGGAGTGCCCAATATTACGA C**A**TCATCCATCACCCGCGGCATTACCTGATTATGGAGTTCAATATGCAAGCCTATTT TGACCAGCTCGATCGCGTTCGTTATGAAGGCTCAAAATCCTCAAACCCGTTAGCATT CCGTCACTACAATCCCGACGAACTGGTGTTGGGTAAGCGTATGTAATCTAGATTTAA GAAGGAGAT

a) Sequences highlighted in gray are from pUA66, indicating where each promoter was inserted into the plasmid. The underlined and bolded base is the previously mapped transcriptional start site.
b) The highlighted yellow text indicates the start codon of lacZ contained within the reporter construct.
c) The highlighted blue text indicates the region mutated for pUA66lacZ-NT3PAM-mutant. Sequence was mutated from `CTT` to `CGT`.

Table 4.S6 Theoretical and experimentally measured masses of individual Cas protein subunits and Cascade complexes. Discrepancies between theoretical and experimental masses are explained by loss of N-terminal methionines from Cse1, Cse2 with tag, and Cas7 during protein expression.

Cascade Complex	**Theoretical Mass (Da)**	**Experimental Mass (Da)**
s1 +0 with 2 tags	405,265.9	405,238.2 ± 45.8
s1 +6 with 2 tags	446,909.1	446,655.1 ± 209.9
s1 +12 with 3 tags	510,232.2	510,127.2 ± 48.9
Cascade Subunit	**Theoretical Mass (Da)**	**Experimental Mass (Da)**
Cse1	55901.1	55,841.0 ± 0.1
Cse2 with tag	21,391.5	21,260.3 ± 0.6
Cse2 without tag	19,204.0	19,204.1 ± 0.1
Cas7	40,025.4	39,894.0 ± 1.0
Cas5e	25,208.9	25,208.3 ± 1.2
Cas6e	22,292.9	22,292.0 ± 0.0

SUPPLEMENTARY REFERENCES

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

2. Brouns,S.J.J., Jore,M.M., Lundgren,M., Westra,E.R., Slijkhuis,R.J.H., Snijders,A.P.L., Dickman,M.J., Makarova,K.S., Koonin,E. V and van der Oost,J. (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, **321**, 960–964.

3. Westra,E.R., van Erp,P.B.G., Künne,T., Wong,S.P., Staals,R.H.J., Seegers,C.L.C., Bollen,S., Jore,M.M., Semenova,E., Severinov,K., *et al.* (2012) CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Mol. Cell*, **46**, 595–605.

4. Afroz,T., Biliouris,K., Kaznessis,Y. and Beisel,C.L. (2014) Bacterial sugar utilization gives rise to distinct single-cell behaviours. *Mol. Microbiol.*, **93**, 1093–1103.

5. Guzman,L.M., Belin,D., Carson,M.J. and Beckwith,J. (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose P_{BAD} promoter. *J. Bacteriol.*, **177**, 4121–30.

CHAPTER 5

Conclusions and future work

5.1 SUMMARY

CRISPR-Cas systems have developed into standard tools for basic research and biotechnology. Cas9 in particular has garnered a significant amount of interest due to the precise and simple way that it can be programmed. CRISPR-Cas9 has enable facile genome editing in single cells and whole organisms, transcriptional control, epigenetic modification, and chromosomal imaging. While Cas9-based technologies have been widely popular, other CRISPR-Cas systems also offer unique engineering opportunities.

This dissertation highlights how Type I CRISPR-Cas systems can be reprogrammed into genetic tools. We demonstrated that this naturally occurring adaptive immune system in *Escherichia coli* can be readily converted into a transcriptional regulator, capable of potent multiplexed gene silencing. This is accomplished by expressing the DNA-binding protein complex, Cascade, in the absence of the Cas3 nuclease.

