The Diverse Impact of Bacteriophages on the Bacterial Host

by

Joseph W. Bondy-Denomy

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy

Graduate Department of Molecular Genetics
University of Toronto

The Diverse Impact of Bacteriophages on the Bacterial Host

Joseph W. Bondy-Denomy

Ph.D. Thesis, 2014

Department of Molecular Genetics
University of Toronto

Abstract

The viruses that infect bacteria, bacteriophages (or phages) are the most abundant biological entities on the planet. Phages are constantly causing bacterial population turnover through infection and lysis, a process that results in the replication of phages and the death of the bacterium. Alternately, phages can integrate into the bacterial genome, at which point the phage is referred to as a prophage. Prophages are abundant in bacterial genomes and are not merely inert entities, as some genes are expressed from the prophage, which can increase host and prophage fitness. Clinical and environmental isolates of the opportunistic human pathogen, *Pseudomonas aeruginosa* were investigated and found to possess prophages in their genomes which were inducible. I identified many novel phenotypes emerging after prophage acquisition, particularly an array of superinfection exclusion traits that protect the lysogenic host from subsequent phage infection. These traits are likely beneficial in the environment, protecting the bacterium (and the resident prophage) from infection and lysis.

In addition to prophages protecting the host from infection, bacteria encode many systems to protect themselves, including the adaptive immune system known as CRISPR-Cas. I demonstrated that the CRISPR-Cas system of *P. aeruginosa* strain PA14 is active and targets phages and plasmids for degradation. Interestingly, the phages are not passive bystanders in this

process and encode genes which inhibit CRISPR-Cas function, called “anti-CRISPRs”. Five different anti-CRISPR proteins were discovered which all inactivate the CRISPR-Cas system of *P. aeruginosa*, thus allowing phages to infect the host and either lyse the bacterium or integrate in the genome. These anti-CRISPRs share no common sequence element and only share the phage genomic position in which they were found. Three of the different anti-CRISPRs were investigated and I show that they utilize completely distinct mechanisms to inactivate the CRISPR-Cas system. These studies shed new light on the arms race between phages and their bacterial hosts and emphasize the complex ecological relationships between them.

Acknowledgments

I would like to acknowledge my supervisor, Dr. Alan Davidson, for his excellent support and mentorship over the last six years. What started out as a fairly serendipitous meeting between us in the spring of 2008 has turned into an excellent and productive relationship. I think that Alan has fostered the perfect "hands-on: hands-off" ratio in the lab, as I have certainly benefited from his advice, but also from the opportunity to figure things out on my own. I would also like to thank Dr. Karen Maxwell for her assistance and expertise over the years from suggesting and troubleshooting experiments to her invaluable assistance moving manuscripts along!

I have had the benefit of working with an excellent supervisory committee comprised of Drs. John Brumell, David Hwang, and David Guttman who have provided excellent suggestions and asked tough questions throughout my many committee meetings. They were certainly not shy to tell me when I was barking up the wrong tree and I appreciate that. I also had the good fortune of working in the lab of Dr. Stephen Lory at Harvard Medical School for four months, which was an eye-opening experience for me, and I thank him for that opportunity.

Early in my days in the Davidson lab, I had a lot of assistance from two very kind and helpful former PhD students in the lab, Drs. Lisa Pell and Lia Cardarelli. Dr. Bianca Garcia and Yurima Hidalgo-Reyes have also been instrumental, both in keeping the lab running smoothly as well as more recently, working together on various aspects of the anti-CRISPR project. They are too numerous to name, but I would also like to thank all previous and current members of the Davidson lab who make the lab a fun place to work everyday. Outside of the lab, the MSB 4th floor and Molecular Genetic community in general has been an amazing place to make friends and connections with both students and faculty alike. I thank everyone who I have met along the way, for both scientific and social reasons!

During my undergraduate education at the University of Waterloo, I was motivated to pursue courses and studies in microbiology by some excellent professors, specifically Drs. Barb Butler and Christine Dupont without whom, I may have ended up as an accountant (*shudder*). I would also like to thank my undergraduate research supervisors at McGill University, Drs. Mary Stevenson and Danielle Malo, as well as my deepest thanks to the man who gave me my first lab job at the University of Western Ontario, Dr. Grant MacFadden. Without these people

reinforcing my love for working in a microbiology lab, I certainly would not be in the position I am today.

On a personal note, I would like to acknowledge my parents, Michael Denomy and Jo Ann Bondy who have been as supportive of all my endeavors as any parents could be. As an only child I certainly received plenty of attention and help over the years with constant support every step of the way.

Last but certainly not least, I would like to thank my beautiful and talented wife, Teresa MacLean, who I met the same day I started graduate school in September of 2008 in the auspicious MSB 4171. My defense date of September 11, 2014 will mark six years (almost to the day) that we have known each other. In addition to all of the personal support and happiness she has brought me over the last six years, I have to acknowledge her amazing support for me in generating the work in this thesis. We have spent many nights and weekends together at the lab, which made everything a lot more enjoyable. She has also been an excellent critic (when needed) for me to discuss ideas with and practice presentations. As we each conclude our time as students, it only seems appropriate that we were recently married here at the beautiful University of Toronto, a place that will always be special to us.

Table of Contents

List of Tables

List of Abbreviations

a.a.	amino acids
ACR	anti-CRISPR
AT-rich	adenine-thymine rich
ATP	adenosine triphosphate
BLAST	basic local alignment search tool
bp	base pairs
cDNA	complementary DNA
CF	cystic fibrosis
CRISPR-Cas	clustered regularly interspaced short palindromic repeats – CRISPR associated
crRNA	CRISPR RNA
CR1_sp1	CRISPR locus 1, spacer 1
Csy complex	CRISPR system *Yersinia*
DNA	deoxyribonucleic acid
dsDNA	double-stranded DNA
ssDNA	single-stranded DNA
EDTA	ethylene diamine tetraacetic acid
EOP	efficiency of plaquing
E. coli	*Escherichia coli*
EMSA	electrophoretic mobility shift assay
e.v.	empty vector
fs	frameshift
gp	gene product
H-NS	histone-like nucleoid-structuring protein
IgG	immunoglobulin G
IPTG	isopropyl β-D-1-thiogalactopyranoside
ITC	isothermal titration calorimetry
kDa	kilo Dalton
λ	lambda phage
LB	lysogeny broth
LPS	lipopolysaccharide
MOI	multiplicity of infection

mV	millivolt
Ni-NTA	nickel-nitrilotriacetic acid
NMR	nuclear magnetic resonance
nt	nucleotide
OD	optical density
ONPG	ortho-Nitrophenyl-β-galactosidase
ORF	open reading frame
P. aeruginosa	*Pseudomonas aeruginosa*
PAM	protospacer adjacent motif
PEG	polyethylene glycol
PFU/mL	plaque forming units per mL
pI	isoelectric point
PMSF	phenylmethylsulphonyl fluoride
pre-crRNA	precursor crRNA
RNA	ribonucleic acid
RNAi	RNA interference
RT-qPCR	reverse transcriptase-quantitative polymerase chain reaction
RRM	RNA-recognition motif
SAXS	small-angle X-ray scattering
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	size exclusion chromatography
SKM	slow killing medium
Srp	signal recognition particle
SSC	saline sodium citrate
T4P	type IV pilus
TBE	tris borate EDTA
TEM	transmission electron microscope
TEV	tobacco etch virus
tracrRNA	*trans*-activating crRNA
WT	wild type

List of Figures

Chapter 1
Introduction

Most, if not all, organisms have viruses that infect them. Infection may lead to viral replication and cell death, a persisting parasitism, or to a mutualistic relationship where the genomes of both virus and host replicate together. In the state of mutualism, both virus and cell experience increased fitness as a result of the other. The viruses that infect bacteria, called bacteriophages (or phages), well exemplify the range of relationships that can exist between viruses and their hosts. The binary interactions between a bacterial host and either a lytic phage (potential parasite) or temperate phage (potential mutualist) are frequent in nature and are a fascinating aspect of phage biology. The tripartite interactions between a bacterial host, its mutualistic and parasitic phages are, however, also likely to be abundant in nature. Having a thorough understanding of the phage-bacteria interactions is essential for understanding how microbes live and die and this will be a focus of the work outlined here.

This thesis will focus on the temperate phages infecting the bacterial pathogen *Pseudomonas aeruginosa*, specifically on the ability of phages to integrate into the bacterial genome and modify host properties as a prophage. I will outline my work to characterize the abundance and role of prophages found in diverse isolates of *P. aeruginosa*, focusing on their ability to exclude superinfection by other phages, thus increasing bacterial fitness. Below, I will review known roles of prophages in bacterial phenotypes, demonstrating the important role that phages have as mobile elements. Despite the potential for prophages to increase bacterial fitness, phage infection poses a threat to bacterial survival given the ability of a phage to replicate and lyse their host. Bacteria can defend themselves from phage attack using the CRISPR-Cas immune system. This introduction will also describe many aspects of the CRISPR-Cas immune system such as the mechanism of action as well as some of the evolutionary and physiological implications of this powerful adaptive immune system. Previously there was significant confusion about the role and function of the *P. aeruginosa* CRISPR-Cas system, but I will demonstrate here that it is functional and able to target phages and plasmids, but that some phages are resistant to CRISPR-Cas activity. I will also describe the first reported examples of phage-encoded genes which antagonize the CRISPR-Cas system known as “anti-CRISPRs.” These genes are important for phages both during infection and for the maintenance of a stable prophage. I will also outline my

work to determine anti-CRISPR mechanism at the molecular level, where I demonstrate diverse modes of action for different anti-CRISPR proteins. Many researches had speculated that anti-CRISPR genes should be found on phage genomes and these data are the first description of their role and mechanism, further solidifying the CRISPR-Cas as a phage defense system.

1.1 Bacteriophages

1.1.1 History

Phages were first discovered in 1915 by Frederick Twort and independently in 1917 by Felix d'Herelle. Although Twort made the initial discovery of bacteriophages, d'Herelle exhaustively characterized them and continued to work on them throughout his career, even being nominated for a Nobel Prize for the discovery. Phages were identified by d'Herelle in the feces of patients suffering from dysentery and he discovered that phages could indeed cause the death of the dysentery bacillus (*Shigella dysenteriae*) (d'Herelle, 1917), although their viral nature was not understood until later. During his thorough characterization of these entities, it was clear that they were indeed mysterious, as he described them as "an antagonistic microbe… anti-Shiga microbe… invisible microbe… a living germ [as opposed to a chemical entity]… filterable germ" and finally concluded: "…the anti-dysentery microbe is an obligate bacteriophage." With this new tool in hand, d'Herelle and others explored developing phage therapeutics to combat human and animal pathogens. Prior to the discovery and commercialization of antibiotics, many clinical studies were examining these new phage therapeutics (Fruciano and Bourne, 2007). A lack of a thorough understanding of their mode of action for killing bacteria hampered this development however, as success was sporadic and difficult to predict (Fruciano and Bourne, 2007). This was likely due to the inherent specificity phages have for a given bacterium, which often extends beyond the species level to individual strain variability. In the mid-20th century, most efforts to utilize phage therapy were abandoned in the west while continuing to be developed in Eastern Europe and Russia.

In the west, the utility of phages throughout the mid-20th century came out of their development as the leading model system in the molecular biology revolution. Fundamental findings such as DNA being the genetic element, the triplet nature of the genetic code, and the discovery of messenger RNA are owed to phages, particularly those infecting *E. coli* (Hershey and Chase, 1952; Brenner *et al.*, 1961; Crick *et al.*, 1961). Furthermore, phages provided and continue to

provide key reagents for bacterial genetics experiments and strain typing which has led to much work and characterization of these viruses (Pitt, 1986). Commercial enterprises such as the dairy industry have also always had a keen interest in phage biology as contaminating phage infection can lead to the destruction of starter cultures in cheese and yogurt production (Marcó *et al.*, 2012). Phages or phage proteins (e.g. phage lysins) are currently being explored extensively as novel therapeutics to fight antibiotic resistant infections in humans and animals (Schuch *et al.*, 2002; Chan *et al.*, 2013). Such an approach is not new, as Felix d'Herelle himself pioneered these ideas in the early 1900's, however interest is resurging as desperation over the treatment of life threatening infections grows.

The ecological implications of phage predation and lysogeny in nature are now also being extensively investigated. Despite original predictions that phages were not abundant in soil and water, a landmark study in 1989 showed that there are ~10^8 viruses/mL in natural waters (Bergh *et al.*, 1989). Indeed, we now appreciate that there are 10^{30}-10^{32} phages globally, making them the most abundant biological entity, outnumbering their bacterial counterparts by a factor of ten-to-one (Wommack and Colwell, 2000). Additionally, phages have a massive role in bacterial turnover, with an estimated 10^{23} infections occurring per second (Suttle, 2005; 2007). Much characterization of the role of phages on microbial ecology has followed these discoveries, for example, the recent discovery that phages prey on the most abundant bacterial clade in the ocean, SAR11 (Zhao *et al.*, 2013). Their potential role in human health is just starting to be investigated with studies characterizing and ascribing functions to the human virome, examining both phages and prophages (phage genomes integrated in bacterial genomes) (Reyes *et al.*, 2010). Whether the goal is to facilitate phage infection, prevent it, or understand its ecological role, much of the work in phage biology aims to identify the "arms" that phages and bacteria bring to battle. By discovering and characterizing the elements of this arms race, one can utilize this information to guide and predict the outcome of phage-bacteria interactions.

1.1.2 Phage Life Cycle

For a successful phage infection to occur, many factors must align properly, beginning at the cell surface. After a phage particle contacts the bacterial cell in the correct orientation, a number of specificity determinants are required for a productive infection to result. The presence and accessibility of an appropriate cellular receptor is required as well as the necessary components

for the entry of the phage genome into the bacterial cytoplasm. This thesis focuses on tailed phages with a dsDNA genome, which make up ~95% of known phages (Ackermann, 2007). For a tailed phage infecting a Gram-negative bacterium, this requires the phage genome to pass through the phage tail, bacterial outer membrane, peptidoglycan-containing periplasmic space, and inner membrane in a process that is poorly understood. There are likely factors of both phage and host origin, which are required to facilitate this process (Boyd and Brüssow, 2002). Once inside the cell, the maintained integrity of this foreign DNA is threatened by DNA degradation systems such as restriction enzymes and the CRISPR-Cas system. At this point, a phage may enter the lytic cycle, in which the phage genome is replicated and packaged into phage particles, and the cell is ultimately lysed by phage-encoded proteins to release these particles.

Alternately, for some phages, the lysogenic cycle can be entered, in which the phage genome is integrated into the bacterial genome, or forms an extrachromosomal plasmid. The phage is then described as a prophage and the cell, a lysogen. Most prophage genes are repressed; thus, the lysogenic cell can survive and replicate without the production of phage particles or deleterious phage proteins. Recent interest in prophage biology has been influenced by an influx of bacterial genome sequences, revealing an average bacterial genome contains three prophages, with extreme examples of up to 20% of the genome being prophage (Canchaya *et al.*, 2003; Casjens, 2003). Based on prophage abundance in sequenced bacterial genomes, it is estimated that approximately 25% of phage genomes on the planet exist in the form of prophages (Casjens, 2005). These prophages can also provide a substrate for recombination with superinfecting lytic or temperate phages, which contributes to much of the mosaicism observed in phage genomes (Casjens, 2003). Given the prevalence of lysogeny, it is not surprising that phages carry a wide variety of genes that provide a fitness advantage to their host when expressed from the prophage. In fact, to truly appreciate the role of phages in bacterial physiology, phages must be regarded not only as bacterial parasites, but also as mutualists. A further crucial aspect of prophages is their ability to readily excise from their host genome, enter the lytic cycle, and infect other cells, thereby spreading their genes to other bacteria. The established capacity of phages to influence bacterial behaviour and disseminate potentially pathogenic genes has spurred great interest in investigations of prophage functions. In this thesis, I will outline my investigation of prophages in environmental and clinical strains of *Pseudomonas aeruginosa*. Such a systematic investigation has not previously been conducted. In CHAPTER 2, I present the phenotypic

outcomes of prophage acquisition in a single strain of *P. aeruginosa* with a particular focus on acquired superinfection exclusion properties. In the sections below, I describe previously characterized prophage-mediated phenotypes and elaborate on some of these studies, particularly those focusing on the contributions of prophages to superinfection exclusion and virulence (summarized in Table 1).

Table 1: Summary of prophage-mediated phenotypes

Bacterial Species	Prophage	Gene(s)	Effect	Reference
Superinfection Exclusion				
Escherichia coli	phi80, N15	*cor*	Inactivation of membrane receptor FhuA inhibits phage adsorption	Vostrov et al. 1996
Pseudomonas aeruginosa	D3	*oac, wzy, iap*	Serotype conversion prevents phage adsorption	Newton et al. 2001
Escherichia coli	HK97	*15*	Phage entry inhibited by inner membrane protein	Cumby et al. 2012b
Salmonella Typhimurium	P22	*sieA*	Phage entry inhibited by inner membrane protein	Hofer et al. 1995
Vibrio cholerae	K139	*glo*	Phage entry inhibited by periplasmic protein	Nesper et al. 1999
Escherichia coli	P1	*sim*	Phage blocked downstream of adsorption	Maillou and Dreiseikelmann 1990
Escherichia coli	T4	*imm*	Phage entry inhibited	Lu and Henning 1994
Escherichia coli	T4	*sp*	Phage lysozyme activity inhibited	Lu and Henning 1994
Lactococcus lactis	Tuc2009	*sie2009*	DNA injection inhibited	McGrath et al. 2002
Streptococcus thermophilus	TP-J34	*ltp*	Phage entry inhibited by lipoprotein	Sun et al. 2006
Streptococcus thermophilus	ΦSfi21	*orf203*	Unknown mechanism protects against heterologous phages	Bruttin et al. 1997
Mycoplasma sp.	MAV1	*vir*	Phage entry inhibited by lipoprotein	Clapper et al. 2004
Escherichia coli	λ	*rexA-rexB*	Superinfecting phage replication is inhibited through induced cell death	Shinedling et al. 1987
Bacillus subtilis	SPβ	*nonA*	Superinfection induces *nonA* expression, aborting the infection	Yamamoto et al. 2014
Pathogenesis				
Escherichia coli	λ	*lom*	Outer membrane protein increases adherance to epithelial cells	Vica Pacheco et al. 1997
Escherichia coli	λ	*bor*	Lipoprotein improves survival in human serum	Barondess and Beckwith, 1990
Pseudomonas aeruginosa	FIZ15	Unknown	Lysogenic cells are more resistant to phagocytosis and human serum	Vaca-Pacheco et al. 1999
Neisseria meningiditis	Unnamed prophage	*orf6*	Encoded TspB protein binds human immunoglobulin in biofilm matrix	Muller et al. 2013
Vibrio cholerae	K139	*glo*	Required for virulence in mouse model of cholera	Reidl and Mekalanos 1995
Vibrio cholerae	VPIΦ	*tcpA*	Filamentous phage encodes colonization factor and receptor for toxin producing phage, CTXΦ	Karaolis et al. 1999
Vibrio cholerae	CTXΦ	*ctx*	Cholera toxin produced from a prophage	Waldor et al. 1996
Escherichia coli	933W	*stx2*	Prophage induction via repressor cleavage produces shiga-like toxin	Tyler et al. 2004
Salmonella Typhimurium	SopEΦ	*sopE*	Prophage-encoded type III effector secreted into human cells	Mirold et al. 1999
Salmonella Typhimurium	Gifsy-2	*sodC*	Prophage-encoded superoxide dismutase produced to neutralized reactive oxygen species	Figueroa-Bossi and Bossi, 1999

Hamiltonella defensa	APSE-3	*Unknown*	Aphid symbiotic bacterium protects host from parasitism through putative toxin	Oliver et al. 2009
Streptococcus mitis	SM1	*pblA/B*	Phage structural proteins and phage lysins allow the bacterium to bind to platelets	Mitchell et al. 2007, Seo et al. 2010
Streptococcus pyogenes	H4489A	*hylP*	Phage-associated hyaluronidase facilitates infection through capsule and is produced from prophage	Benchetrit et al. 1977
CRISPR-Cas				
Pseudomonas aeruginosa	DMS3	*42*	An interaction between the CRISPR-Cas system and gene *42* inhibits biofilm production	Zegans et al. 2009, Cady and O'Toole 2011
Vibrio cholerae	ICPI	CRISPR-Cas	A phage-encoded CRISPR-Cas system inhibits an anti-phage island during infection	Seed et al. 2013
Clostridium difficile	Many	CRISPR-Cas	Prophages possess transcriptionally active CRISPR loci	Soutourina et al. 2013
Metagenome	Metagenome	CRISPR-Cas	Metagenomic analysis of the human gut revealed prophage-encoded CRISPR-Cas systems	Minot et al. 2011, 2013
Microbial Ecology				
Unknown	Unknown	Ab resistance	Ab treatment of mice caused the production of phages carrying various Ab resistance genes	Modi et al. 2013
Bacteroides cellulosilyticus	Unamed	IG region	Prophage provides fitness advantage to host in murine gut	Reyes et al. 2013
Cyanobacteria	Unamed	Metabolic	Prophage-encoded genes provide metabolic and photosynthetic capaciy	Rohwer and Thurber 2009
Bacillus anthracis	Many	σ factors	Prophage-mediated control of sporulation and exopolysaccharide production	Shuch and Fischetti, 2009
Escherichia coli	9 prophages	Unknown	Deletion of nine prophages compromised host fitness	Wang et al. 2010
Pseudomonas aeruginosa	5 prophages	Unknown	Multiple prophages present in epidemic strain which play a role in virulence	Winstanley et al. 2009
Enterococcus faecalis	Many	*pblA/B*	Prophages induced from clinical isolates encode PblA/B, platelet binding proteins	Yasmin et al. 2010
Flavobacterium psychrophilum	Many	-	Eighty per cent of of strains contained a prophage related to phage 6H	Castillo et al. 2013
Unknown	Many	-	Mitomycin C treatment of unculturable bacteria in soil revealed many lysogenic strains	Ghosh et al. 2008
Ab- Antibiotic IG- Intergenic				

1.2 Prophages

1.2.1 General Aspects

When a phage genome integrates into its host genome, most phage genes must be repressed to maintain normal cell viability. This general repression is achieved through the action of phage repressor proteins, and the mechanisms of action of many phage repressors have been studied in great detail (Dodd *et al.*, 2005). The action of the repressor proteins expressed by prophages leads to resistance to superinfection by the same phage. This "immunity" to subsequent infection by homologous phages is a hallmark of all lysogens. Prophage-mediated phenotypic changes that are not mediated simply through repressor activity have been referred to as "lysogenic conversion." To decipher the genetic basis for such phenotypes, two general features of a

prophage need to be examined; a) the insertion site in the bacterial genome, and b) genes that may be expressed from an otherwise repressed prophage. Many phages, such as *Escherichia coli* phage Mu, integrate into the host genome at random positions through a transposition mechanism (Morgan *et al.*, 2002). These integration events can cause phenotypic changes by interrupting host genes. However, if the same phenotypes are observed in every lysogenic isolate of a given phage, prophage-encoded genes are likely responsible. Other phages, such as *E. coli* phage λ, always integrate at the same position in their host genome using a site-specific integrase (Kotewicz *et al.*, 1977). For these phages, the insertion event could cause a consistent phenotype, so this possibility must be investigated. In most cases, however, genes expressed from the prophage are responsible for lysogenic conversion phenomena (see below).

Prophage-expressed genes are usually not essential for the phage life cycle and often comprise more recently acquired genetic elements with the characteristic AT-richness of foreign DNA (Juhala *et al.*, 2000). These elements can increase host fitness and independent transcriptional promoters and terminators often control their expression. These elements are often recognized as extra genes when comparing closely related phage genomes. Thus, they have been referred to as "morons" to indicate that "when one is present in the genome there is more DNA than when it is not present" (Juhala *et al.*, 2000). Many of the phenotypic alterations caused by prophages are mediated by genes contained within moron elements (Cumby *et al.*, 2012a).

1.2.2 Effects of Prophages on the Cell Envelope: Superinfection Exclusion

The most intensively studied effect of prophages is their inhibition of other phages including themselves. This "superinfection exclusion" is expected to be highly adaptive since phages are by far the most abundant predators of bacteria. Superinfection exclusion can be achieved through a wide variety of mechanisms, but most involve alterations to the cell surface or other cell envelope components. For example, the prophage-expressed *cor* gene of *E. coli* phages Φ80 and N15 blocks the cell surface adsorption of superinfecting phages T1, Φ80 and N15 (Vostrov *et al.*, 1996). The ferrichrome uptake protein FhuA is the receptor for all of these phages. The replication of many other phages that require FhuA as their receptor, as well as the uptake of ferrichrome is also blocked by expression of the *cor* gene, demonstrating the inactivation of FhuA (Uc-Mass *et al.*, 2004). In another example, the *Pseudomonas aeruginosa* D3 prophage blocks superinfection through an entirely different mechanism. A three-gene operon expressed

from this prophage modifies the O-antigen of the cell surface lipopolysaccharide (LPS) of its host, thus changing its serotype from O5 to O16. Since the replication of many phages requires specific binding to the O-antigen, this change in serotype prevents superinfection by blocking the adsorption of phages including phage D3 itself (Newton *et al.*, 2001.

Following the adsorption of a superinfecting phage, the phage genome must be injected into the host. This step is a common target for prophage-mediated superinfection exclusion. For example, prophage-expressed gene *15* of *E. coli* phage HK97 produces a small and likely inner membrane protein that inhibits superinfection by HK97. The product of gene *15*, gp15, prevents DNA entry into the cytoplasm, but does not block the replication of other phages using the same cell surface receptor as HK97. Thus, HK97 inhibition by gp15 must occur at a step after surface adsorption, likely via an interaction with the tail tube or tape measure proteins of the superinfecting phage (Cumby *et al.*, 2012b). Similarly, the *Salmonella* Typhimurium prophage P22 produces SieA, an inner membrane protein that blocks DNA entry of phage P22 without affecting its cell surface adsorption (Hofer *et al.*, 1995). *Vibrio cholerae* prophage K139 expresses Glo, a periplasmic protein which also appears to inhibit phage genome entry (Nesper *et al.*, 1999). The *sim* system of the *E. coli* P1 prophage also inhibits superinfection by self and other phages through a poorly characterized process operating downstream of adsorption (Maillou and Dreiseikelmann, 1990). These mechanisms are not unique to prophages, as the well studied lytic phage T4 produces the Imm protein which blocks DNA entry as well as the protein Sp which inhibits activity of the phage lysozyme (Lu and Henning, 1994). This is presumably a way of preventing superinfection when a lytic cycle is already underway, to ensure the fidelity of this process.

Superinfection exclusion systems have also been well studied in Gram-positive species, with much of the work aimed at resisting phage infection in the dairy industry. Many of these systems also block the DNA entry step. The Tuc2009 prophage moron gene, *sie2009*, encodes a protein that associates with the *Lactococcus lactis* cell membrane and blocks the DNA injection step of superinfecting phages (McGrath *et al.*, 2002). Distinct exclusion systems inhibiting DNA entry have also been found in other *L. lactis* prophages, such as two systems that act against the 936 phage group (Mahony *et al.*, 2008). The *Streptococcus thermophilus* TP-J34 prophage expresses the *ltp* gene, which encodes a lipoprotein that can block TP-J34 DNA entry as well as other members of the lactococcal 936 phage group (Sun *et al.*, 2006). Phage ΦSfi21, which infects *S. thermophilus*, possesses *orf203*, a gene that leads to resistance to superinfection by heterologous

phages (Bruttin *et al.*, 1997). Interestingly, this gene is found in the same genomic position (i.e. next to the integrase gene) as the *V. cholerae* phage K139 *glo* and *L. lactis* phage Tuc2009 *sie2009* genes discussed above. Phages infecting *Mycoplasma* sp. (a genus that lacks peptidoglycan but is related to Gram-positive bacteria) demonstrate similar superinfection exclusion effects. The prophage-expressed *vir* gene of phage MAV1 produces a lipoprotein localized to the outer surface of the cell membrane that blocks entry of MAV1 (Clapper *et al.*, 2004).

In summary, there are a wide variety of prophage-produced proteins both in Gram-negative and positive organisms that block phage infection by altering components of the cell envelope. Interestingly, most of these proteins do not block adsorption of phage particles to the cell surface, but inhibit the subsequent step of DNA entry. This phenomenon could indicate that there are generally fewer means by which phages can inject their DNA through the cell membrane than there are for adsorption to the cell surface. Thus, blocking DNA entry can potentially inhibit more phages than blocking one surface receptor. In CHAPTER 2, I will assess how different prophages in *P. aeruginosa* prevent superinfection, motivated by some of the approaches outlined above. I will present experiments conducted to determine whether exclusion is mediated by inhibiting adsorption, entry, and/or processes downstream of these events.

1.2.3 Effects of Prophages on the Cell Envelope: Increasing Pathogenicity

A number of prophage-expressed genes affect bacterial pathogenicity. One of the first prophage-expressed genes identified was *lom* of *E. coli* phage λ, which encodes an outer membrane protein (Reeve and Shaw, 1979). The Lom protein increases the ability of λ lysogens to adhere to human buccal epithelial cells (Vica Pacheco *et al.*, 1997). The *bor* gene, which is also expressed from a λ prophage, encodes a lipoprotein that improves the survival of lysogenic cells in animal serum (Barondess and Beckwith, 1990). Similar effects have been observed in *P. aeruginosa*, where prophage FIZ15 converts strain PAO1 to being more resistant to phagocytosis and human serum, while increasing adherence to epithelial cells (Vaca-Pacheco *et al.*, 1999). A protein called TspB, which is expressed from prophages in invasive *Neisseria meningiditis* strains, is present on the bacterial cell surface and binds to human IgG, leading to formation of large bacterial aggregates in a biofilm (Müller *et al.*, 2013). This reaction may protect the bacteria from immune responses.

In some cases, the alteration of bacterial pathogenic properties caused by prophage-expressed genes may be related to phage superinfection exclusion. Both of these phenomena are often the result of bacterial cell envelope alterations; thus, changes affecting one property could certainly affect the other. For example, the *P. aeruginosa* prophage FIZ15 mentioned above also causes cells to become resistant to phage D3 and likely alters the O-antigen in a similar manner as does the D3 prophage (see section above) (Vaca-Pacheco *et al.*, 1999). Consistent with this theme, the *glo* gene, which mediates superinfection exclusion by *V. cholerae* phage K139, is also required for full virulence in a mouse model of cholera infection (Reidl and Mekalanos, 1995). These examples suggest that resistance to phage superinfection may confer two benefits for a "converted" bacterial host: first, the ability to survive longer in the environment/host by fending of phage infection; and second, changes to the cell envelope that resist phage may have beneficial effects with respect to bacterial virulence. Although papers often focus on only one such aspect of conversion, it seems possible that any prophage-induced envelope modification could display a phenotype in both contexts if the appropriate phage and virulence assays were used.

1.2.4 Other Types of Prophage-Induced Phage Inhibition

One of the longest known and most extensively studied prophage-induced superinfection exclusion systems is the *rexA-rexB* system of *E. coli* phage λ. These genes are expressed in λ lysogens in the same operon as the *cI* gene encoding the λ repressor protein, and they prevent replication of phage T4 *rII* and other phage mutants (Shinedling *et al.*, 1987). The Rex proteins do not prevent phage DNA from entering the cell, but do cause a severe drop in membrane potential during phage DNA replication, which leads to a drop in ATP levels, and eventual "altruistic" cell death. The *rex* genes have a number of other intriguing effects, such as inducing stationary phase-like properties and inhibiting toxin-antitoxin systems (Engelberg-Kulka *et al.*, 1998; Slavcev and Hayes, 2003). In *Bacillus subtilis*, when cells with an SPβ prophage are infected with phage SP10, the prophage gene *nonA* is induced. Expression of *nonA* during the late stages of phage infection inhibits growth of *B. subtilis* and blocks the synthesis of virion proteins, suggesting that this process aborts the infection (Yamamoto *et al.*, 2014). Abortive infection systems operating in a similar manner are common in *L. lactis*, but many are present on plasmids rather than prophages (Chopin *et al.*, 2005). A variety of other inhibition mechanisms have been observed for prophage-expressed proteins (Snyder, 1995; Samson *et al.*, 2013).

1.2.5 Prophage-Encoded Virulence Factors

Many bacterial pathogens rely on prophage-encoded genes for toxin production. In some cases, the presence of a single toxin gene acquired via phage is the difference between a harmless and harmful bacterium. The mobility of phage-borne toxin genes makes them particularly dangerous and facilitates the emergence of novel pathogens. Famously, phages encoding a Shiga-like toxin were identified when *E. coli* K12 acquired the ability to produce the toxin after being lysogenized with phages from the highly virulent *E. coli* O157:H7 strain (O'Brien *et al.*, 1984). The Shiga-like toxin gene is not a traditionally defined moron since prophage induction was required for production of the toxin (Tyler *et al.*, 2004), explaining the clinical observation that treatment with antibiotics capable of causing phage induction exacerbated infections (Wong *et al.*, 2000). *V. cholerae* was converted from a harmless water-dwelling bacterium to a significant pathogen upon the acquisition of prophage VPIΦ, which encodes the toxin co-regulated pilus (TCP). This protein is a colonization factor in humans, and also served as a receptor for subsequent phage CTXΦ infection (Karaolis *et al.*, 1999). Upon lysogeny, the cholera toxin was produced from the CTXΦ prophage (Waldor and Mekalanos, 1996). Type III secretion systems in some organisms have acquired effectors via a prophage. In *Salmonella* Typhimurium, for example, phage SopEΦ provides the secreted effector gene *sopE* (Mirold *et al.*, 1999), and the Gifsy-2 prophage provides the superoxide dismutase *sodC* which is required for virulence (Figueroa-Bossi and Bossi, 1999). Many other prophage-encoded toxins have been discovered, such as the cytotoxin from *P. aeruginosa* (Nakayama *et al.*, 1999), diphtheria toxin from *Corynebacterium diphtheriae* (Freeman, 1951), and the *Clostridium botulinum* neurotoxin (Eklund *et al.*, 1971). Prophage encoded toxins have been extensively studied with excellent reviews on the subject (Boyd and Brüssow, 2002; Brüssow *et al.*, 2004).

Prophage-encoded toxins are not always detrimental to the eukaryotic host of a bacterium. For example, the gamma-proteobacteria symbiont *Hamiltonella defensa* protects its aphid host from attack by a parasitoid wasp. This protection is dependent on an *H. defensa* prophage-encoded toxin (Oliver *et al.*, 2009). Interestingly, the prophage is spontaneously lost when cultured in lab conditions, which leads to reduced aphid reproductive capacity (Weldon *et al.*, 2013). The widespread symbiotic bacterium *Wolbachia* is estimated to be present in 66% of all arthropod species and many isolates harbor prophage WO. *Wolbachia* is found as a parasite or a mutualist

in many different organisms with putative toxin genes in WO being identified that could contribute to these various interactions (Kent and Bordenstein, 2010).

In addition to traditional toxins, which appear to serve no role in the actual production and dissemination of phage particles, some phage proteins have a role in bacterial virulence as well as being part of the virion. Proteins PblA and PblB of *Streptococcus mitis* are encoded on a prophage and promote binding of the bacterium to human platelets. Similar to the Shiga-like toxin, phage induction increases the levels of PblA and PblB protein, but these proteins are also present in the phage virion (Bensing *et al.*, 2001). Interestingly the phage lysin and holin were also required for the platelet binding activity and it was later shown that cytoplasmic PblA/B are released from the bacterium due to lysin and holin activity, allowing PblA/B to bind to the platelets (Mitchell *et al.*, 2007). It was later shown that the lysin protein itself can also interact with fibrinogen in the platelets (Seo *et al.*, 2010), thus ascribing multiple roles to the phage proteins. The capsule produced by group A streptococci consists of hyaluronic acid, and hyaluronidase is detectable in purified phages, presumably to allow phage penetration through the capsule during infection. Interestingly, lysates from some temperate phages infecting this organism have been found to contain very high levels of hyaluronidase, much of it not being phage associated (Benchetrit *et al.*, 1977). In addition to its role during infection, this overproduction of hyaluronidase may help the dissemination of the streptococci through hyaluronic acid-containing human tissue. These somewhat unorthodox examples of phage-encoded virulence traits demonstrate the unpredictability of new phenotypes that emerge upon prophage acquisition and present a rationale for exploring prophage-mediated phenotypes in different organisms.

1.2.6 Prophage Ecology

Advances in metagenomics and sequencing technology have led to the conclusion that not only are phages abundant, but their genetic and protein diversity is massive (Pedulla *et al.*, 2003; Jacobs-Sera *et al.*, 2012). Additionally, the role phages play in human health and their contributions to the gut microbiome are starting to be investigated. For example, antibiotic treatment in a mouse model led to enrichment of temperate phages with the ability to transfer antibiotic resistance genes of all classes, not just to resist the challenge antibiotic (Modi *et al.*, 2013). In another study, an artificial microbial community comprised of common human gut

microbiome members was used to colonize a mouse. Many of these organisms possessed prophages which were frequently induced *in vivo*, and through the mutagenesis of one strain, *Bacteroides cellulosilyticus* WH2, a strong fitness advantage was shown to be provided by a resident prophage (Reyes *et al.*, 2013). Outside of the traditional examples of prophage-mediated changes (i.e. toxin production, increased virulence, superinfection exclusion, etc.), other prophage-controlled traits are becoming apparent due to metagenomic studies. For example, prophages in marine organisms significantly expand the metabolic capacities of their host and cyanobacteria phages carry many key photosynthesis genes (Rohwer and Thurber, 2009).

Few studies have addressed the effects of multiple different prophages within a single bacterial strain, but with the current capability to sequence many phage genomes and conduct rapid phenotype profiling, such studies should yield insightful data. One such study was undertaken in *Bacillus anthracis* where individual prophages were introduced into a strain of interest. Among the observations were the ability of a prophage to block or promote sporulation, induce exopolysaccharide production, and increase long term *B. anthracis* survival in soil and in an earth worm intestine (Schuch and Fischetti, 2009). A phage-encoded sigma factor was identified, which could explain some phenotypes. Using an inverse approach, all of the 9 prophages in an *E. coli* strain were deleted and a significant fitness loss was observed with increased sensitivity to antibiotics, osmotic and oxidative stress (Wang *et al.*, 2010). Furthermore, the Liverpool Epidemic Strain of *P. aeruginosa* caused a significant outbreak in a children's cystic fibrosis unit. Upon genome sequencing, the strain was found to have 5 prophages and mutagenesis of genes in different prophages impaired virulence in a rat infection model, suggesting an important and poorly understood link between novel virulence phenotypes and prophage acquisition (Winstanley *et al.*, 2009).

The characterization of the abundance, distribution, and induction potential of natural prophage populations provides insight about how encoded genes will spread. In *Enterococcus faecalis*, 47 bacteremia isolates were induced with various agents, revealing 34 phages of 4 different groups. Sequencing of 8 phages revealed homologs of the *S. mitis* platelet binding proteins (PblA/B) in each phage. Interestingly, lysogens made with each of the 8 phages showed differential survival in a *Galleria* wax moth larvae infection model (Yasmin *et al.*, 2010). Similar observations of prophage abundance have been made in *Flavobacterium psychrophilum* where prophages similar to the sequenced phage 6H were found in a 80% of a collection of 49 strains (Castillo *et al.*,

2013). Finally, using novel *in situ* phage induction and collection methods, uncultured soil bacteria were induced with mitomycin C, revealing that ~80% of the strains were lysogenic (Ghosh *et al.*, 2008).

The multiple studies described above, particularly those characterizing prophage-encoded genes which prevent entry and/or replication of superinfecting phages (summarized in Table 1) are relevant to the work I describe in CHAPTER 2 where I outline a systematic screen I conducted to determine the role of *P. aeruginosa* prophages. Furthermore, studies assessing the frequency of lysogeny (Ghosh *et al.*, 2008; Winstanley *et al.*, 2009; Castillo *et al.*, 2013), inducible prophages (Schuch and Fischetti, 2009), and phenotypes emerging from prophage acquisition (Schuch and Fischetti, 2009; Winstanley *et al.*, 2009) in other organisms have similarities to the thesis work presented here. Given the previously discovered roles for prophages in manipulating bacterial properties and their yet uncharacterized role in *P. aeruginosa* epidemics (Winstanley *et al.*, 2009), assessing the abundance and properties of different prophages in *P. aeruginosa* is necessary.

Genomic and metagenomic studies have highlighted the prevalence of phages in the environment and prophages within bacterial genomes. Importantly, it is now becoming clear that the interplay between phages and bacteria within the human microbiome have a significant impact on human health. The variety of studies reviewed here demonstrate the multiple and complex mechanisms by which prophages can influence the behaviour and pathogenicity of bacterial species. Thus, it is clear that further work in this area will be crucial for our understanding of the human microbiome and for developing effective new anti-bacterial therapies.

1.3 Phage-Bacteria Arms Race

1.3.1 Cell Surface

Aside from the prophage derived resistance to superinfection discussed above, bacteria also encode many other ways to protect themselves from phages, and a few will be briefly discussed here. Importantly, the ability for phages to “fight back” and evolve counter resistance measures is what powers the arms race between phages and bacteria. For example, many bacteria produce capsules or a variety of extracellular polymeric substance (EPS) often formed by polysaccharides. Although the EPS is well-studied for its role in infectious disease, it also

presents a significant physical barrier to phages as membrane-bound receptors can be blocked. For example, *E. coli* K1 strains produce a capsule containing polymers of sialic acid which can only be penetrated by K1-phages which have acquired endosialidase enzymes on their tail spikes (Stummeyer *et al.*, 2006).

Biofilms are another example of an EPS substance and are often composed of a complex matrix containing polysaccharide, DNA, and/or protein. Similar to capsules, these dynamic surface attached communities of microbes and their secreted products present a challenge for phages, as well as many antibiotics and components of the human immune system (Mah *et al.*, 2003). For example, one major constituent of many *P. aeruginosa* biofilms is alginate, a heteropolymeric polysaccharide, which is often found to be overproduced by *P. aeruginosa* isolated from cystic fibrosis patients. Phage F116, a podophage that infects *P. aeruginosa* possess an alginate lysate which is able to reduce the viscosity of this polysaccharide, presumably enabling access to the cell surface (Hanlon *et al.*, 2001).

1.3.2 Restriction-Modification Systems

Restriction enzymes generally serve to degrade foreign DNA as it enters a bacterial cell while the modification system protects host DNA from this fate. These enzymes function by recognizing sequence specific motifs along with inappropriate or absent modification signatures on foreign DNA such as methylation of adenine or cytosine bases (Tock and Dryden, 2005). Restriction-modification (R-M) systems are widespread in bacteria and have long been appreciated for their role in phage defense and as a commonly used molecular biology tool (Labrie *et al.*, 2010). One way that phages avoid these systems is through DNA sequence changes in the phage genome leading to a reduction or elimination of recognized restriction sites. For example, *Eco*RII requires two copies of its target site for cleavage to proceed, but T3 and T7 phage genomes have large gaps between these sites, preventing cleavage (Krüger *et al.*, 1988). Other phages, like *S. aureus* phage K have don't have any 'GATC' sites in the genome, thus not being recognized by many restriction enzymes, including *S. aureas* derived *Sau*3A1 (O'Flaherty *et al.*, 2004). The lack of these restriction sites is likely due to selective pressures since this four base motif should occur many times by chance.

Some phages incorporate alternate bases into their genomes which can result in the inhibition of R-M function. For example, *E. coli* phage Mu encodes a gene called *mom* which modifies

adenines to N6-(1-acetamido) adenine when also in the presence of host DNA adenine methylases (Hattman, 1982). Another *E. coli* phage, T4, has both hydroxymethyl cytosine as well as glucosylated hydroxymethyl cytosine residues which provide resistance to some restriction enzymes (Krüger and Bickle, 1983). In a further arms race development, however, *E. coli* has restriction enzymes which are able to recognize and cleave DNA with each of these modifications (Sutherland *et al.*, 1992; Bair and Black, 2007).

A more "active" approach to resisting R-M systems has also been discovered where phages inject proteins into the cell to directly or indirectly interfere with restriction enzyme activity. For example, in an extension of the T4 examples presented above, the T4 phage internal protein I* is injected into the cell with the genome and can directly bind to and inhibit the restriction enzyme which recognizes the glucosylated hydroxylmethyl cytosine residues (Bair and Black, 2007). In a more indirect mechanism, phage P1 injects accessory proteins DarA and DarB into the cell which physically interact with the DNA and block the activity of Type I restriction enzymes (Iida *et al.*, 1987). Finally, phage-encoded methylases have been identified which likely facilitate methylation of phage DNA to avoid degradation by restriction enzymes in their next host (McGrath *et al*., 1999). This ability for anti-restriction systems to function through a variety of mechanisms will be paralleled in my mechanistic analysis of anti-CRISPR function in CHAPTER 5.

Although this is not meant to extensively review the subject, these examples of the arms race between phages and their bacterial hosts demonstrates the "back-and-forth dance" between them and the significant evolutionary pressure imposed by each entity on the other. Much of the remainder of this introduction and thesis will focus on the action of the CRISPR-Cas bacterial immune system, where there was previously very little knowledge regarding the arms race between phages and CRISPR-Cas. After demonstrating active CRISPR-Cas systems in *P. aeruginosa* in CHAPTER 3, I will outline my discovery of "anti-CRISPR" genes in CHAPTERS 4 and 5 which represent the first phage genes discovered to inhibit CRISPR-Cas activity, and thus the first identified "active" step in the arms race between phages and the CRISPR-Cas immune system.

1.4 CRISPR-Cas

1.4.1 General Aspects

The successful entry of phage DNA into the cell is necessary for an infection to commence, however the maintained stability of that DNA is essential for a productive lytic infection or the generation of a stable lysogen. By analogy to eukaryotic immune systems, restriction-modification systems could be thought of as an innate immune system, recognizing a foreign signal (i.e. small sequence motif, unmethylated DNA). More recently, an adaptive immune system has been discovered in prokaryotes, known as CRISPR-Cas, which allows bacteria to acquire resistance to foreign elements and specifically degrade them upon subsequent exposure.

Clustered regularly interspaced short palindromic repeats (CRISPRs) and CRISPR associated (*cas*) genes are found in a large number of diverse prokaryotic species. First observed in 1987 in *Escherichia coli* K12 (Ishino *et al.*, 1987) as an array of alternating repeats with interspersed spacers, this region was later given the CRISPR acronym (Jansen *et al.*, 2002). The possibility that CRISPRs might form an immune system against phages and plasmids was raised in 2005 when three groups independently reported that spacer sequences possessed homology with foreign DNA elements (Mojica *et al.*, 2005; Pourcel *et al.*, 2005; Bolotin *et al.*, 2005). The first experimental evidence of CRISPR-Cas-mediated adaptive immunity emerged in 2007 with the isolation of phage resistant *Streptococcus thermophilus* cells possessing new CRISPR spacers after a phage challenge. This novel acquisition from the phage genome into the bacterial CRISPR locus was shown to be causative of the phage resistance phenotype (Barrangou *et al.*, 2007). A notable aspect of CRISPR-Cas systems is that the spacers are incorporated from previously encountered foreign DNA elements, so that resistance to these elements is an acquired trait, similar to adaptive immunity seen in higher eukaryotes. The increased fitness provided by CRISPR-Cas systems is illustrated by the occurrence of CRISPR loci in 46% of bacteria and 84% of archaea (CRISPRdb) (Grissa *et al.*, 2007).

Since the first demonstration of CRISPR-Cas-mediated phage resistance, a number of CRISPR-Cas systems have been found in diverse prokaryotes with differing repeat sequences, Cas proteins, and modes of action. These systems have been grouped into three broad types (Types I-III), along with subtypes (e.g. Type I-F) based on the *cas* genes they possess and their mode of action (Makarova *et al.*, 2011b). Despite the diversity of the *cas* genes and the organisms

possessing CRISPRs, the CRISPR loci are generally composed of multiple repeated sequences ranging from 21-48 base pairs, separated by 26-72 base pair variable spacer sequences (Bhaya *et al.*, 2011) with *cas* genes often located adjacent to the CRISPR locus. CRISPR-Cas systems have been the subject of intense investigation due to their intriguing RNA-based mechanism of action. Parallels exist between CRISPR-Cas function and RNAi systems in eukaryotes, however no homologous proteins have been identified between CRISPR-Cas and RNAi machinery.

For CRISPR-Cas system function, the CRISPR locus is transcribed, yielding a single precursor RNA that is processed within the repeat regions by a non-Cas protein (Type II) or Cas protein (Type I and III) into individual units of CRISPR RNAs (crRNAs) (Brouns *et al.*, 2008; Haurwitz *et al.*, 2010; Deltcheva *et al.*, 2011). The mature crRNA subsequently nucleates the formation of a complex with multiple Cas proteins (Type I and III) or a single Cas protein, Cas9 (Type II) that will survey the cell for invading nucleic acid (Brouns *et al.*, 2008; Hale *et al.*, 2009; Deltcheva *et al.*, 2011). The crRNA-Cas complex recognizes and cleaves foreign DNA (Type I, II, III-A) or RNA (Type III-B) molecules at sites called protospacers with complementarity to the crRNA (Marraffini and Sontheimer, 2008; Hale *et al.*, 2009; Garneau *et al.*, 2010) (Figure 1).

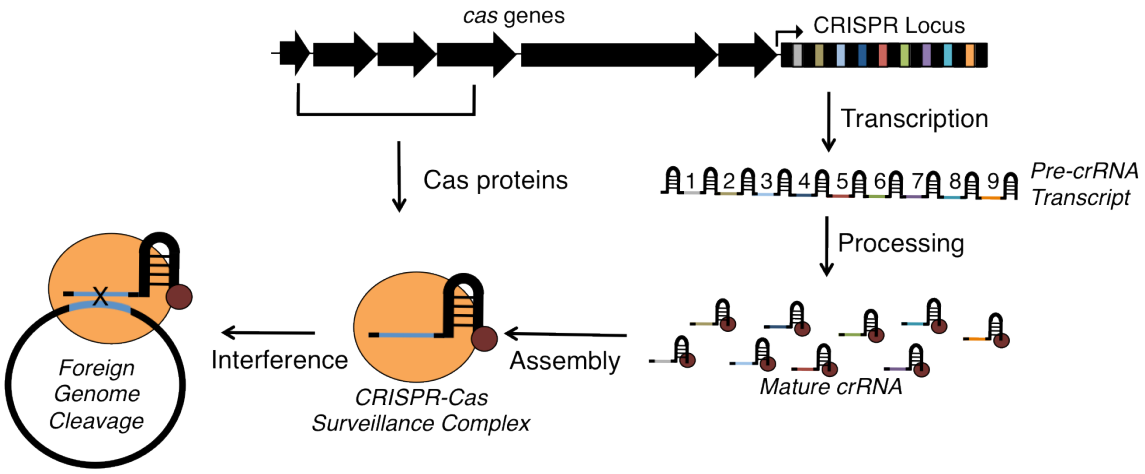


Figure 1: Schematic of the Type I CRISPR-Cas system.

Thick black arrows represent the *cas* genes which encode the members of the CRISPR-Cas surveillance complex as well as the genes for degradation of the target and spacer acquisition. The CRISPR locus is shown on the right with repeats as black boxes and spacers as colored boxes. The CRISPR locus is transcribed to produce the pre-crRNA molecule which is processed by a *cas* gene (brown circle) to produce single mature crRNAs. These are each assembled with other Cas proteins into a complex which will scan the cell for invading DNA. The spacer-derived crRNA sequence can recognize the protospacer sequence through complementary binding (blue line) and guide a cleavage event. The protospacer adjacent motif (PAM, not shown) is also required in Type I and II systems to prevent self targeting (i.e. cleavage of the CRISPR locus). CRISPR-Cas type specific differences are not shown here.