While this engineered nuclease-null Type I-E system has several applications such as metabolic engineering, it can also be applied to probe fundamental biology questions. Using our platform, we interrogated how the CRISPR RNA (crRNA) could be modified. We discovered that the spacer portion of the crRNA can be extended while maintaining the system's function. Extending the crRNA leads to predictable changes in the stoichiometry of the Cascade complex. Interestingly, when directed to certain regions within a transcriptional unit, increasing the spacer length leads to a corresponding increase in gene silencing.

Altogether, this dissertation demonstrates the utility of studying Type I CRISPR-Cas systems and the interesting ways that they can be engineered. Below, we propose some additional work for the continued development of Type I CRISPR-Cas systems as genetic tools.

5.2 FUTURE WORK

5.2.1 Engineering Altered PAM Specificities of the Escherichia coli Type I-E CRISPR-Cas System

While Cas effector proteins are remarkably diverse, they commonly rely on protospacer-adjacent motifs (PAMs) as the first step in DNA targeting. The PAM consists of 2-5 nucleotides that must be present in the targeted region of interest. While the PAM is an essential feature for CRISPR-based DNA recognition, it also restricts which sequences can be targeted and impacts the probability of off-target effects. For any natural CRISPR-Cas system, the PAM is predefined where recognition of the PAM is inherently built into the protein structure. As a result, there have been a number of efforts to modify Cas proteins to generate an expanded set of proteins that are not limited to their natural PAM (1–4). These engineering practices have chiefly concentrated on Type II Cas9 proteins, so PAM engineering in other systems is rife with opportunity.

Across the six currently classified types of CRISPR-Cas systems, Type I systems are by far the most abundant (5). Interestingly, the Type I-E system from *Escherichia coli* has one of the most promiscuous PAM recognition capabilities, with at least nine recognized PAM sequences (AAG, AGG, ATG, GAG, TAG, AAC, AAA, AAT, ATA) (6).

For the *E. coli* Type I-E system, the effector complex is a 405-kDa multiprotein complex termed Cascade comprised of five Cas proteins in the following stoichiometric ratio: $Cse1_1Cse2_2Cas7_6Cas5_1Cas6e_1$ (7). PAM recognition is achieved by the Cse1 protein, located in the tail of the protein complex (8). We propose that directed evolution can be applied to systematically amass a collection of Cse1 proteins that individually exhibit high specificity but collectively can enable Cascade to target any desired sequence. Generating protein variants capable of recognizing diverse PAM sequences will allow for increased flexibility in choosing DNA targets.

This could be accomplished using a positive and tunable screen termed PAM-SCANR (PAM screen achieved by NOT-gate repression) (6). This screen utilizes gene repression as a

basis of CRISPR function and PAM identification (Figure 5.1). A simple genetic NOT gate produces a positive signal in the presence of functional gene repression. The screening platform consists of a library of potential PAM sequences cloned upstream of the *lacI* promoter. Immediately downstream is the LacI-dependent *lacZ* promoter driving expression of the green fluorescent protein (GFP). A catalytically dead CRISPR-Cas system is targeted to a protospacer within the *lacI* promoter, resulting in GFP fluorescence only in the presence of a functional PAM. Fluorescent cells can then be isolated through fluorescence-activated cell sorting.

For this project, instead of screening a library of PAM sequences with a fixed set of Cas protein(s), this project would screen a library of Cas proteins based on a fixed PAM sequence. To generate a diverse set of unnatural Cse1 proteins, the *cse1* gene would be PCR-amplified off of the *E. coli* genome and cloned into a vector. This plasmid-based *cse1* gene would then be subject to mutagenesis to produce a varied library of Cse1 mutants. This library would then be transformed into an *E. coli* strain that is devoid of the *cse1* gene but expresses the remainder of the Cascade operon (*cse2-cas7-cas5-cas6e*). In this strain, Cascade would not operate unless complemented with a functional Cse1 protein.

Using the PAM-SCANR platform, we would design a targeted reporter plasmid with an unnatural PAM and enrich the library for Cse1 proteins that recognize the unnatural PAM. For example, the nucleotide makeup of the native PAM for the *E. coli* Type I-E system is A-rich and C-poor. We could design a reporter plasmid requiring recognition of a C-rich PAM sequence.

Once a number of Cse1 mutants have been pulled out of the library using the PAM-SCANR screen, the selected proteins would require additional characterization. For example, one could identify mutations critical to altering PAM recognition by comparing different proteins collected from the cell sorting. This would require sequencing to identify functionally significant mutations. While crystal structures of the Type I-E Cascade complex bound to dsDNA have isolated key structural features that are implicated in contacting the PAM (9), it remains unknown how these features could be modified.

Since PAM binding is the first step in CRISPR-Cas activity, it would be important to evaluate off-target effects. The *E. coli* Type I-E Cascade naturally tolerates nine PAM sequences, and the goal is that Cse1 mutants would confer novel PAM recognition capabilities. Do these Cse1 mutants have relaxed or constricted PAM requirements relative to the wildtype? Do they recognize more PAMs or fewer? Has recognition of the wildtype PAM been ablated and replaced with a new sequence altogether? Has the length of the consensus sequence changed? All these questions require substantial characterization and are impossible to predict *a priori*.

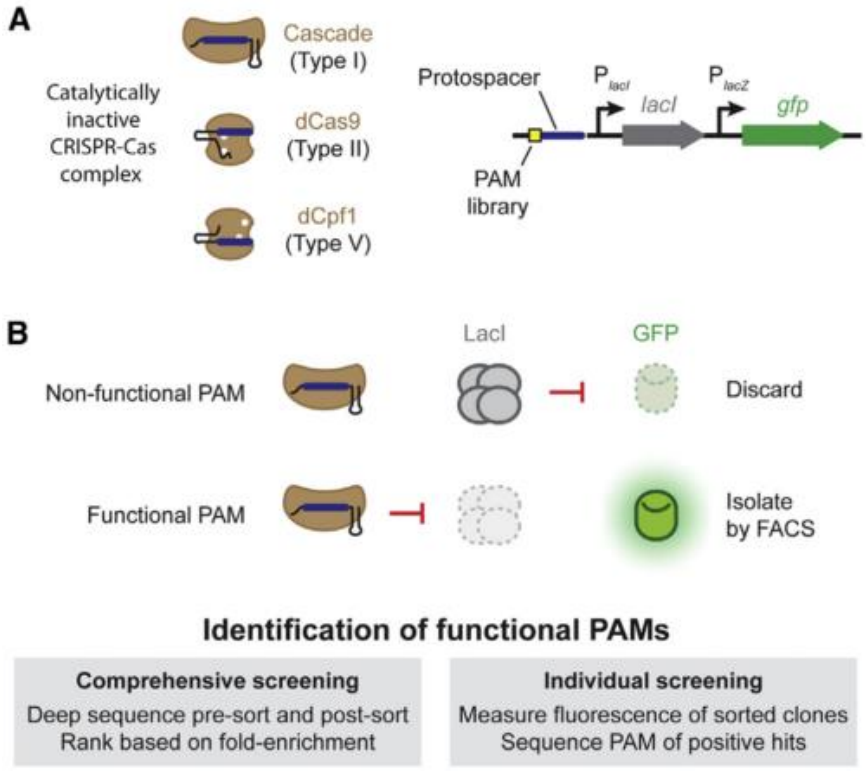

Figure 5.1 Overview of PAM-SCANR. (A) The screening platform consists of a library of potential PAM sequences cloned upstream of the *lacI* promoter. Immediately downstream of *lacI* is the LacI-dependent *lacZ* promoter controlling expression of GFP. A catalytically dead CRISPR-Cas system is targeted to a protospacer within the *lacI* promoter, resulting in GFP fluorescence only in the presence of a functional PAM. (B) Cells harboring a functional PAM can be isolated by fluorescence-activated cell sorting. Figure adapted from (6).