In addition to identity between the spacer and protospacer, Type I and II CRISPR-Cas systems also require a 2-5 nucleotide motif next to the protospacer, called the protospacer adjacent motif (PAM) (Mojica *et al.*, 2009).

1.4.2 Type I-F CRISPR-Cas Mechanism

The Type I system is the focus of the work to be presented in this thesis (CHAPTERS 3, 4 and 5), particularly the Type I-F system of *P. aeruginosa*. In addition to work directly on this system, a number of studies have also analyzed the Type I-E system of *E. coli*, from which functional inferences can be drawn. It should be emphasized, however that the proteins which make up these two systems are distinct with no two proteins sharing more than 25% identity and most possessing no detectable sequence similarity. Despite both being part of the Type I classification (because of the presence of the nuclease/helicase *cas3*) it is not entirely clear whether the other Cas proteins which make up the CRISPR-Cas surveillance complex are homologous (Makarova *et al.*, 2011b).

The Type I-F CRISPR-Cas system of *P. aeruginosa* strain PA14 contains two convergently expressed CRISPR loci (CRISPR1 and CRISPR2) possessing 14 and 21 spacers, respectively. Between these two arrays are the six *cas* genes that comprise this system, *cas1*, *cas3*, *csy1*, *csy2*, *csy3*, and *csy4*. The *cas1* gene is the only one that is conserved across all CRISPR-Cas types and is considered the hallmark to identify these systems genomically, as well as to construct phylogenies (Makarova *et al.*, 2011b). The Cas1 protein is required for the acquisition of novel CRISPR spacers from foreign genomic elements in a process which has been mostly studied in the *E.* coli type I-E system (Yosef *et al.*, 2012). Although the acquisition process is not yet fully understood, it requires Cas1 DNAse activity, in complex with Cas2, playing a non-enzymatic role (Nuñez *et al.*, 2014). The *cas2* gene in some Type I-F systems, including PA14, exists as a *cas2-3* fusion (Richter *et al.*, 2012). Cas1 and Cas2 play no role in the interference process (Barrangou *et al.*, 2007; Brouns *et al.*, 2008) but are necessary and sufficient for spacer acquisition (Yosef *et al.*, 2012), at least in the Type I-E system.

Csy1, Csy2, Csy3 and Csy4 form the mature surveillance complex which scans the cell for foreign invaders via the guide CRISPR RNA (crRNA). The CRISPR locus is first transcribed as a precursor crRNA which contains all repeats and spacers from a given locus. The semi-palindromic repeats form stem-loop structures which provide the substrate for endoribonuclease

Csy4 (Haurwitz *et al.*, 2010; Sternberg *et al.*, 2012), which cleaves the RNA and stays bound to its product. This leaves a 60nt crRNA comprised of 28nt of repeat and 32nt of spacer. The repeat is non-contiguous in the crRNA with 8nt on the 5' end (5' handle) of the spacer and 20nt on the 3' end, containing the stem loop. The crRNA-Csy4 complex nucleates the assembly of the hexameric Csy3 protein and the heterodimer Csy1-2 which all form at a final stoichiometry of $Csy1_1$: $Csy2_1$: $Csy3_6$: $Csy4_1$: $crRNA_1$ (Wiedenheft *et al.*, 2011b). By analogy to the Type I-E CRISPR-Cas complex (Cascade), it is likely that the Csy1-2 heterodimer rests at the 5' end of the crRNA with Csy3 spanning the backbone of the RNA and Csy4 sitting at the 3' end (Figure 2) (Wiedenheft *et al.*, 2011a; Wiedenheft *et al.*, 2011b).

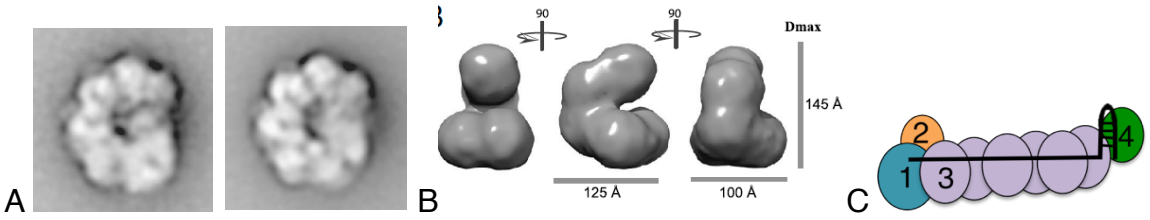


Figure 2: The observed and schematisized structure of the Csy complex.

(A) two-dimensional EM projection, (B) SAXS reconstruction, (Adapted from Wiedenheft *et al.*, 2011b) (C) schematic showing predicted placement of proteins Csy1, 2, 3, 4, and the crRNA

The recognition of a foreign target has been the subject of extensive studies in the Type I-E system and to a lesser extent in the Type I-F system. The CRISPR-Cas complex scans dsDNA templates via non-specific protein-DNA interactions and specific RNA-DNA interactions looking for the protospacer adjacent motif (PAM) and the first 8 nucleotides of the target DNA, respectively. The PAM in the Type I-F system is a GG dinucleotide motif (Mojica *et al.*, 2009). Work in the Type I-E system has shown that the PAM is essential for CRISPR-Cas complex binding (Westra *et al.*, 2012b) by a Cas protein-DNA interaction, not by base pairing with the crRNA (Sashital *et al.*, 2012; Westra *et al.*, 2013; Hochstrasser *et al.*, 2014). In the Type I-F system, the interaction with the PAM is likely mediated by Csy1-2 which then facilitates the destabilization of the dsDNA, allowing the DNA to be interrogated by the crRNA. Downstream of the PAM on the protospacer is the 32nt of complementarity that would exist with a perfect match between spacer and protospacer. Perfect matches at positions 1-5, 7, and 8 have been shown to be essential for targeting and denoted the 'seed' region (Semenova *et al.*, 2011). Mismatches are tolerated in some of the remaining sites along the spacer:protospacer hybrid. Upon a strong interaction mediated by the crRNA:DNA hybrid, the nuclease/helicase protein

Cas3 is recruited via an interaction with the large subunit protein (CasA in *E. coli* Type I-E CRISPR-Cas which is analogous to Csy1), leading to the cleavage and subsequent unwinding of the DNA target (Westra *et al.*, 2012b; Hochstrasser *et al.*, 2014). In the case of a phage infection, this would presumably halt the infection and the production of new phage particles would not proceed.

The structural and biochemical studies, such as those outlined above, have provided extensive insight into the molecular mechanisms of CRISPR-Cas function (reviewed further in Westra *et al.*, 2012a; Reeks *et al.*, 2013) and will be relevant for my analysis of anti-CRISPR function in CHAPTER 5. Considerably fewer studies, however, have addressed the endogenous operation of these systems and the physiological ramifications of CRISPR-Cas activity. In the following sections, I will describe a number of studies which address some examples of CRISPR-Cas systems functioning in various hosts, primarily focusing on *in vivo* work. These facets of CRISPR-Cas systems are most relevant to the work outlined in CHAPTER 3 and 4 of this thesis where I outline the discovery of a naturally active CRISPR-Cas system in *P. aeruginosa* and subsequently describe phage anti-CRISPR genes, which inactivate the CRISPR-Cas system.

1.4.3 The costs and benefits of CRISPR-Cas systems

Phage predation and horizontal transfer of DNA between bacterial species have massive effects on bacterial evolution, virulence, and physiology (Labrie *et al.*, 2010; Samson *et al.*, 2013). Due to the widespread occurrence of CRISPR-Cas systems in prokaryotes and their proven role in inhibition of phage replication and foreign DNA uptake, there is no doubt that these systems play a crucial role in shaping phage genomes and bacterial populations both in the environment in general and in the human microbiome (Rho *et al.*, 2012; Stern *et al.*, 2012). Outlined in Table 2 are a number of examples of naturally active CRISPR-Cas systems with proven roles in resisting foreign DNA.

Table 2: CRISPR-Cas systems with characterized *in vivo* effects

Bacterial Species	CRISPR-Cas Results	Type	Reference
Lab Experiments			
Streptococcus thermophilus	Phage targeting, spacer acquisition	II-A	Barrangou, 2007
Staphylococcus epidermidis	Plasmid targeting	III-A	Marraffini, 2008
Pectobacterium atrosepticum	crRNA targeting self not tolerated, spacers matching mobile genetic elements	I-F	Vercoe, 2013
Streptococcus agalactiae	CRISPR blocked conjugation, 40% of 949 spacers target mobile genetic elements	II-A	Lopez-Sanchez, 2012

Sulfolobus sp.	Phage and plasmid targeting, spacer acquisition	I-A	Manica, 2011, Gudbergsdottir 2011, Rezzonico, 2011
Neisseria meningitidis	Natural transformation blocked	II-C	Zhang, 2013
Francisella tularensis subsp. *Novicida*	Endogenous gene regulation	II-B	Sampson, 2013
Vibrio cholerae	Phage encoded CRISPR-Cas system	I-F	Seed, 2013
Lactococcus lactis	Plasmid-encoded CRISPR-Cas targets phage	III-A	Millen, 2012
Haloferax volcanii	Plasmid targeting, spacers matching viruses	I-B	Cui, 2008
Streptococcus pyogenes	Plasmid targeting	II-A	Deltcheva, 2011
Thermococcus kodakarensis	Plasmid targeting	I-A, I-B	Held, 2013
Escherichia coli	Plasmid targeting	I-F	Watanabe, 2013
Inference/Natural spacer matches			
Streptococcus pyogenes	Many spacers matching phages, inverse correlation with prophage	II-A, I-C	Nozawa, 2011
Enterococcus sp.	Spacers matching horizontal elements, inverse correlation with plasmid	II-A	Palmer, 2010
Campylobacter jejuni	crRNA expression and processing	II-C	Dugar, 2013
Mycoplasma gallisepticum	Related strains which infect different hosts possess different and rapidly evolving spacers	II	Fischer, 2012
Xanthomonas oryzae	139 out of 203 (68%) unique spacers match phage	I-C	Elmore, 2013
Leptospirillum group II	Lateral transfer of CRISPR loci and subsequent locus expansion over time	III	Almendros, 2012, Delaney, 2012
Yersinia pestis	High CRISPR locus diversity among isolates, spacers with matches to prophages	I-F	Semenova, 2009
Francisella tularensis subsp. *novicida*	CRISPR spacers with matches to phage and prophage	II	Schunder, 2013
Clostridium difficile	crRNA expression and processing	I-B	Soutourina, 2013
Sulfolobus islandicus	Extensive spacer reassortment and diversity among related strains	I-A	Andersson, 2008
Erwinia amylovora	Spacers matching a plasmid correlate with the absence of that plasmid	I-E, I-F	Cady, 2011
Porphyromonas gingivalis	1,187 diverse spacers in 60 strains, matches to insertion sequences	I-C, III-B	Erdmann, 2012
Repressed			
Escherichia coli	Repressed by H-NS, does not block plasmid acquisition	I-E,	Touchon, 2011, Westra, 2010
Salmonella enterica Typhi	Repressed	I-E	Medina-Aparicio, 2011

The mere presence of an intact CRISPR-Cas system in a genome does not necessarily mean that it is functioning as a defense system in all conditions, however as some systems are inactive, repressed, or performing alternative functions. The CRISPR-Cas community is just beginning to understand the full extent of these systems, despite their seemingly straight forward role in phage defense (Table 2).

Since resisting lytic phage growth is expected to always benefit a bacterium, the evolutionary advantage of possessing an active CRISPR-Cas system to destroy the genomes of these phages is very clear. Indeed, models of the relationship between CRISPR-Cas and lytic phage have confirmed this assertion (Levin, 2010). Further, indiscriminate foreign DNA insertions within the

genome (e.g. genomic islands, prophages) can often be detrimental due to misregulation of the new genes or interruption of essential genes. Bacteria have developed multiple mechanisms to prevent invasions by such detrimental DNA in addition to CRISPR-Cas, such as restriction endonucleases, abortive infection systems and phage adsorption/entry inhibitors (discussed above and reviewed in (Labrie *et al.*, 2010)). However, invasion by DNA encoded on temperate phages, plasmids, and conjugative elements could, in some cases, be beneficial due to advantageous genes from these sources being incorporated into the bacterial genome. Supporting this idea, many bacterial species acquire foreign DNA through natural competence and/or have evolved strategies to safely incorporate foreign DNA, including H-NS-mediated silencing of foreign DNA (Navarre *et al.*, 2006). Horizontally acquired traits, such as antibiotic resistance and virulence factors, increase the fitness of many bacterial species. This presents a potential cost of CRISPR-Cas (i.e. destroying beneficial foreign DNA) in addition to the energetic costs of maintaining and producing the CRISPR-Cas surveillance system. Thus, due to the unpredictable effects of foreign DNA uptake, evaluating the evolutionary pressure for the maintenance or loss of a CRISPR-Cas system is complicated (Weinberger and Gilmore, 2012).

Experiments utilizing temperate phages, which are able to integrate their genomes into the bacterial genome, have demonstrated that CRISPR-Cas systems can block the uptake of foreign DNA into bacterial genomes. CRISPR-Cas systems can inhibit temperate phages during infection or after integration has taken place (Edgar and Qimron, 2010). The demonstration that bacterial genomic DNA (i.e. a lambda prophage) is not intrinsically protected from the CRISPR-Cas system provided a key differentiation between CRISPR-Cas and restriction enzymes. Several other experimental approaches have subsequently confirmed that a CRISPR-Cas system can kill the cell when a spacer co-exists with a chromosomal protospacer (Manica *et al.*, 2011; Yosef *et al.*, 2012; Vercoe *et al.*, 2013). The ability of the CRISPR-Cas systems to target the host genome explains why, in general, few perfect matches (0.4% of examined spacers) between CRISPR spacers and the host genome are observed (Stern *et al.*, 2010). At the outset of my thesis work, however, examples of perfect matches between a bacterial genome and its CRISPR system or between a permissive infecting phage and the host CRISPR system were difficult to explain and considered to represent inactive or alternately functioning CRISPR-Cas systems (Stern *et al.*, 2010; Cady *et al.*, 2011; Cady and O'Toole, 2011). With my work presented in CHAPTER 4, I provide a novel explanation for these observations.

As prophages can be beneficial to the host bacterium, CRISPR-Cas systems could reduce fitness when they eliminate or prevent the acquisition of beneficial prophages. For example, an organism with a large prophage population, *Streptococcus pyogenes*, has many virulence factors that are prophage-encoded. It has been observed that *S. pyogenes* strains with large numbers of prophages (up to 8 in one genome) generally possess fewer spacers in their Type II and Type I-C CRISPR loci (Nozawa *et al.*, 2011). Although 27 out of the 41 different CRISPR spacers present in these strains matched streptococcal phage genomes, no single strain possessed a spacer that matched a resident prophage within the same strain. These data imply that CRISPR-Cas systems prevent integration of phage genomes in this species even though the phage DNA may be beneficial. Thus, the net outcome on fitness is a balance between the selective pressures that favor the presence of a specific prophage and those that favor the presence of a given CRISPR spacer, given the inability of the CRISPR-Cas system to distinguish a beneficial phage from a detrimental one.

The role of CRISPR-Cas is not limited to resisting phage infection, as plasmids can also be targeted. For example, conjugation efficiency into *Staphylococcus epidermidis* was reduced by greater than 10^4-fold when the conjugated plasmid possessed a protospacer matching a spacer in the Type III-A CRISPR locus present in this species (Marraffini and Sontheimer, 2008). To assess the role of naturally occurring CRISPR-Cas systems in preventing the acquisition of plasmids, retrospective analyses have been conducted in the human pathogens, *Enterococcus faecalis* and *Enterococcus faecium*. The presence of a Type II CRISPR-Cas system was found to have a significant inverse correlation with the presence of horizontally acquired antibiotic resistance genes (Palmer and Gilmore, 2010), suggesting that CRISPR-Cas systems may be functioning in what would appear to be a non-beneficial manner by preventing acquisition of useful genes. Further, RNA sequencing studies on four *Campylobacter jejuni* strains revealed that two had non-functional Type II-C systems (i.e. obvious *cas* gene mutations or deletions) while the other two strains possessed intact CRISPR-Cas systems which produced mature crRNAs. The strains with defective CRISPR-Cas each possessed a prophage or virulence-conferring plasmid not present in the other two strains, where there was a CRISPR spacer matching the virulence plasmid (Dugar *et al.*, 2013). These data indicate that this system is likely excluding potentially beneficial genes from *C. jejuni*. Recently, an experimental approach designed to make the cell 'choose' between an antibiotic resistance-bearing plasmid and its Type

III-A CRISPR-Cas system showed that *S. epidermidis* lost CRISPR-Cas function through a number of different mechanisms to acquire the plasmid under times of selection (Jiang *et al.*, 2013). No mutations of the plasmid protospacer were seen, which is in contrast to lytic phage experiments which find evasion mutations in the phage protospacer and PAM (Deveau *et al.*, 2008).

In addition to phages and plasmids, CRISPR loci can also interfere with competence and natural transformation. Naturally competent bacteria such as *Streptococcus pneumoniae* can take up foreign DNA and be transformed, for example, to produce a capsule, thereby increasing the virulence of the strain in a mouse infection (Avery *et al.*, 1944). Interestingly, *S. pneumoniae* is naturally devoid of CRISPR-Cas systems, while related streptococci possess active systems of different types. When a Type II system from *S. thermophilus* was artificially introduced into *S. pneumoniae* and targeted towards capsule genes, the process of transformation no longer occurred, and mice were protected during infection (Bikard *et al.*, 2012). At low frequencies, the introduced CRISPR-Cas system was lost, leading to DNA acquisition and a lethal mouse infection. Although in an artificial set up, these results suggest that the absence of a CRISPR-Cas system has been adaptive for this organism; the cost of detrimental foreign DNA invasion (i.e. lytic phage) has perhaps been balanced by the benefit of natural competence. This is contrasted with a species from the same genus, *S. thermophilus*, which has an active CRISPR-Cas system as well as natural competence mechanisms, but with most spacers targeting phage and plasmids (Horvath *et al.,* 2008). Conversely, in *Neisseria meningitidis*, the Type II-C CRISPR-Cas system of this naturally competent species appears to be highly active, with isolates possesses a diverse collection of spacers, with ~97% of all database matches being to other *N. meningitidis* and *N. gonorrhoeae* genomes. Despite this, most spacers do not have a match within their own genome and those that do, come with PAM mutations which likely eliminates self-targeting. Further, it has been shown experimentally that this CRISPR-Cas system can indeed block the transformation process. Thus, the CRISPR-Cas system appears to be able to limit interstrain and interspecies genetic exchange within this genus, but is maintained (Zhang *et al.*, 2013). Although the selective pressures that drive these variable outcomes relating to competency are not entirely clear, horizontal transfer and CRISPR-Cas systems certainly have a complex relationship which elicits different phenotypes over the course of evolution. The presence of genes on mobile

elements which interfere with CRISPR-Cas function would presumably be advantageous for these elements, although at the outset of my work no such examples were known.

1.4.4 Mobile CRISPR-Cas systems

CRISPR-Cas systems are a potent method to a way to resist foreign DNA invasion in the cell, but these systems are found on plasmids and megaplasmids (Godde and Bickerton, 2006). In the cyanobacterium *Synechocystis*, three CRISPR-Cas loci belonging to Type I-D and Type III were found on a single 103 kb plasmid that produced highly transcribed and processed crRNAs (Scholz *et al.*, 2013). A Type I-C CRISPR-Cas system comprising 48 spacers was also found on a linear plasmid in *Streptomyces rochei* with no matches to any putative targets (Yang *et al.*, 2011). Despite the absence of chromosomally encoded CRISPRs in *Lactococcus lactis*, this organism appears to have an active plasmid-encoded Type III-A system that is self transmissible and contains many spacers that match phage targets (Millen *et al.*, 2012). This sharing of plasmid-encoded CRISPR-Cas systems can result in the ability of many different related strains to exclude detrimental foreign DNA without the need to each independently acquire CRISPR spacers against common phage targets. These plasmids also enable horizontal transfer of CRISPR-Cas systems which likely explains the distribution of these systems in a way that does not necessarily match the phylogeny of their host (Makarova *et al.*, 2011b).

In addition to plasmid-encoded CRISPR-Cas systems, examples have emerged of phage-encoded systems as well. Before the role of CRISPR-Cas systems were fully appreciated, several CRISPR-Cas loci were found in mobile elements in *Clostridium difficile,* including two in prophages (Sebaihia *et al.*, 2006). Later studies revealed that *C. difficile* isolates have many CRISPR arrays (up to 34 in one isolate), along with *cas* genes, and expression was detected from all 12 of the Type I-B CRISPR loci in one strain, 5 of which were found in prophages (Soutourina *et al.*, 2013). In addition, metagenomic studies of the human gut have revealed examples of prophage-encoded CRISPR arrays, representing a large diversity of CRISPR types and spacers with matches to co-existing viral populations (Minot *et al.*, 2011; Minot *et al.*, 2013). *Vibrio cholera* phage ICP1 encodes a functional Type I-F CRISPR-Cas system that it utilizes to neutralize a phage-inducible chromosomal island-like element that would otherwise mediate phage resistance. The phage CRISPR array has spacers matching the island and protospacer mutations on the island evade CRISPR targeting, thus preventing phage infection (Seed *et al.*,

2013). The phage CRISPR locus can acquire new spacers against this island to shift the balance back to favour the phage. In these cases outlined above, either a mobile element uses its own functional CRISPR-Cas system as a means to invade the host or a newly acquired and established CRISPR-Cas system provides a fitness advantage to the recipient in times of detrimental DNA exposure. Overall, the occurrence of CRISPR-Cas systems encoded by mobile DNA again emphasizes the difficulty in distinguishing between 'good' and 'bad' foreign DNA.

1.4.5 Inactivation of CRISPR-Cas systems

To gauge the impact of the CRISPR-Cas system in any particular species, it is essential to determine whether the system is active under natural and/or laboratory conditions. This is a relevant concern since few species with CRISPR-Cas systems have been shown to be active through experimental challenge with phage or plasmids (Table 2). Most notably, the Type I-E system of *E. coli* is one of the most thoroughly studied systems, yet it is repressed under laboratory conditions (Pul *et al.*, 2010). It may also be repressed in natural conditions because among the many spacers in the CRISPR loci of *E. coli* strains, there is little interstrain diversity observed (Touchon *et al.*, 2011) and few matches to sequenced phages or plasmids. These observations are consistent with a CRISPR-Cas system that has been inactive for more than >200,000 years in this organism (Touchon and Rocha, 2010; Touchon *et al.*, 2012). Further, no CRISPR-Cas mediated exclusion of antibiotic resistance-encoding plasmids was observed among 263 *E. coli* isolates (Touchon *et al.*, 2012). This situation contrasts with the case of *E. faecalis* where many matches between plasmids and CRISPR spacers were observed (Palmer and Gilmore, 2010). To elicit anti-phage activity from the *E. coli* system, it must be activated either by overexpression of transcriptional activator LeuO or elimination of the repressor H-NS (Westra *et al.*, 2010). The maintenance of regulated functionality of this system suggests that it may perform an alternative function. A Type I-E CRISPR-Cas system of *Salmonella enterica* serovar Typhi was shown to be repressed in a similar manner to the *E. coli* Type I-E system (Medina-Aparicio *et al.*, 2011). The inactivity of the CRISPR-Cas system in the leading Gram-negative model organism *E. coli* has presented some barriers to studying the activity of this system. Thus, my description in CHAPTER 3 of a naturally active CRISPR-Cas system that targets phages and plasmids in *P. aeruginosa* is significant as it represented the first description of such a functional system in a Gram-negative organism. Finally, even when a bacterial strain possesses a CRISPR-Cas system with many spacers matching existing phages, this system may

still be inactive towards a given target. The most commonly observed evasion mechanism for phages has been via mutation in the protospacer or PAM region (Deveau *et al.*, 2008). As described in CHAPTER 4, this is not the only mechanism for this phage-mediated evasion of CRISPR-Cas, a finding which will likely extend to other mobile elements.

1.4.6 Alternative CRISPR-Cas functions

Upon the discovery of CRISPR loci through bioinformatic means, it was hypothesized that they may be involved in gene regulation, analogous to functions of RNAi in eukaryotes (Makarova *et al.*, 2006). This hypothesis was strengthened in 2009 when it was shown that the Type III-B CRISPR-Cas system of *Pyrococcus furiosus* cleaves RNA *in vitro* (Hale *et al.*, 2009). While the major role of most CRISPR-Cas systems appears to be resisting invasion by foreign DNA, novel roles of CRISPR-Cas systems in gene regulation have emerged in recent years. For example, *Francisella tularensis* subsp. *novicida* appears to have an active Type II CRISPR-Cas system with a full suite of *cas* genes and arrays with multiple spacers matching phages (Schunder *et al.*, 2013). In addition, this system mediates repression of an endogenous lipoprotein-encoding gene through imperfect base pairing between a *trans*-activating crRNA (tracrRNA) and the target transcript, in a region spanning the start codon (Sampson *et al.*, 2013). This endogenous gene regulation is necessary for full virulence of *F. novicida* in mice and represents the first characterized demonstration of CRISPR-Cas-mediated gene regulation along with a role in virulence.

In *P. aeruginosa*, a few studies had been conducted on the CRISPR-Cas system, primarily with phage DMS3. When DMS3 is present as a prophage, it mediates the inhibition of biofilm formation and swarming motility (Zegans *et al.*, 2009). This inhibition is dependent on full activity of the *P. aeruginosa* CRISPR-Cas system and also required a CRISPR spacer with complementarity to a region of DMS3 gene *42* with five mismatches (Cady and O'Toole, 2011). Not only did this work demonstrate an alternative outcome of CRISPR-Cas function, these data also showed that not all mismatches will necessarily abolish recognition by the CRISPR-Cas system, consistent with work in the Type I-E CRISPR-Cas system showing that mismatches do not eliminate target DNA binding (Semenova *et al.*, 2011). Although cleavage of the DMS3 prophage is unlikely, as it would kill the cell, there is likely a still undetermined effect on transcript production from this region due to CRISPR-Cas complex recruitment. The knowledge

that protospacers with up to five mismatches may still mediate *in vivo* function presents a challenge when attempting to identify and fully understand CRISPR-Cas targets bioinformatically, although excellent tools exist to approach this problem, such as CRISPRTarget (Biswas *et al.*, 2013).

Furthermore, another CRISPR-Cas study in *P. aeruginosa* presented that phages with protospacers containing perfect matches to the CRISPR-Cas system were seemingly recalcitrant to targeting, being able to infect wild type cells and ΔCRISPR-Cas mutants with similar efficiencies (Cady *et al.*, 2011). Together with the studies above, it was suggested that the Type I-F CRISPR-Cas system could be performing alternate functions (i.e. biofilm inhibition), supported by the apparent lack of targeting towards infecting phages. As I will demonstrate in CHAPTER 3, the type I-F CRISPR-Cas system of *P. aeruginosa* does in fact target phages during infection, however some phages are certainly resistant to this targeting. The ability of phages to elude the CRISPR-Cas system will be explained in CHAPTERS 4 and 5, with the discovery and mechanistic characterization of the first examples of phage-encoded anti-CRISPR proteins.

1.4.7 CRISPR-Cas Summary

CRISPR-Cas systems provide a powerful means for bacteria to destroy potentially harmful foreign DNA, and the very common occurrence of these systems within bacterial genomes emphasizes the positive influence that these systems must have on bacterial fitness. However, as summarized above (section 1.4.3), the current literature provides many examples where acquisition of foreign DNA may be advantageous for an organism and the possession of an active CRISPR-Cas system could be non-adaptive. This likely explains the observation that <50% of sequenced bacteria possess these systems despite their ability to be horizontally transferred. The principal conclusion of this section is that the net biological outcome of a CRISPR-Cas system within a given organism in a given environment is difficult to predict. This uncertainty arises because the balance between the beneficial and detrimental effects of foreign DNA depends on the nature of the DNA being acquired and this property varies among species. In addition, some systems may be inactive (i.e. repressed or defective) or be performing alternative roles. For improvement of our understanding of the complex biological outcomes that can result from the presence of CRISPR-Cas systems, future studies must focus on *in vivo*

characterization of more systems operating in diverse species. As shown in Table 2, only a small fraction of systems have been analyzed *in vivo*. Future *in vivo* work would address the important questions of what percentage of seemingly intact CRISPR-Cas systems are actually able to resist the invasion of foreign DNA, and also whether CRISPR-Cas systems are commonly performing alternative functions. Increased accumulation of data pertaining to the *in vivo* functioning of CRISPR-Cas systems will allow accurate interpretation of the roles that these systems are playing in various bacterial species. Since CRISPR-Cas systems can provide a unique "fossil record" of encounters with foreign DNA within bacterial species, this knowledge will greatly improve our understanding of bacterial evolution and the impact of horizontal gene transfer on the environment and human health.

1.5 Thesis Objectives

The abundance of prophages in bacterial genome sequences, specifically in *Pseudomonas aeruginosa*, is striking and yet their role is poorly defined. The objective of this thesis work is to characterize this prophage population and determine its role in bacterial fitness, specifically the way in which they impart a phenotype of superinfection exclusion. Furthermore, the CRISPR-Cas system has been established as a powerful prokaryotic immune system, but few studies have addressed the *in vivo* activity of these systems. By generating a library of *P. aeruginosa*-infecting phages, I hope to address some of these intricacies, as well as assessing what phages do to counteract the CRISPR-Cas system. With the potential that prophages and other mobile elements have to be beneficial to their host, CRISPR-Cas mediated inhibition of foreign DNA acquisition could be an evolutionary downside of these systems. These topics will be addressed in this thesis.

1.6 Thesis Outline

In CHAPTER 2, I will focus on a screen which I conducted to induce prophages from genomes of diverse *Pseudomonas aeruginosa* isolates to assess their role and abundance. By inserting these prophages into a single strain background, I show that superinfection exclusion beyond repressor activity is imparted by nearly every prophage. By utilizing this library of temperate phages, I will demonstrate in CHAPTER 3 that the CRISPR-Cas system of *P. aeruginosa* is active and can target phages and plasmids. At the time, this work represented the first demonstration of a naturally active CRISPR-Cas system in a Gram-negative bacterium. Despite

the ability of the CRISPR-Cas system to target phages, a number of phage isolates appeared to be recalcitrant to targeting, an observation that led to my discovery of the first phage-encoded anti-CRISPR genes. These data will be presented in CHAPTER 4, resolving a significant question in the CRISPR-Cas field about whether such entities exist and demonstrating the potent selective force CRISPR-Cas systems have imposed upon phages. Finally, in CHAPTER 5 I will outline work to characterize the molecular mechanisms of three distinct anti-CRISPR proteins. These data demonstrate diverse mechanism employed by anti-CRISPR proteins, which all serve to inhibit CRISPR-Cas targeting and protect phage genomes. A summary and future directions will be presented in CHAPTER 6.

Parts of this chapter were adapted from:

Bondy-Denomy J, Davidson AR. (2014) When a virus is not a parasite: the beneficial effects of prophages on bacterial fitness. *Journal of Microbiology,* **52**(3):234-42 (Review)

Bondy-Denomy J, Davidson AR. (2014) To acquire or resist: The complex biological effects of CRISPR-Cas systems. *Trends in Microbiology*, **22**(4):218-25 (Review)

Chapter 2
Prophage acquistion in *Pseudomonas aeruginosa* imparts a diversity of phenotypes

2 Overview

Pseudomonas aeruginosa is an opportunistic pathogen and an environmentally ubiquitous organism. As genes expressed from prophages present in bacterial genomes have been shown to have a profound influence on bacterial physiology and pathogenesis, I surveyed a collection of 88 clinical and environmental isolates of *P. aeruginosa* for the presence of inducible prophages and examined the phenotypic effects mediated by this group of phages. I identified phages produced from 66% of these strains and determined that they display a very diverse host range. The majority of phages isolated depend on the type IV pilus (T4P) for infection, with a smaller group using the O-antigen as a receptor, and some depending on both. I created a collection of lysogens in a single strain, PA14, with each containing a single prophage, and examined the ability of these strains to resist phage superinfection. A range of resistance patterns were observed, including very strong multi-phage resistance to all T4P-specific phages in the collection. The resistance mechanisms included repressor-mediated immunity and superinfection exclusion by putative modifications of the T4P and the O-antigen. Phage genome anaylsis revealed a number of small, unique genes in otherwise highly-related phage genomes which likely play a role in prophage-mediated phenotypes. The prophage insertion site specificity was also examined with hot-spots identified for randomly integrating Mu-like phages. I also identified two insertion site hot spots which negatively affect a novel, uncharacterized virulence gene as well as the production of pyocyanin, a compound that is toxic to eukaryotic cells. This study represents the first systematic approach to address how a population of prophages influences bacterial phenotypes in *Pseudomonas aeruginosa*.

2.1 Acknowledgements

The *Pseudomonas aeurginosa* isolates were acquired from Dr. David Guttman and the multi-locus sequence typing to assess strain diversity was performed in his lab. Jason Qian, an undergraduate 4th year and summer student I supervised performed some of the twitching and swimming motility assays and cloned some of the phage repressor genes.

2.2 Materials and Methods

2.2.1 Strains and Growth Conditions

Pseudomonas aeruginosa strains were grown on LB agar or broth at 37 °C. The panel of 88 clinical and environmental strains which all phages were isolated from were acquired from Dr. David Guttman. All plaque assays were conducted at 30 °C on LB agar plates and 0.7% LB top agar, both supplemented with a final concentration of 10 mM $MgSO_4$. To perform plaque assays, overnight cultures were mixed with top agar and poured onto an LB agar plate onto which phage suspensions were spotted. For plaque purification, plate assays were conducted where phage and bacteria were mixed, preadsorbed for 15 minutes at 37 °C before adding to top agar and pouring.

2.2.2 Phage Induction and Isolation

To induce prophages, strains were grown in 5 mL LB to early-log phase (OD_{600}=0.5), when mitomycin C was added to the culture at a final concentration of 3 μg/mL. Incubation was allowed to continue until lysis was visible (3-4 hours after induction), then 200 μL of chloroform was added for 15 minutes, at which point the mixture was centrifuged at 10,000xg for 10 minutes. The ~4-5 mL of supernatant were kept and stored with 200 μL of chloroform at 4 °C. All inductions were conducted at 37 °C. These lysates were then spotted in serial dilutions on lawns of various indicator strains from the initial collection. Visible plaques determined the presence of phage and this was confirmed by a plate assay at the appropriate dilution. From the plate assay, plaques were picked and resuspended in buffer SM. All phages were subjected to three rounds of purification before use. High titer stocks of phage were precipitated with polyethylene glycol, and subjected to a cesium chloride equilibrium centrifugation gradient for purification. These stocks were used for DNA extraction and electron microscopy.

2.2.3 Lysogen Construction

Lysogens were constructed by spotting serial dilutions of a purified phage on the appropriate host and streaking out resistant bacteria from the inside of a plaque or clearing. Colonies were then screened to confirm phage resistance by cross-streaking the colony over top of a line of phage suspension which had been run down a plate. After confirming resistance, putative

lysogen colonies were grown in liquid culture, spontaneous production of the phage which could plaque on the original wildtype strain confirmed lysogeny.

2.2.4 Motility

Twitching motility was determined by stabbing a single colony through the agar layer of an LB plate with 1% agar. Plates were incubated, inverted, for 48 hours at 37 °C. The agar was then carefully peeled off and the dish was stained with 1% (w/v) crystal violet for 1 minute and washed three times with water. The diameter of the twitch zone was then measured.

Swimming motility was determined by stabbing a single colony through the agar of a plate containing 0.3% agar, 10 g/L Tryptone, and 5 g/L NaCl. After 24 or 48 hours of incubation the diameter of the swim zone was measured.

2.2.5 Electron Microscopy

Cesium chloride pure phage stocks were used for most electron microscopy examination. Phage samples were applied to carbon coated grids which were then washed with ddH_2O and negatively stained with 2% uranyl acetate. EM pictures were taken with a Transmission Electron Microscope, typically at magnification levels of 50,000-100,000x.

2.2.6 Potassium Efflux

Bacteria were grown in LB to an OD600 of 0.5 and washed twice in potassium-free SM buffer and stored on ice to prevent potassium leakage. 5 mL of cells were then brought to 37 °C and the potassium-selective electrode inserted, allowing the reading to stabilize. Cesium chloride purified phage in SM buffer were then added and mixed thoroughly and readings (in mV) taken every 5s. MOI for these experiments is approximately 100.

2.2.7 Adsorption Assay

Cells were grown for three hours in LB + 10 mM $MgSO_4$ to an OD_{600}=0.5. Phage were added to cells at an MOI=0.1 and incubated for 0, 4, or 8 minutes. At those time points, an aliquot was removed from the sample and diluted in cold buffer containing chloroform. This mixture was centrifuged to pellet cells and attached phages. The supernatant was titrated on a sensitive strain to quantify the unattached phage.

2.2.8 Southern Blot

Genomic DNA was purified from PA14 or indicated lysogens and digested with NcoI, then electrophoresed on a 0.8% TAE-agarose gel. The gel was washed in 0.25 M HCl for ten minutes, rinsed in water, and washed in 0.4 N NaOH/0.6 M NaCl for 30 minutes. The DNA was transferred to a GeneScreen plus nylon membrane with a vacuum blotter while for 60-120 minutes while adding a 20X SSC solution. Finally, the nylon membrane was washed in blocking buffer (50% formamide, 5X Denhardts solution, 0.5% SDS, 6X SSC, and 100 μg/mL herring sperm DNA) at 42 °C for 2 hours. The blocking buffer was then replaced with a fresh solution of the same buffer but with herring sperm DNA omitted and 5 x 10^6 counts per minute of ^{32}P-labelled phage DNA added. The probe was generated from a random hexamer extension of purified phage genome. After probing overnight at 42 °C, the membrane was washed two times for ten minutes are room temperature (2X SSC, 1% SDS), two times for 30 minutes at 65 °C (2X SSC, 1% SDS), and once for ten minutes at room temperature (0.2X SSC, 0.1% SDS). The membrane wrapped in saran wrap and exposed to a phosphoscreen for, then developed on a Typhoon imager.

2.2.9 Pyocyanin Quantification

Five mL overnight cultures of *P. aeruginosa* were centrifuged to clear the cells and the supernatant was removed and mixed with 3ml of chloroform. After vortexing, the lower phase was mixed with 1ml of 0.2M HCl in a glass tube until the upper phase became a pink to red colour. The absorbance of this product was measured at 520nm. The amount of pyocyanin produced (in μg/mL of supernatant) can be calculated by multiplying the OD_{520} value by 17.07 (Pérez-Martínez and Haas, 2011).

2.2.10 Reverse transcriptase quantitative PCR

Total RNA extracts were treated with DNAse (Ambion) to remove contaminating DNA and 1ng of total RNA was used in a series of RT-qPCR reactions. Reactions were conducted in an Eppendorf qPCR cycler, using VWR white plates with the SensiFAST No-ROX One-step Kit (Bioline). For the purposes of absolute quantification, a PCR reaction was conducted with primers indicated below (PCR1/2), amplifying genomic DNA with ‘external’ primers. This product was gel extracted, quantified and diluted to generate a standard curve. All RT-qPCR reactions were done with internal primers (RT1/2) which were designed to anneal inside of the

external primers for the purified PCR product. Expression of a housekeeping gene, *rpsL*, was used for relative quantification. For RT-qPCR reactions, 1 ng of total RNA was used in each reaction, performed in duplicate. Reverse transcription was conducted using a gene specific primer to generate cDNA in a one-step reaction. The lack of contaminating DNA was confirmed by inclusion of controls without reverse transcriptase added.

phzS	
PCR 1	GATCCTGCAGTACCCGATG
PCR 2	CGCGCTCTTCTCAGTCTTC
RT 1	CAAGATCATCCTGGCCAAC
RT2	CGCGCTCTTCTCAGTCTTC
psiF	
PCR 1	TATCCTGCGTATCCCGATG
PCR 2	CTTCTTCAGGCAGGTGCTC
RT 1	GCGTATCCCGATGTTCGTA
RT2	CAGGCGGTCATCTTCTCC
rpsL	
PCR 1	GCAAGCGCATGGTCGACAAGA
PCR 2	CACGGATCAGCACTACGC
RT 1	GTCGACAAGAGCGACGTG
RT 2	AACCGTTGGTCAGACGTACA
phzM	
PCR 1	TGCTGCGCGTAATTTGATA
PCR 2	AACTCCTCGCCGTAGAACA
RT 1	GCGACGGCTACGCTAATAC
RT 2	AACTCCTCGCCGTAGAACA

2.2.11 Inverse PCR

Genomic DNA of a lysogenized strain was extracted and digested with NcoI. The DNA was then self-ligated to circularize digested products. A PCR reaction was carried out, using primers (see below) which bind to the terminus of the phage genome and amplify outwards into the bacterial genome. PCR products were sequenced to identify the junction between phage and bacterial DNA.

JBD16CinvF	CAGCTTTCTGACTTGGATATCATTG
JBD16CinvR	GAATCACATAGATCGCTTCATCC

2.2.12 *Caenorhabditis elegans* virulence assay

This protocol was followed directly from (Tan *et al.*, 1999a). Indicated PA14 derived lysogen overnight cultures were plated on slow killing medium (SKM) and incubated for 24 hours at 37

°C and then at room temperature for 24 hours. Approximately 40 L4/young adult stage worms were transferred to the *P. aeruginosa* plates in duplicate. Live and dead worms were scored manually over the time course of the experiment. Dead worms were removed from the plates to ensure accurate counting.

2.2.13 Phage DNA Extraction

Cesium chloride purified phages were treated with 20 mM EDTA, 50 μg/mL proteinase K and 0.5% SDS for 1 hour at 56 °C. The DNA was then extracted with subsequent steps using equal volumes of phenol, phenol/chloroform, and chloroform with centrifugation at 3,000 x g for 5 minutes at each step coupled with extraction of the top aqueous layer. DNA was precipitated with 0.3 M sodium acetate (pH 7.0) and 2 volumes of ice-cold 100% ethanol. This mixture was incubated at room temperature for 30 minutes and centrifuged at 10,000 rpm for 2 minutes. The pellet was washed once in 70% ethanol and air dried. DNA was resuspended in water and frozen at -20 °C for future use.

2.2.14 Phage Genome Sequencing and Analysis

Phage genome sequencing was conducted on the Illumina platform, with paired-end reads from a multiplexed run, sequencing 12 phage genomes at once. Assembly was run on Velvet and coverage was >400X for each phage. Sample preparation, sequencing and assembly were done in the Guttman lab, with help from Pauline Wang, Pauline Fung, and Yunchen Gong. Genome analysis was conducted primarily with BLAST (Altschul *et al.*, 1990) and RAST (Aziz *et al.*, 2008) programs.

2.2.15 Phage Repressor Cloning

A shuttle vector that replicates in both *E. coli* and *P. aeruginosa* called pHERD30T was used to clone phage repressors either at the EcoRI (GAATTC) and HindIII (AAGCTT) restriction sites or NcoI (CCATGG) and HindIII. A PCR reaction was conducted with appropriate primers (shown below, restriction sites underlined and bolded) and phage lysate as the template. PCR products were gel extracted, subject to restriction digest and ligated to a digested and gel extracted pHERD30T vector. DH5α cells were transformed with the ligation mixture and clones confirmed by sequencing. Plasmids were isolated from *E. coli* using a BioBasic Miniprep Kit. *P. aeruginosa* cells were made electrocompetent by washing overnight cultures twice in an equal

volume of 300 mM sucrose, and finally re-suspending the cells in 1/10 volume of 300 mM sucrose. One hundred μL of cells were electroporated with 300-400 ng of the indicated repressor construct, recovered in 1.0mL of LB medium for 1 hour at 37 ^{o}C. Following incubation, cells were spread on an LB plate containing 50 μg/mL gentamicin. Each plate was incubated at 37 ^{o}C for 18 hours and the resulting colonies were picked and phages spotted to assess the effect of repressor expression.

JBD88a-1_F	CCCGGG**CCATGG**GTTACGAACAAACGGATG
JBD88a-1_R	CCCGGG**AAGCTT**CTAAGCCATCCAGCGGCCTG
JBD24-1_F	CCCGGG**CCATGG**CCACAGACAGCCTTGCGGCCCG
JBD24-1_R	CCCGGG**AAGCTT**TCAAACCATCCAATAGCTGG
JBD16C-1_F	CCCGGG**GAATTC**GAAAACGCGTAGTGAGCGCACC
JBD16C-1_R	CCCGGG**AAGCTT**CTAAACCATCCATCGCCCAA
JBD30-1_F	GGAGATATACATAC**CCATGG**CAATGACCGAAGGTCTGGCTGC
JBD30-1_R	GGCCAGTGCC**AAGCTT**CTAAGTCATCCAGCGGCCTG
JBD26-1_F	GGAGATATACATAC**CCATGG**CAGTGAAATCAGACACTTACGG
JBD26-1_R	GGCCAGTGCC**AAGCTT**CTAAACCATCCAGCGGCTAG
JBD44a-rep_F	GGAGATATACATAC**CCATGG**AACTCAAAGACCGCAT
JBD44a-rep_R	GGCCAGTGCC**AAGCTT**TTAGGAACCGATTCCGCCGC
JBD88b-rep_F	GGAGATATACATAC**CCATGG**ACTTTTCAGACAGACT
JBD88b-rep_R	GGCCAGTGCC**AAGCTT**CTACATCGCGCCGCCACGCC

2.3 Results

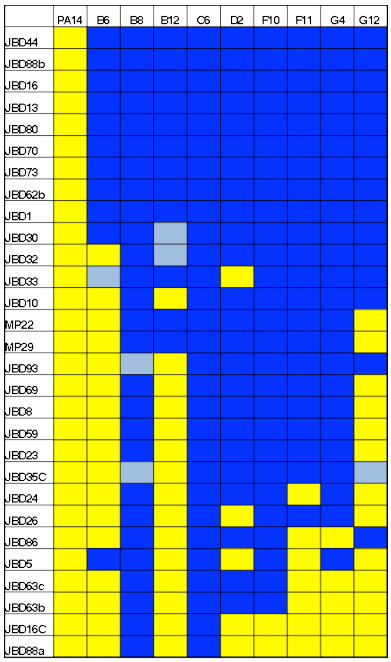
2.3.1 *Pseudomonas aeruginosa* strains frequently harbour inducible prophages

To assess the diversity of prophages present in a collection of clinical and environmental isolates of *P. aeruginosa* and gather an unbiased collection of phages for further analysis, I acquired a set of 88 strains from Dr. David Guttman (University of Toronto). These strains were isolated from diverse geographical locations as well as sites (~2/3 are clinical, ~1/3 are environmental). They were also genotyped at up to seven loci by multi-locus sequence typing in the Guttman lab and selected from a larger group to represent a diverse set of strains. I used mitomycin C to induce production of phages that reside in their genomes. In order to detect phages, a subset of indicator strains were selected from the total collection. This was done by conducting a preliminary screen with eight previously characterized *P. aeruginosa* phages (D3112, F8, D3, 16 Lindberg, F10, M6, phiKZ, 73 Lindberg) acquired from the d'Herelle Centre phage repository against all 88 *P. aeruginosa* strains (data not shown). Nineteen total indicator strains were selected based on a diverse susceptibility and resistance patterns to these eight phages. The indicator panel included

the commonly used laboratory strains PA14, PAO1, and PAK. Serial dilutions of the 88 mitomycin C-induced lysates were spotted onto lawns of the 19 different *P. aeruginosa* indicator strains, and plaque formation was observed. I found that 58/88 (66%) of the strains produced at least one phage that could form plaques on the indicator panel. I used plaque morphology to discern the induction of multiple phages from a single strain; five strains produced two distinct plaque morphologies and one strain produced three. A total of 70 temperate phages were isolated and host ranges were determined by spotting serial dilutions of each phage on each of the 19 indicator strains. A subset of these data are shown in Table 3.The ability of the phages to plaque on the indicator strains was highly variable, with some phages only able to produce plaques on one strain and others able to plaque on eight different strains.

Table 3: Host range of selected phages on isolates of *P. aeruginosa*

Phages shown along the y-axis of the table were used to infect the strains along the top. Yellow indicates plaquing and dark blue indicates resistance. An intermediate blue colour indicates an intermediate resistance of ≥10-fold reduction in plaquing compared to PA14.



2.3.2 Prophages confer a wide variety of phage resistance phenotypes

To determine the bacterial phenotypes influenced by this group of prophages, I created single lysogens in *P. aeruginosa* strain PA14. I chose this commonly used laboratory strain as it has no detectable prophages based on sequence analysis, plaque assays, or examination of induced lysates by negatively stained electron microscopy. The 30 phages used were those shown in

Table 3, which are all able to infect PA14. The natural levels of spontaneous induction of these phages were assessed by growth of overnight cultures and enumeration of the phages present in the supernatant via plaque assays. There was high variability observed, with greater than 10^4-fold differences observed between phages that were spontaneously produced with high and low frequencies (Figure 3).

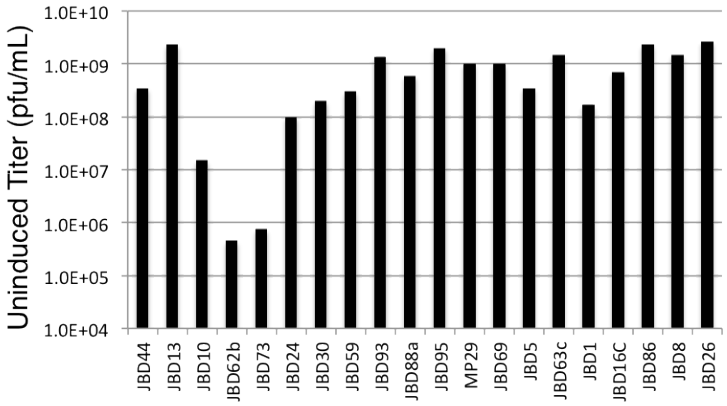


Figure 3: Titer of phages spontaneously produced by constructed lysogenic strains

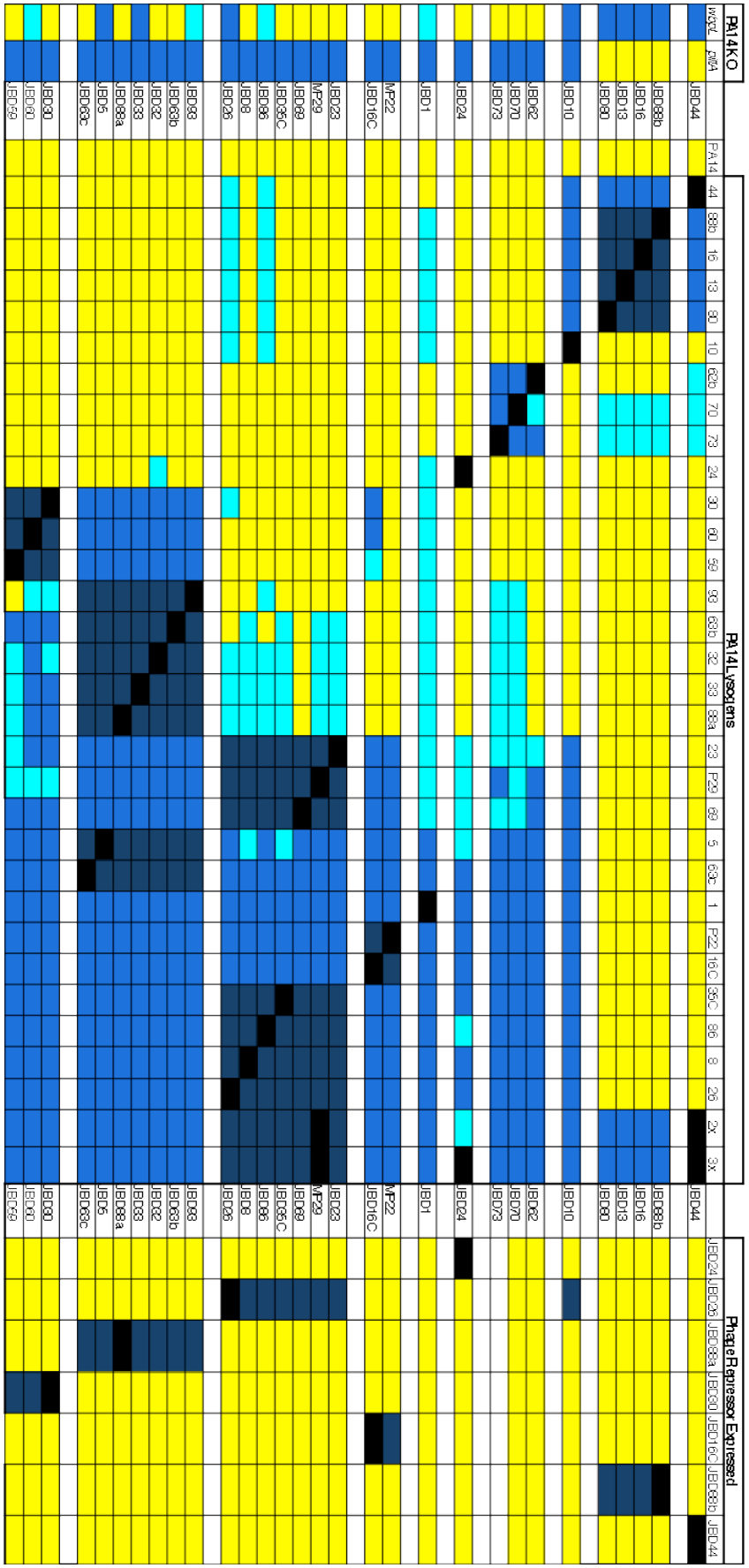
Overnight cultures of indicated PA14 lysogens were centrifuged to clear cells and the supernatant titrated on PA14.

I produced high titre lysates of each of the 30 phages used to create the lysogens, and challenged the set of lysogens with this collection of phages. As *P. aeruginosa* phages frequently use the type IV pilus or O-antigen as a receptor, I also examined the abilities of these phages to form plaques on strains with single gene deletions of *pilA*, which lacks the type IV pilus (T4P); *wbpM*, which lacks the O-antigen; and *wbpL*, which lacks both the heteropolymeric O-antigen specific polysaccharide and the homopolymeric common polysaccharide antigen (Köhler *et al.*, 2010). The T4P was required by 25 of the 30 phages tested and the remaining five phages were unable to form plaques on the *wbpM* and *wbpL* mutants, illustrating a requirement for the O-antigen (Table 4). Interestingly, eight of the phages that required the pilus for infection were also unable to propagate on a *wbpL* or *wbpM* mutant, revealing a dual requirement for the pilus and O-antigen for successful infection (Table 4). Examples of plaquing on lysogens and mutant strains are shown in Figure 4. Note the strong resistance imparted by the indicated PA14 mutant strains as well as by some lysogens. Other lysogens provide a partial resistance effect which results in a lower titer, or smaller/more turbid plaques.

Table 4: Lysogens display a wide variety of superinfection exclusion

Thirty phages were used to challenge PA14, indicated mutants in the PA14 background, PA14 lysogens, or PA14 strains expressing the indicated phage repressor proteins. Strains listed as '2x' and '3x' represent double and triple lysogens, respectively. Resistance is shown in two shades of blue, at >10-fold or >1000-fold reduction in plaque forming ability relative to wtPA14 (see legend)

Legend	
	Sensitive
	Partially resistant (>10-fold)
	Resistant (>1000-fold)
	Repressor group
	Resistant to self



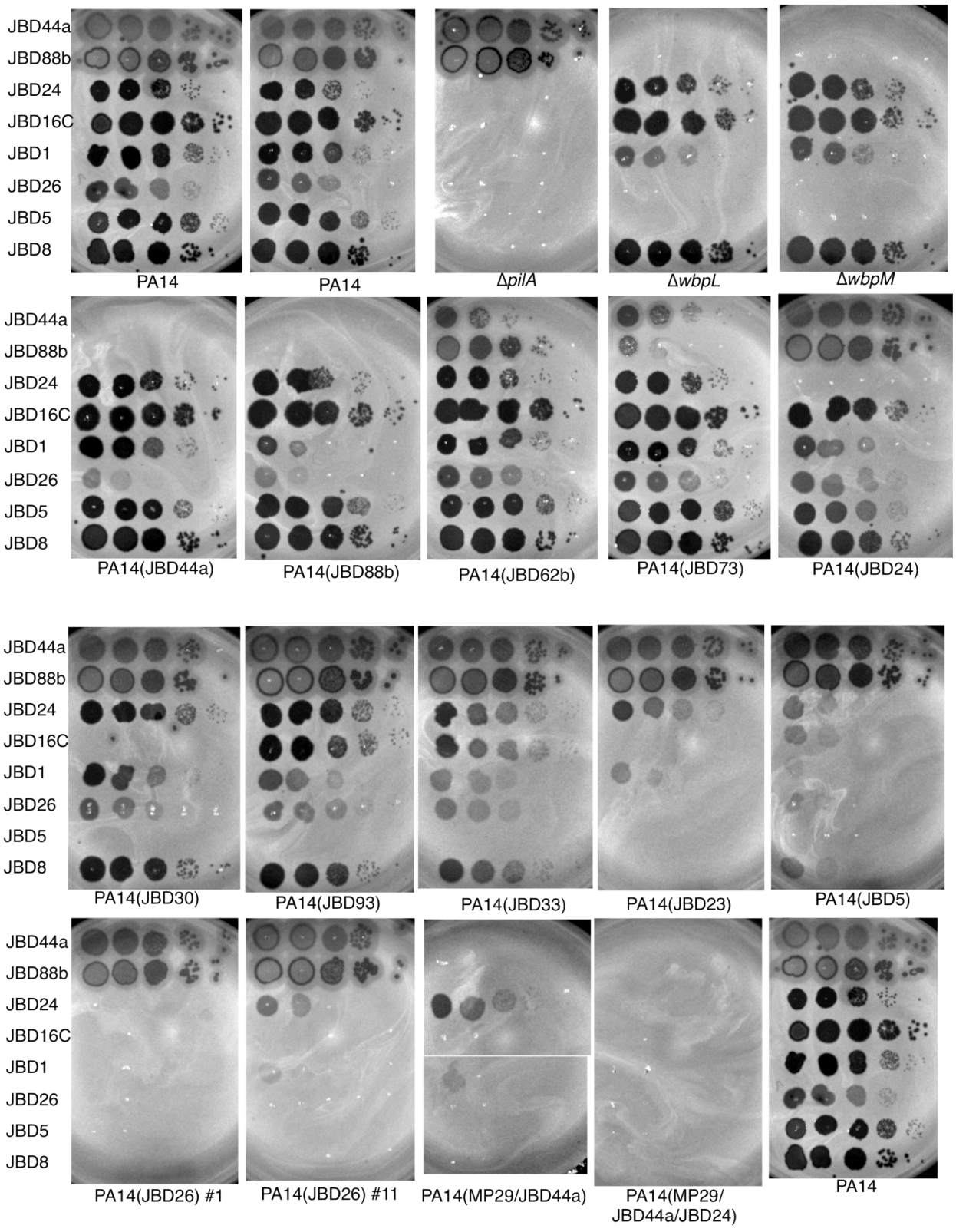


Figure 4: Pictures of 8 selected phages on PA14 mutant strains and lysogens

Ten-fold dilutions of indicated phages were applied to the indicated lawns.

I observed significant variability in the abilities of the phages to form plaques on the set of 30 PA14 lysogens. In general, the phages that required the O-antigen for infection were unable to form plaques on lysogens of other O-antigen-specific phages, but could propagate on lysogens of phages that use the pilus for infection (Table 4). Similarly, the phages that utilize the pilus for infection were inhibited by lysogens of other pilus-specific phages but could form plaques on lysogens of O-antigen requiring phages. To confirm that these effects were not somehow unique to phages isolated in my screen, I also examined the host range of three *P. aeruginosa* siphophages (long non-contractile tails) isolated and characterized by others on the lysogen collection (Table 5), in addition to the two included in the collection (MP22 and MP29). Also, four distinct myophages (long non-contractile tails) were acquired from the d'Herelle phage repository and tested on a subset of the lysogen collection (Table 6). The host ranges observed are comparable to the phages isolated in this study, demonstrating that the superinfection exclusion is broadly applicable to very different phages.

Table 5: Host range of three previously characterized siphophages on lysogen collection

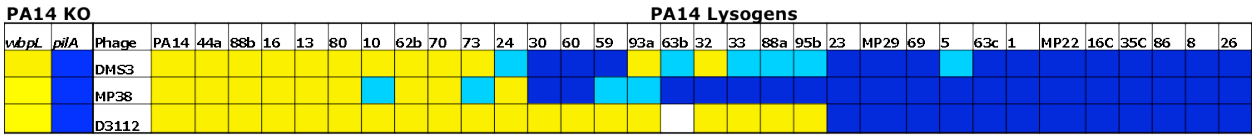
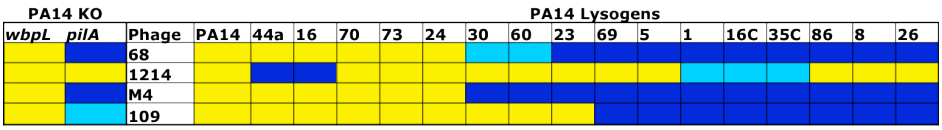


Table 6: Host range of four previously characterized myophages on lyosgen collection



I detected no direct correlation between the number of phages a prophage provided resistance to and the ability of that particular phage to form plaques on other lysogens. Some phages that provide significant protection when present as a prophage are themselves not able to efficiently infect the lysogen collection (e.g. JBD26). Conversely, many prophages that maintain sensitivity to a large number of invading phages are able to infect a significant proportion of the lysogens (e.g. JBD24).

As most bacterial genomes harbour multiple prophages, I also characterized lysogens that contained more than one prophage to assess the additivity of the phenotypic effects. For example, the double lysogen PA14(JBD44/MP29) contains both an O-antigen (JBD44) and

pilus-specific (MP29) prophage, and displayed wide resistance to both groups of phages (shown as '2x' in Table 4). A single phage from my collection, JBD24, was the only phage tested that was able to propagate in this strain; a triple lysogen (3x) was made, PA14(JBD44/MP29/JBD24), and this strain was resistant to every phage in the collection (Table 4).

2.3.3 Closely related phage genomes display localized diversity

To gain insight into the genetic backgrounds leading to the prophage-mediated bacterial phenotypes, I sequenced the genomes of twelve phages that formed lysogens in PA14. The genomes showed that the phages were all siphophages, a phage family which possess a dsDNA genome and a long, non-contractile tail. Transmission electron microscopy was used to assess the morphology of some of the sequenced phages and confirmed the siphophage morphology (Figure 5).

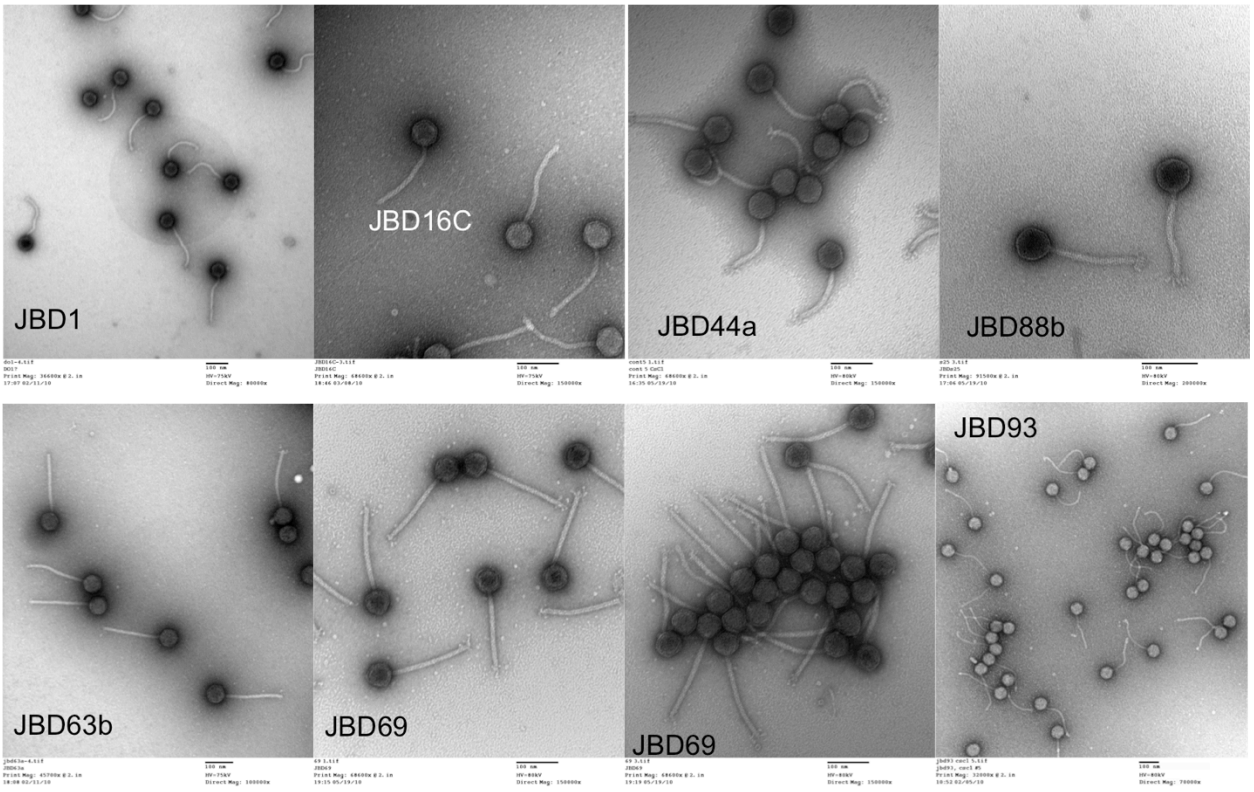


Figure 5: Phages with long, non-conctracile tails observed by TEM

Cesium chloride purified phage preparations were negatively stained with uranyl acetate and examined by transmission electron microscopy. Magnification ranges from 70,000-200,000x with the scale bar at the bottom right representing 100nm.

Ten of these phages had very closely related genomes; the majority of genes encoded highly conserved proteins (generally 80-100% identical), the genes were present in a conserved order, and many of the precise functional annotations could be assigned. This group of phages displayed significant sequence similarity with previously characterized phages MP22 and MP29. As the phage resistance patterns displayed by these phages when present as a lysogen were very different, these highly conserved genes are unlikely to be the cause. Together with Dr. Karen Maxwell, we examined non-conserved genes and identified 20 unrelated genes that encode proteins of unknown function inserted sporadically along the length of the genomes (Figure 6).

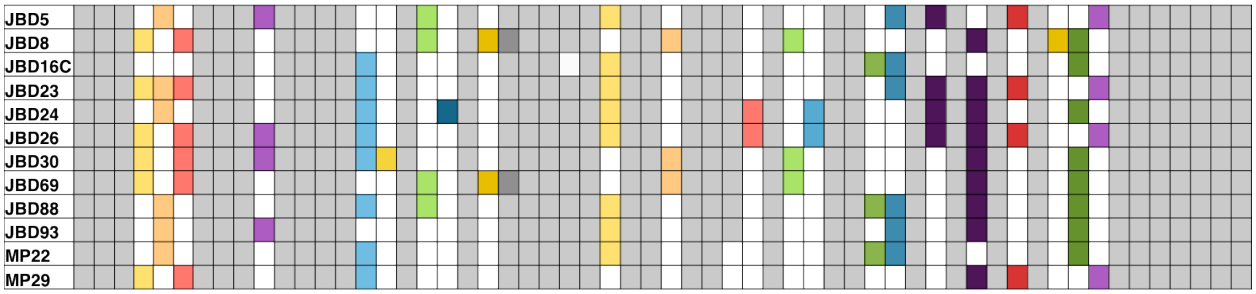


Figure 6: Related phage genomes display localized diversity.

Phage genome schematic, with each conserved gene represented as grey box and each non-conserved gene as a coloured box. Different coloured genes are not related to each other. Figure made by Dr. Alan Davidson and Dr. Karen Maxwell.

These genes, which are conserved among subsets of the phage genomes, represent an abundance of genetic diversity possessed by phages with high overall synteny. These genes could be imparting novel functions to the prophage and lysogenized strain such as excluding superinfection or other uncharacterized phenotypes.

Each of the ten phages with closely related genomes used the T4P for infection and contain Mu-like transpose genes, which are required for phage replication and the random insertion of the phage genome to into the host chromosome upon lysogen formation. The remaining two sequenced phages (JBD44 and JBD88b) use the O-antigen for infection and were found to be similar to each other, both in gene order and sequence identity, but unrelated to the group of T4P-specific phages. In contrast to the T4P-specific phages, these phages possess integrase genes with which they integrate as prophages at specific sites in the bacterial genome. The insertion site mechanisms (i.e. random vs. specific) of a transposase-containing phage (JBD26) and an integrase-containing phage (JBD44) were confirmed by Southern blots (Figure 7). The changes

in restriction digest patterns of each lysogen illustrates that JBD26 inserts randomly into the host genome, while the JBD44 restriction pattern is consistent.

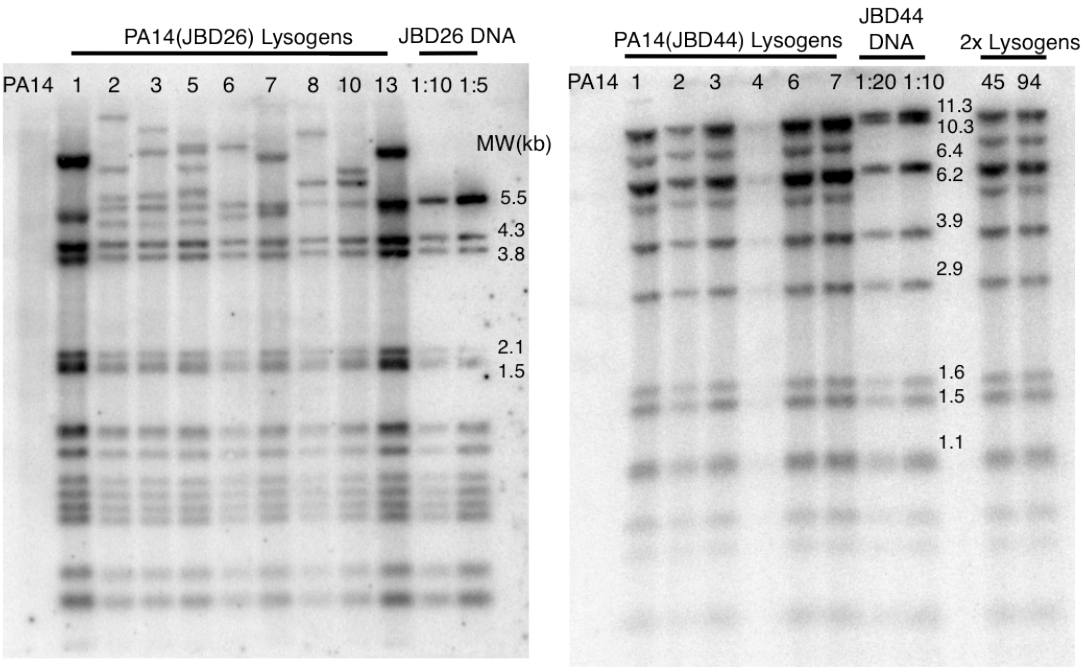


Figure 7: Southern blot reveals JBD26 inserts randomly and JBD44 inserts specifically.

Genomic DNA extracted from wtPA14 or PA14 lysogens and digested with NcoI was separated on an agarose gel, transferred to a membrane where it was probed with ^{32}P-labelled DNA generated from purified phage JBD26 or JBD44 DNA. Purified phage DNA was also probed for comparison to the bacterial genomic DNA, revealing new bands in either variable positions (JBD26) or consistent positions (JBD44). PA14(JBD44) "lysogen"#4 screened negative for the production of the phage.

I sequenced five additional phages that did not infect strain PA14: JBD68, JBD90, JBD18, JBD25, and JBD67b. Within this set, phages JBD18, JBD25, and JBD67b were found to be similar to the B3 group of phages infecting *P. aeruginosa*, a related but diverged family of Mu-like phages compared to the set presented above. These phages will be featured prominently in the Chapters 3 and 4 of this thesis and were the focus of a separate publication (Cady *et al.*, 2012). Phage JBD90 is a member of the podophage family (short non-contractile tail), related to *P. aeruginosa* phage F116 while JBD68 is similar to phage F10, a siphophage which is distinct from the other families of phages discussed above. Given the phage families identified via sequencing, it seems that there was a bias applied when working with phages that just infect PA14, whereby two distinct families of phages were selected for. A random selection of five other phages not infecting this strain revealed three more very distinct sequence families. This

illustrates the importance of using many indicator strains when attempting to assess the phage population infecting a given organism.

2.3.4 Most phage resistance is repressor-independent

Many of the lysogens examined showed non-reciprocal resistance with at least one infecting phage (Table 4). For example, phage JBD24 is unable to form plaques on a lysogen of JBD26, while phage JBD26 is able to plaque on a lysogen of JBD24. This demonstrates that the resistance observed is not a result of classical immunity mediated by phage repressor proteins, which silence incoming phage genomes that share a common repressor (Dodd *et al.*, 2005). To more fully examine the profiles of resistance arising from the presence of specific phage repressor proteins, we (myself and Jason Qian) cloned the repressors from two highly resistant prophages (JBD26, JBD16C), as well as two of intermediate resistance (JBD88a and JBD30) and three susceptible prophages (JBD24, JBD44, JBD88b) into expression plasmids in PA14 and compared the profiles of plaque formation by my collection of 30 phages (See Table 4, far right columns). These phages have repressor proteins which have different sequences in the predicted DNA-binding region (alpha-2, alpha-3 region highlighted), suggesting that they may have different specificity (Figure 8).

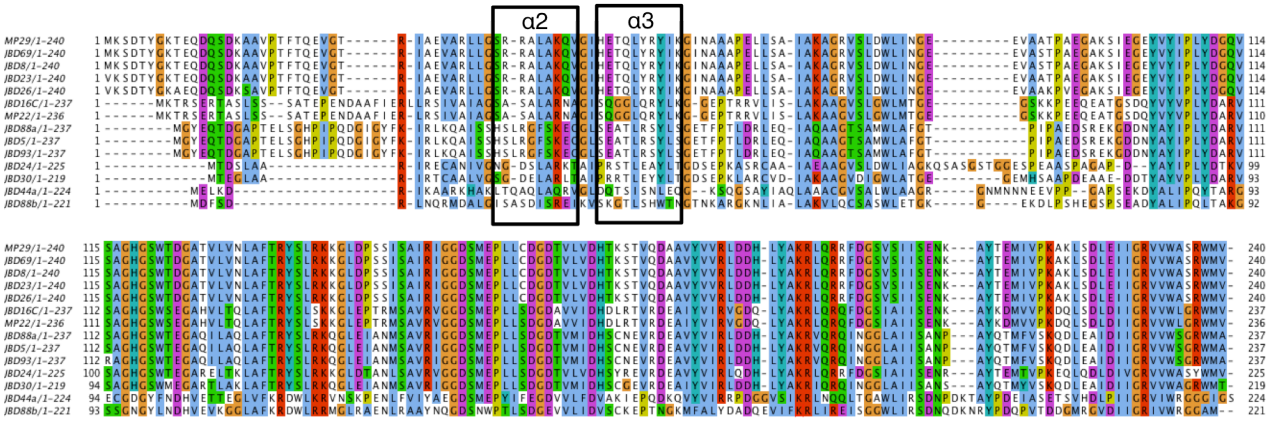


Figure 8: Alignment of the phage repressor proteins

The repressor proteins for all phages sequenced here are shown, with the alpha2-3 helix region highlighted. This region is the site which is responsible for differentiating between different phage DNA-binding sites. Representative non-redundant members of these repressors were cloned and expressed in PA14 (Table 4, right-most columns).

I found that the majority of phage resistance arises from mechanisms independent of repressor-mediated immunity. For example, while the highly resistant JBD16C lysogen could prevent

propagation of all pilus-specific phages, the JBD16C repressor protein confers immunity only against itself and MP22. Similarly, the JBD30 lysogen provided complete resistance to 10 and partial resistance to 2 of the 25 pilus-specific phages while its repressor protein was only active against two other phages (JBD60 and JBD59). The prophage with the weakest resistance profile, JBD24, possessed a repressor which was unique and only inhibited JBD24, as expected. Finally the repressors from the two O-antigen specific phages (JBD44 and JBD88b) did not inhibit each other, demonstrating that the resistance seen by these lysogens is non-repressor mediated. These data demonstrate that the prophages in this collection provide much of the phage resistance via repressor-independent mechanisms.

Phage propagation can be inhibited by a wide variety of mechanisms beyond repressor-mediated immunity. These mechanisms can be grouped into early (e.g. attachment, adsorption, DNA entry) or late stages (e.g. restriction, repression, CRISPR-Cas immunity) of the phage life cycle. To determine if the timing of the block to phage propagation is shared between highly resistant lysogens, I performed a plaque-independent assay to measure the ability of the phage genome to enter the cell. The assay uses an ion selective electrode to monitor the efflux of K^+ ions from bacterial cells, which is directly correlated with injection of the phage DNA into the host cell cytoplasm (Boulanger and Letellier, 1992). Wild type PA14 showed robust potassium efflux when challenged with a phages from my collection such as JBD26, JBD88a or JBD93 (Figure 9).

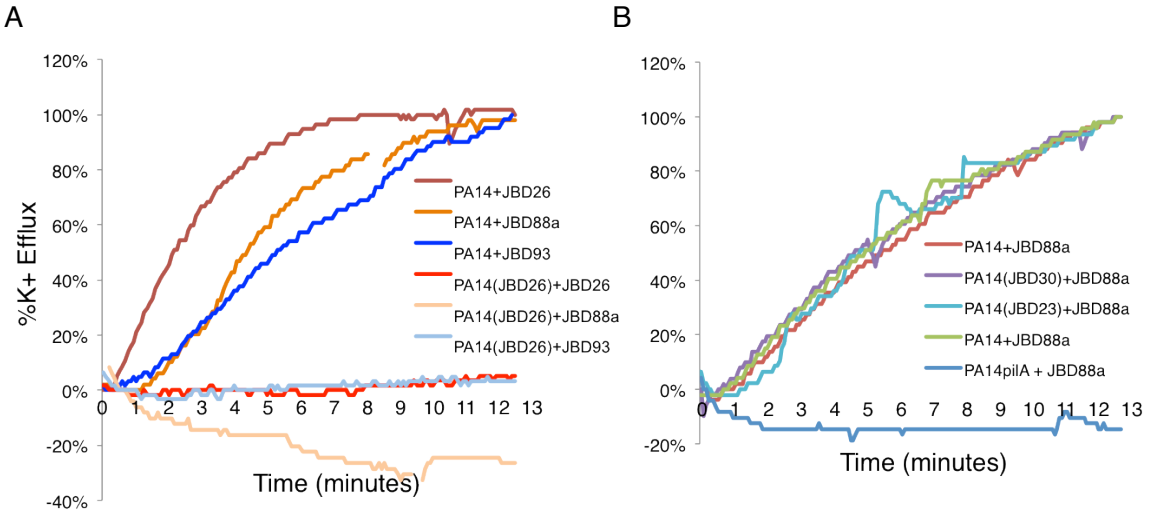


Figure 9: The JBD26 lysogen is resistant to phage entry, while other lysogens are not

K+ efflux assays are shown for (A) PA14 or a PA14(JBD26) lysogen infected with the indicated phages, (B) PA14, indicated PA14 lysogens, or a negative control PA14Δ*pilA* infected with phage JBD88a

The highly resistant PA14(JBD26) lysogen blocked potassium efflux when mixed with phages JBD26, JBD88a or JBD93 (Figure 9), showing that the phage resistance of this lysogen arises at the cell surface. By contrast, the highly resistant PA14(JBD23) lysogen, which is also insensitive to plaque formation by JBD88a, showed robust potassium efflux upon infection, indicating that the block to phage propagation occurs after DNA injection. Infection with JBD88a also caused efflux from a JBD30 lysogen, but as shown above, both JBD23 and JBD30 are in different repressor groups than JBD88a, implicating a non-repressor-mediated intracellular block to a productive infection in both cases. JBD88a infection also caused potassium efflux from cells in which it was present as a lysogen, which would be expected for repressor-mediated inhibition. These examples contrast with the resistance of the PA14(JBD26) to demonstrate both extra and intracellular mechanisms at play to block productive superinfection.

To determine if the lysogens (like JBD26) that block genome entry also prevent superinfecting phages from attaching to the cell surface, I examined phage adsorption. I performed adsorption assays with PA14(JBD26) and a number of other highly resistant lysogens. During infection of wtPA14, Approximately 80% of phages in a sample bound within eight minutes, while <5% bound to the *pilA* mutant. PA14(JBD26), a highly resistant lysogen that did not allow phage genome entry as monitored by K^+ efflux assays, was found to block adsorption of superinfecting phage JBD24 (Figure 10). Similarly, other lysogenic strains with prophages that were found to confer a strong resistance profile, including JBD63c and JBD69 also inhibited phage adsorption (Figure 10). A PA14(JBD30) lysogen which was not resistant to the phages used, showed adsorption comparable to wtPA14.

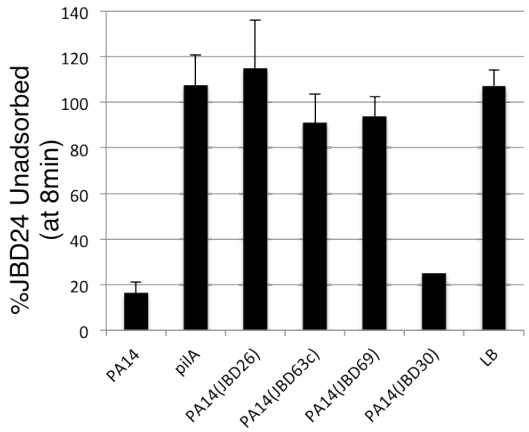


Figure 10: The most resistant lysogens prevent adsorption of a superinfecting phage

Adsorption of phage JBD24 to PA14, PA14 lysogens or PA14Δ*pilA* is shown, represented as the percentage of phage which are unadsorbed to cells after eight minutes.

Different infecting phages were used here from K^+ efflux experiments as each assay had its own technical limitations. All phages used, however, rely on the T4P for adsorption and subsequent DNA entry. Taken together, the superexclusion provided by JBD26 and other 'adosrption inhibitors' tested here provide the strongest resistance profile, preventing infection by all phages that require the T4P, while other prophages (i.e. JBD30, JBD23) provide intracellular inhibition of superinfection.

Another distinct mechanism of resistance emerged from the group of five phages utilizing the O-antigen to infect PA14. Each of these five phages, when present as prophages, excludes the others from superinfection (Table 4). The JBD44 repressor does not recapitulate this exclusion phenotype, demonstrating that a repressor-independent mechanism is at play. This lysogen-mediated resistance extends to some of the phages which require both the pilus and the O-antigen to infect PA14, but not any that utilize just the pilus. This demonstrates that only phages which require the O-antigen (i.e. are inhibited by a *wbpL* or *wbpM* gene mutation) are also blocked by a lysogen of JBD44, in a non-repressor-dependent manner. Interestingly, not all O-antigen requiring phages (i.e. JBD5, JBD33) were inhibited by these lysogens. These data suggest a subtle modification which affects certain phages which is not simply a complete removal/inactivation of the O-antigen. I tested these lysogens for serotype conversion by a slide agglutination assay, since a related *P. aeruginosa* phage (D3) has previously been shown to mediate this effect (Newton *et al.*, 2001). Notably, no serotype conversion was detected here and

all lysogens still reacted positively as the O10 serotype, confirming that the O-antigen has not been removed. For technical reasons, neither the adsorption nor efflux assays can be applied to these phages so although this hypothesis has not been directly tested, it seems likely that these prophages induce a modification of the bacterial envelope, imparting resistance to some O-antigen-requiring phages.

2.3.5 Prophages alter Type IV pilus function

As the phages in my collection that utilize the T4P for infection were found to be blocked at the cell surface, I examined pilus function in each of the 30 lysogens. *P. aeruginosa* strains use the T4P for twitching motility, whereby the bacteria move along a solid surface through a series of extensions and retractions of the pilus. Nine lysogens showed a 20-40% decrease in twitching motility diameter, and in several cases the morphology of the twitching zone was also altered (Figure 11).

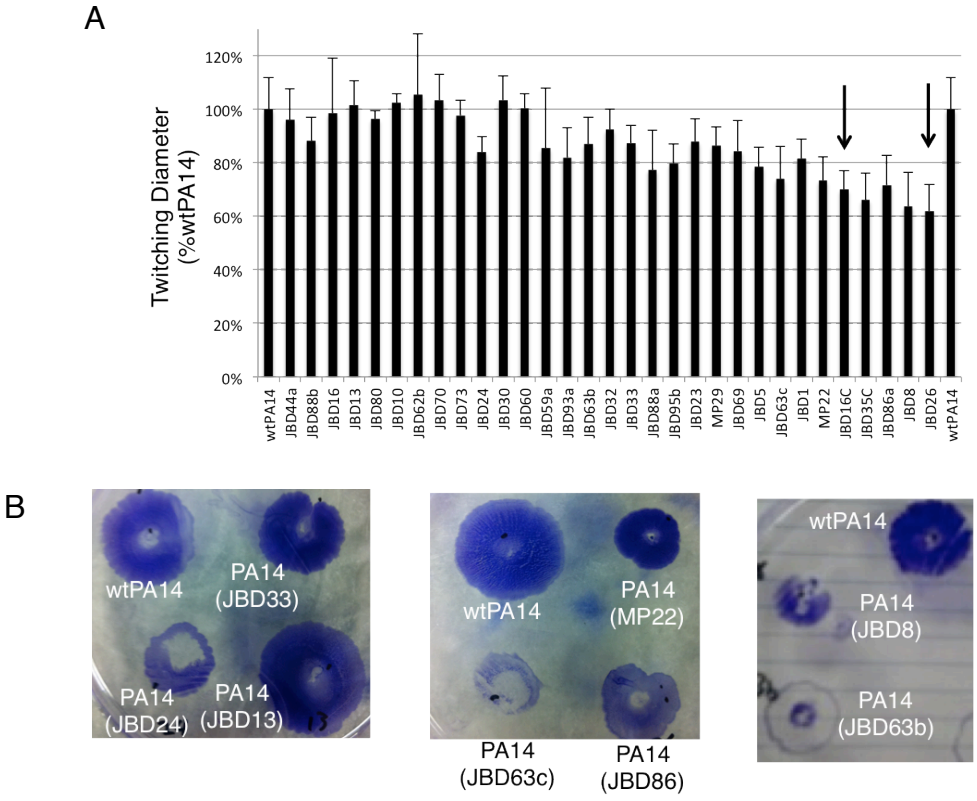


Figure 11: Twitching motility inhibited by prophages

(A) Twitching motility assays were conducted for wtPA14 and indicated PA14 lysogens. The diameter of the twitching zone, normalized to wtPA14 is shown with arrows indicating the discussed lysogens JBD16C and JBD26, (B) Examples of the crystal violet stained twitching zone produced by different lysogens.

These twitching defects correlated with the prophages that displayed the strongest phage resistance patterns. For example, the integration of a JBD26 prophage decreased twitching motility of PA14 by 40% compared to wild type cells and provided resistance to all phages that rely on the pilus for infection, including the previously characterized pilus-specific phages MP22 and D3112 (Heo *et al.*, 2007). Notably, after my work had been conducted this same group published similar observations, noting that phage D3112 inhibited twitching motility as a lysogen in PA14, although no link to phage resistance was made (Chung *et al.*, 2012). The group of phages that were still able to form plaques on the JBD26 lysogen required the O-antigen for infection, suggesting that the attenuation of twitching motility is linked to the observed phage resistance. Although the JBD26 lysogen phage resistance profile was identical to the *pilA* single gene deletion, the *pilA* strain exhibits no twitching, while the JBD26 lysogen exhibits intermediate twitching, an adaptation which may be favorable in nature. This suggests a modification to the pilus, not a complete absence of assembly as in the *pilA* mutant. The lysogens

in which twitching motility was unaffected include the group of lysogens of phages that require the O-antigen for infection, as well as lysogens of a group of T4P-specific phages that provided limited resistance to further phage infection (e.g. JBD70, JBD30, JBD10). In addition to twitching motility, I also examined the flagellum-mediated swimming motility of all 30 lysogens (Figure 12), and found it that it was unaltered for all lysogens except PA14(JBD16C), which showed decreased swimming motility (see insertion site section for comment).

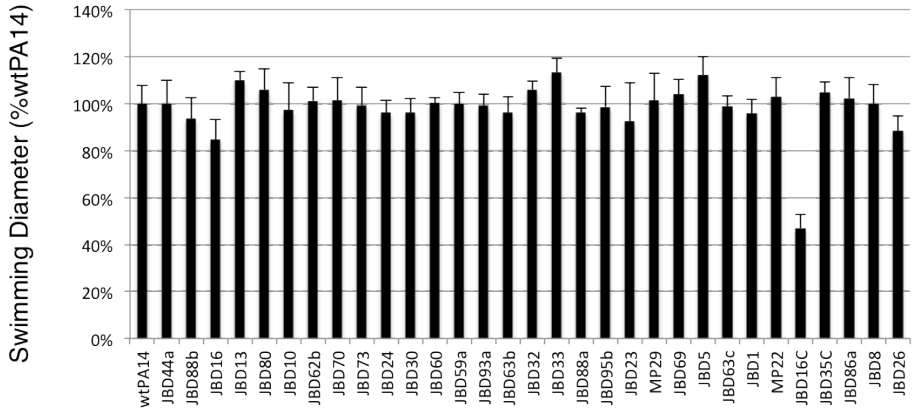


Figure 12: Swimming motility is not generally affected by lysogeny.

The diameter of the zone of flagellar-mediated swimming motility was measured for indicated lysogens and normalized to wtPA14. Note that the only lysogen which is significantly affected is PA14(JBD16C), which is discussed below.

All phages in this collection were able to infect a PA14Δ*fliC* flagellar mutant as well as wtPA14, demonstrating that this is not an essential receptor for these phages. Thus, it seems that the prophages discussed here have mechanisms to directly or indirectly modify their receptors, the T4P and O-antigen but not a non-receptor, the flagellum.

To assess the reproducibility of the twitching and phage resistance phenotypes, and to rule out phenotypic effects resulting from the integration site of the prophage, I examined 17 individual JBD26 lysogens. In each case, I observed identical phage resistance patterns and attenuated twitching motility (Figure 13).

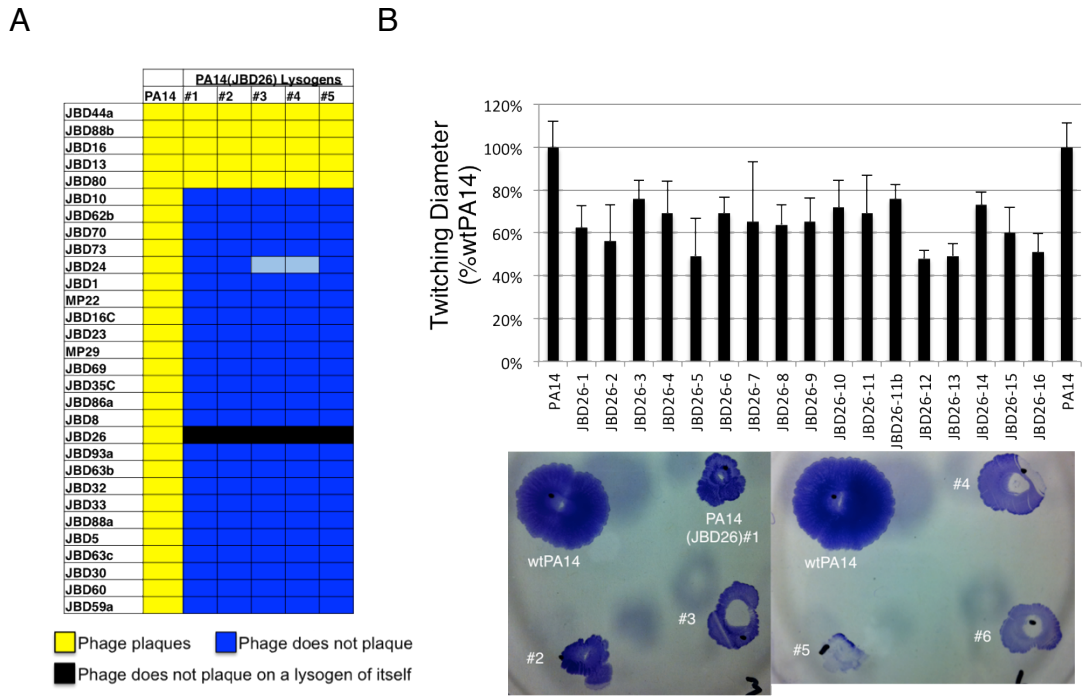


Figure 13: Independent PA14(JBD26) lysogens are phage resistant

(A) Multiple lysogens of JBD26 were constructed in PA14 and all were resistant to all phages utilizing the T4P. All lysogens also had twitching defects although this was variable, with measurements shown in the top panel of (B) with representative pictures of the twitching zones shown in the bottom panel.

Given the Southern blot shown above (Figure 7), demonstrating that JBD26 integrates randomly, it is unlikely that the phage resistance/twitching phenotypes are a result of the interruption of a gene during phage integration, but rather is due to expression of a gene or genes from the JBD26 prophage.

2.3.6 Lyosgeny has minimal impact on virulence in *Caenorhabditis elegans*

As increased pathogenesis resulting from gene expression from prophages has been demonstrated previously (Waldor and Mekalanos, 1996; Vaca-Pacheco *et al.*, 1999; Nesper *et al.*, 1999), I tested my collection of PA14 lysogens for their ability to kill the nematode *Caenorhabditis elegans,* an established model for mammalian bacterial pathogenesis (Tan *et al.*, 1999a; b). I performed slow killing assays, which involves an infection-like process that results from the accumulation of *P. aeruginosa* PA14 within the worm intestines. I examined the effects of 22 different lysogens in these slow killing assays. The killing kinetics of the majority of

lysogens was unchanged compared with wild type PA14, and no lysogens exhibited faster killing ability (Figure 14a,b).

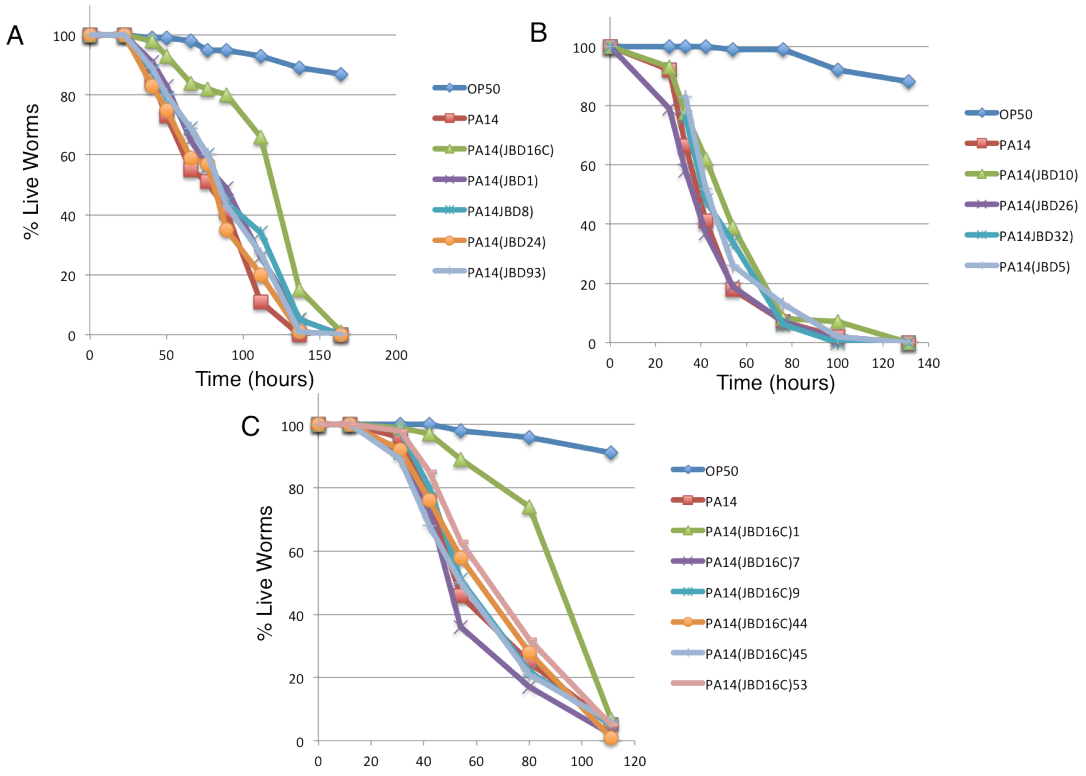


Figure 14: PA14(JBD16C) virulence is attenuated in *C. elegans* killing assay

(A), (B), The virulence of PA14 and indicated lysogens in a *C. elegans* infection model. Comparison to the *C. elegans* food source, *E. coli* strain OP50 is included, (C) independent JBD16C lysogens were tested and only isolate #1 is attenuated.

However, a single lysogen, PA14(JBD16C), was attenuated in its ability to kill *C. elegans*; at ~50 hours, ~50% of the worms infected with wild type PA14 were dead, compared to <10% of the worms fed the PA14(JBD16C) lysogen . At 90 hours, when 100% of the worms inoculated on PA14 were dead, approximately 20% of the worms on PA14(JBD16C) were still alive. In addition, worms seeded directly onto the PA14 lawn quickly vacate it for the surrounding agar and are sluggish for the duration of the experiment, while the worms on PA14(JBD16C) remained in the lawn and were highly motile. I tested five additional lysogens of PA14(JBD16C) in this assay, and discovered that they killed *C. elegans* with similar kinetics to wild type PA14 (Figure 14c), suggesting that the effects observed with the first lysogen were a result of the prophage integration site.

2.3.7 Prophage insertion hot spots effect bacterial phenotypes

Given that the PA14(JBD16C) lysogen was having such a dramatic effect on virulence, but other insertions of the same prophage were not, I mapped the insertion site of the attenuated PA14(JBD16C) strain assay using inverse PCR (Ochman *et al.*, 1988). This revealed that the prophage was inserted six nucleotides after the stop codon of gene PA14_64460 (*psiF*), interrupting a putative transcriptional terminator downstream of this gene (Figure 15).

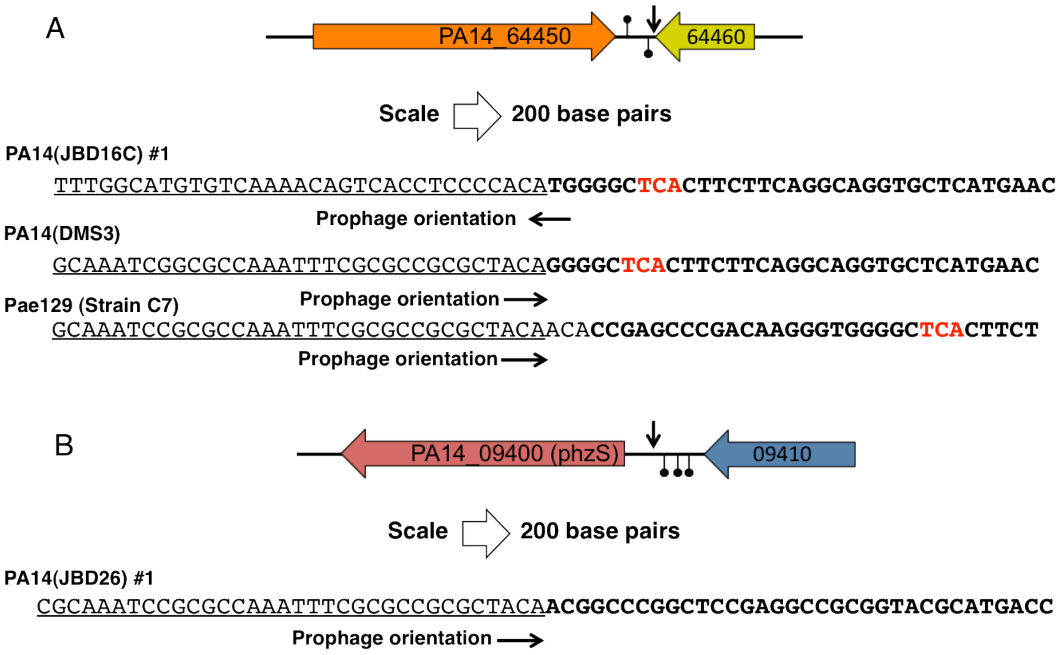


Figure 15: Insertion site hot spots for Mu-like phages

The arrow indicates the insertion site for the prophage and predicted terminators are shown. The genes and intergenic distance are to scale. Stop codons are shown in red (in reverse orientation). (A) PA14_64460 is a hot spot for insertion for phages related to JBD16C and (B) an insertion near *phzS*, a gene involved in pyocyanin production is shown for phage JBD26.

As this insertion spot could lead to destabilization of the transcript, I extracted RNA from this lysogen and used RT-qPCR to examine gene expression and discovered that PA14_64460 (*psiF*) was down-regulated six-fold in PA14(JBD16C) as compared to wild type (Figure 16).

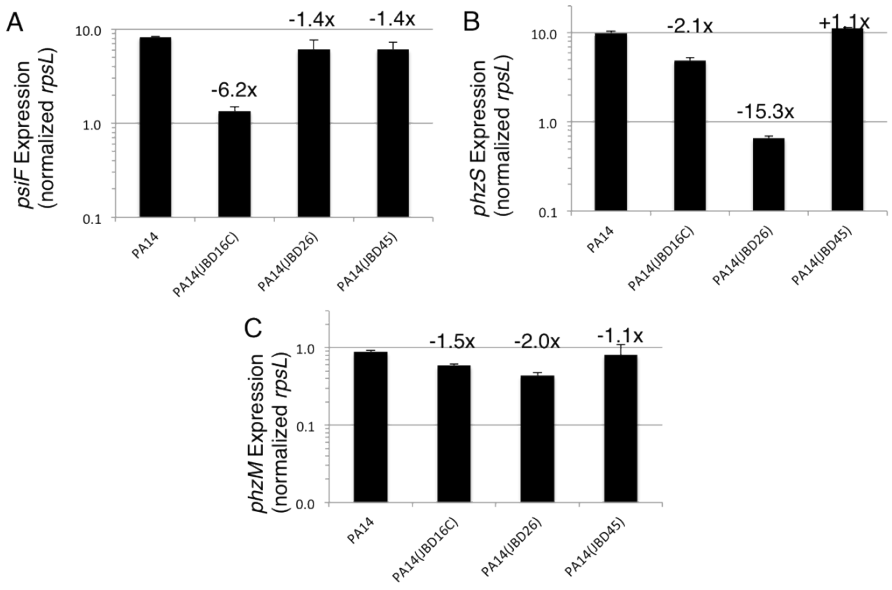


Figure 16: Prophage insertion negatively affects expression of neighbouring genes.