5.2.2 A diverse CRISPR-based toolbox for programmable gene repression in cell-free transcription-translation systems

Cell-free expression systems can theoretically emulate an *in vivo* cellular environment in a controlled *in vitro* platform (Figure 5.2). This has been useful for synthetic biologists, enabling the exploration of gene circuits in a controlled manner and optimizing protein expression (10, 11). Most *in vitro* characterization techniques rely on T7 RNA polymerase transcription. However, cell-free transcription-translation (TX-TL) expression systems that preserve endogenous *Escherichia coli* mechanisms are able to more accurately reflect *in vivo* cellular conditions (12).

We propose using the TX-TL cell-free system as a prototyping environment to characterize CRISPR-Cas systems. To start, an *E. coli* crude lysate is prepared where the cytoplasm is extracted from *E. coli* cells. The endogenous DNA and mRNA in the lysate are then degraded, so that only user-supplied elements will be expressed. The resulting lysate contains basic *E. coli* transcription machinery, including the transcription factor σ^{70} and RNA polymerase. The core translation machinery is also maintained. A number of additives must also be added to the crude lysate such as amino acids, nucleotides, salts, and an energy mixture. The final extract serves as a pseudo-buffer which mimics an intracellular environment, but allows for very controlled addition of user-defined components.

TX-TL can readily be applied to observe binding kinetics of deactivated CRISPR-Cas systems. For this project, Cas protein(s) would be encoded on plasmids and either expressed in *E. coli* cells pre-lysis or added as purified plasmid to the TX-TL extract. Corresponding guide RNAs would be designed and ordered as a linear double-stranded DNA transcriptional unit containing a σ^{70}-promoter, guide RNA sequence, and transcriptional terminator. These guide RNAs would be designed to target a reporter plasmid expressing the green fluorescent protein (GFP), and the fluorescent output would be quantified using a plate reader.

Developing a CRISPR testing platform in an *in vitro* TX-TL environment would allow for more rapid characterization of CRISPR components. Presently, using a new guide RNA *in vivo* requires an extra cloning step. TX-TL bypasses this hurdle because linear double-stranded

DNA can be commercially ordered and added directly into the reaction. Using single-stranded DNA as a template for guide RNAs would further accelerate the process as ssDNA oligonucleotides are often cheaper and quicker to synthesis; however, this has not been tested in TX-TL yet.

Once established, this platform could be used to test the orthogonality of different CRISPR-Cas systems, validate CRISPR-based genetic circuits, rapidly screen guide RNAs, and more.