Reverse transcriptase quantitative PCR of three indicated PA14 lysogens, compared to wtPA14, with all results normalized to *rpsL*, encoding a ribosomal protein, for genes (A) *psiF*, (B) *phzS*, and (C) *phzM*

I examined expression of PA14_64460 in two other lysogens (PA14(JBD26) and PA14(JBD44/MP29), which do not have prophage insertions at this site and found that they maintained wild type levels of expression, illustrating that this effect is specific to this PA14(JBD16C) lysogen. Interestingly, this lysogen was the only one to display decreased swimming motility, suggesting that this insertion site may play a role in modulating swimming motility as well as *in vivo* pathogenesis. Analysis of the integration sites of other Mu-like *P. aeruginosa* phages in strain PA14 revealed a previously published example of phage DMS3 integrating at this site (Zegans *et al.*, 2009) and a prophage in a strain from our collection, Pae129, also integrated into the genome immediately following the stop codon for PA14_64460 (Figure 15). These results suggest that this genomic location is an integration hot spot, and while Mu-like phages integrate randomly into bacterial genomes, integration hot spots have been previously recognized (Manna *et al.*, 2001).

I also observed that a number of the lysogens created in PA14 with the Mu-like phages led to the production of a red pigment, while this strain typically produces a blue-green pigment, the virulence factor pyocyanin (Ran *et al.*, 2003). I found that when I constructed multiple lysogens of the Mu-like phages, approximately 1 in 10 were red in both liquid and solid medium. This red pigment has been observed in mutants of *phzS*, the final gene in the biosynthetic pathway for the

production of pyocyanin as this gene interruption results in accumulation of a red precursor molecule, 5-methyl-phenazine-1-carboxylate (5MPCA) (Gibson *et al.*, 2009). Interestingly, this precursor molecular 5MPCA is highly toxic to the pathogenic fungus, *Candida albicans* (Morales *et al.*, 2010). Quantification of the chloroform-soluble pyocyanin from one such lysogen, PA14(JBD26), confirmed that there was very little pyocyanin produced, as observed for the *phzS* mutant (Figure 17).

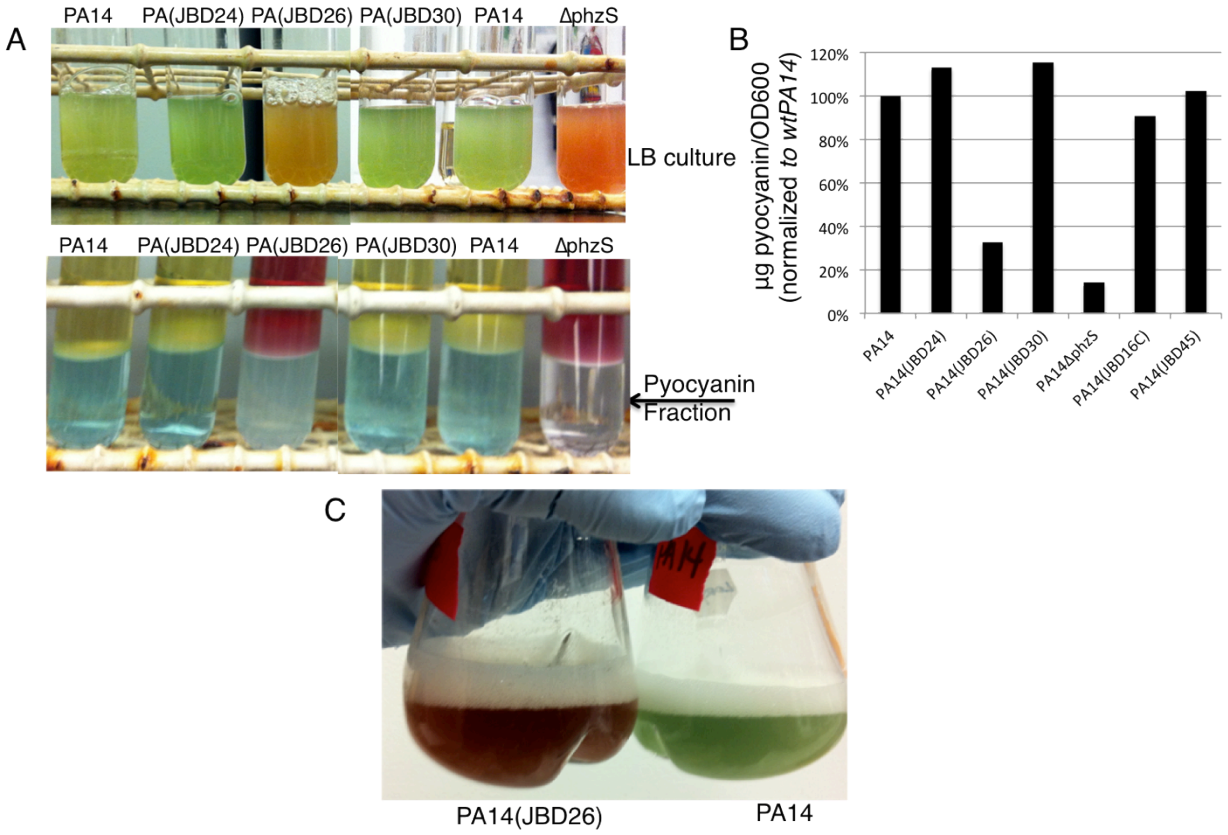


Figure 17: Pyocyanin production inhibited by prophage insertion

(A) Cultures grown in LB (top) were cleared by centrifugation and the pyocyanin in the supernatant was extracted (lower panel) and quantified by a colourometric assay with results shown in (B), (C) shows a direct comparison of overnight cultures of PA14 and PA14(JBD26).

I mapped the insertion site for this prophage and found that it was integrated upstream of the *phzS* gene (Figure 15). Analysis of the gene expression levels from this lysogen using RT-qPCR revealed that phzS was decreased 15-fold as compared with wild type PA14, and expression from the adjacent gene, *phzM*, was decreased 2-fold (Figure 16). Although I didn't map the insertion sites of multiple lysogens, the red pigment was consistently observed with many prophages. The basis for the potential prophage integration hotspot in *phzS* is unclear, but

frequent insertions in this gene could have a profound effect on virulence of a *Pseudomonas* strain.

2.4 Discussion

The presence of prophages within bacterial genomes has been shown to influence virulence, and the movement of phages among bacterial strains plays a significant role in the ongoing evolution of pathogens. In this study I systematically examined bacterial phenotype changes resulting from the acquisition of different prophages in a single *P. aeruginosa* strain background to address the frequency with which phage-mediated phenotypic changes occur. I found that each lysogenic strain I created displayed at least one altered phenotype, such as increased phage resistance, and often had other differences, including changes to twitching motility, pigment production, and attenuation of pathogenesis in *C. elegans*.

I discovered that 2/3 of the bacterial isolates in our diverse collection of *P. aeruginosa* produce at least one phage that could form plaques on at least one strain in the collection of indicator strains. When the prophages were present in a common strain background (PA14) I observed wide variation in the levels of spontaneous induction, with differences greater than 10,000-fold between low and high frequency inducers. The spontaneous induction of phage particles from individual bacterial cells likely has a diverse effect on the population of cells as a whole; while the individual bacterial cell is killed, the resulting release of phage particles can kill or lysogenize other bacteria present in the surrounding environment. For bacteria that are widely distributed in water and soil environments like *P. aeruginosa*, interactions between diverse strains may be a very common occurrence, and the ability of a bacterial strain to resist the activity of phages produced from rival cells would be strongly selected for. This provides a selective advantage for the maintenance and spread of prophages that encode genes that supply a fitness advantage for their bacterial host.

The phages examined in this study were induced from a very diverse collection of clinical and environmental isolates. However, the selection for phages that could propagate in a single strain (PA14) appeared to greatly narrow the range of phages under study. Analysis of the genomes of the twelve phages that could propagate in this host revealed the presence of two groups of phages; ten closely related phages that use the type IV pilus for infection, and a second group of two closely related phages that use the O-antigen for infection. Genomic analyses of five

additional phages that were unable to propagate in PA14 revealed a much more diverse group. These data clearly illustrate that the selection of a host for phage propagation can have a profound influence on the composition of the phage families that are isolated. While the majority of phages selected on PA14 were very closely related across the length of their genomes, there was significant localized genetic diversity observed. These variable genes inserted between highly conserved homologues in phage genomes have been termed "morons" as they add "more on" the phage genome when they are present (Juhala *et al.*, 2000). These genes are likely the cause of the lysogen-specific phenotypes I observed. The unique assortment of moron genes that each phage possesses could lead to variable and potentially overlapping phenotypes in the bacterial lysogen.

In summary, at least five distinct mechanisms of prophage-mediated phage resistance operating in a single strain have been described here; (i) repressor-mediated, (ii) non-repressor mediated, intracellular, weak resistance (i.e. JBD30), (iii) non-repressor mediated, intracellular, strong resistance (i.e. JBD23), (iv) cell surface inhibition of adsorption/entry (i.e. JBD26), (v) non-repressor mediated inhibition of O-antigen requiring phages (i.e. JBD44). These have all been described just in PA14 and perhaps expanding these studies to many strains could reveal novel mechanisms. In addition to resisting phage infection, these modifications to the host could have direct consequences on pathogenesis as well. Previous work has demonstrated prophage-expressed genes that influence both phage superinfection and virulence in *Vibrio cholerae* and *P. aeruginosa* (Waldor and Mekalanos, 1996; Vaca-Pacheco *et al.*, 1999; Nesper *et al.*, 1999). Classical repressor-mediated immunity accounted for less than half of the phage resistance observed. The remaining phage resistance profiles encompass a variety of different mechanisms and act as proxies for changes in the cell physiology mediated by the prophages. Several highly resistant lysogens were shown to block the entry of invading phages at the cell surface as monitored by potassium efflux assays and displayed decreased twitching motility, suggesting that a modification to the T4P machinery might be the cause of the observed resistance. While lack of motility is rare in environmental isolates, *P. aeruginosa* isolated from the lungs of chronically colonized Cystic Fibrosis patients are frequently non-motile, presumably because the pilus is a target for the human immune system (Mahenthiralingam *et al.*, 1994) and phages have been detected in the CF lung both as prophages (Winstanley *et al.*, 2009) and free phage particles (Ojeniyi, 1988). The ability of the phages that use the O-antigen as a receptor to prevent

superinfection by other O-antigen utilizing phages while not inhibiting the phages that use the just the T4P suggests modification to the O-antigen. This has been previously demonstrated with related phage D3 (Newton *et al.*, 2001), via a serotype change, although no serotype change was detected here.

The emergence of new epidemic bacterial strains can reflect the presence of multiple prophages within the strain, each of which may contribute a unique set of fitness factors that, in combination, provide the strain with novel survival/pathogenic properties (Winstanley *et al.*, 2009). In addition, the potential recombination of invading phages with prophages resident in a bacterial genome provides a powerful mechanism for the generation of new phages that could provide novel combinations of bacterial fitness factors.

Chapter 3
The CRISPR-Cas adaptive immune system of *Pseudomonas aeruginosa* mediates resistance to phage infection

3 Overview

The CRISPR-Cas bacterial immune system inhibits foreign genetic elements using an sRNA-guided surveillance complex to detect and cleave foreign DNA. The *in vivo* activity of these systems had only been studied in select model organisms such as *Streptococcus thermophilus* (Type II CRISPR-Cas) (Barrangou *et al.*, 2007) and *Staphylococcus aureus* (Type III CRISPR-Cas) (Marraffini and Sontheimer, 2008) at the time this work was conducted. Much of the work by others on CRISPR-Cas systems has focused on the structural and mechanistic details with a paucity of data regarding *in vivo* function. Studies in the Gram-negative model organism *Escherichia coli,* had revealed that the Type I-E CRISPR-Cas system was repressed under laboratory conditions and appeared to be inactive (Westra *et al.*, 2010; Touchon *et al.*, 2011). In this work, I demonstrate that the Type I-F CRISPR-Cas system of *Pseudomonas aeruginosa* is naturally active and targets phages and plasmids. Six temperate phages isolated from different *P. aeruginosa* strains and constructed plasmids were targeted by the CRISPR-Cas system of strain PA14, limiting phage propagation and transformation efficiency, respectively. I also demonstrated the first examples of the CRISPR-Cas system possessing a gradient of activity, depending on the number of mismatches with the target. This represented the first evidence of a naturally functioning CRISPR-Cas system in a Gram-negative organism, providing an *in vivo* tool with which to dissect the finer details of CRISPR-Cas function.

3.1 Acknowledgements

The publication associated with this chapter was done in collaboration with Dr. Kyle Cady, a former graduate student in Dr. George O'Toole's laboratory. Dr. Cady demonstrated that other aspects of this CRISPR-Cas system were functional, such as the ability to acquire new spacers and that phages escape the CRISPR-Cas system by acquiring point mutations in the protospacer. These data will not be presented here. Together, we demonstrated that protospacer mismatches in

specific locations are tolerated by the CRISPR-Cas system, but lead to a gradient of CRISPR-Cas activity that is, in part, affected by temperature. These data are presented below.

Parts of this chapter were adapted from:

Cady K.C.*, **Bondy-Denomy J**.* *et al.* (2012) The CRISPR/Cas adaptive immune system of *Pseudomonas aeruginosa* mediates resistance to naturally occurring and engineered phages. *J. Bacteriology* **194**(21):5728-38

*Authors contributed equally

3.2 Materials and Methods

3.2.1 Strains and media

The *P. aeruginosa* strain UCBPP-PA14 (abbreviated *P. aeruginosa* PA14) was used in this study. *P. aeruginosa* and *E. coli* strains were routinely cultured in lysogeny broth (LB) at 37 °C. Growth media were supplemented with antibiotics at the following concentrations: ampicillin, 150 μg/ml (*E. coli*), gentamicin, 10 μg/ml (*E. coli*) and 50 μg/ml (*P. aeruginosa*); carbenicillin 50 μg/ml (*E. coli*) and 250 μg/ml (*P. aeruginosa*).

3.2.2 Plaque assay

150 μL of WT *P. aeruginosa* PA14 or mutant was added to 3 mL molten top agar (0.7%) and poured over a LB agar plate supplemented with 10 mM $MgSO_4$. After solidification of top agar lawns, 3 μL of serially diluted, phage lysates were spotted onto the top agar lawn and incubated at 30 ^{o}C or 37 ^{o}C overnight. Plaques were counted and expressed as plaque forming units (PFU)/mL.

3.2.3 Strain construction

Most strains and constructs were created by Kyle Cady using a *Saccharomyces cerevisiae* recombineering technique described previously (Shanks *et al.*, 2006; Cady and O'Toole, 2011). During attempts to mutate the DMS3-42 protospacer, he could only create 3 of the five mutations in a CRISPR-Cas intact background (see Figure 23/24 and associated text for details). The other two mutations (C253G and T255C) were constructed in CRISPR/Cas-deficient backgrounds ΔCRISPR2 (ΔC2) and Δ*csy3*.

3.2.4 Plasmid transformation efficiency assay

A shuttle vector that replicates in both *E. coli* and *P. aeruginosa* called pHERD30T was used to clone predicted protospacers for the purposes of plasmid transformation assays. 42 nt oligonucleotides were synthesized, corresponding to a protospacer of interest (32 nt), along with 5 nt up and downstream. Extra bases were added onto the synthesized oligonucleotides, corresponding to digested NcoI and HindIII restriction sites such that sticky ends were created when the oligonucleotides were annealed. After annealing, these protospacers were ligated to pHERD30T digested with NcoI and HindIII, transformed into DH5α and positive clones confirmed by sequencing. Plasmids containing protospacers matching the sequence of spacer #1 from CRISPR2 (abbreviated CR2_sp1) from WT DMS3 (5 mismatches), T255C (4 mismatches), JBD18 (0 mismatches) as well as protospacers matching CR1_sp1 (from JBD25) or CR1_sp6 (from JBD18) were isolated from *E. coli* using a BioBasic Miniprep Kit. *P. aeruginosa* CRISPR/Cas intact (WT) or CRISPR/Cas deficient (ΔCR/*cas*) cells were grown in 5 mL of LB at 37 ^{o}C in a shaker for 18 hours. Following growth, 1.0 mL of the culture was washed twice in 1.0 mL 300 mM sucrose, re-suspended in 100 µL 300 mM sucrose, and mixed with 300-400 ng of the indicated plasmid DNA. Each sample was electroporated, and recovered in 1.0 mL of LB medium for 1 hour at 37 ^{o}C. Following incubation, each sample was diluted 10 and 100-fold and 100 µl was plated onto a pre-warmed LB plate containing 30 µg/mL gentamicin. Each plate was incubated at 37 ^{o}C for 18 hours and the resulting colonies on each plate were counted and used to determine the transformation efficiency as a function of colony forming units per nanogram of plasmid added to each plate. Transformation efficiency of plasmids containing protospacers were compared to empty vector for each individual recipient strain and shown as a percentage. Three separate transformations were performed for each construct tested.

3.2.5 Nucleotide sequence accession numbers

The full genome sequences of JBD18, JBD25, and JBD67 are available in the NCBI database under accession numbers JX495041, JX495042, and JX495043, respectively.

3.3 Results

3.3.1 Phage replication is inhibited by the *P. aeruginosa* CRISPR-Cas system.

Utilizing the phage collection generated in Chapter 2, I observed that approximately 60 of the isolated temperate phages were able to form plaques on *Pseudomonas aeruginosa* strain PA14, while 30 phage isolates could only form plaques on other *P. aeruginosa* isolates in the collection. To assess the possible role of CRISPR-Cas-mediated immunity in preventing these phages from forming plaques on *P. aeruginosa* PA14, I determined the ability of this collection of phages to form plaques on a *P. aeruginosa* PA14 strain lacking a functional CRISPR-Cas system (ΔCR/*cas*) (Cady *et al.*, 2011). Six of these 30 phages (designated JBD18, JBD25, JBD37, JBD54, JBD55, JBD67) were able to form plaques on the *P. aeruginosa* PA14 ΔCR/*cas* strain, but not on wild type (WT) *P. aeruginosa* PA14 (Figure 18). Upon generation of high titer lysates of these 6 phages by growth on the ΔCR/*cas* mutant, I observed 10^5-10^9-fold decrease in efficiency of plaquing (EOP) on WT PA14 as compared to the ΔCR/*cas* mutant (Figure 18). These data strongly imply that the CRISPR-Cas system in *P. aeruginosa* PA14 strain drastically inhibits the growth of these phages.

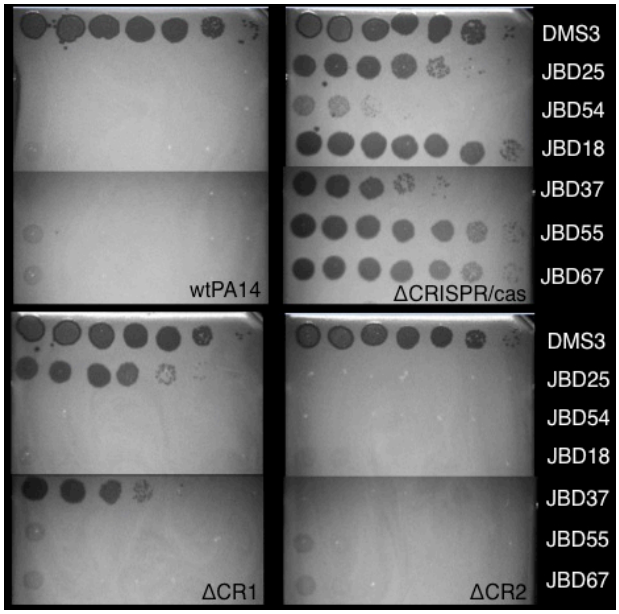


Figure 18: Isolation of phages which are targeted by the CRISPR-Cas system of *P. aeruginosa* strain PA14

CRISPR-insensitive phage DMS3, and CRSPR-sensitive phages JBD25, JBD54, JBD18, JBD37, JBD55, and JBD67 are spotted in ten-fold dilutions on lawns of WT *P. aeruginosa* PA14, and strains lacking the crRNA encoding regions CRISPR1 (ΔCR1), CRISPR2 (ΔCR2), or both CRISPR loci and *cas* genes (ΔCRISPR/cas).

If resistance to the JBD phages were mediated by a CRISPR-Cas system, we would predict that these phages, while unable to replicate, should still be capable of injecting their DNA into the host. To test this idea, K^+ efflux assays were conducted to quantify the phage genome entry kinetics of one of these phages, JBD18. This efflux assay directly measures the increase in extracellular K^+ ions that occurs as a result of cytoplasmic ion leakage as the phage genome passes through the inner bacterial membrane (Boulanger and Letellier, 1992). JBD18 infection resulted in the same kinetics and magnitude of K^+ ion efflux from the WT and ΔCR/*cas* strains (Figure 19), demonstrating that the JBD18 genome is able to enter WT cells at a normal rate. In contrast to the WT strain, infection of a *ΔpilA* mutant, which is resistant to JBD18 because it lacks the Type IV pilus required for cell surface adsorption of this phage, resulted in little K^+ efflux. This result confirms that the observed K^+ efflux requires attachment of phage to cells, and that mutating the CRISPR-Cas system does not alter injection kinetics.

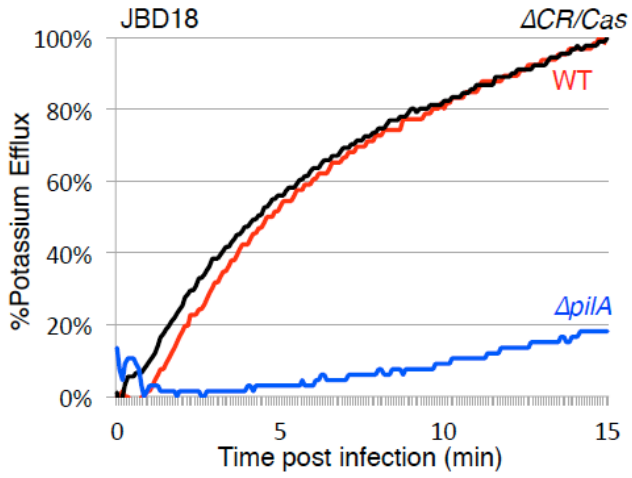


Figure 19: Phage JBD18 causes K^+ efflux with similar kinetics for WT PA14 and ΔCRISPR/cas mutant.

A K^+ efflux assay was performed using CRISPR-sensitive phage JBD18 to infect WT PA14, ΔCR/*cas* and negative control, PA14Δ*pilA P. aeruginosa* strains. The phage was added at t=0 minutes and measurements collected every five seconds for 15 minutes.

To confirm the requirement of *cas* genes for the apparent CRISPR-Cas system mediated inhibition of the phages under investigation here, strains containing single gene deletions of each of the *P. aeruginosa* PA14 *cas* genes were assayed (note: these mutations were constructed by Dr. Kyle Cady, (Cady and O'Toole, 2011)). As shown in Table 7 (first 3 rows), deletion of any of the *cas* genes, with the exception of *cas1*, resulted in sensitivity to infection by JBD18, JBD25 and JBD67. Representative images of the plaque assays are shown in Figure 20 for mutations of *cas1, cas3,* and *csy2* genes. The absence of an effect for the Δ*cas1* strain was expected as the Cas1 protein is believed to play a role in the acquisition of new spacers, but not in CRISPR-mediated interference (Barrangou *et al.*, 2007; Zegans *et al.*, 2009; Wiedenheft *et al.*, 2009). Deletions of *cas3*, *csy1*, *csy2*, *csy3*, and *csy4* in *P. aeruginosa* PA14 have been shown to reduce or eliminate crRNA accumulation while Δ*cas1* strains display normal levels of processed crRNA (Cady and O'Toole, 2011).

Table 7: Summary of plaque assays with *P. aeruginosa* PA14 CRISPR/*cas* mutants.

	Genotype									
Phage[a]	**WT**	**ΔCR/*cas***	**ΔCR1**	**ΔCR2**	**Δ*cas1***	**Δ*cas3***	**Δ*csy1***	**Δ*csy2***	**Δ*csy3***	**Δ*csy4***
JBD18	--	+	--	--	--	+	+	+	+	+
JBD25	--	+	+	--	--	+	+	+	+	+
JBD67	--	+	--	--	--	+	+	+	+	+
DMS3	+	+	+	+	+	+	+	+	+	+
C253G[b]	-	+	-	+	-	+	+	+	+	+
T255C[b]	-	+	-	+	-	+	+	+	+	+
DMS3*m*[c]	--	+	--	+	--	+	+	+	+	+

[a]Plaque assays conducted with the indicated phages used to infect lawns of WT or various mutant strains. '+' Indicates the formation of plaques; '-' represents reduced plaquing efficiency (>10^4 - fold reduction), '--' represents a greater reduction in plaquing efficiency (>10^6 – fold reduction) compared to the ΔCR/cas mutant. [b]Gene *42* of phage DMS3 was mutated at two nucleotide positions, C253G and T255C for reasons discussed in section 3.3.4 and Figure 24. [c]DMS3*m* is a mutant phage generated to possess a protospacer with perfect complementarity to CR2_sp1 of PA14. Wild type DMS3 has five mismatches at this location in gene *42* (shown in Figure 23).

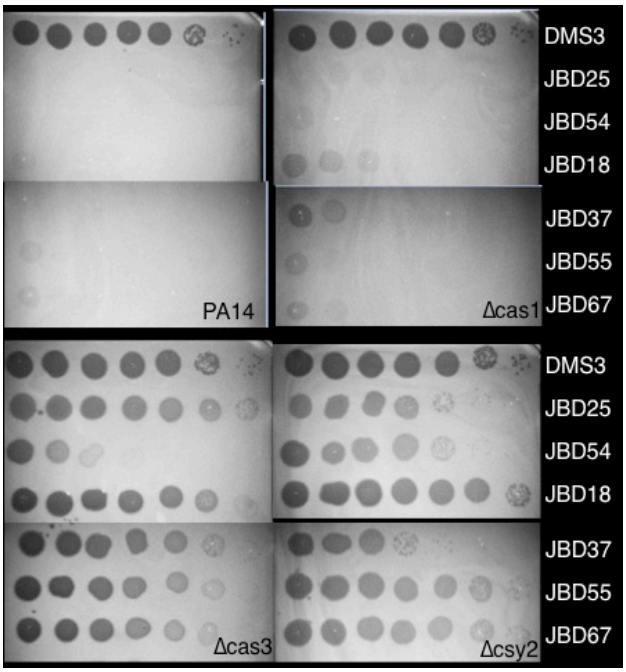


Figure 20: CRISPR targeting is alleviated by *cas* single gene knockouts

CRISPR-insensitive phage DMS3 and CRISPR-sensitive phages JBD25, JBD54, JBD18, JBD37, JBD55, and JBD67 were spotted in ten-fold dilutions on lawns of WT PA14 or single gene knockouts of *cas1, cas3,* and *csy2.* Interference was defective in mutants of the gene encoding the nuclease/helicase protein, *cas3* and CRISPR-Cas complex structural component, *csy2,* but not *cas1*, which is involved in spacer acquisition.

3.3.2 Sequencing the genomes of JBD18, JBD25 and JBD67 reveals protospacer sequences.

To further elucidate the mechanism by which the phages identified here were targeted by the CRISPR-Cas system, I purified the phage DNA and sequenced the genomes of three CRISPR-sensitive phages (JBD18, JBD25 and JBD67). Electron micrographs of the purified phages are shown in Figure 21a. *P. aeruginosa* PA14 possesses two different CRISPR loci designated as CRISPR1 and CRISPR2 with the *cas* genes located between the CRISPRs (Zegans *et al.*, 2009). Phage JBD18 contains regions with 100% matches to spacer 6 in the CRISPR1 locus (CR1_sp6), and spacers 1 and 2 in the CRISPR2 locus (CR2_sp1, and CR2_sp2, Figure 21b,c).

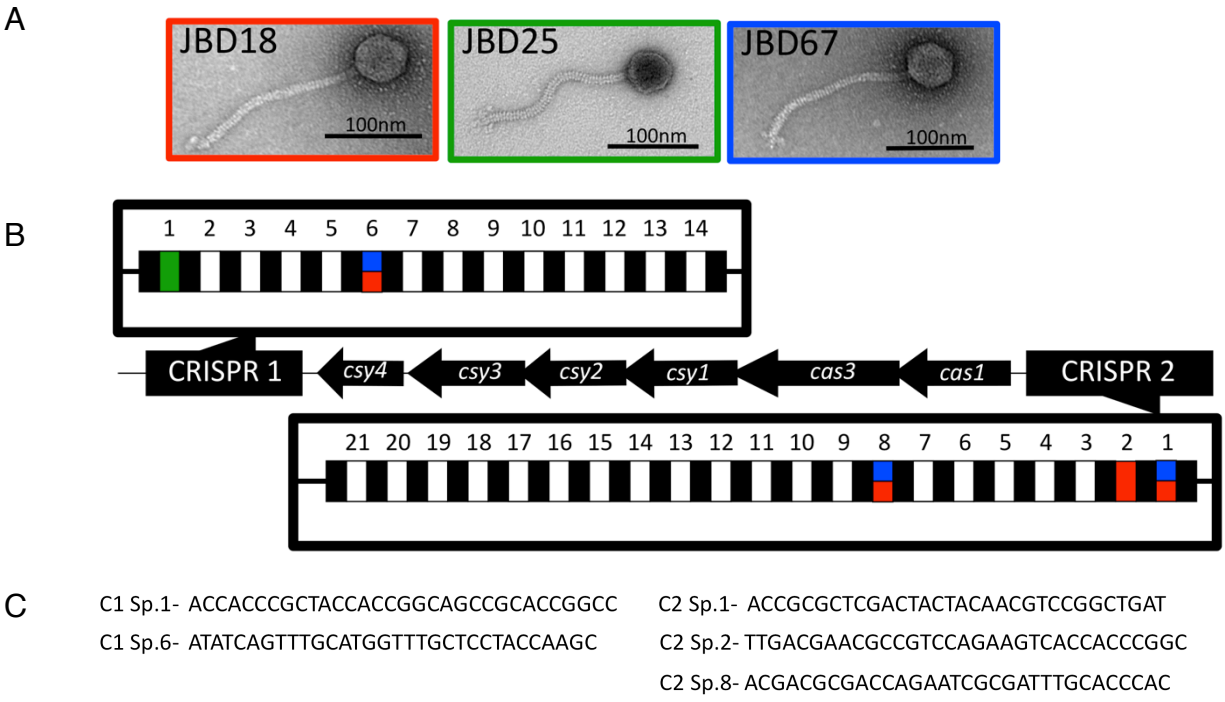


Figure 21: The genomes of CRISPR-sensitive phages JBD18, JBD25, and JBD67 contain protospacer targets.

(A) Representative negative-stained images of phage JBD18, JBD25, and JBD67 obtained using transmission electron microscopy. All three phages display long, noncontractile tails (B) Diagram of the CRISPR and *cas* genes found in *P. aeruginosa* PA14. CRISPR spacer content that is 100% identical over all 32 nucleotides to a region of phage JBD18 are indicated with red boxes, while those identical to regions of JBD25 and JBD67 are depicted in green and blue, respectively. CRISPR1 and 2 are encoded on opposing DNA strands and are numbered in order. (C) The sequences of the five CRISPR spacers which bear 100% identity with phages JBD18, JBD25 and JBD67.

Importantly, these putative protospacers also display the protospacer adjacent motif (PAM), which is GG in the Type I-F CRISPR-Cas system found in *P. aeruginosa* PA14 (Mojica *et al.*,

2009). This motif is required for the recognition of a phage protospacer by the CRISPR-Cas system but is not matched by a complementary region on the CRISPR RNA. JBD67 displays the same putative protospacers as JBD18 with the exception of a single mismatch with CR2_sp2. JBD25 possesses only one predicted spacer match to CR1_sp1, and the PAM is also present. Plaque assays showed that deletion of either CRISPR1 (ΔCRISPR1 strain) or CRISPR2 (ΔCRISPR2 strain) alone did not alleviate the inhibition of JBD18 and JBD67, whereas JBD25 was not inhibited in the ΔCRISPR1 strain (Figure 18). These results are consistent with the protospacer matches found in the genomes of these phages as outlined above.

3.3.3 Transformation is inhibited by the CRISPR-Cas system

To demonstrate that the protospacers found in JBD18, JBD25, and JBD67 were authentic targets of the PA14 CRISPR-Cas system, the putative protospacer sequences and PAMs from each phage genome were synthesized and cloned into a high copy number shuttle vector capable of replication in *E. coli* and *P. aeruginosa*. The protospacer-containing plasmids or the vector with no insert were electroporated into WT *P. aeruginosa* PA14 or ΔCR/*cas* strain, and transformation efficiencies were calculated. The plasmid containing the CR2_sp1 protospacer transformed WT *P. aeruginosa* PA14 with less than 0.1% efficiency compared to the empty vector, but no difference in transformation efficiency was observed in the ΔCR/*cas* strain (Figure 22).

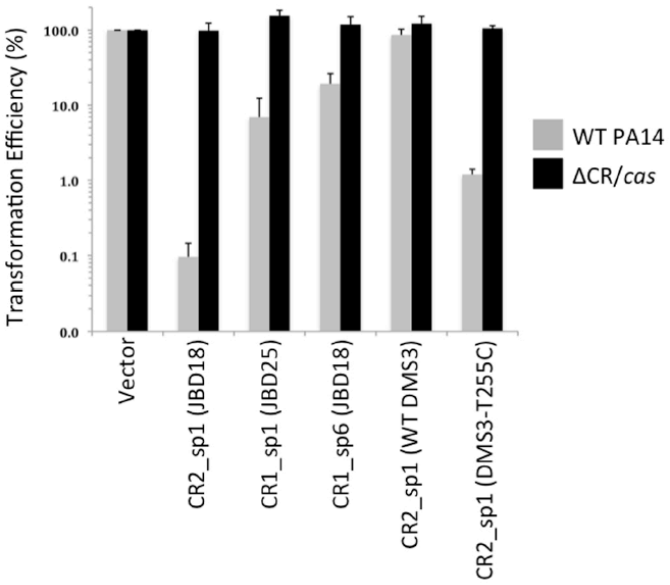


Figure 22: The CRISPR-Cas system of PA14 reduces transformation efficiency

Transformation efficiency of plasmids harboring no protospacer (vector) or the protospacer indicated on the X-axis of the graph, in the presence or absence of the CRISPR-Cas system. The transformation efficiency of each spacer-bearing plasmid was quantified relative to the transformation efficiency of the empty vector introduced into the same strain.

Similarly, transformation of WT *P. aeruginosa* PA14 by plasmids containing either the CR1_sp1 or CR1_sp6 protospacers was reduced by ≥80% compared to empty vector (Figure 22). By contrast, no reduction of transformation efficiency was observed for these plasmids when they were introduced into the ΔCR/*cas* strain. Although, the magnitude of the reduction in transformation efficiency caused by action of the CRISPR-Cas system varied depending on the particular protospacer being tested, these results clearly demonstrate that the *P. aeruginosa* PA14 CRISPR-Cas system is active against protospacers present in phages that are inhibited by the system. Thus, the presence of these protospacers is likely the source of the vulnerability of these phages to the CRISPR-Cas system.

In summary, the results described above demonstrate for the first time that the Type I-F CRISPR-Cas system of *P. aeruginosa* functions in a manner similar to other CRISPR-Cas systems to inhibit the replication of phages bearing matches to spacers contained within the CRISPR loci.

3.3.4 Engineering of phage DMS3 to induce targeting by the CRISPR-Cas system.

Previously, the O'Toole Lab demonstrated that the crRNA encoded by CR2_sp1 recognizes a protospacer with five mismatches in phage DMS3 gene *42* (*DMS3-42*), leading to the inhibition of biofilm formation in DMS3 lysogens (Zegans *et al.*, 2009; Cady and O'Toole, 2011). However, they observed no resistance mediated by this spacer towards phage DMS3 (Zegans *et al.*, 2009; Cady *et al.*, 2011). Since the transformation efficiency of a plasmid bearing the CR2_sp1 protospacer from JBD18, which has no mismatches was strongly inhibited by the *P. aeruginosa* PA14 CRISPR-Cas system (Figure 22), we hypothesized that the lack of inhibition of DMS3 by the CRISPR-Cas system might be due to the lack of perfect complementarity the *DMS3-42* protospacer (Figure 23).

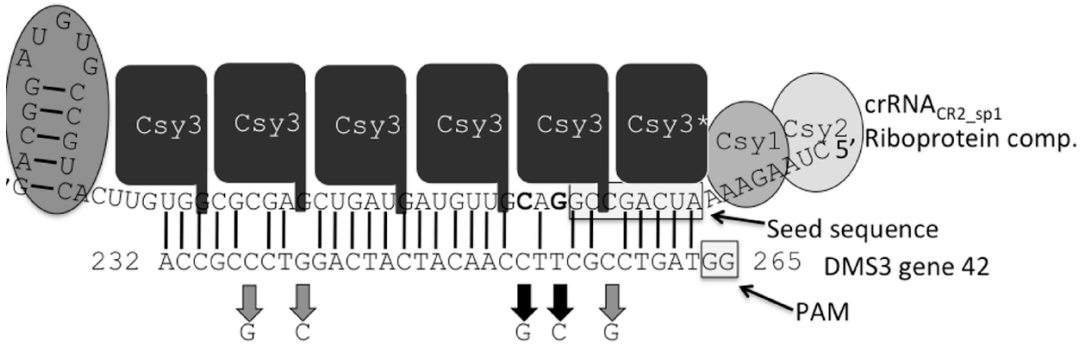


Figure 23: Engineering of DMS3 variants which are targeted by the CRISPR-Cas system

(Figure and mutations made by Dr. Kyle Cady)
Model of Csy-crRNA$_{CR2_sp1}$ complex interacting with *DMS3-42* target sequence. The model is based on previous work performed by Doudna group (Haurwitz *et al.*, 2010; Wiedenheft *et al.*, 2011b). Csy proteins (various shades of grey) are shown coating crRNA$_{CR2_sp1}$, while lines denote base pairing between crRNA$_{CR2_sp1}$ and its phage target sequence in gene 42 of phage DMS3 (*DMS3-42*). The crRNA seed sequence and phage PAM, which is critical for crRNA-target interaction, are shown within shadowed boxes. Thick black arrows show the location of mutant alleles C253G and T255C.

To test the effect of creating complementarity between crRNA $_{CR2\_sp1}$ and the DMS3 protospacer, single base pair mutations in DMS3 were created by Dr. Cady (denoted by thick arrows in Figure 23). Strikingly, the single nucleotide changes in the DMS3 genome, C253G or T255C, led to strong inhibition of DMS3 replication by the CRISPR-Cas system even though 4 mismatches with crRNA $_{CR2\_sp1}$ remained (Figure 24, Table 7). Conversely, other changes in the protospacer (C237G, G240C, and C258G) had no effect on the ability of DMS3 to replicate in WT *P. aeruginosa* PA14.

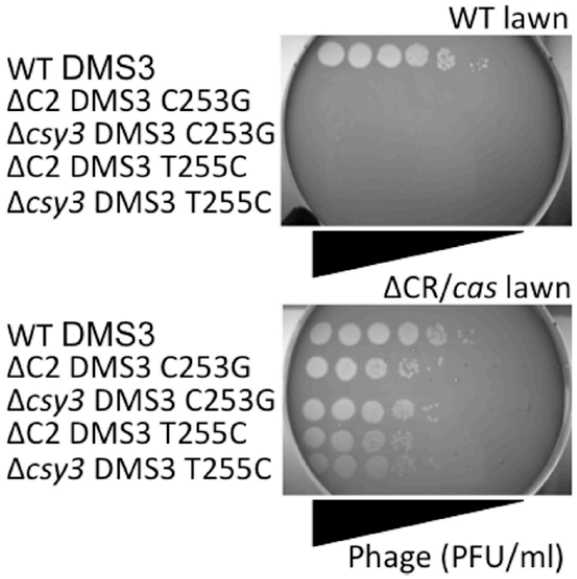


Figure 24: Mutations C253G and T255C reduce the number of spacer: protospacer mismatches, facilitating CRISPR-Cas targeting.

(Figure and mutations made by Dr. Kyle Cady)
Wild type phage DMS3 or phages harboring mutations C253G or T255C in gene *42* were spotted on PA14 or ΔCR/cas. WT DMS3 can replicate in the presence or absence of the CRISPR/Cas system while the mutants are unable to replicate on wtPA14. Phages with either mutation were constructed in ΔCRISPR2 (ΔC2) or Δ*csy3* backgrounds, as indicated.

These data indicate that for crRNA $_{CR2\_sp1}$ to mediate resistance to phage DMS3, complementarity is required between the crRNA and the protospacer at nucleotide positions 253 or 255, but not at positions 237, 240, and 258. To confirm that mutagenesis of this putative tail protein did not affect the phage, I conducted K^+ efflux assays to again demonstrate that this resistance was manifested post-genome injection (Figure 25), as would be expected for resistance mediated by the CRISPR-Cas system.

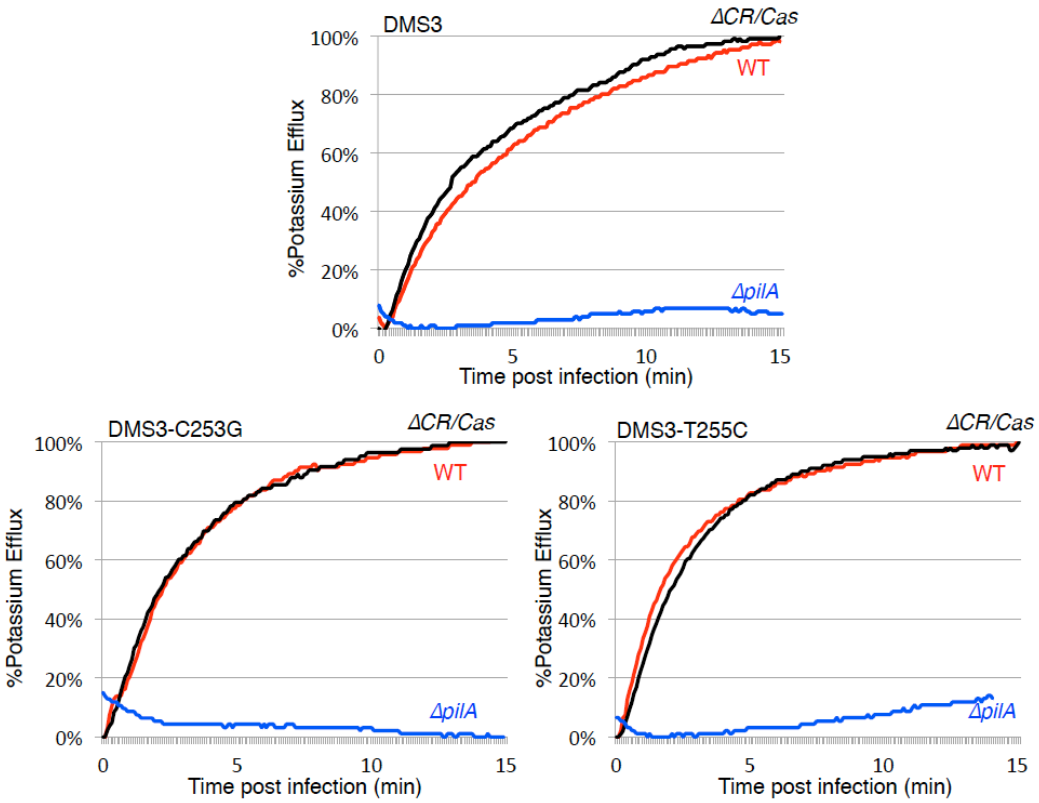


Figure 25: Phage DMS3 mutants cause comparable K^+ efflux from both WT PA14 and ΔCR/Cas strains.

The ability of phages harboring C253G and T255C alleles to inject their genomic DNA into WT PA14, ΔCR/*cas*, or Δ*pilA* (lacks phage receptor) mutants was determined by the K^+ efflux assay.

To further investigate the effects of the point mutations described above, the transformation efficiencies of plasmids harboring the WT DMS3 CR2_sp1 protospacer (5 mismatches), and the T255C mutant protospacer (4 mismatches) were compared. As shown in Figure 22 (right-most two sets of bars), the presence of the T255C mutation in the DMS3 protospacer sequence caused an 80-fold decrease in the transformation efficiency of *P. aeruginosa* PA14 as compared to the plasmid bearing the WT DMS3 protospacer sequence. This difference was not observed when the same plasmids were used to transform the PA14 ΔCR/*cas* mutant strain. These data show that the same point mutation that led to a dramatic CRISPR-Cas-mediated inhibition of phage DMS3 replication (Figure 24) also imparted a reduction in transformation efficiency similar in magnitude to those described for the JBD phage-derived sequences.

3.3.5 Mismatches between crRNA and protospacer results in intermediate resistance.

In analyzing the effects of point mutations in the DMS3 CR2_sp1 protospacer, Dr. Cady created a phage mutant, which I will refer to as DMS3*m*, bearing a protospacer that is 100% complementary to the spacer portion of crRNA $_{CR2\_sp1}$. Interestingly, while the plaquing efficiency of the DMS3 mutant bearing the T255C mutation was inhibited by ~10^4-fold when plated on *P. aeruginosa* PA14 the DMS3*m* phage was inhibited by greater than 10^6-fold (Figure 26, Table 7). This discrepancy was even more apparent when I conducted plaque assays at 30 °C where the plaquing efficiency of DMS3-T255C was reduced by only ~100-fold compared to WT DMS3 while the plaquing efficiency of DMS3*m* was still reduced by ~10^6-fold (Figure 26). The DMS3-C253G mutant also displayed at least a 1000-fold higher plaquing efficiency on *P. aeruginosa* PA14 at 30 °C as compared to DMS3*m* (Figure 26c, and inset).

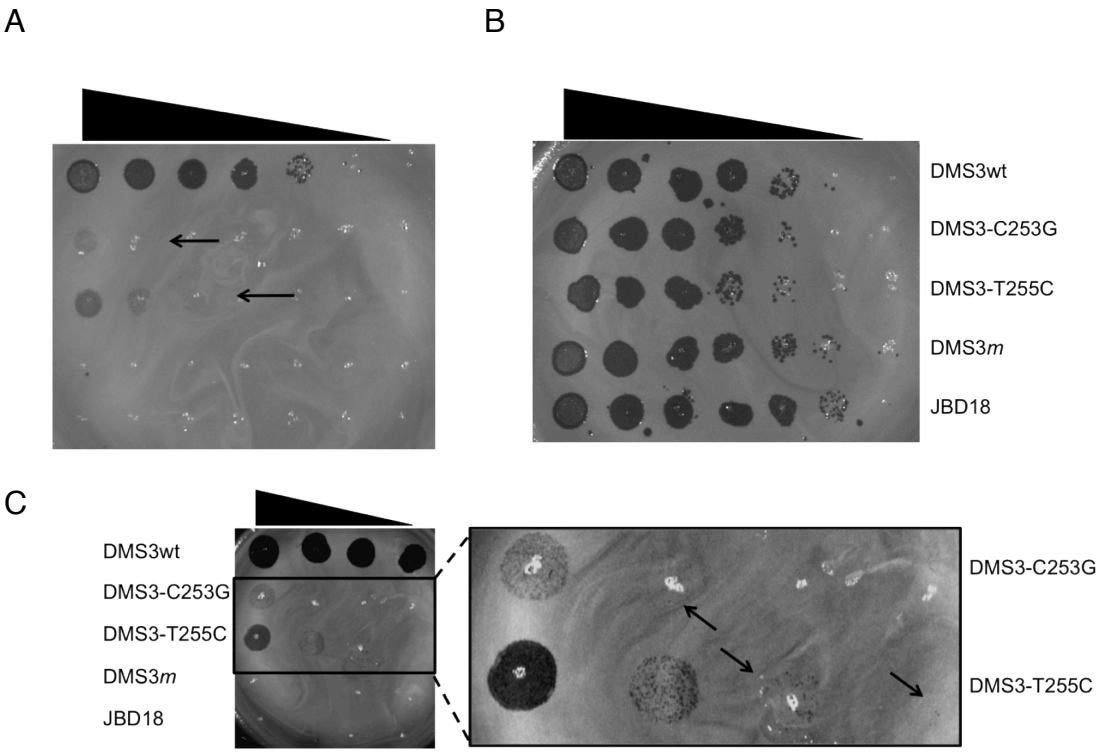


Figure 26: Mismatches lead to partial CRISPR-Cas mediated inhibition of phage replication.

Plaque assays showing the replication of WT DMS3, DMS3-C253G, DMS3-T255C, DMS3*m* and JBD18 in the presence (panel A, C) or absence (panel B) of the CRISPR/Cas system at 30°C. Arrows denote the location of single plaques. Phages were spotted in 10-fold serial dilutions. Panels (A) and (C) show the same experiment, with (C) being zoomed in for single plaque resolution.

These data imply that although this CRISPR-Cas system still functions in the face of crRNA mismatches, the efficiency of the system may be reduced. This finding was mirrored by transformation efficiency assays with *P. aeruginosa* PA14 where the transformation efficiency of a plasmid bearing a protospacer with 100% identity to CR2_sp1 was 10-fold lower than one bearing the T255C mutant protospacer, which still has four mismatches with the crRNA (see Figure 22). It should be noted that I do not have an explanation for the weaker CRISPR-Cas system effect in plaquing assays performed at 30 °C. The results in plasmid transformation efficiency assays were not affected by temperature, suggesting that the CRISPR-Cas system functions with similar efficiency at the two temperatures. I would speculate that phage replication may be more efficient at a lower temperature and, thus, evasion of the CRISPR-Cas system occurs more readily.

3.4 Discussion

Here I demonstrate, in collaboration with Dr. Kyle Cady and Dr. George O'Toole, that the Type I-F CRISPR-Cas system, like that of Type I-E (*E. coli*) (Brouns *et al.*, 2008) and Type II-B (*S. thermophilus*) (Barrangou *et al.*, 2007), can provide sequence specific resistance to phage challenge. Using a diverse temperate phage library isolated from environmental and clinical strains of *P. aeruginosa*, I isolated six phages that are targeted by the endogenous *P. aeruginosa* CRISPR-Cas system. In addition, Dr. Cady used single nucleotide point mutations to engineer a phage, which was not targeted by the CRISPR-Cas system, to become targeted. This work provided the first evidence for an endogenously functioning CRISPR-Cas adaptive immune system in a Gram-negative organism, and opened up a new *in vivo* system for future study. This was an important advance because many key structural and *in vitro* studies on Cas proteins have been performed on the *P. aeruginosa* PA14 Type I-F system (Wiedenheft *et al.*, 2009; Haurwitz *et al.*, 2010; Wiedenheft *et al.*, 2011b).

Using the transformation efficiency assay, I demonstrated that protospacers with 100% identity to the CR1_sp1, CR1_sp6, and CR2_sp1 spacers are targeted by the CRISPR-Cas system. Plasmids containing these sequences display reduced transformation efficiencies. Surprisingly, the degree of inhibition of transformation varied widely depending on the protospacer tested. For example, the CR1_sp6 protospacer from JBD18 and JBD67 when placed on a plasmid resulted in only a 5-fold inhibition of transformation while the CR2_sp1 protospacer from JBD18 and

JBD67 (or DMS3*m*) caused a 1000-fold inhibition. The replication of phages JBD18 and JBD67, which contain the CR1_sp6 protospacer, however, was still inhibited 10^7-fold when plated on a strain with the CRISPR2 locus deleted. In this background the CR1_sp6 protospacer is the only sequence in these phages that displays any significant similarity to a CRISPR spacer; thus, its presence in the JBD18 and JBD67 genomes is almost certainly the cause of the poor plating efficiency of these phages despite the weak effect of this protospacer in the transformation efficiency assay. The varying effects on transformation efficiency of different protospacers are difficult to explain, but these differences were observed consistently and likely reflect subtleties in the functioning of this system that are yet to be elucidated.

Through engineering of phage DMS3, we found that the protospacer within the *DMS3-42* gene could be efficiently targeted by the CRISPR-Cas system even when it possessed 4 mismatches with $crRNA_{CR2\_sp1}$. However, the DMS3*m* phage, which matches $crRNA_{CR2\_sp1}$ with 100% identity, was clearly targeted more efficiently than phage containing mismatches as was observed most noticeably in the plaquing assays performed at 30 °C (Figure 26). This decreased targeting of the mismatched protospacer was also reflected in the transformation efficiency assay where the protospacer with no mismatches caused a 10-fold greater decrease in transformation efficiency as compared to a protospacer with mismatches (Figure 22). These data show that the CRISPR-Cas system is not "all-or-nothing", and that the degree of complementarity between the crRNA and protospacer sequence can affect the efficiency with which the system operates.

Consistent with a gradient of effectiveness for the *P. aeruginosa* CRISPR-Cas system, the WT *DMS3-42* protospacer, which has 5 mismatches with $crRNA_{CR2\_sp1}$, does not cause resistance to phage DMS3. However, the O'Toole lab has previously shown that this protospacer does interact with the CRISPR-Cas system to inhibit biofilm formation in *P. aeruginosa* PA14 when it contains a DMS3 prophage. These data suggest that the effect on biofilm formation (in the absence of detectable resistance) may reflect a weak or altered interaction between the CRISPR-Cas system and phage DMS3.

Previously, the O'Toole lab identified, sequenced, and assayed the function of CRISPR-Cas systems found in a diverse array of clinical *P. aeruginosa* strains (Cady *et al.*, 2011). In that study, they were unable to detect CRISPR-Cas-mediated resistance to phages DMS3, MP22, F116, and D3 even though the strains on which they were tested (*P. aeruginosa* PA14 and 6

other clinical isolates) were shown to express fully processed crRNA and harbor spacers 100% identical to the tested phages (Cady *et al.*, 2011). In the work described here, by screening a large collection of temperate phages isolated from diverse *P. aeruginosa* strains, I identified a group of phages that are inhibited by the CRISPR-Cas system of *P. aeruginosa* PA14. At the point of this publication, it was unclear why some phages are inhibited by the CRISPR-Cas system while others possessing protospacers and intact PAM sequences, and thus would be predicted to be targeted, are not inhibited. More insight will be provided on this phenomenon in the following chapter.

In conclusion, this work presents an important advance for the investigation of CRISPR-Cas *in vivo* function. The results presented here provide the first evidence for phage and plasmid resistance by a Type I-F CRISPR-Cas system, as well as a collection of CRISPR-sensitive phages for future analysis. Furthermore, my demonstration of variable effects of different protospacers in the transformation efficiency assay and the potential modulation of sensitivity to the CRISPR-Cas system through reduction of complementarity between protospacer and crRNA highlight a potential for subtlety in this system that has not been previously recognized.

Chapter 4
Bacteriophage genes that inactivate the CRISPR-Cas bacterial immune system

4 Overview

A widespread system used by bacteria for protection against potentially dangerous foreign DNA molecules consists of the clustered regularly interspaced short palindromic repeats (CRISPR) coupled with *cas* (CRISPR-associated) genes (Makarova *et al.*, 2011b). Similar to RNA interference (RNAi) in eukaryotes (Makarova *et al.*, 2006), these CRISPR-Cas systems utilize small RNAs for sequence-specific detection and neutralization of invading genomes (Brouns *et al.*, 2008). Here, I describe the first examples of genes that mediate the inhibition of a CRISPR-Cas system. Five distinct "anti-CRISPR" genes were found in the genomes of phages infecting *Pseudomonas aeruginosa*. Mutation of a phage anti-CRISPR gene rendered it unable to infect bacteria with a functional CRISPR-Cas system, and the addition of the same gene to the genome of a CRISPR-Cas-targeted phage allowed it to evade the CRISPR-Cas system. Phage-encoded anti-CRISPR genes may represent a widespread mechanism for phages to overcome the highly prevalent CRISPR-Cas systems. The existence of anti-CRISPR genes presents new avenues for the elucidation of CRISPR-Cas functional mechanisms and provides new insight into the co-evolution of phages and bacteria.

4.1 Acknowledgements:

The construct to knock out anti-CRISPR JBD30-35 was generated by Diane Bona, a technician in Dr. Karen Maxwell's lab. April Pawluk, a graduate student in the lab conducted the work to test this anti-CRISPR in other *P. aeruginosa* strains and performed the beta-galactosidase assay of the *csy3*::*lacZ* fusion. April also tested the Type I-F anti-CRISPRs against the I-E CRISPR-Cas system of *E. coli* as well as assigning Type I-E anti-CRISPR function to genes I had cloned which were not Type I-F anti-CRISPRs. I conducted all other experiments outlined in this chapter.

Parts of this chapter were adapted from:

Bondy-Denomy J, Pawluk A, Maxwell KL, Davidson AR. (2013) Bacteriophage genes that inactivate the CRISPR-Cas bacterial immune system. *Nature* **493**:429-32, doi:10.1038/nature11723

Published version can be found at:
http://www.nature.com/nature/journal/v493/n7432/full/nature11723.html

Pawluk A, **Bondy-Denomy J**, Cheung VHW, Maxwell KL, Davidson AR. (2014) A new group of phage anti-CRISPR genes inhibits the Type I-E CRISPR-Cas system of *Pseudomonas aeruginosa*. *mBio* **5**(2):e00896-14

4.2 Materials and Methods

4.2.1 Strains and Growth Conditions

Pseudomonas aeruginosa UCBPP-PA14 (PA14), other *P. aeruginosa* isolates and *E. coli* DH5α were grown on lysogeny broth (LB) agar or liquid medium at 37 °C. LB was supplemented with gentamicin (30 μg/mL for *P. aeruginosa* and 15 μg/mL for *E. coli*) to maintain the pHERD30T plasmid.

4.2.2 Phage Genome Analysis

New phage genome sequences were first analyzed using BLASTn to assess general similarity to previously sequenced phages. To predict open reading frames and align multiple phages for comparison, the RAST program (Aziz *et al.*, 2008) was used. Comparison and analysis of specific phage proteins was done using RAST, CLUSTAL (Larkin *et al.*, 2007) or BLASTp/psi-BLAST (Altschul *et al.*, 1990).

4.2.3 Plasmid Construction

A shuttle vector that replicates in *E. coli* and *P. aeruginosa,* pHERD30T (Qiu *et al.*, 2008), was used for cloning and expression of genes in *P. aeruginosa*. This vector has an arabinose inducible promoter and a selectable gentamicin resistance cassette. Inserts were amplified by PCR. Vector and insert were digested with the appropriate restriction enzymes, ligated, and the ligation mix was used to transform *E. coli* DH5α. All plasmid constructs were verified by sequencing using primers that annealed to sites outside of the multiple cloning site.

To produce versions of JBD30 gene *35* with divergent DNA sequences (JBD30-35varA and varB), sequences containing maximal numbers of silent mutations were designed manually and

synthesized by GenScript USA (Piscataway, NJ). The synthesized genes were subcloned into pHERD30T by digestion with NcoI and HindIII.

4.2.4 DMS3 Recombination

Wild-type phage DMS3 contains a protospacer region in gene *42* with 5 mismatches to the 32 nt CR2_sp1 crRNA produced by PA14. A constructed mutant DMS3 phage (DMS3*m*) described as $DMS3_{100\%}$ in (Cady *et al.*, 2012), contains 5 point mutations in gene *42* creating a 100% match to this crRNA. Due to targeting by the CRISPR-Cas system, this mutant phage is unable to form plaques on wild-type PA14, with the exception of rare ($< 10^{-6}$) escaper mutants. Cells containing a pHERD30T construct containing JBD30 genes *34* to *38* (i.e. the anti-CRISPR gene from JBD30 with large flanking regions) were infected with DMS3*m* and recombinant phages were selected by plating on wild-type PA14. Plaques were picked and purified three times by re-plating on wild-type PA14. The protospacer containing region (gene *42*), as well as the expected site of anti-CRISPR gene recombination (between DMS3 gene *29* and *31*) were amplified by PCR and sequenced. Phages which had maintained 100% complementarity with the CR2_sp1 crRNA had acquired the anti-CRISPR gene and were used in the experiments described here (denoted DMS3*m*+JBD30-35) (Figure 44/45). Escapers were also identified with mutations in the protospacer and no recombination event identified.

4.2.5 Northern Blot

Northern blots were conducted essentially as in (Cady and O'Toole, 2011), with exceptions described below. WT PA14, PA14Δ*csy4*, PA14Δc*r1/cr2* (both CRISPR loci deleted, cas genes intact) and PA14 lysogens were grown in LB. Total RNA was extracted from log phase cultures (OD_{600nm}=0.8) using the *mir*Vana microRNA isolation kit (Ambion) and 5μg were run on a 10% TBE-Urea polyacrylamide gel and stained with SYBR Gold Nucleic Acid Gel Stain (Invitrogen) before transferring RNA to a nylon membrane at 200 mA for 1 h. A radiolabelled probe corresponding to the last 4 spacers and 3 repeats of the PA14 CRISPR locus was generated as in (Cady and O'Toole, 2011).

Prehybridization (blocking) was conducted using 50% formamide, 5x Denhardts, 0.5% SDS, 6x SSC and 100μg/mL ssDNA at 42 °C for 2h. Probing was conducted at 42 °C for 16h using fresh prehybridization buffer, but with the ssDNA omitted and radiolabelled probe added. Wash

solution 1 consisted of 2x SSC and 1% SDS while wash solution 2 consisted of 0.2x SSC and 0.1% SDS. Wash solution 1 was used for two 10 min washes at 25 °C, two 30 min at 65 °C, and wash solution 2 for one 10min wash at 25 °C. Blots were developed using a phosphor screen and imager. A low range ssRNA ladder (NEB) was also used to confirm the location of 5S RNA and crRNA.

4.2.6 Beta-Galactosidase Assays

Overnight cultures were subcultured 1:100 into LB containing 0.1% arabinose to induce anti-CRISPR gene expression from the pHERD30T plasmid, and then grown at 37°C to an OD_{600nm} of 0.3-0.6. Cultures were diluted 1:1 in complete Z buffer, in triplicate. Two drops each of 0.1% SDS and chloroform were added to each sample and after vortexing, 200uL ONPG was added. Samples were vortexed to begin the reaction, and then incubated at 30°C without shaking for 20-30 minutes. Absorbance measurements were taken at 420 and 550nm, and beta-galactosidase activity was calculated using the Miller equation. Data are expressed relative to cells containing the pHERD30T empty vector.

4.3 Results

4.3.1 Prophages inhibit CRISPR-Cas function

Since phage genes have been discovered that can neutralize most of the prevalent bacterial anti-phage defences (Labrie *et al.*, 2010), the failure to identify genes that counteract the widely occurring CRISPR-Cas systems was surprising. To search for such "anti-CRISPR" activity, I investigated the Type I-F CRISPR-Cas system of *Pseudomonas aeruginosa* utilizing a collection of 30 lysogens of *P. aeruginosa* PA14, which each contained a different phage genome (see Chapter 2). In lysogens, phage genomes are integrated into the bacterial genome and are referred to as prophages. Although prophage genes are generally repressed, all prophages have some genes that are actively transcribed. To test whether prophages might express anti-CRISPR activity, I measured the plaquing efficiency of three "CRISPR-sensitive" phages (JBD18, JBD25, and JBD67, see Chapter 3) on my collection of lysogens. The CRISPR-sensitive phages fail to replicate on wild type (WT) PA14 due to the action of the CRISPR-Cas system, but are able to replicate on PA14 ΔCR-*cas*, which contains no CRISPR loci or *cas* genes (Figure 27). I identified three lysogenic strains, PA14(JBD30), PA14(JBD24), and PA14(MP29), on which the

CRISPR-sensitive phages could form plaques robustly as compared to unlysogenized PA14 (Figure 27).

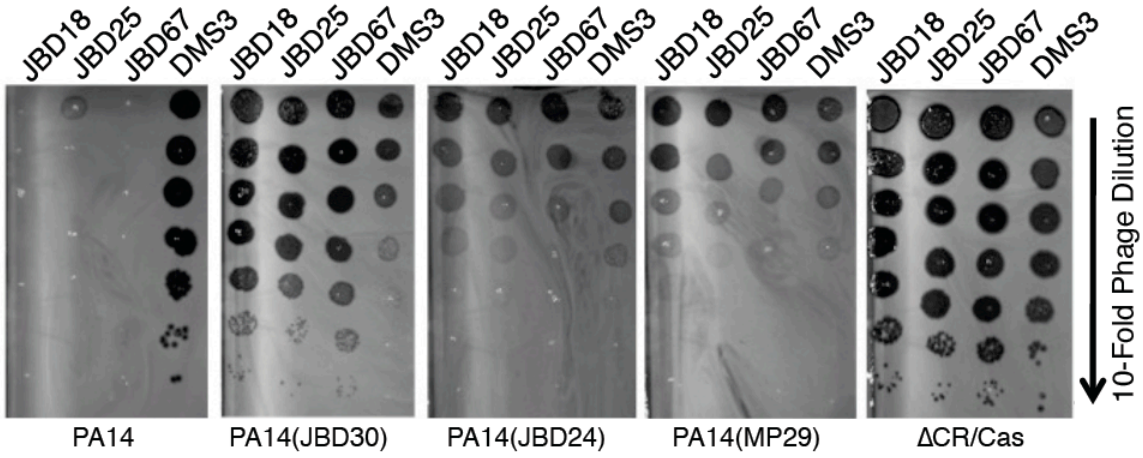


Figure 27: Prophages inhibit the PA14 CRISPR-Cas system

Ten-fold serial dilutions of CRISPR-sensitive phages (from L to R) JBD18, JBD25, JBD67 and CRISPR-insensitive phage DMS3 were spotted on lawns of PA14, ΔCR-*cas* or indicated lysogens made in WT PA14

Notably, the plaquing efficiency of the CRISPR-sensitive phages on PA14(JBD30) was equivalent to that on the ΔCR-*cas* strain, indicating that the JBD30 prophage caused complete inactivation of the CRISPR-Cas system. The somewhat lower plaquing efficiency of the CRISPR-sensitive phages on the other lysogens relative to their plaquing on the ΔCR-*cas* strain may be due to their production of less potent anti-CRISPR activity. However, these prophages also attenuate plaquing through mechanisms independent of the CRISPR-Cas system as is demonstrated by the partial inhibition of plaquing of the control phage DMS3, which is not affected by the CRISPR-Cas system (Figure 27).

To directly assess the anti-CRISPR activity of the PA14 lysogens, I used a plasmid-based transformation efficiency assay. The sequences within phages that are targeted by the CRISPR-Cas system are called protospacers. In order to be targeted, a protospacer sequence must be complementary to a specific spacer sequence within the CRISPR locus and also possess a correct Protospacer Adjacent Motif (PAM) (Mojica *et al.*, 2009) (Figure 28). Protospacer sequences are named according to the spacer sequence that they match in the PA14 genome (Figure 28).

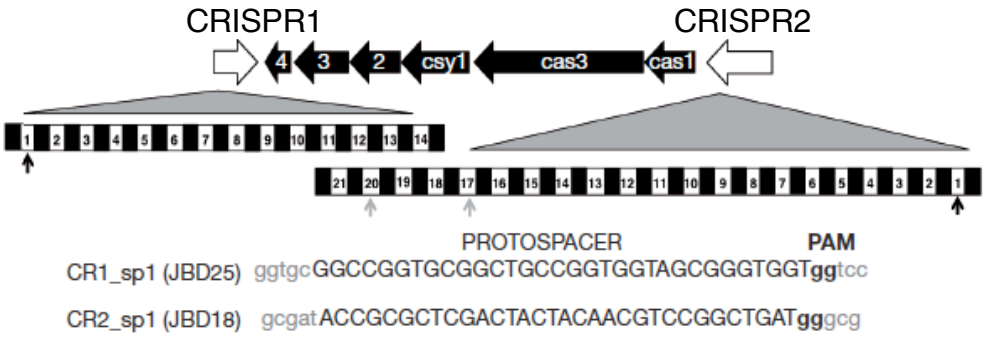


Figure 28: A schematic of the PA14 CRISPR loci and *cas* gene region.

An expanded version of each CRISPR locus indicates the number of spacers in each, shown with white boxes, each of which is flanked by repeats denoted by black boxes. Black arrows indicate the CRISPR spacers corresponding to protospacers tested in Figure 29 and gray arrows indicate the CRISPR spacers corresponding to protospacers tested in Figure 42. The DNA sequences of the protospacers tested in Figure 29 are shown.

I constructed plasmids containing targeted protospacer sequences from phages JBD18 (CRISPR2 locus, spacer 1 or CR2_sp1) and JBD25 (CR1_sp1) (Cady *et al.*, 2012). The transformation efficiencies of the plasmids bearing protospacers into unlysogenized PA14 were reduced by at least 90% compared to an empty vector control, whereas no difference in transformation efficiency was seen for the three strains containing prophages expressing anti-CRISPR activity, or for the ΔCR-*cas* strain (Figure 29). These data confirm that the prophages isolated in our screen inhibit the PA14 CRISPR-Cas system.

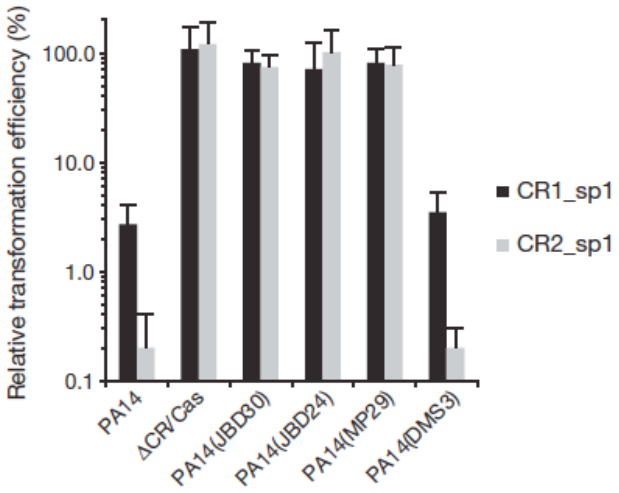


Figure 29: Prophages inhibit CRISPR-Cas targeting of protospacer-containing plasmids.

Plasmids containing protospacers shown in Figure 28 were electroporated into the indicated strains. The relative transformation efficiency was calculated by comparison with the transformation efficiency of the cloning vector containing no protospacer insert. Error bars represent standard deviation from the mean of three biological replicates. See Materials and Methods in Chapter 3 for cloning specifics and protocols.

4.3.2 Diverse anti-CRISPR genes are inserted in morphogenetic region

Genome comparisons revealed that the phages possessing anti-CRISPR activity were closely related to each other, and also displayed high sequence similarity and synteny with the previously characterized *P. aeruginosa* Mu-like phages D3112, DMS3, and MP22. The general genome characteristics and CRISPR protospacer matches are shown in Table 8.

Table 8: Summary of the CRISPR-associated characteristics of phages used in this study.

Phages used in this study are listed along with their relevant characteristics. Phages have been divided into two categories: those that have anti-CRISPR genes (and thus replicate on WT PA14), and those that are CRISPR-targeted (and thus do not replicate on WT PA14). Genome size, predicted number of open reading frames (ORFs), and anti-CRISPR gene identity are shown. The predicted spacer matches for a given phage as well as any mismatches are represented in the last two columns.

Anti-CRISPR Phages	Genome Size (bp)	#ORFs	Anti-CRISPR[2]	PA14 Spacer Matches[3]	#nt Mismatches
JBD5	37740	59	c,d	CR2_sp1	3 [4]
JBD24	37095	58	b	CR2_sp17	0
				CR2_sp20	1
JBD26	37840	61	b	CR2_sp17,20	0
JBD30	36947	56	a	CR2_sp17,20	0
JBD88a	36321	55	c	CR2_sp17,20	0
MP22	36049	51	c	CR2_sp17,20	0
MP29	36632	51	e	CR2_sp17	0
D3112	37611	55	e	CR2_sp17,20	0
DMS3[1]	36415	52	-	CR2_sp17	SND[5]
				CR2_sp20	PAM absent
CRISPR-Targeted Phages					
JBD18	39014	54	-	CR1_sp6, CR2_sp1,2	0
				CR2_sp3	4
				CR2_sp8	PAM absent
				CR2_sp12	2
JBD25	39552	57	-	CR1_sp1	0
JBD67	38084	55	-	CR1_sp6, CR2_sp1	0
				CR2_sp2	1
				CR2_sp8	PAM absent
				CR2_sp12	3

1-DMS3 is most closely related to the phages with anti-CRISPRs but does not have a detected anti-CRISPR gene **2**-Anti-CRISPRs: a-JBD30 gene *35*, b-JBD24 gene *36*, c-JBD88a gene *33*, d-JBD5 gene *36*, e-MP29 gene *29* **3**-If not noted, the PAM (GG) is present in the phage **4**-These mismatches should not prevent targeting (Cady *et al.*, 2012) **5**-Single nucleotide deletion

An unusual feature of each of these genomes compared to more distantly related Mu-like phages infecting *P. aeruginosa* (Braid *et al.*, 2004) and those infecting other hosts, like *E. coli* phage Mu (Morgan *et al.*, 2002) is the presence of diverse atypical genes at a single position within the region encoding phage head proteins. Figure 30 shows a detailed genome map of this region in phage D3112 (Wang *et al.*, 2004) and Figure 31 shows multiple phages with a focus on the novel insertion region. I suspected that some of these genes might encode anti-CRISPR activities; thus, seventeen of them were cloned and expressed from a high copy number plasmid in PA14 under

the control of an arabinose-inducible promoter. Remarkably, expression of eight of these genes led to dramatic increases in the plaquing efficiency of the CRISPR-sensitive phages (Figure 32).

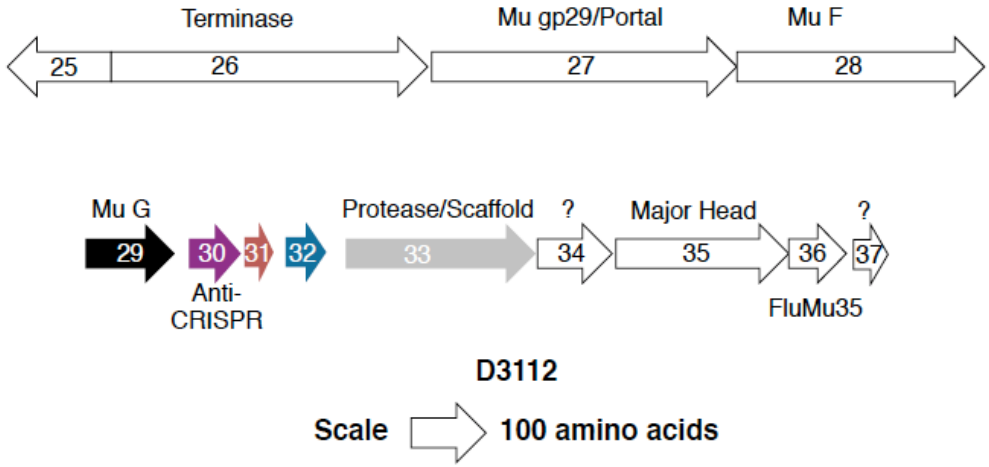


Figure 30: Genomic map of phage D3112

A schematic showing the head structural genes of D3112 and the predicted functions of various genes based on homology. The anti-CRISPR region is also shown, and coloured in accordance with the expanded version of this region in Figure 31 (genes *29-33*). These proteins are shown to scale.

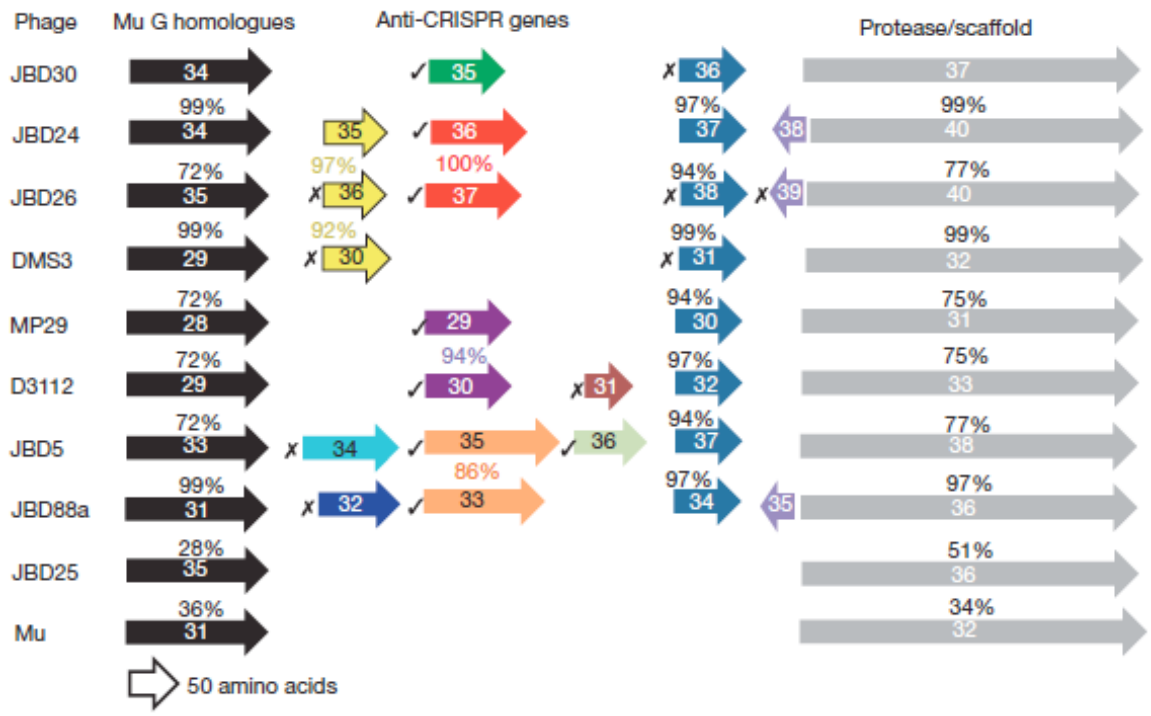


Figure 31: A diverse collection of anti-CRISPR genes are inserted in the morphogenetic region of related Mu-like phages

The anti-CRISPR genes of the indicated phages are located in the head gene regions of these genomes between genes homologous to the *G* gene of *E. coli* phage Mu (black boxes) and genes encoding the protease/scaffold protein of the phage head (gray boxes). The percent identity of the proteins encoded by these genes to representatives from JBD30 are shown. The coloured boxes represent putative anti-CRISPR genes. Boxes of the same colour represent closely related genes and the sequence identity of their encoded proteins is indicated. Genes found to mediate anti-CRISPR activity are indicated by check marks and genes tested but displaying no anti-CRISPR activity are marked with an "X". Unmarked genes were not tested due to their high similarity to tested genes. The gene box sizes are proportional to the sizes of the proteins in question (scale bar is 50 amino acids), but the spacing of the genes is not to scale.

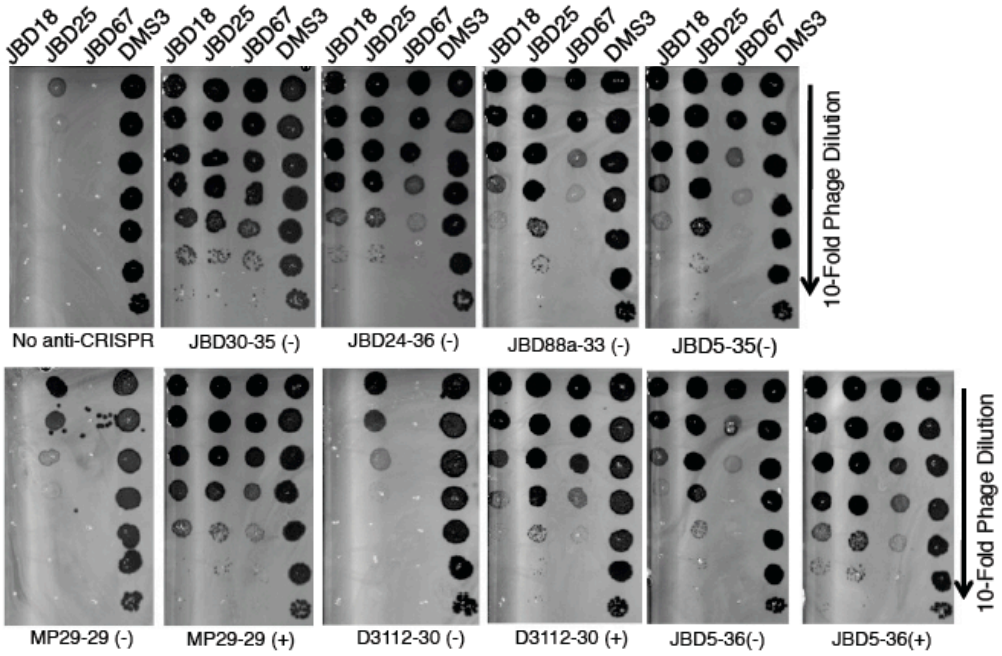


Figure 32: Different phages have distinct anti-CRISPR genes

The same phages from Figure 27 were spot titrated on PA14 containing empty vector or plasmids expressing the indicated anti-CRISPR genes. Induction of the plasmid promoter (+) with arabinose was required to produce a maximal effect for some of the anti-CRISPR genes while others did not (-).

Each anti-CRISPR gene allowed these three different phages to evade the CRISPR-Cas system even though they bear distinct protospacer targets (Table 8), indicating that the anti-CRISPR genes do not function through protection of specific DNA sequences on the phages. April Pawluk, a graduate student in the lab also assayed the activity of JBD30-35 in other *P. aeruginosa* strains possessing active Type I-F systems and she found that this gene acts as an anti-CRISPR in these strains also. This demonstrates that the anti-CRISPR phenomenon is not particular to strain PA14 (Figure 33). The WT and ΔCR deleted versions of these strains were acquired from Dr. Kyle Cady and Dr. George O'Toole.

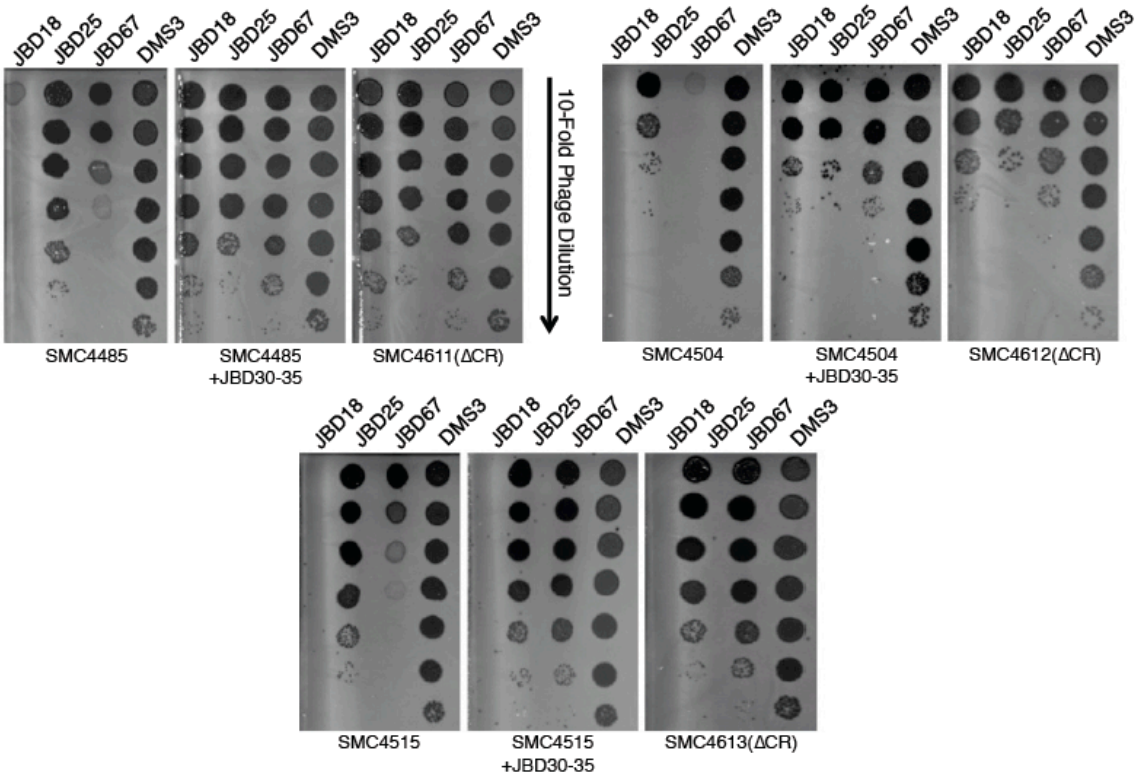


Figure 33: Phage JBD30 anti-CRISPR gene *35* inhibits the Type I-F CRISPR-Cas system in other *P. aeruginosa* isolates

(Experiment conducted by April Pawluk)
Ten-fold dilutions of lysates of CRISPR-sensitive phages (from L to R) JBD18, JBD25, JBD67 and CRISPR-insensitive phage DMS3 were applied to lawns of indicated *P. aeruginosa* strains (Cady *et al.*, 2011). The left column of panels show wild-type (WT) isolates SMC4485, 4504, and 4515, the middle column of panels show the respective WT strains transformed with a plasmid containing anti-CRISPR JBD30-*35*, and the right panels show the respective ΔCRISPR mutants derived from each WT strain.

Additionally, April found that the anti-CRISPR genes did not inhibit the Type I-E CRISPR-Cas system of *E. coli* (Datsenko *et al.*, 2012), while the anti-CRISPR proteins were well expressed (Figure 34a,b). Since Type I-E is the most closely related CRISPR-Cas system to Type I-F (Makarova *et al.*, 2011b) (comparison in Figure 34c), I do not expect that these genes would inhibit the function of any of the other more distantly related CRISPR-Cas systems.

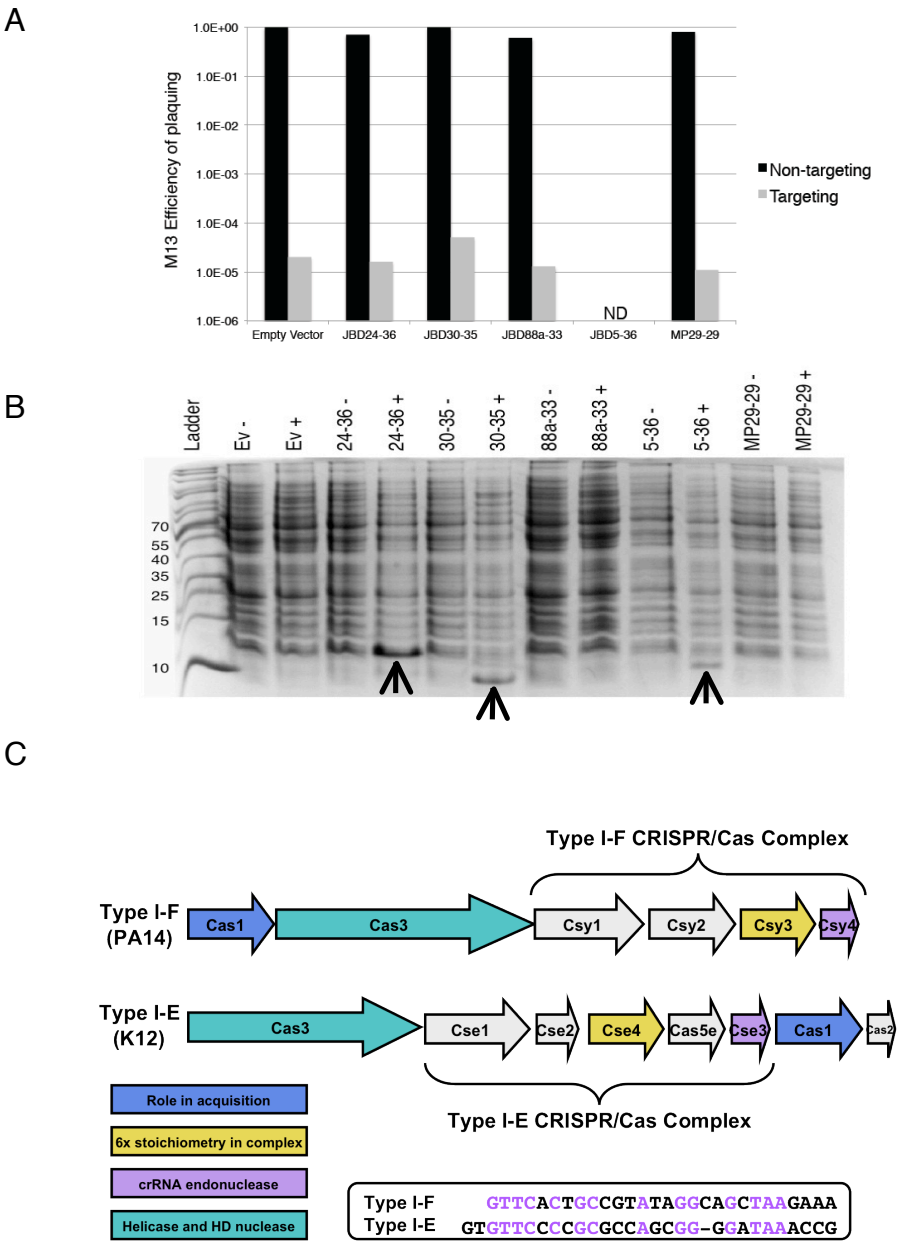


Figure 34: Anti-CRISPRs do not inhibit the Type I-E CRISPR-Cas system of *E. coli*

(Experiments and Figure by April Pawluk).
(A) A non-targeting *E. coli* strain that is permissive for phage M13 was engineered to target M13 (targeting) with the Type I-E CRISPR-Cas system (Datsenko *et al.*, 2012). Plaquing efficiency of phage M13 was assessed in the presence and absence of anti-CRISPR genes and calculated relative to the non-targeting strain. Cells expressing JBD5 gene *36* grew poorly; thus, data were not determined (ND) for this anti-CRISPR. (B), Lysates were generated from the targeting cells containing anti-CRISPR constructs or empty vector (Ev) under promoter repression (-) or induction (+) and analyzed by SDS-PAGE to assess the expression of anti-CRISPR proteins. Proteins corresponding to the expected sizes (see Figure 37) were produced from cells expressing JBD24 gene *36* (JBD24-36), JBD30-*35*, and JBD5-*36* were grown under inducing conditions, as indicated by arrows. JBD88a-*33* and MP29-*29* may also be expressed, but the proteins do not accumulate to a high enough level to be observed by coomassie staining, (C) Despite

being more closely related to each other than other CRISPR-Cas systems (Makarova *et al.*, 2011b), the *cas* genes of the Type I-F and the Type I-E systems (shown here from *P. aeruginosa* PA14 and *E. coli* K12, respectively) still encode proteins that are very distinct from one another. Proteins of analogous function are highlighted in the same colour, but no protein sequences have more than 25% identity between Type I-F and Type I-E. The processing enzymes (csy4 and cse3) have 18% identity and 38% similarity while csy3 and cse4 possess 15% identity and 42% similarity. The other proteins do not display detectable identity to each other. The CRISPR repeat sequences shown at the bottom are also very different between the two systems

4.3.3 Anti-CRISPRs do not inhibit crRNA biogenesis, *cas* gene expression

For the Type I CRISPR-Cas system to function, transcription of pre-crRNA from the CRISPR locus must occur, followed by processing into small crRNAs and incorporation of these RNAs into a complex with Cas proteins. The accumulation of processed crRNA within the cell requires Cas proteins. Thus, the lack of any *P. aeruginosa* Cas protein except Cas1, which is involved in spacer acquisition, causes a marked reduction in crRNA levels (Cady and O'Toole, 2011). As can be seen in Figure 35, expression of five different anti-CRISPR genes within PA14 caused no change in the level of processed crRNA molecules as detected by northern blotting. The normal level of crRNA observed implies that expression of anti-CRISPR genes does not cause a reduction in the expression levels of the CRISPR loci or *cas* genes.

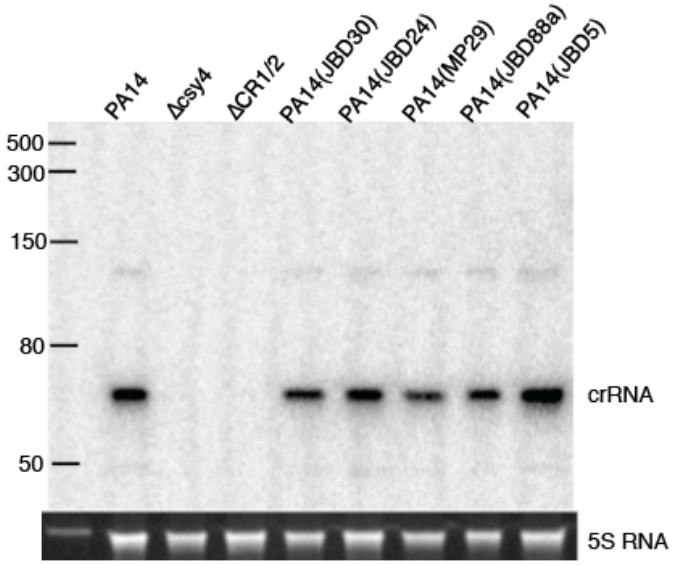


Figure 35: Anti-CRISPR expression does not impact crRNA accumulation

Total RNA extracted from *P. aeruginosa* cell lysates was electrophoresed and a blot was probed with ^{32}P-labelled DNA derived from the last four repeats and three spacers of the CRISPR2 locus (top panel). Below the blot is the same gel stained with SYBR Gold showing the 5S RNA band in each lane. RNA was extracted from wild-type PA14, a strain with one *cas* gene deleted (Δ*csy4*), a strain lacking both CRISPR loci (ΔCR1/ΔCR2), and PA14 lysogenized by JBD30, JBD24, MP29, JBD88a, or JBD5. The position of ssRNA molecular weights are indicated, n=3.

Supporting this finding, I also found that transcription levels of the *cas* genes *cas3* and *csy3* were unaffected by the anti-CRISPR genes as assessed by reverse transcriptase quantitative PCR (RT-qPCR) (Figure 36a,b). Furthermore, April Pawluk assayed β-galactosidase activity produced from a chromosomally located *csy3*::*lacZ* fusion gene and found it was not perturbed by expression of any of the anti-CRISPR genes (Figure 36c).

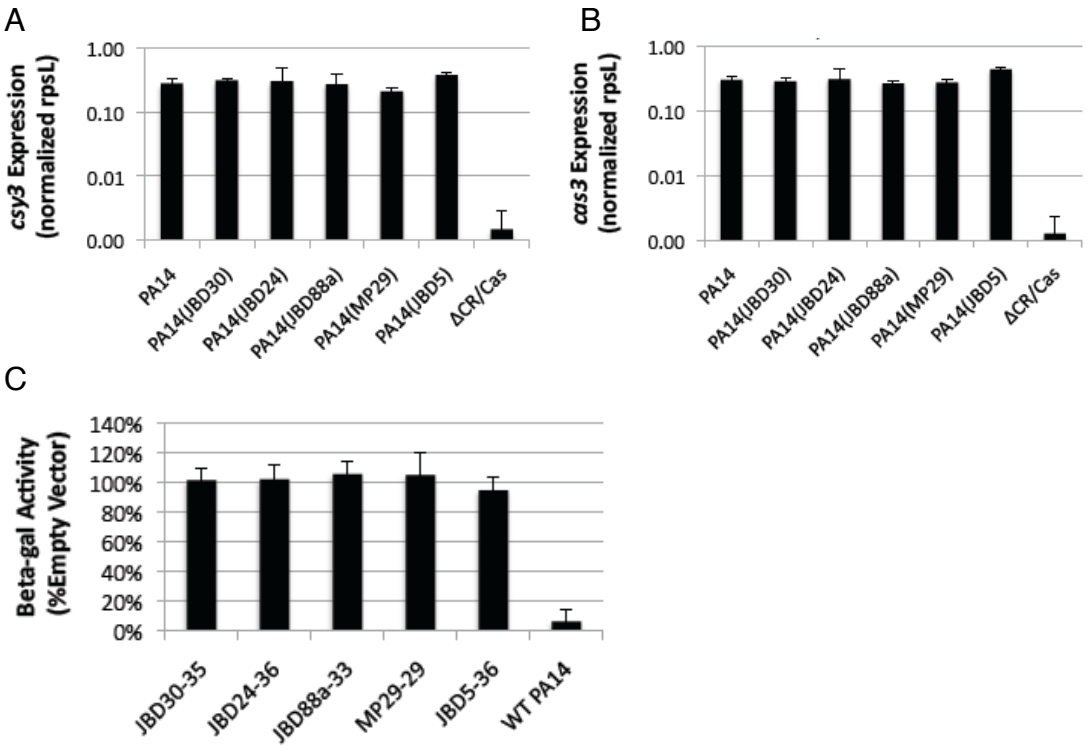


Figure 36: Anti-CRISPR expression does not impact *cas* gene transcription

Reverse transcriptase-quantitative PCR (RT-qPCR) was conducted to determine the transcription levels of *cas* genes, (A) *csy3* and (B) *cas3* in PA14 and indicated lysogens. The transcription levels were normalized to PA14 gene *rpsL*, which encodes ribosomal protein S12. Error bars represent the standard deviation, (C) April Pawluk conducted ß-galactosidase assays on a chromosomal *csy3::lacZ* fusion strain containing empty vector or plasmids expressing the indicated anti-CRISPR genes. Values were normalized to empty vector. Wild-type PA14 (lacking the *lacZ* fusion) is shown as a control for background β-galactosidase activity.

I conclude from these experiments that the anti-CRISPR genes exert their effects at a step occurring after formation of the crRNA-Cas complex, and that there is no effect detected on biogenesis of either the crRNA or Cas proteins.

4.3.4 Anti-CRISPR homologs found in mobile elements

Despite the common genomic positions of the anti-CRISPR genes in very similar *P. aeruginosa* phages, these eight anti-CRISPR genes are predicted to encode five different proteins with completely distinct sequences (Figure 37).

JBD30 gp35, 78 a.a., 9 kDa, pI=8.03
MKFIKYLSTAHLNYMNIAVYENGSKIKARVENVVNGKSVGARDFDSTEQLESWFYGLPGSGLGRIENAMNEISRRENP

JBD24 gp36, 100 a.a., 11 kDa, pI=5.25
MMTISKTDIDCYLQTYVVIDPVSNGWQWGIDENGVGGALHHGRVEMVEGENGYFGLRGATHPTEKEAMAAALGYL
WKCRQDLVAIARNDAIEAEKYRAKA

JBD88a gp33, 140 a.a., 16 kDa, pI=4.61
MSSTISDRIISRSVIEAARFIQSWEDADPDNLTESQVLAASSFAARLHEGLQATVLQRLVDESNRDEYREFQAWEEALL
NADGRVTSNPFADWGWWYRIANVMLATASQNVGVAWGSHVHGRLMAIFQDRFQQHYEDEEC

JBD5 gp35, 139 a.a., 16 kDa, pI=4.9
MSNTISDRIVARSVIEAARFIQSWEDADPDSLTEDQVLAAAGFAARLHEGLQATVLQRLVDESNHEEYREFKAWEEAL
LNADGRVASSPFADWGWWYRIANVMLATASQNVGVTWGSRVHGRLMAIFQDKFKQRYEEQA

JBD5 gp36, 79 a.a., 10 kDa, pI=9.7
MSRPTVVTVTETPRNPGSYEVNVERDGKMVVGRARAGSDPGAAAAKAMQMAMEWGSPNYVILGSNKVLAFIPEQL
RVKM

D3112 gp30, 90a.a., 10 kDa, pI=3.9
MIAQQHKDTVAACEAAEAIAIAKDQVWDGEGYTKYTFDDNSVLIQSGTTQYAMDADDADSIKGYADWLDDEARSAEA
SEIERLLESVEEE

MP29 gp29, 90 a.a., 10 kDa, pI=3.9
MIAQQHKDTVAACEAAEAIAIAKDQVWDGEGYTKYTFDDNSVLIQSGTTQYAMDADDADSIKGYADWLDDEARTAED
SEIKRLLEAVEDEA

Figure 37: Anti-CRISPR protein sequences

The predicted protein sequences of anti-CRISPRs are shown. The length of each protein in amino acids (a.a.) is shown as well as the predicted molecular weight and estimated pI value.

Identity between some proteins is detectable and this is indicated in Figure 31. These proteins do not have any detectable transmembrane domains (TMHMM Server), nor do secondary structure predictions (Jpred 3) yield consistent predictions that point to a common fold or function. Homology searches using HHpred yielded no functional insight while PSI-BLAST searches revealed less than 15 significant hits, of which all but four were proteins encoded in genomes of closely related phages or prophages (Table 9).

Table 9: Anti-CRISPR BLASTp hits.

PSI-BLAST searches were conducted with the protein sequence encoded by each anti-CRISPR gene and the hits are shown along with their respective accession numbers, percent identity to the query and E value.

Query	Hits	Accession	Identity	Expect	Notes
JBD30-35	*P. stutzeri* TS44 YO5_18187	EIK54721	73%	7.0E-32	Type IV Secretion-Related[1]
	Pseudomonas phage MP42_32	AFE86461	100%	2.0E-38	Phage
JBD24-36	*Pseudomonas* phage JBD26gp0036	AEY99442	100%	3.0E-68	Phage
	P. aeruginosa 39016_870056	EFQ38430	100%	3.0E-31	Prophage
JBD88a-33	*Pseudomonas* phage MP22_orf29b	ABH09854	100%	1.0E-97	Phage
	Pseudomonas phage F_gp34	AEY99407	100%	1.0E-97	Phage
	P. aeruginosa LESB58_15761	CAW26304	88%	3.0E-84	Prophage
	Pseudomonas phage PA1/KOR/2010	ADU15525	82%	1.0E-42	Phage
	P. aeruginosa hypotheical protein (PAGI-5)	ABR13384	43%	9.0E-21	Genomic Island
	P. aeruginosa E2_0713	EKA57093	88%	9.0E-84	Type IV Secretion/Conjugation[1]
JBD5-35	*P. aeruginosa* LESB58_15761	CAW26304	96%	3.0E-92	Prophage
	Pseudomonas phage F_gp34	AEY99407	86%	1.0E-97	Phage
	Pseudomonas phage MP22_orf29b	ABH09854	86%	3.0E-83	Phage
	Pseudomonas phage PA1/KOR/2010	ADU15525	98%	5.0E-51	Phage
	P. aeruginosa hypotheical protein (PAGI-5)	ABR13384	43%	9.0E-21	Genomic Island
	P. aeruginosa E2_0713	EKA57093	95%	3.0E-90	Type IV Secretion/Conjugation[1]
JBD5-36	*Pseudomonas* phage PA1/KOR/2010	ADU15526	99%	2.0E-43	Phage
	P. aeruginosa LESB58_15771	CAW26305	92%	7.0E-17	Prophage
	P.aeruginosa E2_0714	EKA57094	92%	2.0E-16	Type IV Secretion/Conjugation[1]
D3112-30	*Pseudomonas* phage MP29_gp29	ACA57674	94%	1.0E-43	Phage
MP29-29	*Pseudomonas* phage D3112p30	AAQ94468	94%	2.0E-43	Phage

1-See Figure 38

One of the non-phage associated anti-CRISPR protein homologues, which is 43% identical to the product of gene *33* of phage JBD88a (JBD88a gp33), is encoded within an active pathogenicity island of a highly virulent *P. aeruginosa* clinical isolate that is likely transferred by conjugation between *P. aeruginosa* strains (Battle *et al.*, 2008). This island contains 4 protospacers with correct PAMs and 100% identity to CRISPR spacers in various *P. aeruginosa* strains (Cady *et al.*, 2011). The three other non-phage associated anti-CRISPR homologues are also found in regions of *Pseudomonas* genomes that may be mobile elements as indicated by presence of genes in these regions encoding homologues of proteins involved in DNA transfer and/or Type IV secretion (Figure 38). Thus, these putative bacterial anti-CRISPR genes may increase the fitness for inter-strain transfer of these mobile elements by inactivating the CRISPR-Cas system of a recipient strain.