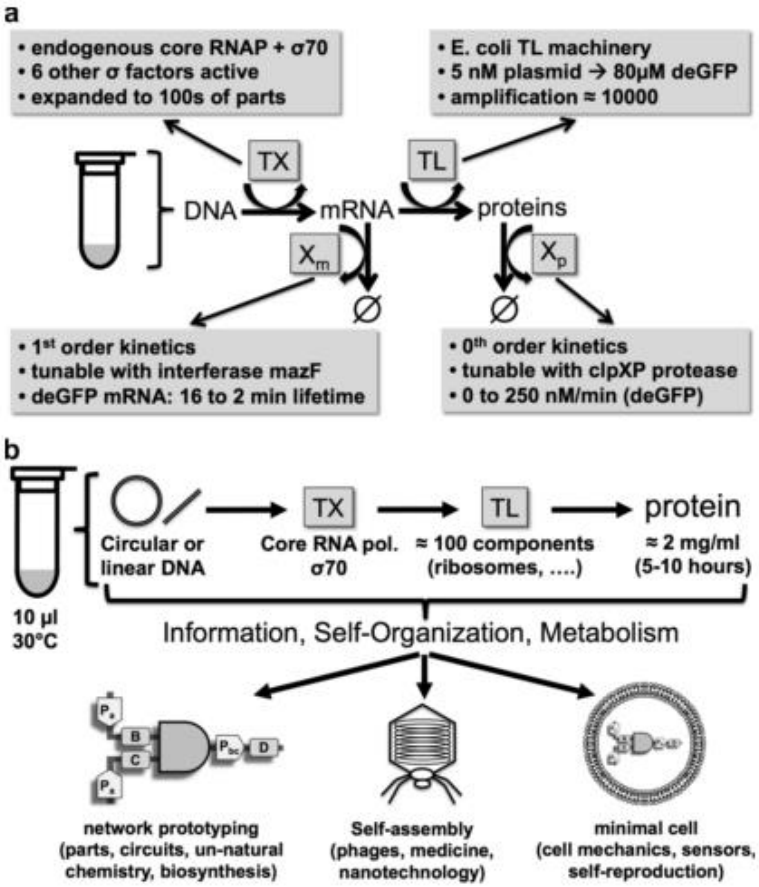

Figure 5.2 Overview of *E. coli* TX-TL characteristics and scope of application. (A) The four parts of the protein synthesis reactions (TX: transcription, TL: translation, X_m: mRNA degradation, X_p: protein degradation) were characterized and engineered for the *E. coli* cell-free toolbox 2.0. (B) The toolbox is a highly flexible, easy-to-use cell-free platform with a multitude of applications. Figure adapted from (12).

5.3 CLOSING THOUGHTS AND REFLECTIONS

With CRISPR-edited food receiving the green light from the US government and the first CRISPR gene therapy clinical trial in the works, researching CRISPR-Cas systems, at this moment in time, has been an exciting and sometimes stressful process. While most applications stem from the Cas9 protein, there are other intriguing opportunities in exploring other CRISPR-Cas systems. This dissertation aims to highlight how studying these other systems can contribute to our general understanding of CRISPR-Cas systems as a whole and how these underutilized systems can be engineered with applications in their own right.

REFERENCES

1. Anders,C., Bargsten,K. and Jinek,M. (2016) Structural Plasticity of PAM Recognition by Engineered Variants of the RNA-Guided Endonuclease Cas9. Mol. Cell, 61, 895–902.

2. Kleinstiver,B.P., Prew,M.S., Tsai,S.Q., Nguyen,N.T., Topkar,V. V, Zheng,Z. and Joung,J.K. (2015) Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. Nat. Biotechnol., 33, 1293–1298.

3. Kleinstiver,B.P., Prew,M.S., Tsai,S.Q., Topkar,V. V, Nguyen,N.T., Zheng,Z., Gonzales,A.P.W., Li,Z., Peterson,R.T., Yeh,J.J., et al. (2015) Engineered CRISPR-Cas9 nucleases with altered PAM specificities. Nature, 523, 481–5.

4. Hirano,S., Nishimasu,H., Ishitani,R. and Nureki,O. (2016) Structural Basis for the Altered PAM Specificities of Engineered CRISPR-Cas9. Mol. Cell, 61, 886–94.

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

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

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

8. Sashital,D.G., Wiedenheft,B. and Doudna,J.A. (2012) Mechanism of foreign DNA selection in a bacterial adaptive immune system. Mol. Cell, 46, 606–615.

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

10. He,M. (2008) Cell-free protein synthesis: applications in proteomics and biotechnology. N. Biotechnol., 25, 126–132.

11. Takahashi,M.K., Hayes,C.A., Chappell,J., Sun,Z.Z., Murray,R.M., Noireaux,V. and Lucks,J.B. (2015) Characterizing and prototyping genetic networks with cell-free transcription-translation reactions. Methods, 86, 60–72.

12. Garamella,J., Marshall,R., Rustad,M. and Noireaux,V. (2016) The All E. coli TX-TL Toolbox 2.0: A Platform for Cell-Free Synthetic Biology. ACS Synth. Biol., 5, 344–355.