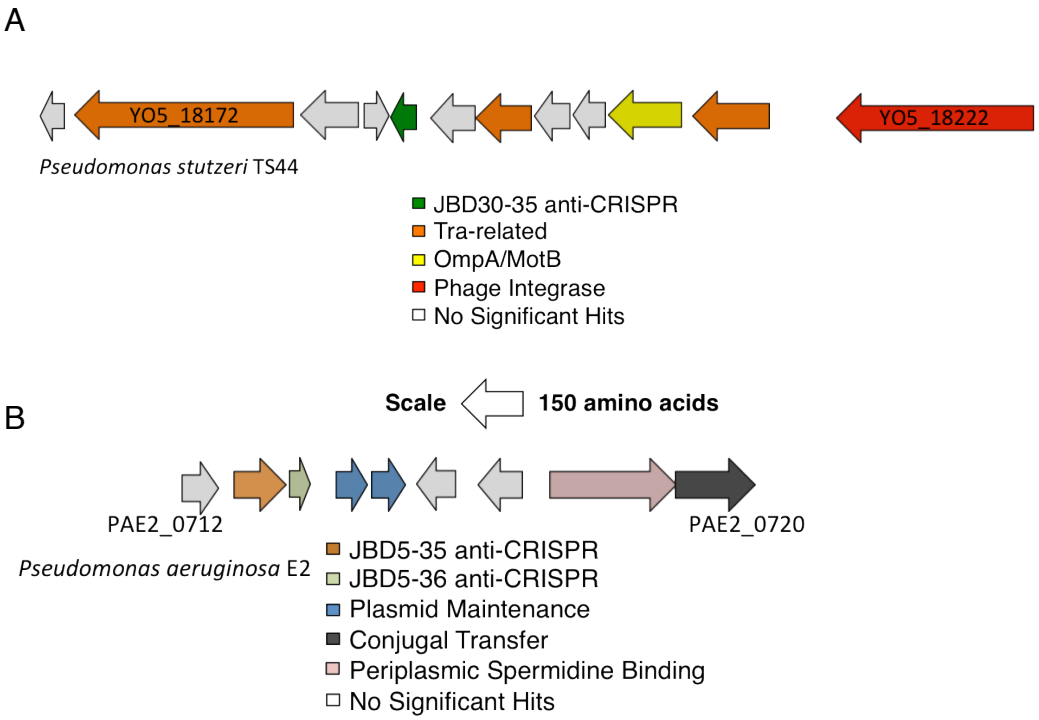


Figure 38: *Pseudomonas stutzeri* TS44 and *P. aeruginosa* E2 have anti-CRISPR homologues in mobile reigons

(A)The *P. stutzeri* TS44 genomic region surrounding a homologue of JBD30 gene *35* is shown. The genomic locations of two genes are shown (YO5_18172, Accession EIK54718 and YO5_18222, Accession EIK54728). Genes with significant hits on BLAST searches are colour coded and their predicted functions indicated. '*tra*' genes are known to be involved in conjugal transfer of plasmids. (B) The *P. aeruginosa* E2 genomic region surrounding homologues of JBD5 genes *35* and *36* which each have anti-CRISPR activity in PA14. Genes involved in plasmid maintenance and conjugation are indicated.

4.3.5 Anti-CRISPR activity is protein-mediated

Since the crRNA-Cas complex is guided by RNA, anti-CRISPR activity might be mediated by a non-coding RNA molecule or a protein encoded by an anti-CRISPR gene. I addressed this issue by performing experiments on JBD30 gene *35*. A nonsense mutation at the third codon and two different frameshift mutations were introduced to the plasmid encoding gene *35*. Each of these mutations abrogated anti-CRISPR activity (Figure 39), implying that translation of this region was required for function.

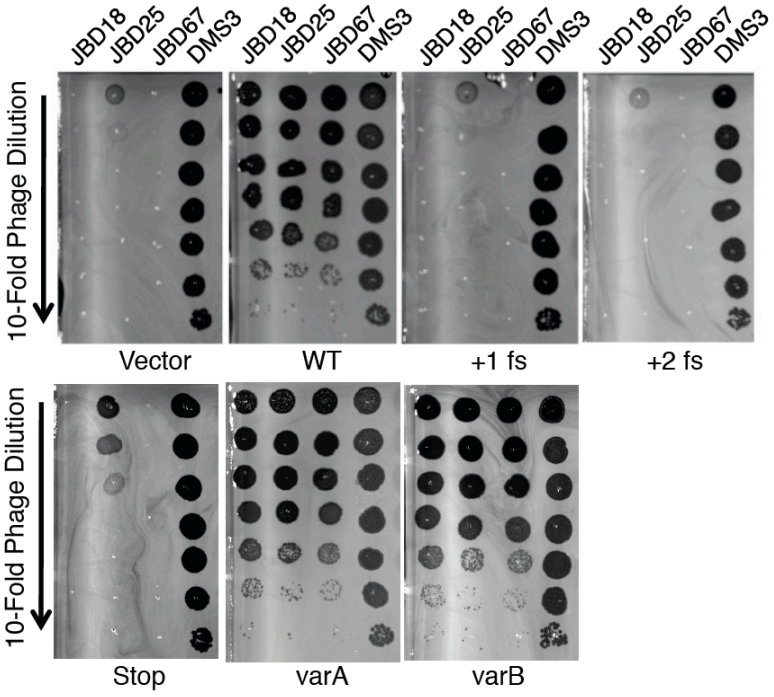


Figure 39: The anti-CRISPR activity of JBD30 gene *35* is mediated by the encoded protein

Ten-fold dilutions of lysates of CRISPR-sensitive phages (from L to R) JBD18, JBD25, JBD67, and CRISPR-insensitive phage DMS3 on lawns of PA14 transformed with mutant forms of JBD30 gene *35* (from L to R): empty vector, WT gene *35*, +1 frameshift (fs), +2 fs, a premature stop codon, variant sequence (var) A, and varB. Frameshift mutations were introduced immediately following the start codon and the introduced stop codon is in the third position (i.e. Met-Ala-STOP). The variant JBD30 gene *35* sequences have silent mutations introduced at every possible codon (shown in Figure 40) to create a gene producing a distinct RNA molecule with ~65% nucleotide identity to wild-type, but one that will be translated to produce wild-type protein.

Since these experiments did not rule out a combined role for anti-CRISPR non-coding RNA and protein, two variant genes were synthesized that encoded the same amino acid sequence as gene *35*, yet had DNA sequences that differed by ~35% through variation of codon wobble positions (Figure 40).

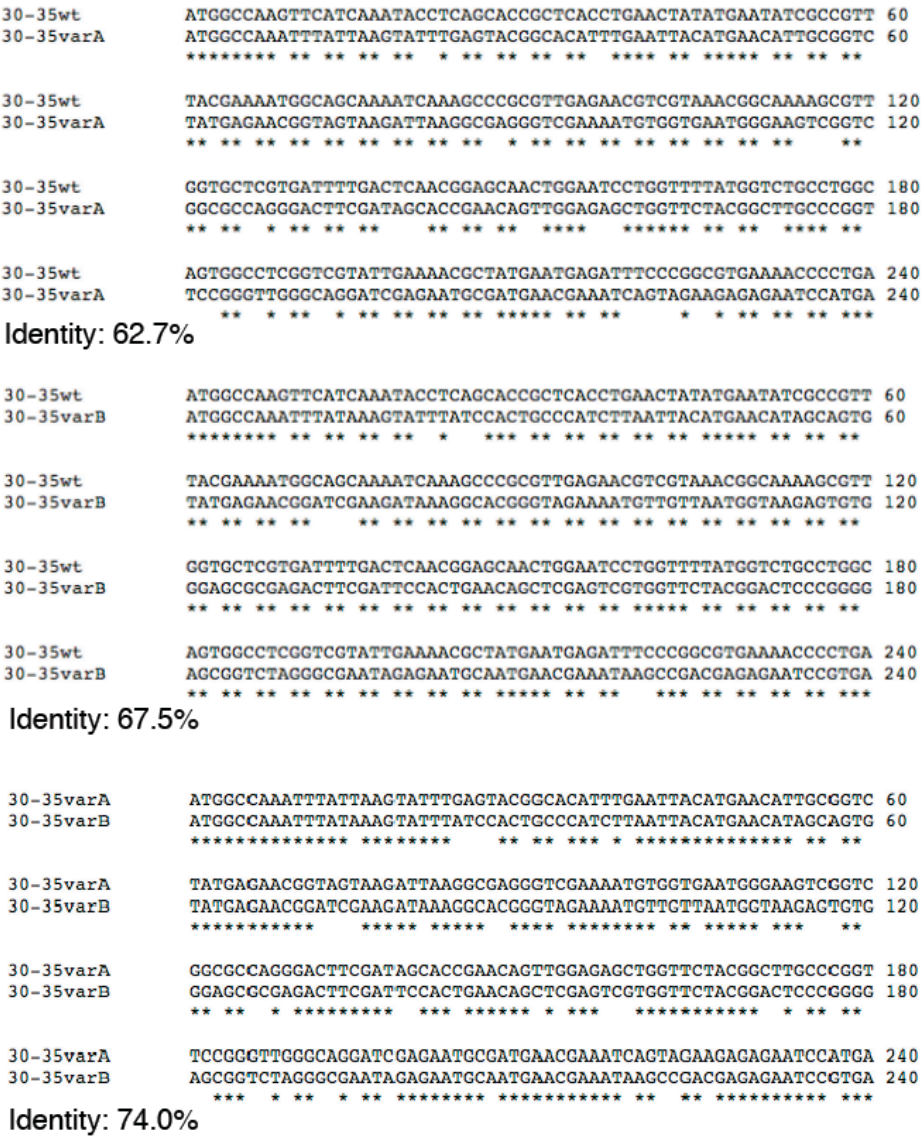


Figure 40: Alignment of synthesized variants of JBD30 gene *35* with silent mutations

Two variants of JBD30 gene *35* were synthesized (varA, varB) with silent mutations introduced to modify the RNA molecule as much as possible without changing the sequence of the translated protein. Alignments of each of these synthesized genes are shown, compared to WT, as well the two newly synthesized genes compared to each other. Nucleotide identity for each alignment is shown, while protein sequence identity is 100% in all cases.

As shown in Figure 39, each of these synthetic versions of gene *35* imparted full anti-CRISPR activity. These data demonstrate that anti-CRISPR protein is required for anti-CRISPR activity and that a direct mechanistic role for a gene *35*-encoded RNA is unlikely.

4.3.6 Anti-CRISPRs are required during phage infection

The genomes of six of the seven "anti-CRISPR phages" (i.e. those phages bearing active anti-CRISPR genes) contain at least one protospacer with a perfect match to the PA14 CRISPR locus (Table 8); thus, their replication should be inhibited by the PA14 CRISPR-Cas system. However, each was able to form plaques on PA14 with similar efficiency compared to the ΔCR-Cas strain (Figure 41).

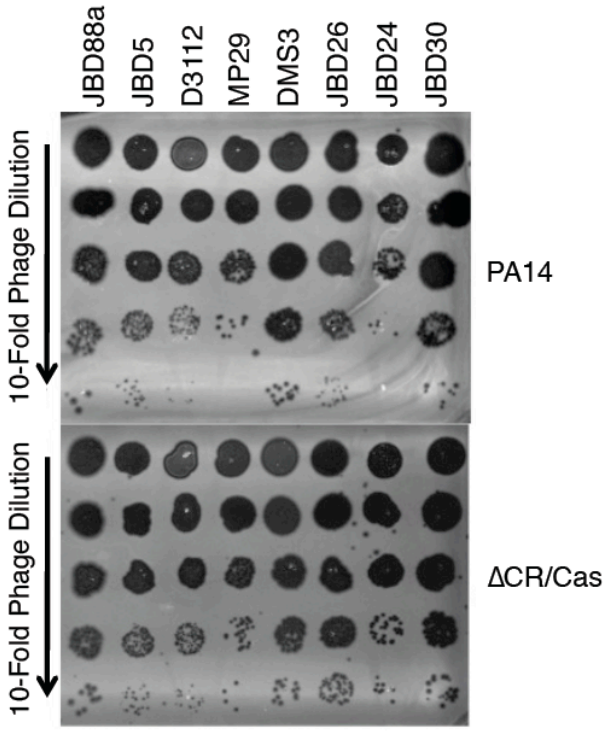


Figure 41: Phages with anti-CRISPRs can infect PA14 despite possessing protospacers

Ten-fold dilutions of lysates of the eight indicated phages are shown in a spot titration assay (as described in Figure 27) on lawns of WT PA14 and ΔCR-*cas*. The specific anti-CRISPR gene possessed by a given phage and predicted protospacer matches to the PA14 CRISPR loci are shown in Table 8. Anti-CRISPR phages plaque equally well on PA14 and ΔCR-Cas or slightly better (i.e. 2-5 fold) on ΔCR-*cas* whereas the CRISPR-sensitive phages plaque on ΔCR-*cas* with 10^6-10^8-fold higher efficiency than on PA14, as demonstrated previously (Figure 27).

Using the transformation efficiency assay, I confirmed that the two protospacers found most commonly in the anti-CRISPR phages were indeed targeted by the PA14 CRISPR-Cas system (Figure 42).

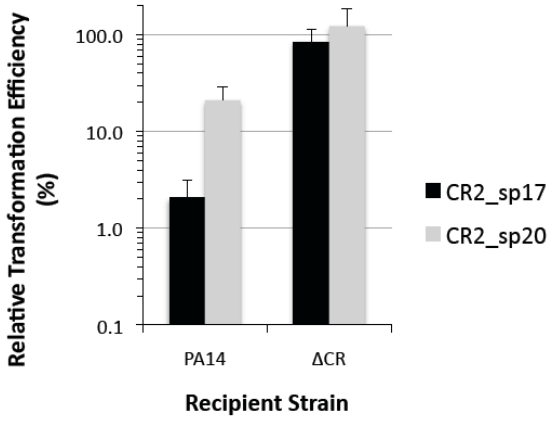


Figure 42: Protospacers found in anti-CRISPR containing phages are CRISPR-Cas targets

The transformation efficiency of plasmids containing protospacers matching CR2_sp17 and CR2_sp20 of the PA14 CRISPR-Cas system (gray arrows in Figure 28) was assessed. These two protospacers are found on most of the phages studied here which contain anti-CRISPR genes (See Table 8). The experiment and calculations were done as in Figure 29.

These results implied that the anti-CRISPR phages are able to replicate on PA14 because they possess anti-CRISPR genes. To address this hypothesis, a frameshift mutation was introduced into the phage JBD30 anti-CRISPR gene (gene *35*). This mutant phage was unable to replicate on wild-type PA14 but still replicated robustly on the ΔCR-*cas* strain, demonstrating the requirement of the anti-CRISPR gene for replication in cells bearing an intact CRISPR-Cas system (Figure 43).

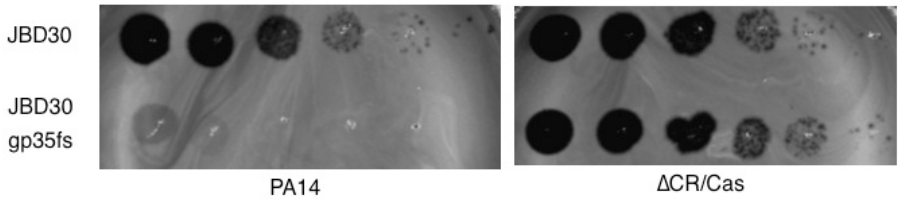


Figure 43: An anti-CRISPR protects phages from the CRISPR-Cas system during infection

Ten-fold dilutions of lysates of anti-CRISPR phage JBD30, and the same phage with a frameshift mutation introduced into the anti-CRISPR gene *35* (gene *35fs*) were applied to lawns of PA14 or PA14 ΔCR-Cas.

To determine whether the introduction of an anti-CRISPR gene into a CRISPR-sensitive phage would allow that phage to evade CRISPR-Cas immunity, I utilized a CRISPR-sensitive mutant of phage DMS3, called DMS3*m* (Cady *et al.*, 2012). This phage possesses a functional protospacer and is very similar in sequence to the anti-CRISPR phages, yet it contains no functional anti-CRISPR gene. Taking advantage of the high DNA sequence identity between

phages DMS3*m* and JBD30, *in vivo* homologous recombination was used to create a version of DMS3*m* bearing JBD30 gene *35* (Figure 44, see methods for details).

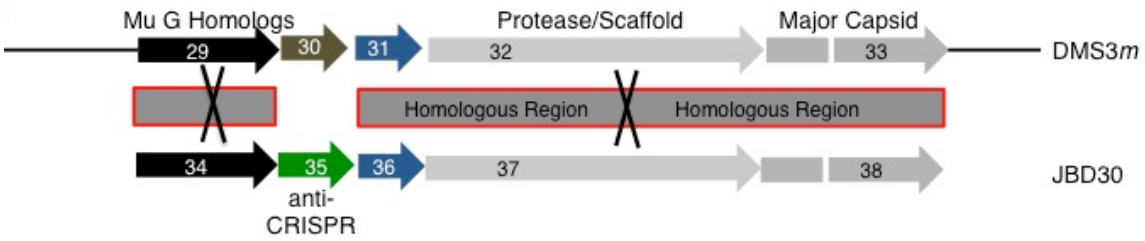


Figure 44: Schematic of anti-CRISPR *in vivo* homologous recombination.

Phage DMS3*m* was used to infect cells with a plasmid containing the anti-CRISPR region from phage JBD30 and recombinants selected. The X marks show the mapped region of recombination, up- and downstream of the anti-CRISPR gene *35* from JBD30.

As shown in Figure 45, the introduction of this gene into DMS3*m* resulted in a 10^6-fold increase in plaquing efficiency on PA14, clearly demonstrating that an anti-CRISPR gene present on an infecting phage allows that phage to overcome the CRISPR-Cas system.

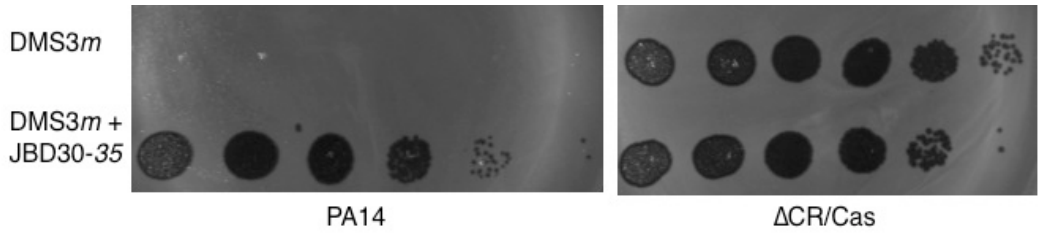


Figure 45: A CRISPR-Cas targeted phage can be rescued by an anti-CRISPR.

Ten-fold dilutions of lysates of a CRISPR-sensitive phage (DMS3*m*) or DMS3*m* with anti-CRISPR gene *35* from JBD30 inserted (DMS3*m* + JBD30-35) were applied to lawns of PA14 or PA14 ΔCR-*cas*.

By testing lysogens of the DMS3*m* and JBD30 mutant prophages I found that a protospacer-bearing plasmid efficiently transformed only those lysogens in which an intact anti-CRISPR gene was present (Figure 46). These assays demonstrate the necessity and sufficiency of the anti-CRISPR gene for inhibition of the CRISPR-Cas system.

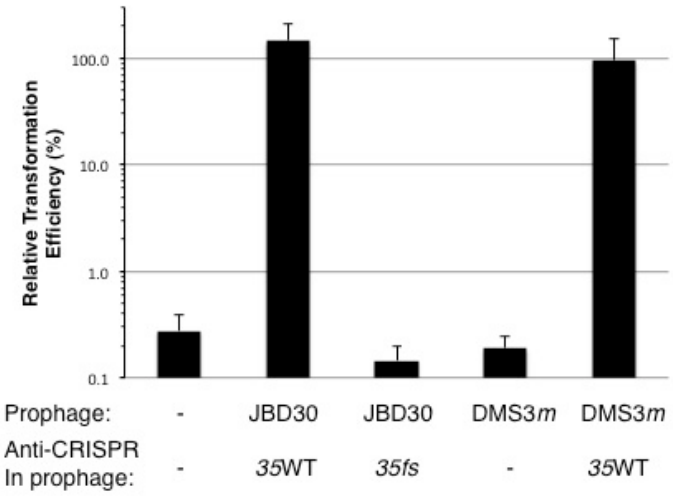


Figure 46: Prophage-expressed anti-CRISPR JBD30 gene *35* inhibits CRISPR-cas targeting of a plasmid

A plasmid containing a protospacer matching CR1_sp1 (shown in Figures 28 and 29) was electroporated into the indicated lysogens or parent strain. As indicated, the prophages within these lysogens contain either a wild-type (WT) version of anti-CRISPR gene *35*, a frameshift mutant of this gene, or no anti-CRISPR gene. Lysogens were made in the PA14ΔCR2 background to allow lysogeny while the plasmid used is targeted by CRISPR locus 1.

4.3.7 Type I-E Anti-CRISPR found in same genomic locus as type I-F

After establishing the identity of five Type I-F anti-CRISPRs in a conserved genomic position of related phages, there were still a number of unidentified genes remaining in this locus (Figure 31, genes with 'X' marks). April Pawluk, a graduate student in our lab identified a *P. aeruginosa* strain with an active Type I-E CRISPR-Cas system. She demonstrated that four of the remaining genes in this region without Type I-F anti-CRISPR activity had Type I-E anti-CRISPR activity. These different families of anti-CRISPRs did not have any activity on the other CRISPR-Cas subtype. This work was recently published, describing the second group of anti-CRISPR genes described (Pawluk *et al.*, 2014). Additionally, a PCR-based screen that was performed by Vivian Cheung, a rotation student whom I supervised in the lab, revealed that other phages in my collection had Type I-F and I-E anti-CRISPRs arranged in different combinations. Although no novel anti-CRISPRs were discovered, these data demonstrate that these genes have likely been independently acquired, as well as mixed and matched with respect to their order. These data are summarized in Figure 47.

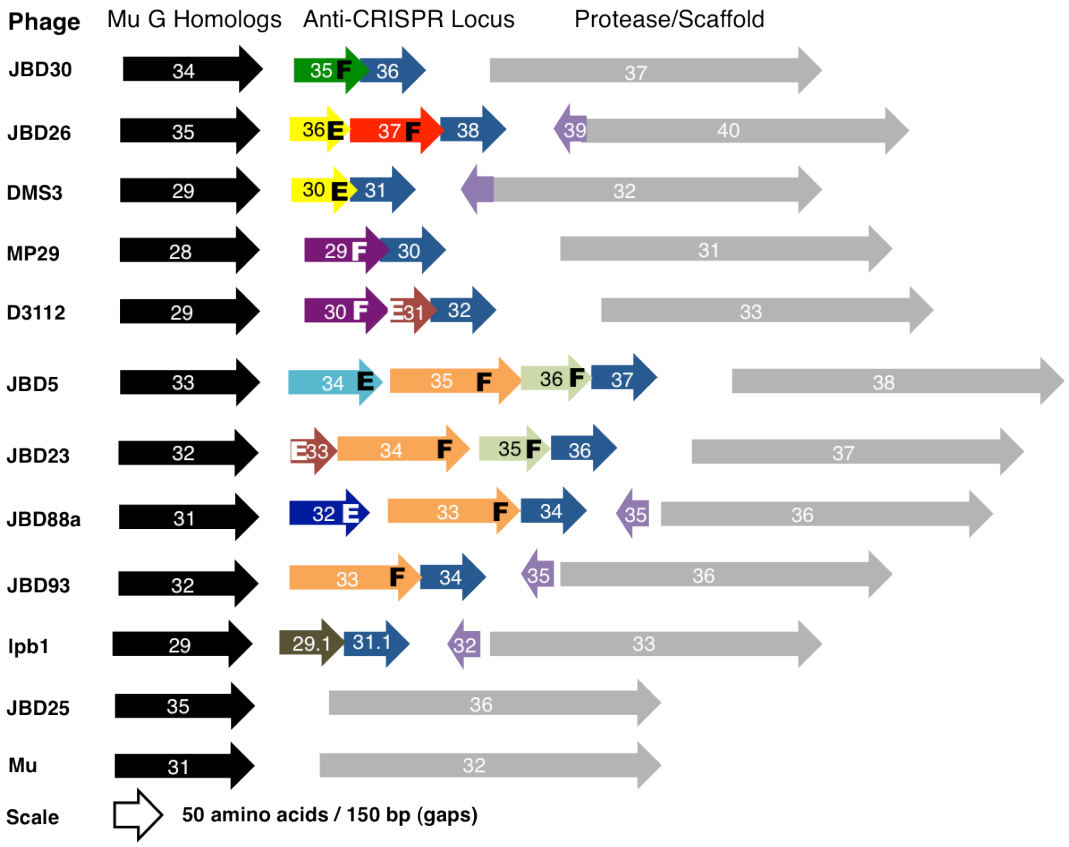


Figure 47: Type I-E anti-CRISPRs are found in the same genomic locus as type I-F anti-CRISPRs

Genes of different colours represent genes which do not possess obvious sequence identity. Type I-F anti-CRISPRs are indicated by a 'F' and Type I-E by an 'E'. Gene size and spacing are to scale. Phage lpb1 is present in NCBI and has a novel gene in this location which may be an anti-CRISPR but has not been tested.

4.4 Discussion

The adaptive nature of the CRISPR-Cas system and the widespread occurrence of CRISPR regions in bacterial genomes suggest that this system could be the most powerful weapon possessed by bacteria to resist invasion by foreign DNA. Prior to my work, the only known mechanism for phages to evade CRISPR-Cas systems was by mutation of the protospacer (Barrangou *et al.*, 2007; Semenova *et al.*, 2011), which is a low frequency event. Mutation also comes with a potential cost, especially in phage genomes which are mostly coding. Here, I have provided the first demonstration that the *in vivo* activity of a CRISPR-Cas system is dramatically inhibited by any one of five different "anti-CRISPR" genes. The presence of anti-CRISPRs on temperate phages and horizontally transferred elements suggests that perhaps their role is not simply one of ensuring a faithful lytic phage infection. Anti-CRISPRs facilitate both the

formation and maintenance of a prophage. Thus, although a CRISPR-Cas system would have rejected this foreign DNA, the potential benefits of the prophage (i.e. superinfection exclusion, fitness factors, etc.) could then manifest whereas they wouldn't have without an anti-CRISPR. This role could be extended to plasmids and other mobile elements, allowing dissemination of virulence factors, drug resistance determinants, etc.

The 'forced' acquisition of a mobile element due to an anti-CRISPR on that mobile element could also come with a cost, for two reasons. One is the misregulation or misintegration of foreign elements which always poses a potential cost to the host. Another cost to consider, however, is that anti-CRISPR expression is necessary to facilitate the ongoing protection of the foreign genetic element after integration into the host chromosome. Thus, the CRISPR-Cas system of the host is essentially inactivated and is therefore more sensitive to further attack from foreign DNA, such as a lytic phage. This makes the phenomena described in Chapter 2 more relevant (i.e. superinfection exclusion) as perhaps these mechanisms can compensate for constitutitve anti-CRISPR activity.

The existence of anti-CRISPR genes is one possible explanation for how phages have continued to proliferate despite the ubiquity and potency of CRISPR-Cas systems. The possibility that anti-CRISPR genes are diverse and widespread among phages and other mobile genetic elements may account for the large diversity of CRISPR-Cas systems and the existence of multiple CRISPR-Cas system types within single bacterial strains. This proliferation of CRISPR-Cas systems may be driven by the concomitant proliferation and diversification of anti-CRISPR genes. This newly discovered arms race may have a profound effect on the evolution of both phage and bacterial genomes and knowledge of anti-CRISPR genes will be crucial for understanding this process. The failure to detect anti-CRISPR genes until now may be due only to a lack of systematic searches using naturally functioning *in vivo* systems. Future studies to discover more anti-CRISPR genes and elucidate the mechanisms of their inhibition of CRISPR-Cas systems will provide new inroads for illumination of CRISPR-Cas function.

Chapter 5
Anti-CRISPRs operate via distinct mechanisms to inhibit CRISPR-Cas function

5 Overview

The battle for survival between organisms and the viruses that infect them is a constantly evolving landscape akin to an arms race. The clustered regularly interspaced short palindromic repeats (CRISPR) and the CRISPR associated (Cas) genes make up an adaptive immune system which bacteria use to fend off the viruses that infect bacteria, bacteriophages (phages). I previously identified five distinct phage-encoded anti-CRISPR genes, which inhibit the Type I-F CRISPR-Cas system of *Pseudomonas aeruginosa*, thus facilitating phage infection. Here, I present the solution structure for the product of anti-CRISPR gene *35* from phage JBD30 (ACR30-35) and identify a small functional epitope on the protein that is essential for *in vivo* activity. ACR30-35 directly interacts with the Type I-F CRISPR-Cas (Csy) complex in a three-to-one ratio, with low pico-molar affinity via the hexameric backbone protein, Csy3. A Csy complex that is bound by ACR30-35 is unable to recognize and bind to any DNA target, thus inhibiting CRISPR-Cas function. Another distinct anti-CRISPR, ACR3112-30, interacts with the Csy complex via the heterodimer Csy1-2 and inhibits the invasion of a double stranded DNA target, thus preventing target recognition. Finally, a third anti-CRISPR ACR5-35 does not interact with the Csy complex, but binds to the trans-acting Cas3 helicase/nuclease, which is required for targeting of foreign DNA. These studies provide the first mechanistic insight into anti-CRISPR structure and function, while providing novel insight into CRISPR-Cas function.

5.1 Acknowledgements

Many different biochemical techniques were utilized to assemble the data discussed below. I have utilized, to varying degrees, the expertise of those in the Davidson lab to try to put together a thorough mechanistic story. Yurima Hidalgo-Reyes, a technician in the lab and Hayoung Yoo, a rotation student I supervised, and myself made mutations to ACR30-*35* and screened them for *in vivo* activity. Yurima also assisted with protein purification. Dr. Bianca Garcia, a research associate in the lab conducted much of the *in vitro* analysis of ACR30-35 mutants including competition assays and protein melt curves. We worked closely together on determining the stoichiometry of ACR30-35. I was involved in coordinating and planning these experiments as

well as analyzing the results. Dr. Karen Maxwell, a senior research associate with Dr. Aled Edwards solved the NMR solution structure of ACR30-35 and Diane Bona, a technician in Dr. Maxwell's lab prepared the sample. Our collaborators at Montana State University, MaryClare Rollins and Dr. Blake Wiedenheft conducted surface plasmon resonance experiments. Finally, Mingjian Du, a summer exchange student whom I supervised performed *in vivo* Csy protein overexpression experiments. All other ACR30-35 experiments and all work with ACR3112-30 and ACR5-35 were done by myself.

5.2 Materials and Methods

5.2.1 NMR spectroscopy

Spectra were collected with 1mM protein in 25mM Na_2HPO_4 (pH 6.8), 200mM NaCl, and a D_2O concentration of 90% or 100% (^{13}C-NOESY). NMR data were obtained on a Bruker 800-MHz at 25 °C. Backbone and aliphatic resonance assignments were attained using a combination of standard triple resonance experiments (Kay, 1995; Kanelis *et al.*, 2001). NMR data processing and analysis was performed using NMRPipe (Delaglio *et al.*, 1995), and spectra were analysed with Sparky (Lee *et al.*, 2009). Structure calculations were performed by CYANA 3.0 (Güntert, 2004) using automatically assigned and manually verified distance restraints from ^{15}N- and ^{13}C-NOESY experiments, and dihedral angle restraints derived by TALOS (Cornilescu *et al.*, 1999). One hundred structures were calculated, and the 20 lowest energy structures were selected and analysed. Structural quality was assessed by PROCHECK (Laskowski *et al.*, 1996).

5.2.2 ACR30-35 mutagenesis

Two complementary primers were designed to mutagenize the desired residue with 15bp of complementarity up and down-stream of the site. The orientation of the primers was such that they would amplify "away" from the site of mutagenesis, in a reaction also known as "around the world PCR." The PCR reaction was conducted using pfu DNA polymerase with 18 cycles and extension times in the range of 6-8 minutes (1kb per minute). Following the reaction, the wild type parent plasmid was digested with DpnI. The resulting product was precipitated and used to transform *E. coli*. Plasmids from antibiotic resistant colonies were isolated and sequenced to confirm the mutation. This construct was then electroporated into PA14 to assay anti-CRISPR activity.

5.2.3 Protein Purification

All proteins used here were affinity purified via 6x His tags using Ni-NTA beads. Generally, cultures of BL21 with the appropriate plasmid were grown to OD_{600}=0.5 and then induced with 1mM IPTG for 3h at 37 °C (anti-CRISPRs) or for 16h at room temperature (Csy complex). Cells were then centrifuged and resuspended in a buffer containing 20mM Tris, pH 7.5, 250mM NaCl, 5mM imidazole and 1mM PMSF was added before sonication. The resulting lysate was centrifuged at 15,000xg for 15 minutes to pellet debris and the supernatant mixed with Ni-NTA beads that had been washed in the same resuspension buffer 5 times. Binding to the beads was allowed to proceed for 1h at 4 °C, at which point the solution was passed through a column, washed ~3-5x with buffer containing 30mM imidazole and ultimately eluted in buffer containing 250mM imidazole. Colourimetric Bradford assays were conducted during the procedure to inform decisions about the number of washes and elution fractions. Purified protein was dialyzed to remove imidazole and visualized on coomassie stained SDS-PAGE gels.

5.2.4 Size exclusion chromatography

Affinity purified proteins were further purified by size exclusion chromatograph (SEC) using the S200 10/30 column with a range up to 600kDa. Fractions were collected monitored by OD_{280} and run on silver stained SDS-PAGE gels. Interactions were also assessed between purified proteins using SEC. The Csy complex and anti-CRISPR protein of interest were generally incubated together for 1h at 4 °C, and then centrifuged at 12,000rpm for 2 minutes to pellet any precipitate. The resulting supernatant was then run on SEC at room temperature. Some of the input was also kept for SDS-PAGE analysis. Fractions were collected and analyzed.

5.2.5 ACR30-35 Stoichiometry

Purified Csy complex and ACR30-35 were incubated together and fractionated on SEC as above. Fractions corresponding to the peak co-elution were run on SDS-PAGE in 2-fold serial dilutions and quantified with a variety of proportional stains; coomassie blue, Seeband (FroggaBio), and Oriole (Bio-Rad).

5.2.6 Surface Plasmon Resonance

The Csy complex was attached to a chip via the Csy3 6x His tags. ACR30-35 had its 6x his tag removed by TEV protease digestion and was flowed over the chip and the binding kinetics

determined. A heterogenous binding model fit the data best shown in this chapter, with two distinct binding events detected.

5.2.7 *In vitro* protein competition assay

Purified Csy complex was first incubated with a mutant ACR30-35 protein of interest and subsequently incubated with wild type protein. Different tags were present on two proteins to ensure they would migrate to different sizes on SDS-PAGE. The mixture was either run on SEC or subject to a Ni-NTA-mediated pull-down of the Csy complex and bound proteins. The co-elution or pull-down was run on SDS-PAGE and silver stained or immunoblotted.

5.2.8 RNAse A treatment of the Csy complex

Pancreatic RNAse A was used to treat the Csy complex for 30 minutes at 37 °C at a final RNAse A concentration of 1 mg/mL (73 μM). After digestion, the treated Csy complex was mixed with either buffer or anti-CRISPR protein and incubated on ice for 1h. This mixture was then fractionated on SEC. Samples from SEC were either run on silver stained SDS-PAGE or SYBR Gold stained TBE-Urea gels.

5.2.9 Isothermal Titration Calorimetry

Purified Csy complex was added to the temperature controlled cell at a concentration of 7.5-10μM. A DNA ligand of interest at a concentration of 75-100 μM was titrated into the cell and evolution of heat measured. The experiment was conducted at 25 °C To assess the role of anti-CRISPRs, the Csy complex was first incubated with an ~8-10 molar excess of anti-CRISPR which was then applied to the cell after the reaction had proceeded for 1h. The DNA titration was then performed.

5.2.10 Electrophoretic Mobility Shift Assay

The DNA oligonucleotide of interest was used at a final concentration of 200nM in a T4 kinase reaction with gamma-^{32}P ATP. The reaction was stopped with EDTA and GE MicroSpin G-25 columns used to remove remaining radiolabelled nucleotides. Csy-DNA binding reactions were conducted in the following binding buffer: 10mM HEPES, pH7.5, 1mM $MgCl_2$, 20mM KCl, 1mM TCEP, bromophenol blue, and 6% glycerol. The amount of Csy complex used varied due to the use of different DNA targets with varying affinities. The Csy complex was incubated with

a DNA target for 15 minutes at 37 °C and separated by a native TBE acrylamide gel. In a 'protection assay', the anti-CRISPR was first bound to the Csy complex for 1h and then mixed with DNA or for a rescue assay, the Csy-DNA complex was preformed and then anti-CRISPR added. Gels were wrapped in seran wrap and visualized with a phosphoscreen and Typhoon imager. Optimal exposures were ~2-3h.

5.2.11 Circular Dichroism

The thermal stability of wild type and mutant ACR30-35 proteins was measured using an Aviv Circular Dichroism Model 202 Spectrometer. Ellipticity of the wild type and mutant proteins were measured at a concentration of 30 μM at wavelengths between 200 nm and 260 nm. The proteins were unfolded by raising the temperature to 95 °C and subsequently refolded by lowering the temperature to 25 °C. Temperature-induced protein unfolding was assessed by heating the proteins from 5 to 95 °C in 2 °C increments with a 1 min equilibration time and a 15 second averaging time and monitoring the change in ellipticity at 220 nm. Temperature values midpoint of the unfolding transition (T_m) values were calculated as previously described (Maxwell and Davidson, 1998).

5.3 Results

5.3.1 CRISPR-Cas mechanism

CRISPR-Cas immune systems have been found in nearly half of all bacteria and ~90% of archaea (Jore *et al.*, 2012), demonstrating their importance for bacterial fitness. These widespread systems have been classified into three types (I, II, and III) as well as subtypes, such as type I-F (Makarova *et al.*, 2011b). In the type I-F system specifically, the CRISPR array is transcribed as a precursor RNA and subsequently processed within the repeats by the Cas protein, Csy4, to produce single CRISPR RNAs (crRNA), 60nt in length (Haurwitz *et al.*, 2010; Sternberg *et al.*, 2012). The mature Csy complex is then formed by the addition of the Csy1, 2, and 3 proteins to the Csy4-bound crRNA with a final stoichiometry of $Csy1_1$:$Csy2_1$:$Csy3_6$:$Csy4_1$:$crRNA_1$ (Wiedenheft *et al.*, 2011b; van Duijn *et al.*, 2012). The 32nt spacer region within the 60nt crRNA serves as a guide for the Csy complex to identify and bind complementary regions on phage genomes, known as protospacers (Figure 48). By inference from work done on the Type I-E system of *E. coli*, binding of target DNA then likely leads to the

recruitment of the nuclease/helicase protein Cas3 and subsequent phage genome cleavage (Westra *et al.*, 2012b; Hochstrasser *et al.*, 2014).

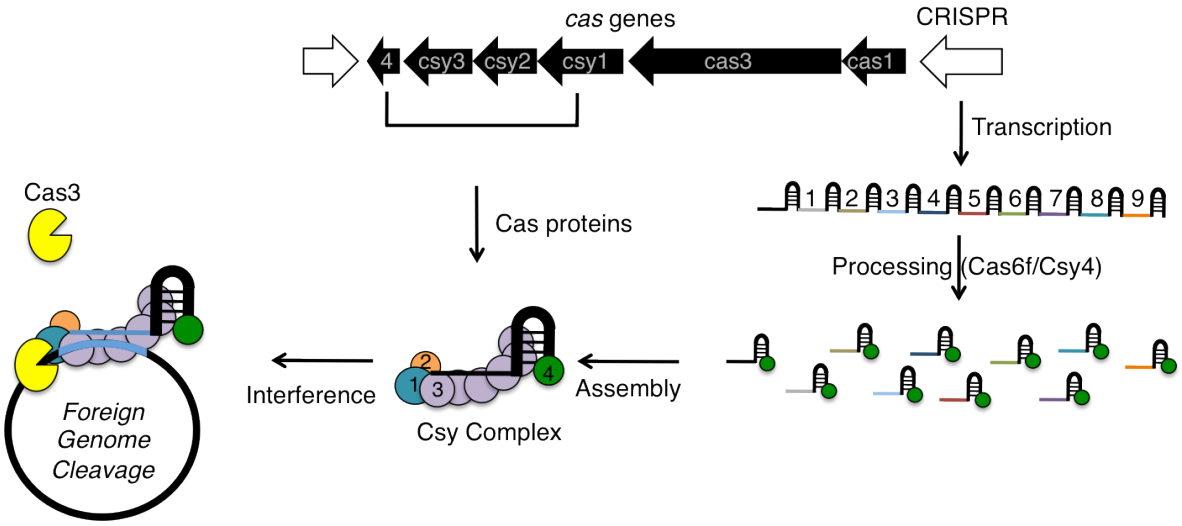


Figure 48: Schematic of the Type I-F CRISPR-Cas system mechanism

See text for description

Together with April Pawluk in the lab, we have reported that phages infecting *Pseudomonas aeruginosa* possess anti-CRISPR genes that specifically inactivate either type I-F or I-E CRISPR-Cas function (Bondy-Denomy *et al.*, 2013; Pawluk *et al.*, 2014). Five distinct anti-CRISPRs with activity against the type I-F system were discovered, along with four distinct type I-E anti-CRISPRs in the same genomic locus of related phages (Bondy-Denomy *et al.*, 2013; Pawluk *et al.*, 2014).

5.3.2 Structure of ACR30-35

Notably, the nine distinct anti-CRISPRs identified in these two papers share no obvious sequence similarity to each other or to any other proteins in the database making their evolutionary origins and mode of action unclear. pBLAST, psi-BLAST and hhPred searches yielded very little functional insight. Homologs of the anti-CRISPR proteins were found mostly in prophages and other mobile elements with no hits to proteins whose functions have been previously studied. Furthermore, secondary structure predictions (jPred3) did not present a common topology between anti-CRISPRs, nor are any common motifs identifiable in these small and novel proteins. Thus, to determine the mode of action for anti-CRISPR proteins (specifically Type I-F), I undertook structural and functional studies.

The type I-F anti-CRISPR protein from phage JBD30, gene *35* (ACR30-35), was previously shown to be necessary and sufficient for successful phage infection in the presence of an active CRISPR-Cas system (Bondy-Denomy *et al.*, 2013). To gain insight into the mechanism of ACR30-35, the NMR solution structure was solved by Dr. Karen Maxwell, revealing a similar tertiary structure to an RNA-recognition motif (RRM) (Figure 49).

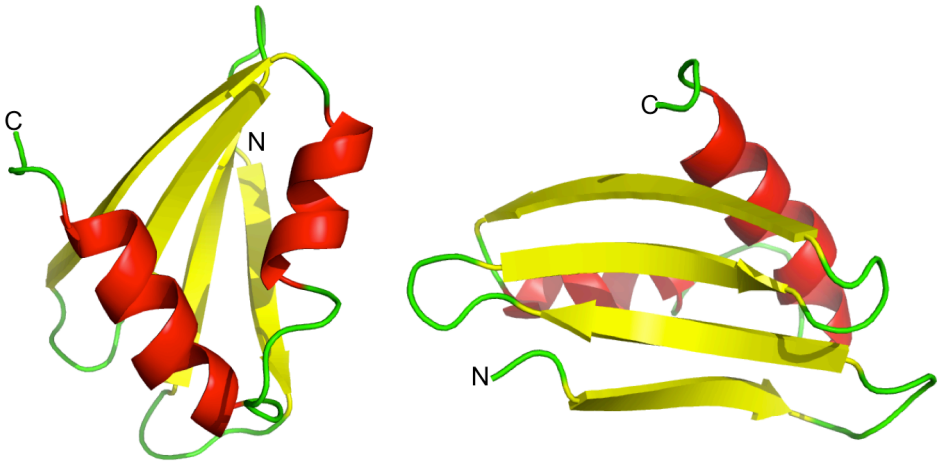


Figure 49: NMR Solution structure of ACR30-35

The structure was solved by Dr. Karen Maxwell. The same structure is shown here at two different angles, showing both the helical face and the β-sheet.

Searches to find proteins with a similar fold using the Dali server (Holm *et al.*, 2006) revealed many hits across the β-sheet. Only one hit was to a protein of comparable size and orientation, a member of the signal recognition particle, Srp14 from *Schizosaccharomyces pombe* (Brooks *et al.*, 2009). This protein is a homo-dimer and interacts with protein and RNA in the macromolecular signal recognition particle. The kind of tertiary structure exhibited by ACR30-35 is also very common among many of the Cas proteins (Haurwitz *et al.*, 2010; Samai *et al.*, 2010; Makarova *et al.*, 2011a) (Figure 50).

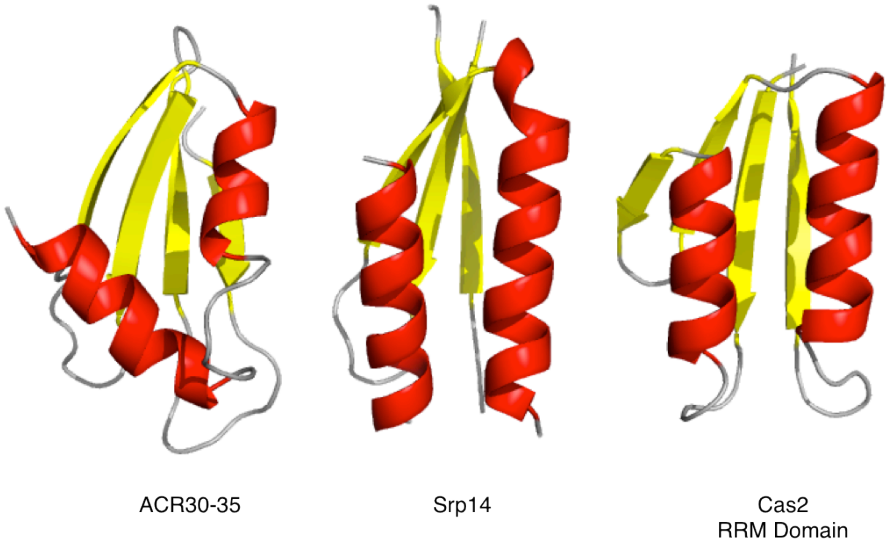


Figure 50: ACR30-35 contains a tertiary structure similar to an RRM

The structure of ACR30-35 is shown along with the structures of two RNA-binding proteins that display similar tertiary structures. The secondary structure topologies of these proteins are different despite their similar tertiary structures.

An explanation for the evolution of anti-CRISPRs could be that they arose originally as Cas proteins but have acquired mutations, which led to the formation of a dominant negative inhibitor of the CRISPR-Cas process. However, this is not obviously the case for ACR30-35, as the topology for the Cas proteins (and Srp14) are all distinct from that of ACR30-35. Thus these similarities may facilitate mechanistic inferences, but do not imply homology.

5.3.3 ACR30-35 requires a single functional interface for *in vivo* function

To determine how ACR30-35 inhibits the CRISPR-Cas immune system, ACR30-35 residues which were exposed in the structure were mutagenized to alanine and screened in *P. aeruginosa* strain PA14. Phage DMS3*m* is a CRISPR-sensitive phage that cannot replicate in wild type (wt) PA14 due to the activity of the CRISPR-Cas system, but with the expression of wtACR30-35 from a plasmid, DMS3*m* is able to replicate (Bondy-Denomy *et al.*, 2013). Two mutants, F3A and Y6A, had no anti-CRISPR activity, while the N16A, Y20A, and E31A mutants had attenuated activity (~100-fold reduction in plaquing), and mutants H11A and L12A displayed subtle changes (~10-fold) in activity (Figure 51).

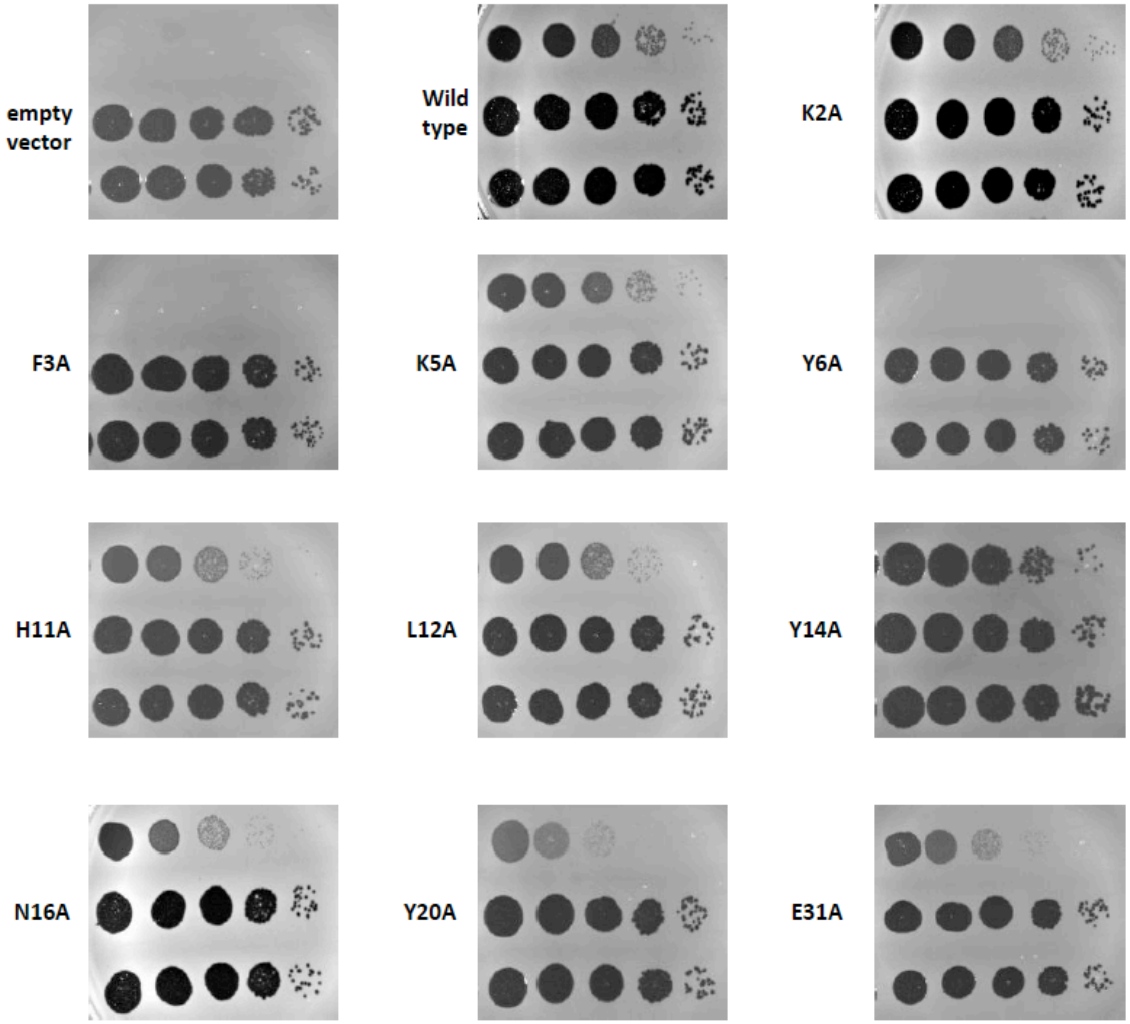


Figure 51: ACR30-35 mutagenesis reveals essential residues for *in vivo* activity

Exposed residues of ACR30-35 were mutated and screened *in vivo* for their ability to inactivate the CRISPR-Cas system and allow CRISPR-sensitive phage DMS3*m* to replicate on PA14. The mutagenized genes were expressed from a plasmid in wtPA14 and a standard plaque assay conducted. Note that mutations were introduced by myself, a rotation student who I supervised (Hayoung Yoo), and a technician in the lab, Yurima Hidalgo-Reyes. YHR made this figure.

All other mutants tested (35 in total) displayed activity similar to wtACR30-35. Notably, no essential amino acids in the C-terminal α-helical region of the protein were identified. All of the mutants with the strongest reduction in activity (Y6A) and intermediate levels (Y20A, N16A, and E31A) mapped to one face on the protein, putatively implicating an important interface for

in vivo interactions which facilitate anti-CRISPR function (Figure 52). The F3A mutation abrogated detectable protein expression, so this mutant was not worked with further.

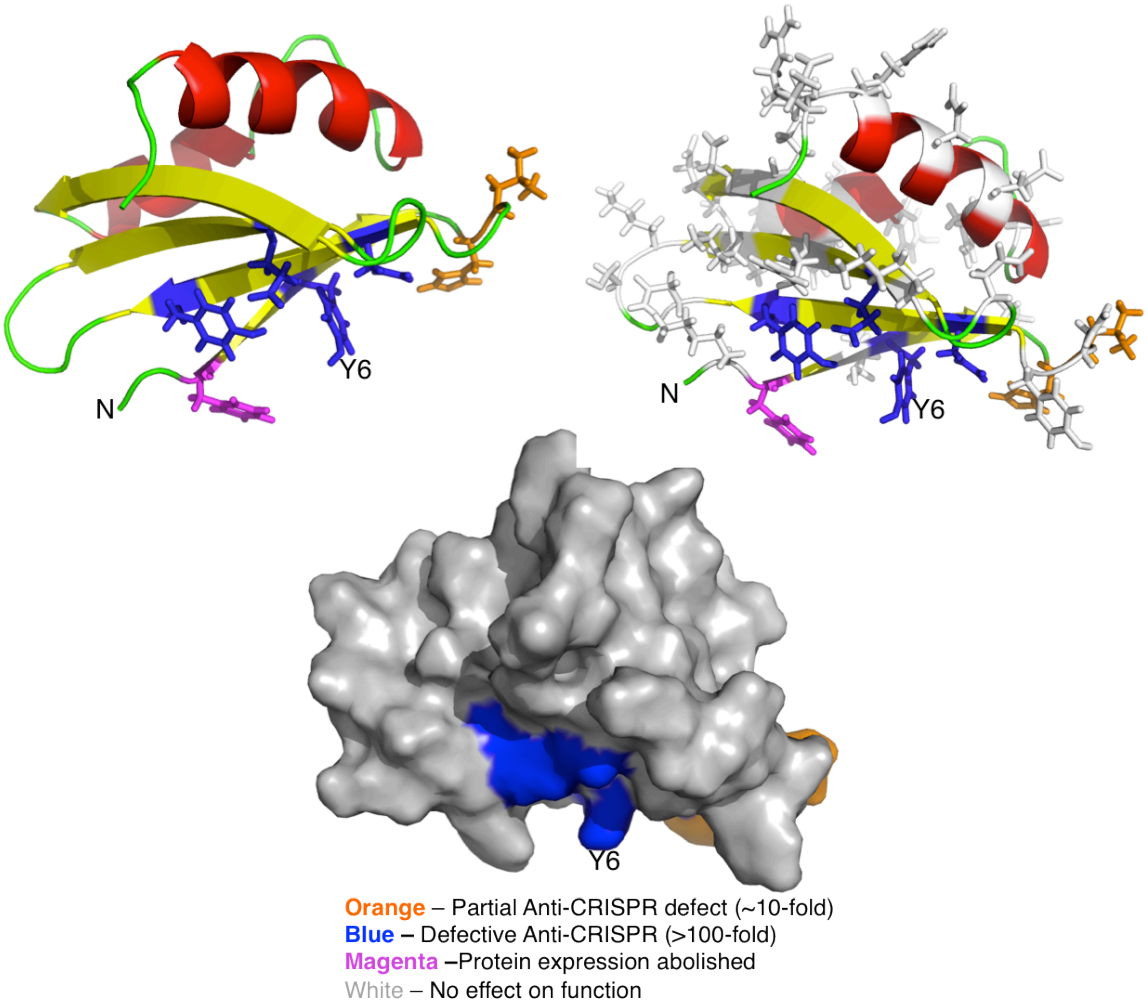


Figure 52: ACR30-35 structure with functional epitope highlighted

Mutated residues described above are shown on the solution structure of ACR30-35. All side chains shown were mutagenized to alanine. A small epitope, characterized by tyrosine 6 (Y6) was identified which is important for *in vivo* anti-CRISPR activity

5.3.4 ACR30-35 interacts with the Csy Complex

I have previously shown that the presence of active ACR30-35 did not influence the expression of cas genes *csy3* and *cas3*, nor did it affect the accumulation of mature crRNAs in the cell (Bondy-Denomy *et al.*, 2013). Given this, I predicted that ACR30-35 could be interacting directly with the mature Csy complex. The *P. aeruginosa* PA14 Csy complex consisting of Csy1-4 and a mature crRNA was overexpressed and affinity purified from *E. coli* as described

previously (Wiedenheft *et al.*, 2011b). The type I-F anti-CRISPRs ACR30-35 and ACR5-35 were independently expressed and purified as well as a negative control, the type I-E anti-CRISPR ACR88a-32. The Csy complex (350kDa) was incubated with ACR30-35 (10.5kDa), ACR5-35 (17kDa), or ACR88a-32 (10.5kDa) and analyzed via size exclusion chromatography (SEC). ACR30-35 formed a stable interaction with the Csy complex in this assay, migrating in the same fraction as the Csy complex, while ACR5-35 and ACR88a-32 did not (Figure 53).

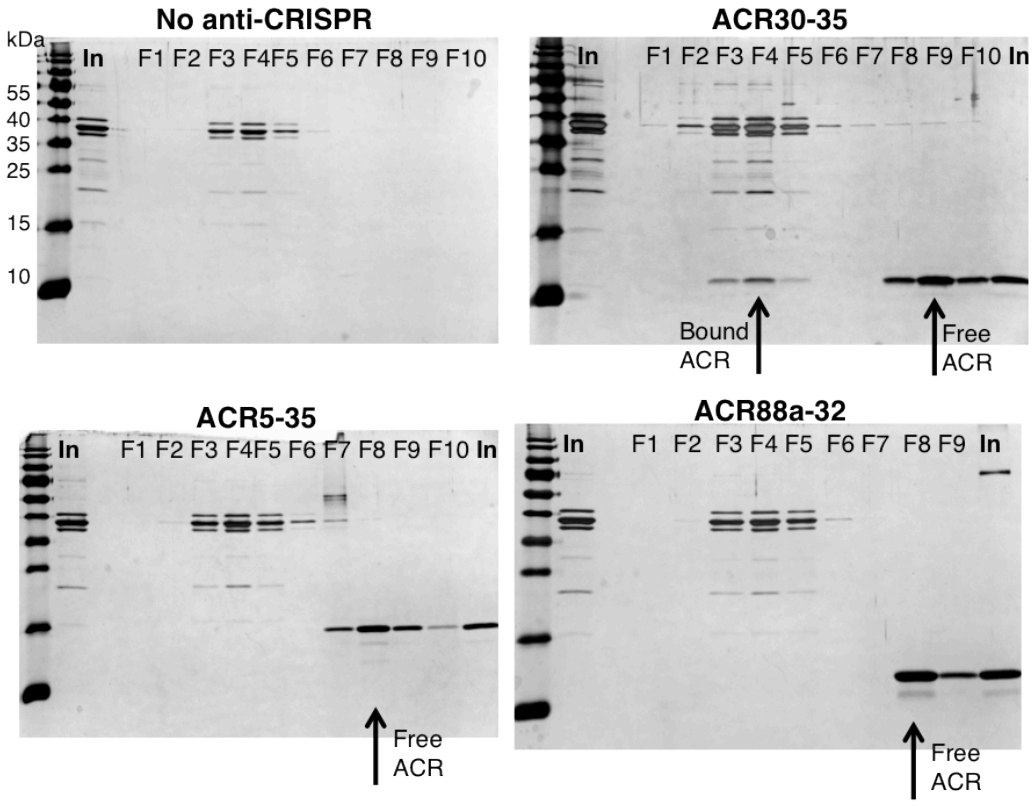


Figure 53: ACR30-35 interacts with the Csy complex

Purified Csy complex was applied to SEC alone or after incubation with purified ACR30-35, ACR5-35, or Type I-E ACR88a-32. Input (In) for each purified component and collected fractions (F1, F2, etc.) were run on SDS-PAGE and silver stained to identify protein. While ACR30-35 showed a robust interaction, co-eluting in the same fraction as the Csy complex, the other two proteins did not. Each anti-CRISPR protein was at an ~10 fold molar excess relative to the Csy complex.

SDS-PAGE coupled with Coomassie staining of Csy complex/ACR30-35 co-elution fractions revealed the stoichiometry of this interaction to be in the range of 2.5-3.5 ACR30-35 per Csy complex. This quantification was done by myself and Dr. Bianca Garcia, a research associate in the lab. Together we tested many different tags on ACR30-35 as well conditions and stains, to ensure accurate quantification of the stoichiometry. ACR30-35 with FLAG, 6x his, or 3x HA

tags were each tested in stoichiometry measurements and confirmed to be active *in vivo*. A representative gel is shown in Figure 54.

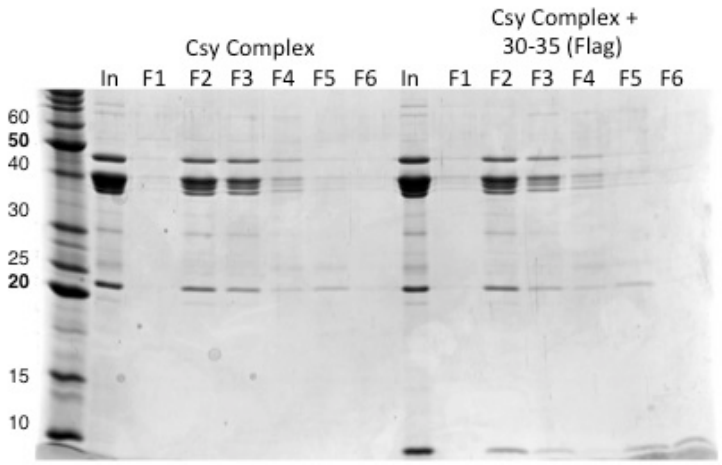


Figure 54: Three molecules of ACR30-35 interact with the Csy complex

The Csy complex was run on SEC with or without ACR30-35. The input (In) and fractions (F1-F6) are shown on coomassie stained SDS-PAGE. Quantification of the ACR30-35 bands and Csy proteins demonstrated a stoichiometry of $_3$ACR30-35:$_1$CsyComplex.

Interestingly, ACR30-35 is a monomer in solution, which suggests that there are either three similar binding sites on the Csy complex or that the protein trimerizes upon binding to the Csy complex. Given the small epitope identified above, I favour the former hypothesis, since one might expect more than one important epitope if the latter hypothesis were the case (i.e. one Csy complex interaction interface and a distinct oligomerization interface).

Through collaboration with Dr. Blake Wiedenheft and his research associate MaryClare Rollins (Montana State University), we were able to understand the kinetics of ACR30-35 binding to the Csy complex. Using surface plasmon resonance (SPR), they determined that when ACR30-35 is flowed into a chip containing Csy complex, they detect a K_D=3 x 10^{-12} M. Using this method, they also detected a binding stoichiometry of ~2.6 as well as two distinct binding events, suggesting cooperativity. Two important functional observations come out of these data. First, after the binding of ACR30-35 to the Csy complex, the off rate is very slow (K_d=8x10^{-8} s^{-1}). This suggests that ACR30-35 inhibits CRISPR-Cas function by staying tightly bound as opposed to inactivating the complex by an enzymatic mechanism. Secondly, they had previously determined that the affinity of the Csy complex for a dsDNA target is K_D= 5 x 10^{-10} M, which is

approximately two orders of magnitude weaker than the ACR30-35 interaction. These data are summarized in Figure 54, a figure made by MaryClare Rollins.

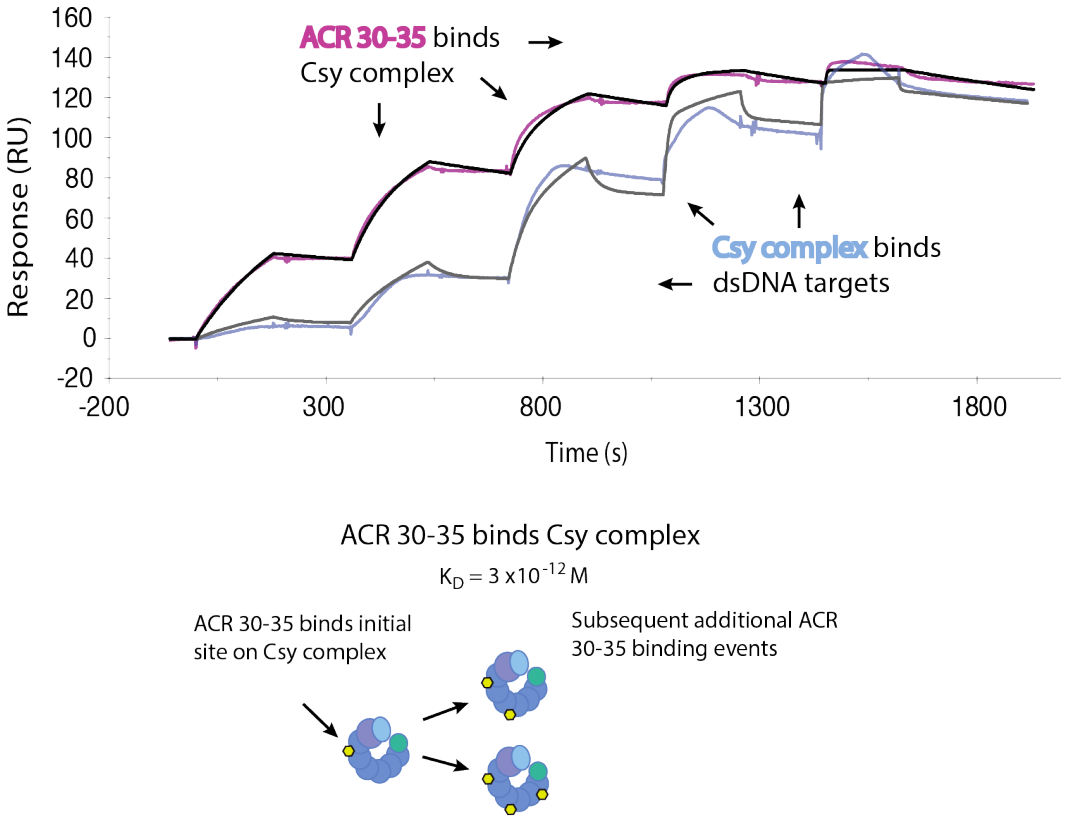


Figure 54: ACR30-35 binds to the Csy complex with an affinity of ~3pM

(Experiments and figure by MaryClare Rollins.)
Surface plasmon resonance experiments were conducted by linking the Csy complex to a chip and assessing the binding of ACR30-35. Multiple ACR30-35 binding events were detected and it bound with a faster on-rate and a slower off-rate compared to a dsDNA target. Two distinct binding events were also identified.

Given the insight that ACR30-35 binds tightly to its target, I then wanted to investigate the mutants of ACR30-35 identified *in vivo* to determine whether these were weak or defective interactors. The ACR30-35 mutants identified above still interacted with the Csy complex by SEC and pull down assays with purified protein. To determine whether these mutants displayed weaker binding to the Csy complex than wtACR30-35, a competition assay with wtACR30-35 protein was set-up. These experiments were conducted by Dr. Bianca Garcia in the Davidson lab. A FLAG-tagged mutant ACR30-35 was prebound to the Csy complex and then competed with 3xHA-tagged wtACR30-35. As a control experiment, Dr. Garcia showed that FLAG-tagged wtACR30-35 prebound to the Csy complex did not dissociate in 1h in the presence of wtACR30-

35(3xHA), and very little dissociation was seen after 16h. These results were consistent with SPR experiments, demonstrating a very slow off-rate (Figure 55).

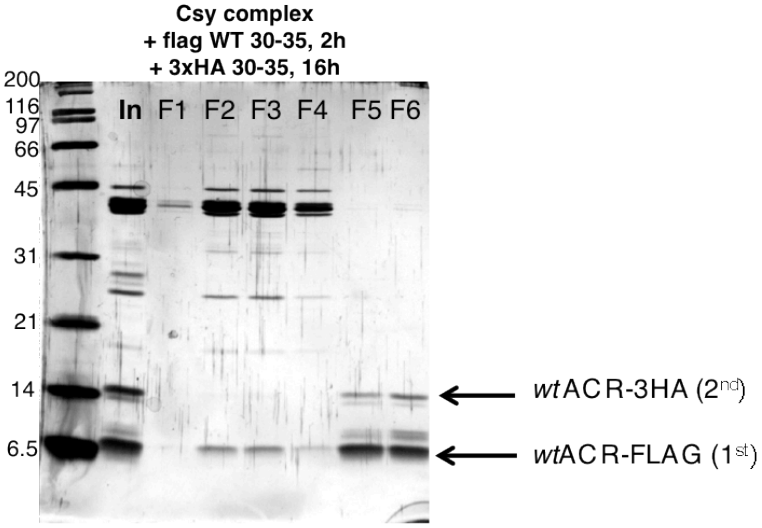


Figure 55: ACR30-35 has a does not dissociate when left for 16 hours

(Experiments by Dr. Bianca Garcia.)
The Csy complex was first incubated with wtACR30-35 (FLAG) for 2h and then wtACR30-35 (3xHA) was added and incubated for 16 hours. Migration of the protein mixture over SEC and subsequent SDS-PAGE with silver staining revealed very little binding of wtACR30-35 (3xHA) which indicated that the FLAG-tagged version stayed bound overnight, in the presence of the competitor.

FLAG tagged mutant proteins with intermediate function *in vivo* (L12A, Y20A, E31A) reached an ~50:50 equilibrium when bound to the Csy complex first and then competed with wtACR30-35 (3xHA), suggesting a weaker interaction than wtACR30-35. The completely defective Y6A mutant was fully out-competed by wtACR30-35 (3xHA) (Figure 56).

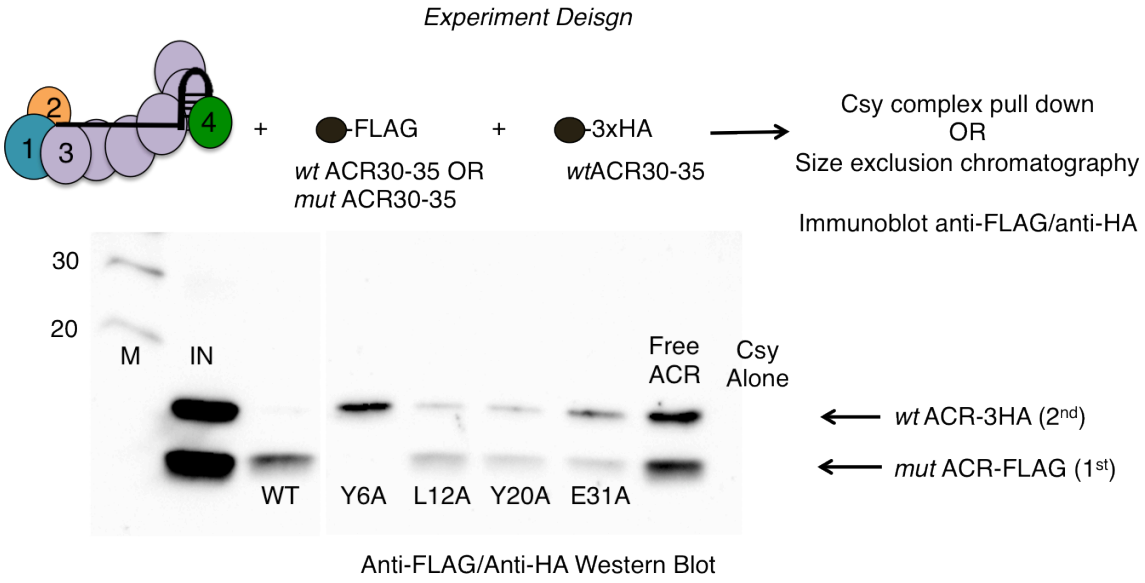


Figure 56: ACR30-35 with a Y6A mutation is outcompeted by wtACR30-35

An indicated mutant ACR30-35 (FLAG) protein was bound to the Csy complex for 1h and then competed with wtACR30-35 (3xHA) for 1h. This mixture was run over SEC and the peak Csy fraction from each sample was run on SDS-PAGE and a western blot conducted to identify both anti-CRISPR proteins. The approximate ratio of *wt* to *mut* ACR30-35 indicates the degree to which the mutant protein was out-competed. Experiments by Dr. Bianca Garcia.

Circular dichroism (CD) studies were performed on wtACR30-35 and the indicated mutants to determine whether the attenuation of *in vivo* function was a result of gross alterations to the structure of the protein. Wild type ACR30-35 and all mutants had similar far UV CD spectra, indicating an overall similar secondary structure between them (data not shown). Wild type ACR30-35 and mutant proteins also behaved in a similar fashion during thermal unfolding with the exception of the Y20A mutation which was not reversible (data not shown). Temperature-induced unfolding experiments comparing wtACR30-35 and the mutants, showed that Y20A was significantly destabilized with T_m values between 51-55 °C for wtACR30-35 and 28 °C for Y20A. E31A was also destabilized by ~5-10 °C with T_m value of 45 °C while all other mutant examined were similar to wtACR30-35 (Figure 57). Most notably, the substitution of residue Y6 did not have a significant effect on protein stability, demonstrating that the inability to function *in vivo* or bind tightly to the Csy complex *in vitro* were not due to the misfolding or decreased stability of this protein. CD studies were conducted by Dr. Garcia in the Davidson lab.

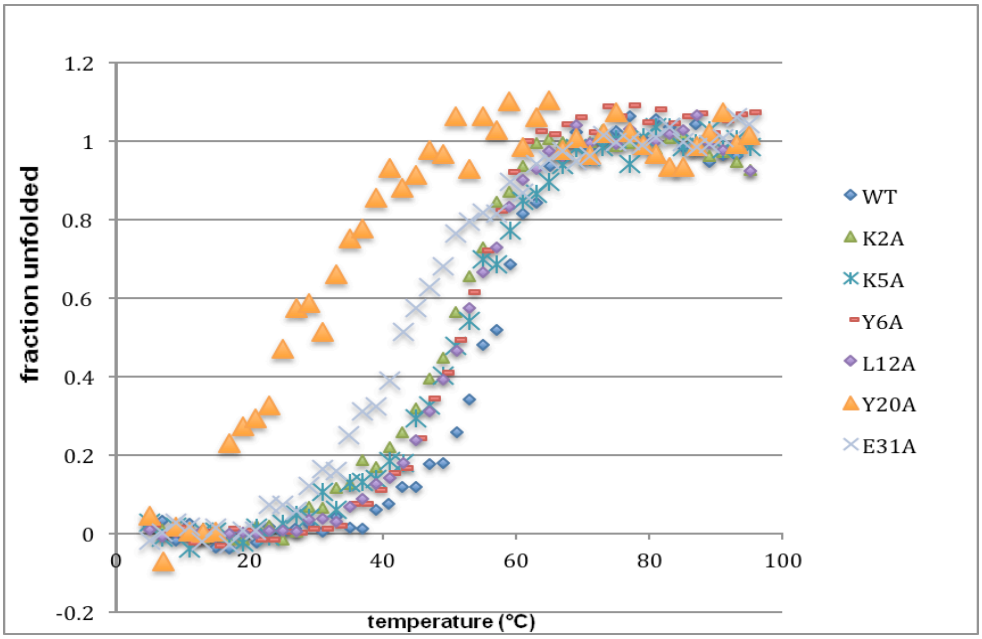


Figure 57: Temperature-induced unfolding of ACR30-35 mutants

Purified ACR30-35 mutant proteins were subjected to protein melts using circular dichroism. Experiments conducted by Dr. Bianca Garcia. Note that K2A and K5A have wild type anti-CRISPR activity and were included as controls.

5.3.5 ACR30-35 interacts with the hexameric Csy complex backbone protein, Csy3

To identify the Csy complex member that interacts with ACR30-35, I first conducted an RNAse A treatment of purified Csy complex to degrade the crRNA and determine whether the ACR30-35 interaction was maintained. The treatment of purified Csy complex with up to 1 mg/mL RNAse A (73 µM) and subsequent fractionation by SEC revealed partial degradation of the crRNA (Figure 58a) and the complete displacement of crRNA-associated Csy4 (Figure 58b).

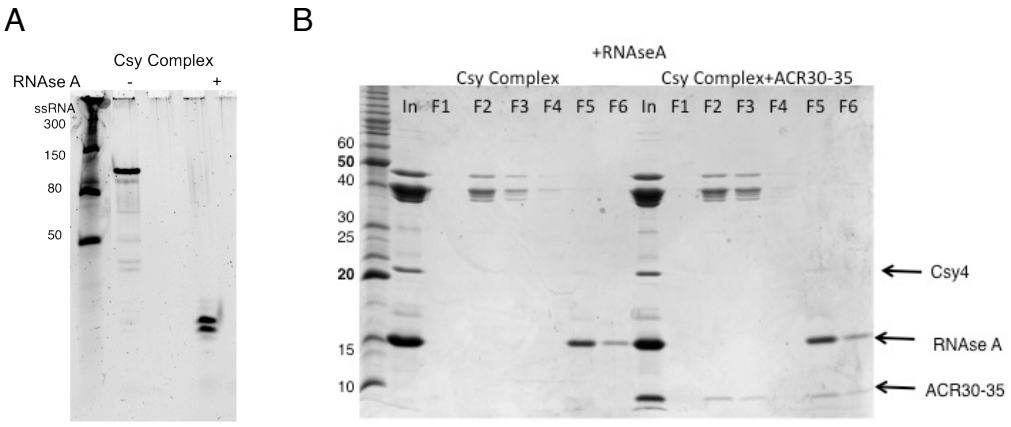


Figure 58: ACR30-35 interacts with RNAse A treated Csy complex

(A) The Csy complex was pretreated with 1 mg/mL RNAse A and fractionated by SEC. The RNA from the peak Csy complex fraction was visualized on a TBE-Urea gel and stained with SYBR gold, revealing some protected RNA with an approximate size of 20 nt, (B) the RNAse A treated Csy complex was incubated with ACR30-35 and fractionated by SEC. Input (In) and fractions (F1-F6) are shown on coomassie stained SDS-PAGE. Note the absence of Csy4 in the F2 fractions and the presence of ACR30-35 the same fraction.

This result was quite surprising, as I expected complete degradation of the crRNA and subsequent disassembly of the complex. Interestingly, the stable "pseudo-complex" consisting of a partial crRNA species, Csy1, 2, and 3 still interacted with ACR30-35. Given the stoichiometry of the ACR30-35 interaction (3 units), it seemed that the hexameric Csy3 subunit could be the binding target. This protein was overexpressed and purified and then assessed for ACR30-35 binding. While Csy3 forms a hexamer in the Csy complex, the purified protein eluted in both monomeric and oligomeric fractions on SEC. Calculations based on standard curves place the oligomeric fraction at an ~11-12-mer. ACR30-35 showed a preferential interaction with the oligomeric Csy3 fraction, as opposed to the monomeric fraction (Figure 59a). In contrast, negative control protein, ACR5-35, did not interact with Csy3. The selectivity that ACR30-35 showed for oligomeric Csy3 may indicate that monomeric Csy3 does not have the appropriate binding site for ACR30-35 but that the junctions between the Csy3 proteins generate a binding site (see model in Figure 59b).

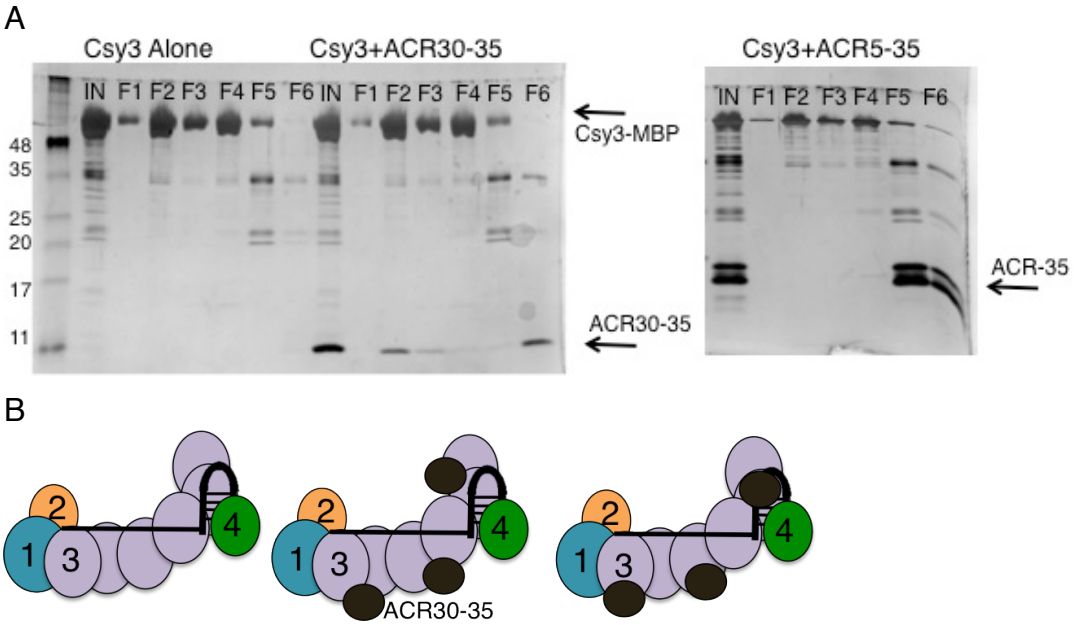


Figure 59: ACR30-35 interacts with purified oligomeric Csy3

(A) Purified Csy3 was fractionated by SEC with or without ACR30-35 or ACR5-35. Oligomeric Csy3 ran at 'F2' while monomeric ran at 'F4.' Input and fractions were run on SDS-PAGE and silver stained, (B) a hypothetical model showing two possibilities for ACR30-35 binding to the Csy complex via Csy3; either at an interface only available at Csy3 junctions or on an interface only present on alternating Csy3 proteins when it exists as an oligomer.

Taken together, ACR30-35 uses a single epitope to bind tightly to the Csy complex via the Csy3 backbone protein, suggesting that the sole mechanism of ACR30-35 lies in the ability to bind and stay bound to the Csy complex. This is contrasted with an alternate situation where a protein would have both an "interaction" interface/residue and a "functional" interface/residue. I did not identify any mutants that bound to the Csy complex as well as wtACR30-35 but did not function to inhibit it. I hypothesize that this binding event is sufficient to inactivate the complex, and the experiments below outline the testing of this hypothesis.

5.3.6 ACR30-35 inhibits target recognition by the Csy complex

The function of the CRISPR-Cas immune system relies on the identification and binding of an invading DNA target (i.e. a phage genome bearing a protospacer). Isothermal titration calorimetry (ITC) was conducted (with the assistance of Scott Strum) to measure the interactions between the Csy complex and target DNA in the presence or absence of ACR30-35. An eight nucleotide ssDNA target was used which represents the important seed region (nucleotide positions 1-6, 8 of the spacer) bound to the Csy complex with a K_d value of 90nM (in the

absence of ACR30-35), which is consistent with previous work (Wiedenheft *et al.*, 2011b). No binding was detected with non-target DNA. Preincubation of the Csy complex with ACR30-35, but not ACR5-35, completely eliminated all detectable binding between the Csy complex and its target DNA (Figure 60).

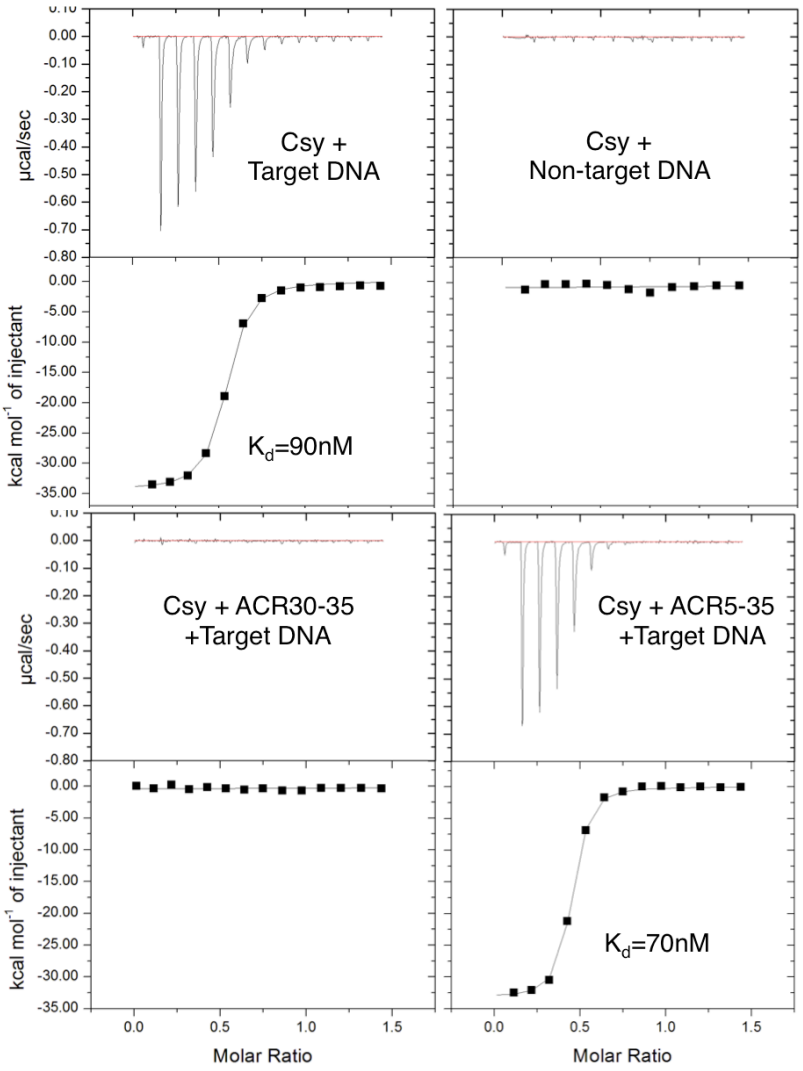


Figure 60: ACR30-35 inhibits the recognition of the targeted seed region

Isothermal titration calorimetry experiments measure the heat evolved when an 8nt ssDNA ligand (target or non-target) is titrated into a cell containing purified Csy complex with or without ACR30-35 or ACR5-35. The K_d value for two interactions is shown in the figure.

The complete absence of any detectable interaction in ITC suggested that the DNA binding site is inaccessible when ACR30-35 is bound. In order to assay numerous different ss and dsDNA targets of varying lengths and positions on the crRNA, I turned to electrophoretic mobility shift assays (EMSAs). This experimental set-up allowed me to 'scan' along the 32nt crRNA and

assess whether any area of the crRNA was accessible to DNA binding or if ACR30-35 completely blocks this whole region (see Figure 61 for a schematic and list of all targets tested).

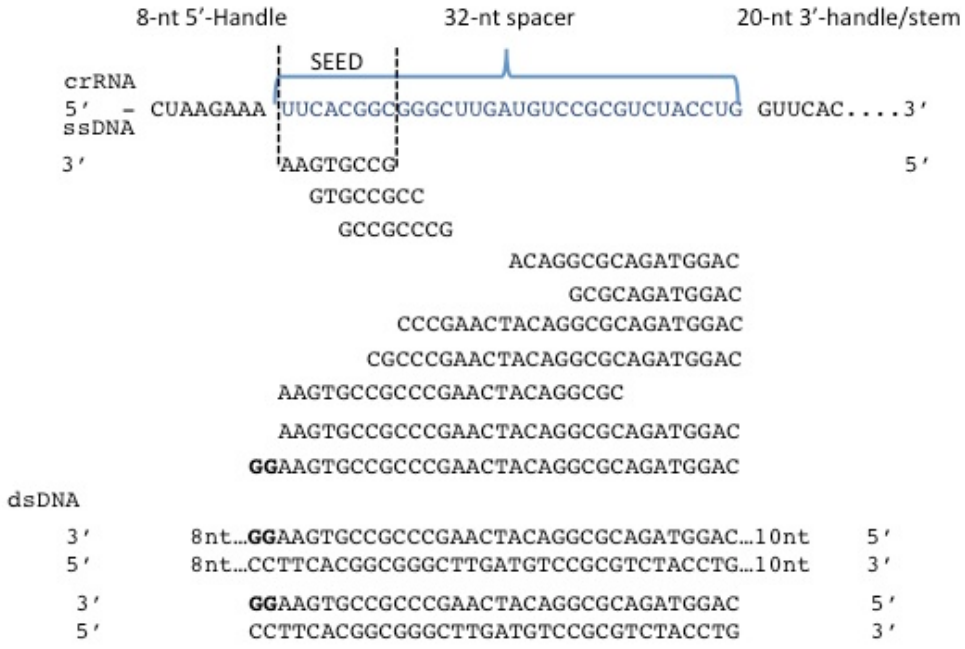


Figure 61: Schematic representation of DNA targets tested in EMSA experiments

The 60nt crRNA is comprised of 32nt of spacer-derived sequence (shown in blue) with 28nt of repeat-derived sequence flanking the spacer. The seed region is highlighted, which must be complementary to the target for *in vivo* function. The protospacer adjacent motif (PAM) is shown in bold, which is also required *in vivo*. Ten ssDNA targets and two dsDNA targets were tested. All DNA targets shown were 5' end-labeled with ^{32}P and were bound by the Csy complex in EMSA experiments shown in Figure 62.

The Csy complex bound DNA targets with complementarity to the crRNA but incubation of the Csy complex with ACR30-35 (but not ACR5-35), completely inhibited the ability of the Csy complex to interact with all DNA targets shown in Figure 61, demonstrating that the entire DNA binding site (i.e. the crRNA) is occluded by this protein. A representative gel with a ssDNA target of 32bp and a dsDNA target of 50bp is shown in Figure 62.

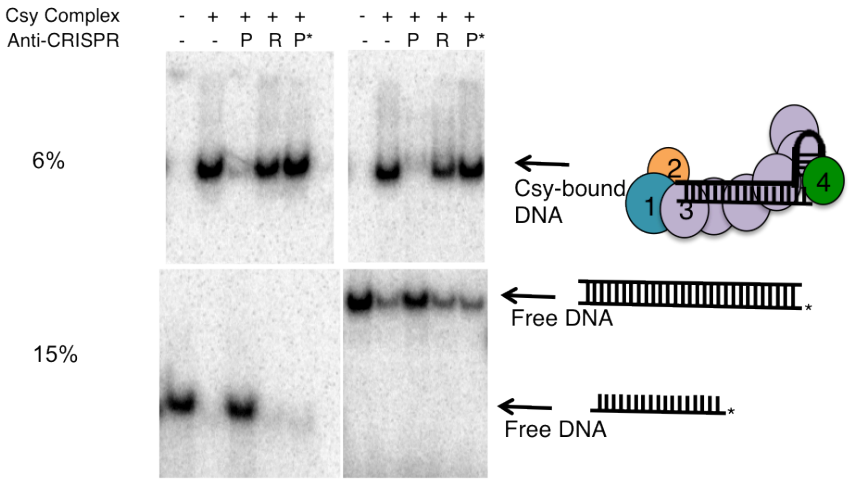


Figure 62: ACR30-35 prevents target recognition by the Csy complex.

The Csy complex was incubated with a radiolabeled DNA target in a DNA-binding buffer. This mixture was then separated by a native TBE-acrylamide gel and imaged. The Csy complex was first incubated with ACR30-35 to determine whether the anti-CRISPR can "protect" (P) DNA from being bound. The negative control ACR5-35 was also tested (P*). Conversely, the Csy complex was prebound to DNA and then ACR30-35 added to determined if DNA could be rescued (R). 6% gels were better for imaging the bound DNA due to the large size of the Csy complex, while 15% gels were used to visualized the unbound DNA. The same sample was run in parallel on both gels. The shift of the band to the upper panels indicates binding by the Csy complex.

ACR30-35 mutant Y6A was unable to inhibit the DNA-binding activity of the Csy complex, while partially active mutants could inhibit DNA binding only at ~3-fold higher concentrations than used for wtACR30-35 (Figure 63). Y6A mutant protein up to 1000-fold excess relative to the Csy complex was unable to inhibit any DNA binding at all.

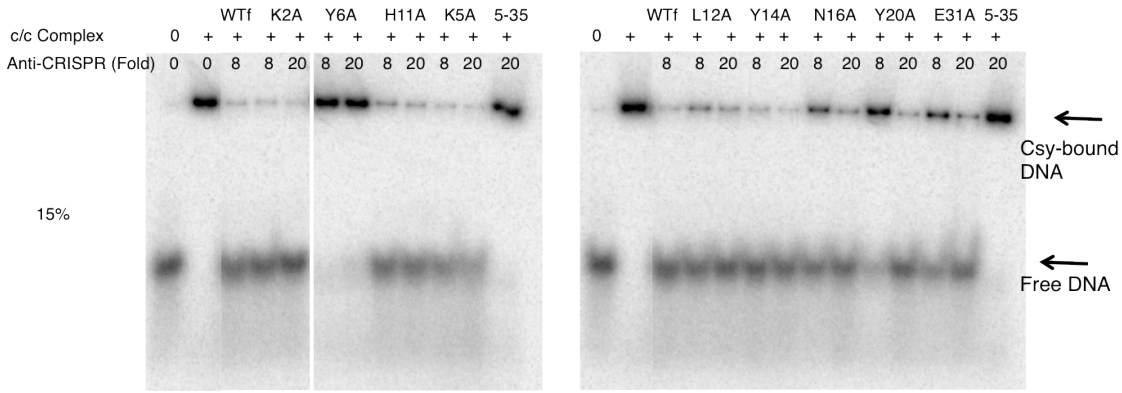


Figure 63: Mutant ACR30-35 proteins are defective in protecting the DNA target

Csy complex was incubated with indicated ACR30-35 mutants at an 8 or 20-fold molar excess relative to the Csy complex or negative control ACR5-35 at a 20-fold molar excess. A 32nt ssDNA target was then added and the mixture separated on a 15% native TBE-acrylamide gel. Free DNA indicates a functioning anti-CRISPR (i.e. wt, K2A) while bound DNA indicates a poorly functioning anti-CRISPR (i.e. Y6A). Some mutants possess intermediate activity and inhibit DNA binding activity only at a 20 fold molar excess (i.e. N16A, Y20A, E31A).

Together, EMSAs and ITC experiments demonstrate that the Csy complex bound with ACR30-35 is unable to detect and/or bind target DNA over the entire length of the crRNA. Broadly speaking, this demonstrates that ACR30-35 either occludes this entire crRNA through a steric hindrance of the DNA binding site or by modifying the complex in such a way that the DNA binding site is no longer accessible.

To differentiate between these two mechanisms, the order of the EMSA experiment was reversed. If the ACR30-35 mechanism were steric, then DNA and ACR30-35 would share the same binding site. Thus, if DNA were bound first to the Csy complex, the ACR30-35 binding site would then be occluded. When a 50bp dsDNA target was first bound to the Csy complex, it was not displaced by then adding ACR30-35. This would be an analogous process to "rescuing" phage DNA in the cell (Figure 62, see 'R' lane, Figure 64).

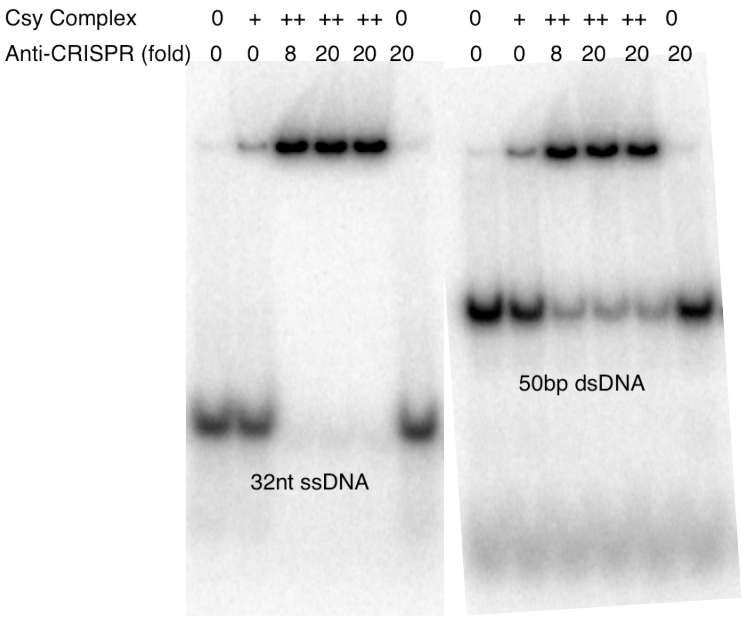


Figure 64: ACR30-35 does not rescue DNA prebound to the Csy complex

The Csy complex was incubated with targeted 32nt ssDNA or 50bp dsDNA and then ACR30-35 was added at the indicated molar excess. Csy complex is at 10nM (+) or 100nM (++), while the DNA target is <1nM. Csy complex was omitted from the last lane to shown that ACR30-35 does not interact with DNA.

Since the target DNA remained bound to the Csy complex in the presence of ACR30-35, I then determined whether the ACR30-35 binding site was being occluded or not, in the presence of DNA. The Csy complex was incubated first with a 50 bp dsDNA target. After this binding reaction was complete, the mixture was fractionated on SEC (Figure 65a, panel ‘i’) to reveal the Csy complex eluting in the peak fraction (F2). This same fraction was separated on a denaturing acrylamide gel and stained for nucleic acid, with or without pretreatment with RNAse A (Figure 65b, panel ‘i’). Note the presence of RNAse A-resistant 50bp dsDNA co-eluting with the Csy complex. Next, after binding the Csy complex with dsDNA, ACR30-35 was added and the entire mixture was fractionated by SEC. The inability of ACR30-35 to “rescue” bound DNA was not due to an absent ACR30-35 binding site, as ACR30-35 bound and co-eluted with the Csy-DNA complex as a single fraction (F2) containing co-eluted Csy complex, bound ACR30-35 (Figure 65a, panel ‘ii’) and bound target DNA (Figure 65b, panel ‘ii’). Csy complex exposed to non-target DNA also bound ACR30-35 (Figure 65a, panel ‘iii’) and no DNA co-eluted in the peak Csy complex fraction (Figure 65b, panel ‘iii’). Note that unbound DNA and ACR30-35 elute later in the fractionation (fractions not shown).

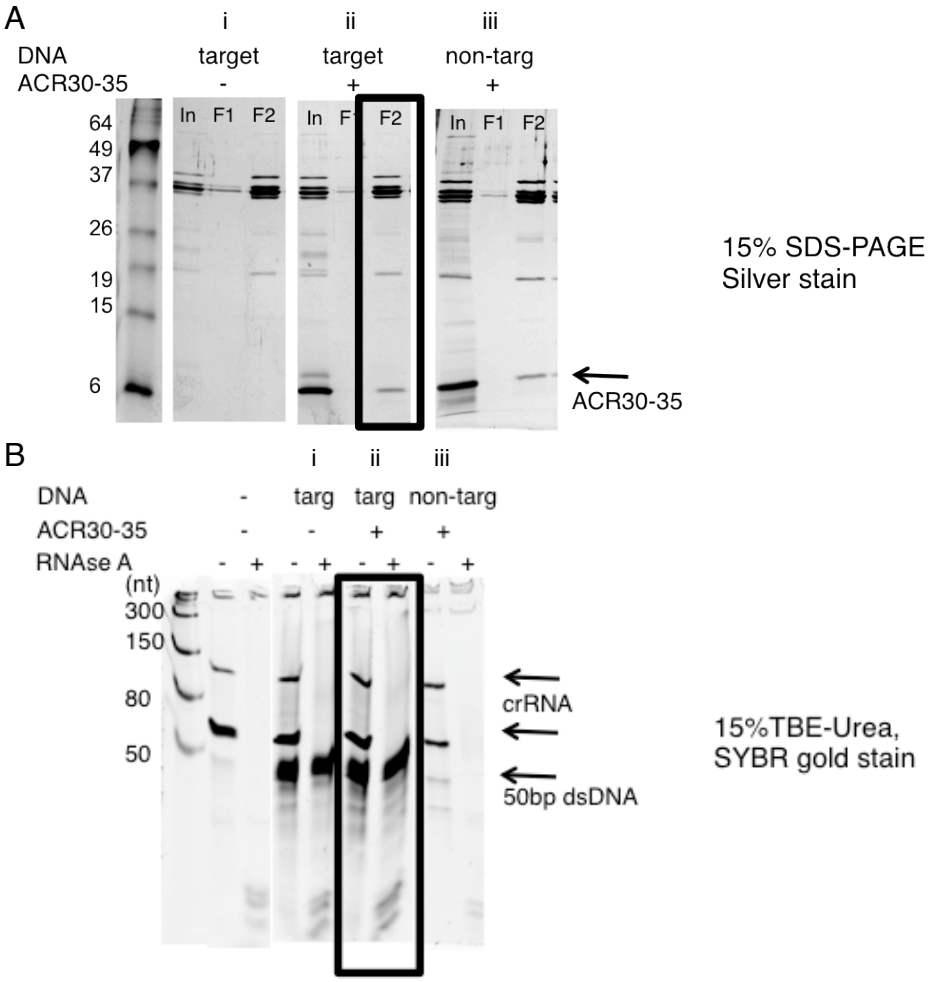


Figure 65: ACR30-35 can form a ternary complex with Csy-DNA

The Csy complex incubated with an excess of target (panel i, ii) or non-target (panel iii) 50bp dsDNA and then incubated with ACR30-35 (panel ii, iii). The mixture was separated via SEC with the input and peak fraction (F2) run on an (A) SDS-PAGE silver stained gel, or (B) on a TBE-Urea acrylamide gel and stained for nucleic acid with SYBR Gold with or without RNAse A treatment after SEC, to delineate crRNA from bound DNA. The boxed in lanes demonstrate a single sample in which the Csy complex, a dsDNA target and ACR30-35 co-eluted.

In conclusion, ACR30-35 could simultaneously bind the Csy complex with 50bp dsDNA previously bound. Similar results were obtained with 32nt of ssDNA prebound to the Csy complex. These data indicate that ACR30-35 will likely not rescue phage DNA *in vivo*, but that DNA and ACR30-35 do not share the same binding site. Therefore, when ACR30-35 is bound to the Csy complex, the occlusion of the DNA binding site is likely occurring via an allosteric mechanism, possibly a conformational change of the Csy complex.

5.3.7 ACR3112-30 interacts with Csy1-2 and prevents dsDNA binding

After gaining mechanistic insight into the activity of ACR30-35, I utilized similar assays to investigate other anti-CRISPRs. This mechanistic comparison is important because it could shed light on whether anti-CRISPR proteins all diverged from a common ancestor, as the expectation would be that they would act via the same or similar mechanisms. If anti-CRISPRs originated as distinct proteins, then unique binding partners and mechanisms of action would be expected.

I next investigated ACR3112-30, the anti-CRISPR from phage D3112 and determined that this protein also directly interacted with the Csy complex, much like ACR30-35 (Figure 66). Additionally, the RNAse A-treated pseudo-complex of Csy1-2-3 also presented a binding partner for ACR3112-30 (Figure 66). In each case, however, the stoichiometry of the interaction was one ACR3112-30 to one Csy Complex.

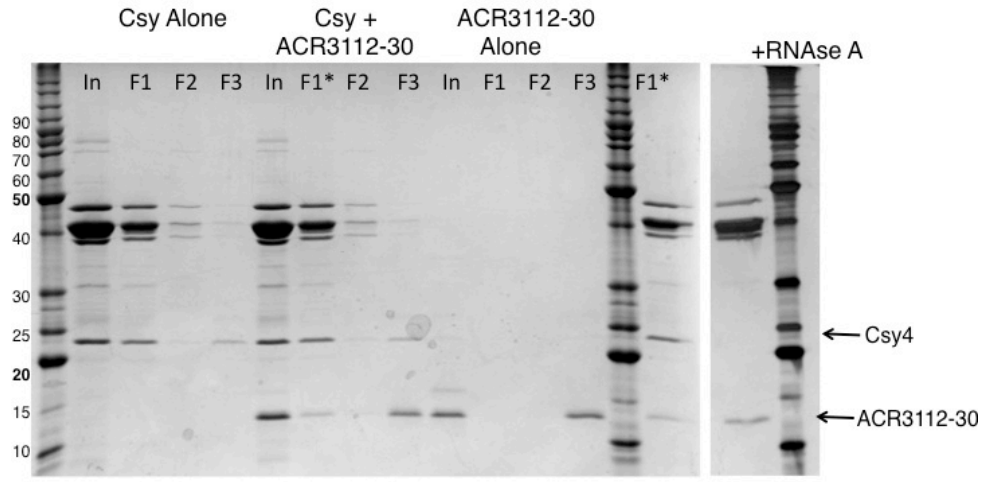


Figure 66: ACR3112-30 interacts with the Csy complex and Csy1, 2, 3

Purified Csy complex was incubated with purified ACR3112-30 and fractionated by SEC with input (In) and fractions F1-F3 shown on coomassie stained SDS-PAGE. F1* shows co-eluting ACR3112-30 with the Csy complex and is run again at the end of the gel. The RNAse A treated Csy complex also shows an interaction with ACR3112-30 in the right panel, on silver stained SDS-PAGE. Note the absence of Csy4.

Furthermore, unlike ACR30-35, ACR3112-30 formed a robust interaction with the purified Csy1-2 heterodimer, demonstrating a distinct interaction partner, (Figure 67).

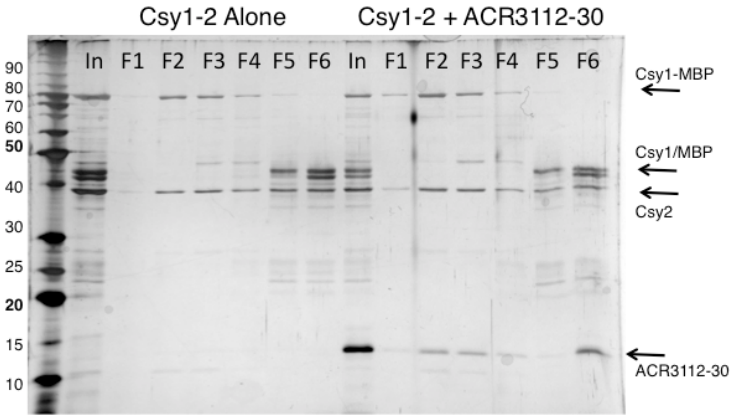


Figure 67: ACR3112-30 interacts with purified Csy1-2

Csy1-2 was purified via an MBP-6xHis tag on Csy1, co-expressed with Csy2. Csy1-2 was incubated with ACR3112-30 and fractionated by SEC. Input (In) and fractions (F1-F6) are shown. F2 shows the position where ACR3112-30 co-elutes with Csy1-2. 15% SDS-PAGE was run and silver stained.

When EMSA experiments were conducted in the presence of ACR3112-30, target binding was also inhibited but a striking distinction with ACR30-35 was observed. ACR3112-30 was a very poor inhibitor of the interaction with a ssDNA target but robustly blocked the interaction between the Csy complex and a target dsDNA substrate (Figure 68) This contrasted with ACR30-35, which inhibits Csy-mediated binding of any DNA target.

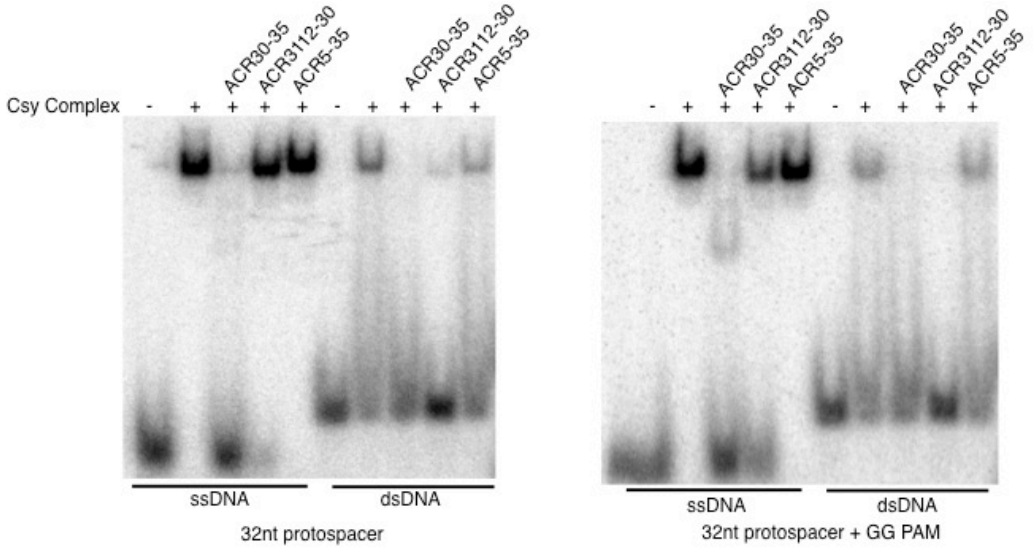


Figure 68: ACR3112-30 inhibits Csy-mediated interaction with target dsDNA

The indicated anti-CRISPRs were incubated with the Csy complex before being mixed with a labeled ss or dsDNA of 32 or 34bp in length. While ACR3112-30 shows weak function with ssDNA, it prevents a Csy-mediated interaction with dsDNA. ACR5-35 is a negative control, which does not bind the Csy complex.

This distinct mode of action from ACR30-35, which inactivates target DNA binding by the Csy complex to both single-stranded and double-stranded DNA targets, indicates subtleties in how the Csy complex functions. Interestingly, the analagous proteins to Csy1-2 in the Type I-E CRISPR-Cas system, CasA (Csy1) and CasD (Csy2) have been shown to be required for interacting non-specifically with dsDNA and destabilizing the double helix for invasion of the crRNA (Sashital *et al.*, 2012). Thus perhaps the presence of the ACR3112-30 inhibits the ability of Csy1-2 to mediate the displacement of the non-target strand, which prevents the crRNA from interrogating its binding partner (i.e. complementary DNA), whereas ssDNA has a fully accessible binding partner.

5.3.8 ACR5-35 interacts with the nuclease/helicase, Cas3

Anti-CRISPR gene 35 from phage JBD5 (ACR5-35) does not interact with the Csy complex, nor does it have an effect on target DNA recognition (see negative controls above). The final protein involved in targeting of foreign DNA in the Type I-F system is Cas3. Cas3 is a helicase/nuclease protein which is recruited to the CRISPR-Cas complex after target DNA binding and is responsible for its degradation in Type I CRISPR-Cas systems (Westra *et al.*, 2012b; Hochstrasser *et al.*, 2014). The ability of ACR5-35 to function without interacting with the Csy complex suggested that perhaps it operates through an interaction with Cas3. Cas3 was purified and fractionated on SEC alone or with ACR5-35 or ACR30-35. Although ACR30-35 did not interact with Cas3, ACR5-35 co-eluted with Cas3, indicating a direct interaction (Figure 69).

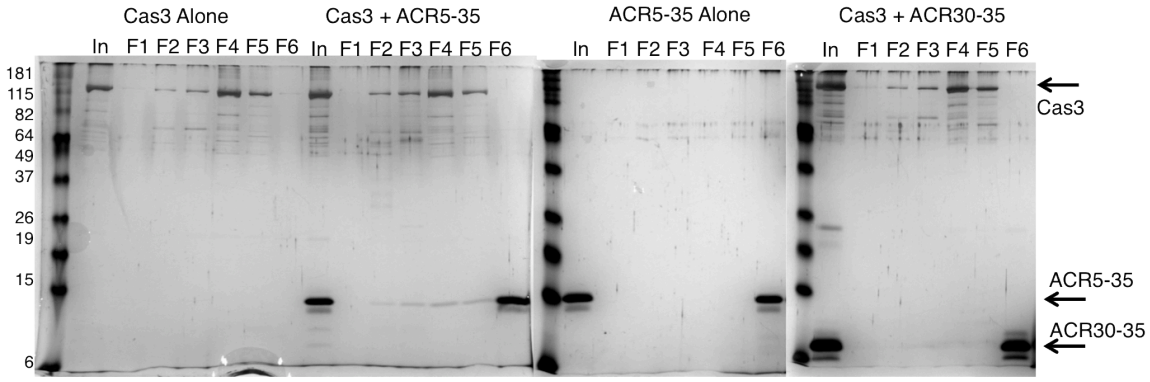


Figure 69: ACR5-35 interacts with helicase/nuclease protein Cas3

Purified Cas3 was fractionated on SEC alone or with ACR5-35 or ACR30-35. ACR5-35 was also run alone for comparison. Note the co-elution of Cas3 and ACR5-35. Proteins were separated on SDS-PAGE, and visualized by silver stain.

Although the outcome of this interaction is not yet clear, after being bound to ACR5-35, Cas3 maybe unable to interact with the Csy complex or its enzymatic activity may be inhibited.

These *in vitro* experiments demonstrate three distinct interaction partners for three anti-CRISPR proteins, indicating distinct modes of action. These data imply unique evolutionary ancestors for these three anti-CRISPRs and suggests that other anti-CRISPRs identified to date with distinct sequences may be functioning using other mechanisms.

5.3.9 Csy gene overexpression inhibits anti-CRISPR function *in vivo*

To assess the relevance of the Cas protein interactions *in vivo*, *P. aeruginosa* strain PA14 overexpressing *csy* genes were infected with various phages that require an anti-CRISPR for their replication. I hypothesized that the overexpression of a relevant Csy protein target would titrate anti-CRISPR protein away from that target in the endogenous Csy complex thus allowing the CRISPR-Cas system to inhibit phage replication. The genes that encode all proteins in the Csy complex (*csy1-4*) were cloned and expressed in PA14. To confirm gene activity, plaque assays with CRISPR-sensitive phage DMS3*m* were conducted, showing that the *csy1-4* expression complemented four different mutant strains PA14Δ*csy1*, PA14Δ*csy2*, PA14Δ*csy3*, and PA14Δ*csy4* independently. An example is shown in Figure 70 with PA14Δ*csy3*. Note the complementation of the mutation (i.e. inhibition of CRISPR-sensitive phage DMS3*m*). Complementation was not observed when an amber stop codon was introduced in the csy1-4 construct (Figure 70, left panels).

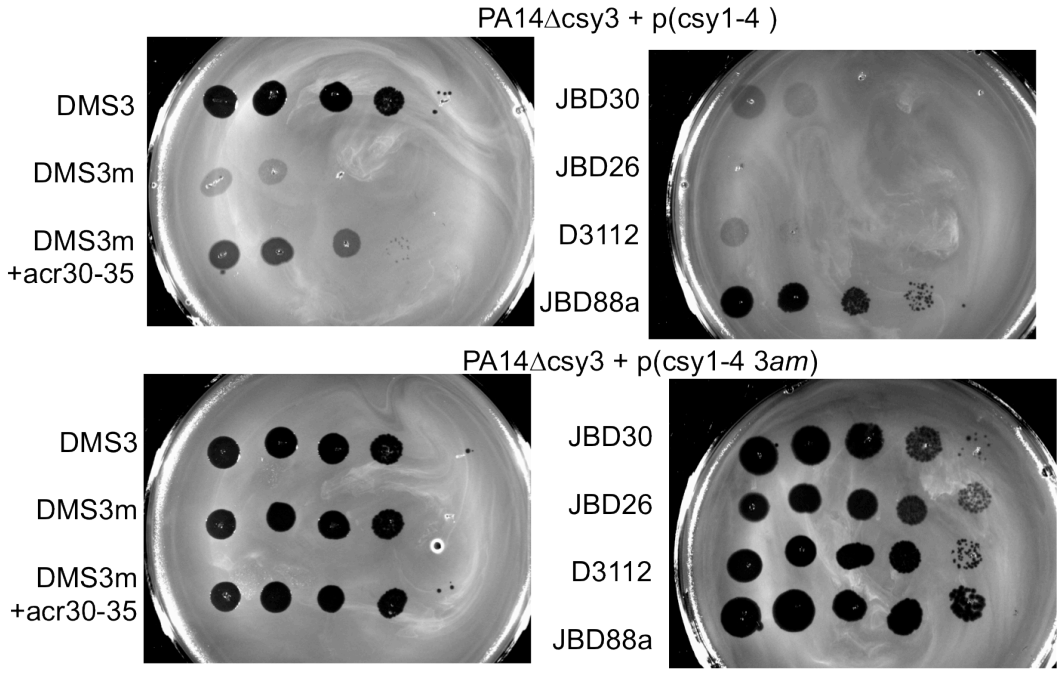


Figure 70: Csy1-4 expression complements a chromosomal *csy3* mutation and inhibits anti-CRISPR phages

A plasmid expressing *csy1-4* or the same plasmid with an amber stop codon in *csy3* (3*am*) was introduced into PA14Δ*csy3* and a CRISPR-insensitive phage (DMS3), a CRISPR-sensitive phage (DMS3*m*), or a CRISPR-sensitive phage relying on ACR30-35 (DMS3*m*+ACR30-35) was tested. Also, natural anti-CRISPR phages JBD30, JBD26, D3112, and JBD88a were assayed on these strains.

Interestingly, the overexpression of Csy proteins inhibited phages DMS3*m* +ACR30-35, JBD30, JBD26, and D3112 (Figure 70, right panels), which rely on ACR30-35, ACR30-35, ACR26-37, and ACR3112-30, respectively. Phage JBD88a was not inhibited by the overexpression of *csy1-4* nor was CRISPR-insensitive phage DMS3. Phage JBD88a relies on a homologue of ACR5-35 (ACR88a-33), which is the anti-CRISPR that interacts with Cas3 and thus I did not expect it to be inhibited by the overexpression of *csy1-4*. Attempts to transform wtPA14 or PA14Δ*cas3*with a plasmid expressing the helicase/nuclease Cas3 were unsuccessful, presumably because it is a toxic when overexpressed (data not shown).

Upon introduction of the *csy1-4* construct into wtPA14, I saw a similar result. Phages JBD30, D3112 and JBD26, but not JBD88a or JBD5, were inhibited. This effect was not seen in cells lacking CRISPR loci (Δ*CR1/2*), demonstrating that overexpression of *csy1-4* does not inhibit phage replication *per se* but inhibits phage replication by interfering with anti-CRISPR function (Figure 71).

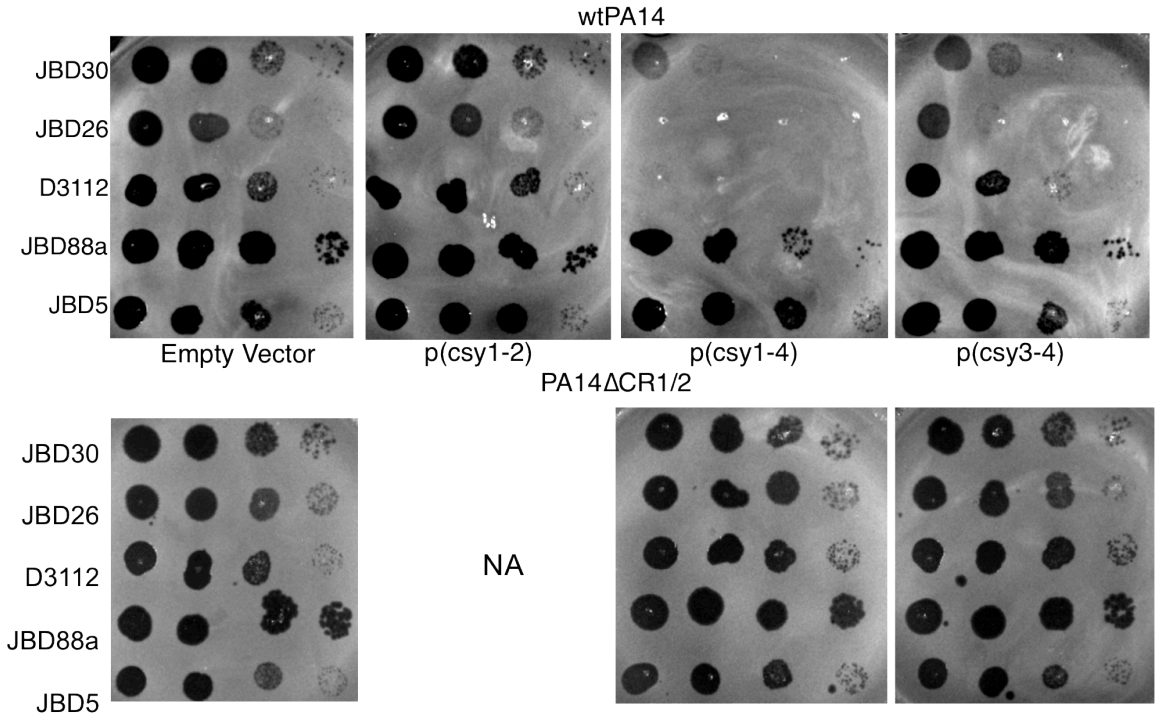


Figure 71: Overexpression of Csy target proteins decrease anti-CRISPR efficacy

wtPA14 or PA14ΔCR1/2 were transformed with empty vector or a plasmid expressing the indicated genes. JBD30 and JBD26 were inhibited by *csy3-4* while these phages and D3112 were inhibited by *csy1-4*. No inhibition was seen in the PA14ΔCR1/2 background. 'NA' – not applicable, since the *csy1-2* construct showed no inhibition in wtPA14.

Expression of *csy3-4* also inhibited phages JBD30 and JBD26. Despite having identified a direct interaction *in vitro* between ACR30-35 and Csy3, *csy3* overexpression alone was not sufficient to inhibit phage JBD30 (data not shown), but needed to be co-expressed with *csy4*, suggesting that these proteins form a "pseudo-complex" with endogenous crRNA that is able to titrate ACR30-35 . The expression of *csy4* alone also had no inhibitory effect (data not shown). D3112 was only inhibited by *csy1-4* expression but not *csy3-4*, suggesting that *csy1* and/or *csy2* overexpression inhibits this phage. The expression of *csy1*, *csy2*, or *csy1-2* had no inhibitory effect on this phage, suggesting that perhaps stable overexpression of these proteins in *P. aeruginosa* requires interaction partners. These *csy1-2* expressing constructs also provided weak or no complementation to chromosomal mutations in the corresponding genes.

5.4 Discussion

In summary, this chapter outlines the detailed structural and functional analysis of ACR30-35 where I have demonstrated that this protein utilizes a single interface to bind tightly to the Csy

complex and inactivate it. This inactivation prevents the Csy complex from binding to a DNA target at any location along the crRNA. The binding of three ACR30-35 proteins along the Csy3 backbone could indicate that ACR30-35 functions via a steric mechanism, blocking the crRNA. However, the Csy complex can be simultaneously bound by ACR30-35 and DNA (Figure 65), which demonstrates distinct binding sites and implies that ACR30-35 blocks target recognition by an allosteric mechanism. For example, ACR30-35 binding could induce a conformational change in the Csy complex which occludes the DNA-binding site, thus preventing target binding and phage DNA cleavage. The other two anti-CRISPRs studied have unique binding partners in the CRISPR-Cas system of PA14. Specifically, ACR3112-30 binds to Csy1-2 and may inhibit those proteins from destabilizing dsDNA, which is necessary for strand invasion by the crRNA. The inability of this anti-CRISPR to effectively prevent binding to ssDNA is consistent with this model. Finally, the absence of any interaction between ACR5-35 and the Csy complex suggested a distinct mode of action for this anti-CRISPR. The direct physical interaction with Cas3 confirms this prediction and although the consequence of ACR5-35 binding Cas3 is not yet known, it is likely that this prevents its recruitment to the Csy complex or its enzymatic activity.

The ability of phages to evade the CRISPR-Cas system via anti-CRISPR genes provides a novel tool to enhance our understanding of this bacterial adaptive immune system and the arms race between bacteria and phages. Here, I describe three distinct mechanisms behind the function of three different anti-CRISPR proteins. As the first CRISPR-Cas inhibitors reported, and now the first mechanistic insight into these inhibitors, these studies provide information not only into the function of anti-CRISPRs but also to the overall activity of a functioning CRISPR-Cas system. This work informs us about the important *in vivo* roles of individual component proteins of the Csy complex as anti-CRISPR proteins likely inhibit essential steps in CRISPR-Cas function. Interestingly, the anti-CRISPR genes were first discovered in the same genomic locus of related phages, and display completely distinct sequences and modes of action, suggesting independent acquisition of non-homologous genes. Anti-CRISPR homologues were also identified in mobile elements and completely novel anti-CRISPRs have recently been identified in unrelated phages and mobile elements by April Pawluk in the Davidson lab. This diversity in sequence and mechanism of both anti-CRISPRs and CRISPR-Cas systems demonstrates the widely diverse tools and selective pressures employed by phages and their bacterial hosts during this seemingly never-ending arms race.

Chapter 6
Summary and Future Directions

6 Summary

The prophages that reside inside bacterial genomes are not simply inert collections of genes waiting to kill their host. Two general observations come from this work, which applies, to all chapters presented here. One is that prophages are active entities, expressing genes which can have an effect on the host (e.g. superinfection exclusion or CRISPR-Cas system inactivation). Secondly, upon induction, these traits can be passed onto a new host rapidly, utilizing the genes which phages possess to ensure their transfer from one host to the next. Although this kills the 'induced cell' the bacterial strain has presumably multiplied while carrying the prophage, thus producing a lineage of lysogenic cells with prophage-derived traits.

In Chapter 2, I demonstrated that isolates of *P. aeruginosa* often possess prophages which are frequently spontaneously induced to produce infectious phage particles. This process can lead to the ubiquitous spread of these prophages among different strains. The purification and sequencing of many phages revealed a diverse collection consisting of at least five different sequence families. By lysogenizing a single host, a number of new traits were imparted to PA14, particularly an array of superinfection exclusion properties. Interestingly, one theme was conserved by the two broadly different families of phages studied in PA14. The pilus-specific and O-antigen-specific phages could both lysogenize the PA14 host and make them appear as receptor mutants to superinfecting phages. These lysogens were not completely defective in their respective receptors, however, as the lysogens still had some intermediate twitching activity, and were positively serotyped. This suggests that some of the diverse collections of novel genes possessed by these prophages are manipulating their host in ways which impart superinfection exclusion.

In Chapter 3, I reported that some of the phages in the collection of temperate phages were inhibited by the PA14 CRISPR-Cas system which partially explains why not all phages in my collection could infect this strain. I also demonstrated that the CRISPR-Cas system targets phages with a gradient of activity depending on the number of mismatches. Together with Dr.

Kyle Cady, we reported the first example of a naturally active Gram-negative CRISPR-Cas system, which has provided an excellent tool for further studies.

In Chapter 4, I reported the first examples of phage encoded anti-CRISPR genes. These anti-CRISPRs are essential for both stages of the phage life cycle (lytic growth and lysogeny), for phages which possess a protospacer that matches the host CRISPR locus. Without anti-CRISPRs, prophages and other mobile elements may be far more restricted in their ability to transfer from one strain to another. This work identified five novel and unique proteins which all mediate the same phenotype of inhibiting the CRISPR-Cas system. None of the genes have an obvious evolutionary path as homologues are limited and there is no identity to previously studied proteins.

In Chapter 5, I explored the mechanistic basis for anti-CRISPR function, largely through a series of *in vitro* experiments, reconstituting various aspects of CRISPR-Cas function. This also served to test whether anti-CRISPRs are all homologous or whether they have distinct ancestors. Three different anti-CRISPRs were investigated and all three have different binding partners and mechanisms to reach the same end, CRISPR-Cas inactivation. This provides evidence that anti-CRISPR genes have been acquired independently from different origins, and were co-opted for this purpose.

6.1 Future Directions

6.1.1 Lysogeny

In the introductory sections on prophages, I outlined a number of phenotypic outcomes that have been ascribed to prophage genes. Most work has focused on superinfection exclusion and the provision of novel virulence traits. The ability to resist superinfection would likely benefit a strain in the environment and inside a host organism due to the omnipresence of phages. Furthermore, although the exact modifications have not been worked out here, the bacterial envelope has been implicated with respect to both the type IV pilus and O-antigen which could each have a strong effect on survival and virulence outside the *C. elegans* model. Also, the high number of unique ‘moron’ genes identified in these related phage genomes (the anti-CRISPRs are included as morons), are quite intriguing and mysterious. These moron genes could be regulated in ways that would perhaps not see them expressed in PA14 but only in other

backgrounds, which could lead to an array of new phenotypes in other hosts. Work is underway in the lab to assess what phenotypes arise when expression of some of these moron elements are driven from a plasmid, to determine what role they could have in a different regulatory framework. This will allow us to also test these morons in other *P. aeruginosa* isolates. Monitoring the prophage expression levels of various moron genes in different growth conditions and strain backgrounds will also be necessary to truly understand the role(s) of these genes.

The ability of these prophages to impart strong phenotypes of superinfection exclusion has certainly masked some anti-CRISPR activity which was presented in CHAPTER 3. Although the strains of PA14(JBD30), and PA14(JBD24) allowed the superinfection of CRISPR-sensitive phages (leading to the anti-CRISPR discovery), these lysogens provide very weak resistance in general (Table 4). Most of the lysogens tested actually resisted CRISPR-sensitive phages, despite also expressing anti-CRISPR genes. This means that the superinfection exclusion mechanisms employed by these prophages potentially serves as a "backup plan" to allow the expression of an anti-CRISPR gene, thus facilitating the survival of the prophage but then also provide resistance at a level which prevents superinfecting phage adsorption. Indeed, without the 'weak resistance' provided by phage JBD30, it is possible that the anti-CRISPR discovery would have never been made.

6.1.2 Anti-CRISPR

The most common question (for which I have no great answer) I receive is why the anti-CRISPRs have all inserted in one genomic locus. Indeed, the additional finding that Type I-E anti-CRISPRs are also located in this region means that this region is purely an anti-CRISPR locus as every gene is accounted for as an anti-CRISPR (see Figure 47). The only remaining gene is the highly conserved (>90% sequence identity in Mu-like phages) JBD30 gene *36* homologue, which is found in each anti-CRISPR containing phage. This gene is likely a transcription factor which regulates anti-CRISPR production (this is the project of Sabrina Stanley in the Davidson lab). Therefore the only genes which have been inserted (and maintained) in this region are anti-CRISPRs and I predict the reason is one of regulation. For an unknown reason, the combination of the upstream promoter and the JBD30 gene *36* transcription factor makes this a perfect location for anti-CRISPR genes to get expressed at the opportune time (presumably early in infection). Measuring the timing of anti-CRISPR expression upon phage

injection into the cell is an ongoing project, also being worked on by Sabrina Stanley. I was unable to detect packaged anti-CRISPR protein in the phage head, suggesting that *de novo* anti-CRISPR expression will be required. The mystery which is not resolved by this explanation though, is why other early genes have not been inserted in this same locus, although we are limited by a small number of phage genomes. What makes this region exclusive to anti-CRISPR genes only is something that will require further understanding of anti-CRISPR function and regulation. The observation that anti-CRISPRs exist in non-phage elements demonstrates that the insertion in this phage region is not absolutely required for function.

The largest questions remaining in my mind with respect to anti-CRISPR biology lies in the breadth and abundance of similar genes in other CRISPR-Cas types and organisms. Given the data presented here that suggests that the anti-CRISPR genes discovered so far have come from distinct evolutionary origins, it is difficult to design bioinformatic methods to find such genes in other systems, with the exception of using the conserved transcription factor as a query. The main approach may be a similar one to mine, looking for phages which seem to be recalcitrant to expected targeting and carrying out a screen from there. This approach would be quite low throughput, however. Other approaches could include generating gene libraries from metagenomic samples of interest (e.g. soil, sewage, dairy, etc.), and expressing them in a strain with an active CRISPR-Cas system. One could then transform the library with a construct that would produce a self-genome targeting crRNA to screen for constructs from the library that alleviate self-targeting. Regardless of the approach, it seems likely that anti-CRISPR discovery will take wet lab experiments. However, April Pawluk in the lab has recently identified novel anti-CRISPRs through bioinformatic means, conducting searches utilizing the conserved transcription factor (JBD30 gene *36* homologues) as well as the conserved promoter upstream of the anti-CRISPR genes. This approach has yielded new type I-E and I-F anti-CRISPRs. Of high interest, would be discovering Type II CRISPR-Cas system inhibitors given the current explosion of applications of CRISPR/Cas9 genome editing. Anti-CRISPRs targeting this type could provide further mechanistic insight with benefits to CRISPR/Cas9 technology.

Further to finding more anti-CRISPRs, their exact role is of significant interest to me. Will they be found exclusively on virulent phages or again on mobile integrative elements such as conjugative islands and temperate phages? This could shed light on the primary selective pressure driving CRISPR-Cas maintenance in some organisms, whether anti-CRISPR genes

emerge on parasitic foreign DNA or on mobile elements that could provide a beneficial role for the host. Another intriguing aspect would be whether there are anti-CRISPRs which function in entirely novel ways, not just by interacting with the CRISPR-Cas machinery. Other general themes that come to mind include: Transcriptional repression (particularly in CRISPR-Cas systems which are induced upon phage infection (Young *et al.*, 2012)), post-transcriptional interference (i.e. processing inhibition), enzymatic degradation of CRISPR-Cas components (i.e. crRNA), or even a phage-encoded gene that modifies DNA, rendering it resistant to CRISPR-Cas targeting.

Also, given the strong effects of the anti-CRISPRs discovered here, it seems logical to expect that this arms race isn't over, and that bacterial (either *cas* or non-*cas* genes) may exist which inhibit the anti-CRISPRs, or "anti-anti-CRISPRs," as this type of back-and-forth phenomenon is well documented with restriction-modification systems (Labrie *et al.*, 2010). A foreseeable mechanism for this was demonstrated in Figures 70 and 71, which is the overexpression of an anti-CRISPR target protein which titrates away anti-CRISPR protein and blocks function.

Finally, the application of this work is of high importance. First, the possibility of revisiting phage therapeutics for treating antibiotic resistant infections adds to the relevance of all data discussed here. Prophages mediating superinfection exclusion properties is something that must be fully understood before we can effectively kill *P. aeruginosa* clinical strains as many strains are lysogenic. Host-encoded properties like CRISPR-Cas also present significant barriers to phage therapy, especially since this is an adaptive immune system. Thus, the discovery of anti-CRISPRs is certainly important in this context, as one would ideally engineer the 'perfect' phage therapeutic to include one or more anti-CRISPRs to remove this potent adaptive immune system from the equation.

These data also shed light on how prophages and mobile elements are acquired in *P. aeruginosa* and what role they have after acquisition on the survival of *P. aeruginosa*. With many *P. aeruginosa* strains possessing CRISPR-Cas systems and Mu-like prophages, it is reasonable to predict two things. First is that many of these systems may be inactivated by resident anti-CRISPRs and second is that if they are not, anti-CRISPRs on horizontal elements could increase the fidelity of transfer. Being able to identify anti-CRISPRs in the genomes of newly sequenced

clinical isolates could identify strains with a high propensity to acquire foreign DNA in spite of a genomically intact CRISPR-Cas system.

6.1.3 Anti-CRISPR Mechanism

A number of outstanding questions remain from the mechanistic work. First, for ACR30-35 it seems likely that a conformational change is induced by ACR30-35 binding, but this is something that is difficult to directly assess. Cryo-EM and negative stained EM studies are currently underway with our collaborator, Dr. Blake Wiedenheft (Montana State University) to try to image bound ACR30-35 to the Csy complex as well as observe any possible changes in Csy complex shape. Further to this, I have setup crystal trials (with Dr. Trevor Moraes, University of Toronto) with Csy bound to anti-CRISPRs, to determine whether anti-CRISPRs enhance crystallization. Attempts to crystallize this complex have been underway for many years in other labs and anti-CRISPRs may provide a key biochemical reagent.

The ability of ACR5-35 to bind Cas3 is also intriguing and obvious experiments here are to assess whether this inhibits enzymatic activity of Cas3. If this is not the case, it seems likely that ACR5-35 inhibits Cas3 recruitment to the Csy complex. Recently, researchers in Jennifer Doudna's (UC-Berkeley) lab have been able to crosslink Cas3 to DNA-bound Cascade (Type I-E CRISPR-Cas complex) (Hochstrasser *et al.*, 2014). This approach could be utilized for the Csy complex, to see if ACR5-35 inhibits this process.

Finally, new insight into the process of spacer acquisition in the Type I-F CRISPR-Cas system has revealed a method which relies on the Csy complex as well as Cas3. Canonically the spacer acquisition process relies only on Cas1/2 but this second method, called "priming" results from a mismatched DNA target, guiding the CRISPR-Cas complex to the site but with no cleavage resulting (Datsenko *et al.*, 2012). The complex is then thought to "slide" away from the mismatched target and extract a new spacer into the CRISPR locus. This was discovered in the Type I-E CRISPR-Cas system of *E. coli* but recent unpublished work from Dr. Konstantin Severinov's lab (Rutgers University) indicates that this is also happening in the Type I-F system of *P. aeruginosa.* The dependence of this process on the Csy complex and Cas3, suggests that anti-CRISPRs may also be able to inhibit spacer acquisition, a completely new and exciting aspect of anti-CRISPR function. In collaboration with the Severinov lab, experiments to test the anti-CRISPRs in this system are underway.

With each unique anti-CRISPR discovered, it is likely that novel mechanisms are to be expected. The biochemical assays outlined in this thesis that have been established by myself and others provide very handy tools to rapidly dissect the mode of action of any new anti-CRISPRs and gain rapid and novel insight into CRISPR-Cas function.

References

Ackermann, H.-W. (2007). 5500 Phages examined in the electron microscope. *Archives of virology* **152**, 227–243. doi:10.1007/s00705-006-0849-1.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403–410. doi:10.1016/S0022-2836(05)80360-2.

Avery, O. T., Macleod, C. M. and McCarty, M. (1944). Studies on the chemical nature of the substance inducing transformation of pneumococcal types: induction of transformation by a desoxyribonucleic acid fraction isolated from Pneumococcus Type III. *The Journal of experimental medicine* **79**, 137–158.

Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M., Kubal, M., Meyer, F., Olsen, G. J., Olson, R., Osterman, A. L., Overbeek, R. A., McNeil, L. K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G. D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A. and Zagnitko, O. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC genomics* **9**, 75. doi:10.1186/1471-2164-9-75.

Bair, C. L. and Black, L. W. (2007). A type IV modification dependent restriction nuclease that targets glucosylated hydroxymethyl cytosine modified DNAs. *Journal of Molecular Biology* **366**, 768–778. doi:10.1016/j.jmb.2006.11.051.

Barondess, J. J. and Beckwith, J. (1990). A bacterial virulence determinant encoded by lysogenic coliphage lambda. *Nature* **346**, 871–874. doi:10.1038/346871a0.

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. doi:10.1126/science.1138140.

Battle, S. E., Meyer, F., Rello, J., Kung, V. L. and Hauser, A. R. (2008). Hybrid pathogenicity island PAGI-5 contributes to the highly virulent phenotype of a Pseudomonas aeruginosa isolate in mammals. *Journal of Bacteriology* **190**, 7130–7140. doi:10.1128/JB.00785-08.

Benchetrit, L. C., Gray, E. D. and Wannamaker, L. W. (1977). Hyaluronidase activity of bacteriophages of group A streptococci. *Infection and Immunity* **15**, 527–532.

Bensing, B. A., Siboo, I. R. and Sullam, P. M. (2001). Proteins PblA and PblB of Streptococcus mitis, which promote binding to human platelets, are encoded within a lysogenic bacteriophage. *Infection and Immunity* **69**, 6186–6192. doi:10.1128/IAI.69.10.6186-6192.2001.

Bergh, O., Børsheim, K. Y., Bratbak, G. and Heldal, M. (1989). High abundance of viruses found in aquatic environments. *Nature* **340**, 467–468. doi:10.1038/340467a0.

Bhaya, D., Davison, M. and Barrangou, R. (2011). CRISPR-Cas Systems in Bacteria and Archaea: Versatile Small RNAs for Adaptive Defense and Regulation. *Annual Review of Genetics* **45**, 273–297. doi:10.1146/annurev-genet-110410-132430.

Bikard, D., Hatoum-Aslan, A., Mucida, D. and Marraffini, L. A. (2012). CRISPR interference can prevent natural transformation and virulence acquisition during in vivo bacterial infection. *Cell host & microbe* **12**, 177–186. doi:10.1016/j.chom.2012.06.003.

Biswas, A., Gagnon, J. N., Brouns, S. J. J., Fineran, P. C. and Brown, C. M. (2013). CRISPRTarget: bioinformatic prediction and analysis of crRNA targets. *RNA biology* **10**, 817–827. doi:10.4161/rna.24046.

Bolotin, A., Quinquis, B., Sorokin, A. and Ehrlich, S. D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology (Reading, England)* **151**, 2551–2561. doi:10.1099/mic.0.28048-0.

Bondy-Denomy, J., Pawluk, A., Maxwell, K. L. and Davidson, A. R. (2013). Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature* **493**, 429–432. doi:10.1038/nature11723.

Boulanger, P. and Letellier, L. (1992). Ion channels are likely to be involved in the two steps of phage T5 DNA penetration into Escherichia coli cells. *The Journal of biological chemistry* **267**, 3168–3172.

Boyd, E. F. and Brüssow, H. (2002). Common themes among bacteriophage-encoded virulence factors and diversity among the bacteriophages involved. *Trends in microbiology* **10**, 521–529.

Braid, M. D., Silhavy, J. L., Kitts, C. L., Cano, R. J. and Howe, M. M. (2004). Complete genomic sequence of bacteriophage B3, a Mu-like phage of Pseudomonas aeruginosa. *Journal of Bacteriology* **186**, 6560–6574. doi:10.1128/JB.186.19.6560-6574.2004.

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

Brooks, M. A., Ravelli, R. B. G., McCarthy, A. A., Strub, K. and Cusack, S. (2009). Structure of SRP14 from the Schizosaccharomyces pombe signal recognition particle. *Acta crystallographica. Section D, Biological crystallography* **65**, 421–433. doi:10.1107/S0907444909005484.

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. doi:10.1126/science.1159689.

Bruttin, A., Desiere, F., Lucchini, S., Foley, S. and Brüssow, H. (1997). Characterization of the lysogeny DNA module from the temperate Streptococcus thermophilus bacteriophage phi Sfi21. *Virology* **233**, 136–148. doi:10.1006/viro.1997.8603.

Brüssow, H., Canchaya, C. and Hardt, W.-D. (2004). Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiology and molecular biology reviews : MMBR* **68**, 560–602– table of contents. doi:10.1128/MMBR.68.3.560-602.2004.

Cady, K. C. and O'Toole, G. (2011). Non-Identity-Mediated CRISPR-Bacteriophage Interaction Mediated via the Csy and Cas3 Proteins. *Journal of Bacteriology* **193**, 3433–3445. doi:10.1128/JB.01411-10.

Cady, K. C., Bondy-Denomy, J., Heussler, G. E., Davidson, A. R. and O'Toole, G. A. (2012). The CRISPR/Cas adaptive immune system of Pseudomonas aeruginosa mediates resistance to naturally occurring and engineered phages. *Journal of Bacteriology* **194**, 5728–5738. doi:10.1128/JB.01184-12.

Cady, K. C., White, A. S., Hammond, J. H., Abendroth, M. D., Karthikeyan, R. S. G., Lalitha, P., Zegans, M. E. and O'Toole, G. (2011). Prevalence, conservation and functional analysis of Yersinia and Escherichia CRISPR regions in clinical Pseudomonas aeruginosa isolates. *Microbiology* **157**, 430–437. doi:10.1099/mic.0.045732-0.

Canchaya, C., Proux, C., Fournous, G., Bruttin, A. and Brüssow, H. (2003). Prophage genomics. *Microbiology and molecular biology reviews : MMBR* **67**, 238–76, table of contents.

Casjens, S. (2003). Prophages and bacterial genomics: what have we learned so far? *Molecular Microbiology* **49**, 277–300. doi:10.1046/j.1365-2958.2003.03580.x.

Casjens, S. R. (2005). Comparative genomics and evolution of the tailed-bacteriophages. *Current opinion in microbiology* **8**, 451–458. doi:10.1016/j.mib.2005.06.014.

Castillo, D., Espejo, R. and Middelboe, M. (2013). Genomic structure of bacteriophage 6H and its distribution as prophage in Flavobacterium psychrophilum strains. *FEMS Microbiology Letters*. doi:10.1111/1574-6968.12342.

Chan, B. K., Abedon, S. T. and Loc-Carrillo, C. (2013). Phage cocktails and the future of phage therapy. *Future microbiology* **8**, 769–783. doi:10.2217/fmb.13.47.

Chopin, M.-C., Chopin, A. and Bidnenko, E. (2005). Phage abortive infection in lactococci: variations on a theme. *Current opinion in microbiology* **8**, 473–479. doi:10.1016/j.mib.2005.06.006.

Chung, I.-Y., Sim, N. and Cho, Y.-H. (2012). Antibacterial efficacy of temperate phage-mediated inhibition of bacterial group motilities. *Antimicrobial Agents and Chemotherapy* **56**, 5612–5617. doi:10.1128/AAC.00504-12.

Clapper, B., Tu, A.-H. T., Elgavish, A. and Dybvig, K. (2004). The vir gene of bacteriophage MAV1 confers resistance to phage infection on Mycoplasma arthritidis. *Journal of Bacteriology* **186**, 5715–5720. doi:10.1128/JB.186.17.5715-5720.2004.

Cornilescu, G., Delaglio, F. and Bax, A. (1999). Protein backbone angle restraints from

searching a database for chemical shift and sequence homology. *Journal of biomolecular NMR* **13**, 289–302.

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

Cumby, N., Davidson, A. R. and Maxwell, K. L. (2012a). The moron comes of age. *Bacteriophage* **2**, 225–228. doi:10.4161/bact.23146.

Cumby, N., Edwards, A. M., Davidson, A. R. and Maxwell, K. L. (2012b). The bacteriophage HK97 gp15 moron element encodes a novel superinfection exclusion protein. *Journal of Bacteriology* **194**, 5012–5019. doi:10.1128/JB.00843-12.

d'Herelle, F. (1917). An invisible microbe that is antagonistic to the dysentery bacillus. *Comptes rendus Acad. Sciences* **165**, 373–375.

Datsenko, K. A., Pougach, K., Tikhonov, A., Wanner, B. L., Severinov, K. and Semenova, E. (2012). Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nature Communications* **3**, 945. doi:10.1038/ncomms1937.

Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *Journal of biomolecular NMR* **6**, 277–293.

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. doi:10.1038/nature09886.

Deveau, H., Barrangou, R., Garneau, J. E., Labonté, J., Fremaux, C., Boyaval, P., Romero, D. A., Horvath, P. and Moineau, S. (2008). Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. *Journal of Bacteriology* **190**, 1390–1400. doi:10.1128/JB.01412-07.

Dodd, I. B., Shearwin, K. E. and Egan, J. B. (2005). Revisited gene regulation in bacteriophage lambda. *Current opinion in genetics & development* **15**, 145–152. doi:10.1016/j.gde.2005.02.001.

Dugar, G., Herbig, A., Förstner, K. U., Heidrich, N., Reinhardt, R., Nieselt, K. and Sharma, C. M. (2013). High-resolution transcriptome maps reveal strain-specific regulatory features of multiple Campylobacter jejuni isolates. *PLoS genetics* **9**, e1003495. doi:10.1371/journal.pgen.1003495.

Edgar, R. and Qimron, U. (2010). The Escherichia coli CRISPR System Protects from Lysogenization, Lysogens, and Prophage Induction. *Journal of Bacteriology* **192**, 6291–6294. doi:10.1128/JB.00644-10.

Eklund, M. W., Poysky, F. T., Reed, S. M. and Smith, C. A. (1971). Bacteriophage and the toxigenicity of Clostridium botulinum type C. *Science* **172**, 480–482.

Engelberg-Kulka, H., Reches, M., Narasimhan, S., Schoulaker-Schwarz, R., Klemes, Y., Aizenman, E. and Glaser, G. (1998). rexB of bacteriophage lambda is an anti-cell death gene. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 15481–15486.

Figueroa-Bossi, N. and Bossi, L. (1999). Inducible prophages contribute to Salmonella virulence in mice. *Molecular Microbiology* **33**, 167–176.

Freeman, V. J. (1951). Studies on the virulence of bacteriophage-infected strains of Corynebacterium diphtheriae. *Journal of Bacteriology* **61**, 675–688.

Fruciano, D. E. and Bourne, S. (2007). Phage as an antimicrobial agent: d"Herelle"s heretical theories and their role in the decline of phage prophylaxis in the West. *The Canadian journal of infectious diseases & medical microbiology = Journal canadien des maladies infectieuses et de la microbiologie médicale / AMMI Canada* **18**, 19–26.

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. doi:10.1038/nature09523.

Ghosh, D., Roy, K., Williamson, K. E., White, D. C., Wommack, K. E., Sublette, K. L. and Radosevich, M. (2008). Prevalence of lysogeny among soil bacteria and presence of 16S rRNA and trzN genes in viral-community DNA. *Applied and Environmental Microbiology* **74**, 495–502. doi:10.1128/AEM.01435-07.

Gibson, J., Sood, A. and Hogan, D. A. (2009). Pseudomonas aeruginosa-Candida albicans interactions: localization and fungal toxicity of a phenazine derivative. *Applied and Environmental Microbiology* **75**, 504–513. doi:10.1128/AEM.01037-08.

Godde, J. S. and Bickerton, A. (2006). The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes. *Journal of molecular evolution* **62**, 718–729. doi:10.1007/s00239-005-0223-z.

Grissa, I., Vergnaud, G. and Pourcel, C. (2007). CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Research* **35**, W52–7. doi:10.1093/nar/gkm360.

Güntert, P. (2004). Automated NMR structure calculation with CYANA. *Methods in molecular biology (Clifton, N.J.)* **278**, 353–378. doi:10.1385/1-59259-809-9:353.

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. doi:10.1016/j.cell.2009.07.040.

Hanlon, G. W., Denyer, S. P., Olliff, C. J. and Ibrahim, L. J. (2001). Reduction in exopolysaccharide viscosity as an aid to bacteriophage penetration through Pseudomonas aeruginosa biofilms. *Applied and Environmental Microbiology* **67**, 2746–2753. doi:10.1128/AEM.67.6.2746-2753.2001.

Hattman, S. (1982). DNA methyltransferase-dependent transcription of the phage Mu mom gene. *Proceedings of the National Academy of Sciences of the United States of America* **79**, 5518–5521.

Haurwitz, R. E., Jinek, M., Wiedenheft, B., Zhou, K. and Doudna, J. A. (2010). Sequence- and Structure-Specific RNA Processing by a CRISPR Endonuclease. *Science* **329**, 1355–1358. doi:10.1126/science.1192272.

Heo, Y.-J., Chung, I.-Y., Choi, K. B., Lau, G. W. and Cho, Y.-H. (2007). Genome sequence comparison and superinfection between two related Pseudomonas aeruginosa phages, D3112 and MP22. *Microbiology (Reading, England)* **153**, 2885–2895. doi:10.1099/mic.0.2007/007260-0.

Hershey, A. D. and Chase, M. (1952). Independent functions of viral protein and nucleic acid in growth of bacteriophage. *The Journal of general physiology* **36**, 39–56.

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. *Proceedings of the National Academy of Sciences* **111**, 6618–6623. doi:10.1073/pnas.1405079111.

Hofer, B., Ruge, M. and Dreiseikelmann, B. (1995). The superinfection exclusion gene (sieA) of bacteriophage P22: identification and overexpression of the gene and localization of the gene product. *Journal of Bacteriology* **177**, 3080–3086.

Holm, L., Kääriäinen, S., Wilton, C. and Plewczynski, D. (2006). Using Dali for structural comparison of proteins. *Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis ... [et al.]* **Chapter 5**, Unit 5.5. doi:10.1002/0471250953.bi0505s14.

Horvath, P., Romero, D. A., Coûté-Monvoisin, A.-C., Richards, M., Deveau, H., Moineau, S., Boyaval, P., Fremaux, C. and Barrangou, R. (2008). Diversity, activity, and evolution of CRISPR loci in Streptococcus thermophilus. *Journal of Bacteriology* **190**, 1401–1412. doi:10.1128/JB.01415-07.

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

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

Jacobs-Sera, D., Marinelli, L. J., Bowman, C., Broussard, G. W., Guerrero Bustamante, C., Boyle, M. M., Petrova, Z. O., Dedrick, R. M., Pope, W. H., Science Education Alliance Phage Hunters Advancing Genomics And Evolutionary Science Sea-Phages Program, Modlin, R. L., Hendrix, R. W. and Hatfull, G. F. (2012). On the nature of mycobacteriophage diversity and host preference. *Virology* **434**, 187–201. doi:10.1016/j.virol.2012.09.026.

Jansen, R., Embden, J. D. A. V., Gaastra, W. and Schouls, L. M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular Microbiology* **43**, 1565–1575.

Jiang, W., Maniv, I., Arain, F., Wang, Y., Levin, B. R. and Marraffini, L. A. (2013). Dealing with the evolutionary downside of CRISPR immunity: bacteria and beneficial plasmids. *PLoS genetics* **9**, e1003844. doi:10.1371/journal.pgen.1003844.

Jore, M. M., Brouns, S. J. J. and van der Oost, J. (2012). RNA in Defense: CRISPRs Protect Prokaryotes against Mobile Genetic Elements. *Cold Spring Harbor perspectives in biology* **4**. doi:10.1101/cshperspect.a003657.

Juhala, R. J., Ford, M. E., Duda, R. L., Youlton, A., Hatfull, G. F. and Hendrix, R. W. (2000). Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. *Journal of Molecular Biology* **299**, 27–51. doi:10.1006/jmbi.2000.3729.

Kanelis, V., Forman-Kay, J. D. and Kay, L. E. (2001). Multidimensional NMR methods for protein structure determination. *IUBMB life* **52**, 291–302. doi:10.1080/152165401317291147.

Karaolis, D. K., Somara, S., Maneval, D. R., Johnson, J. A. and Kaper, J. B. (1999). A bacteriophage encoding a pathogenicity island, a type-IV pilus and a phage receptor in cholera bacteria. *Nature* **399**, 375–379. doi:10.1038/20715.

Kay, L. E. (1995). Pulsed field gradient multi-dimensional NMR methods for the study of protein structure and dynamics in solution. *Progress in biophysics and molecular biology* **63**, 277–299.

Kent, B. N. and Bordenstein, S. R. (2010). Phage WO of Wolbachia: lambda of the endosymbiont world. *Trends in microbiology* **18**, 173–181. doi:10.1016/j.tim.2009.12.011.

Kotewicz, M., Chung, S., Takeda, Y. and Echols, H. (1977). Characterization of the integration protein of bacteriophage lambda as a site-specific DNA-binding protein. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 1511–1515.

Köhler, T., Donner, V. and van Delden, C. (2010). Lipopolysaccharide as shield and receptor for R-pyocin-mediated killing in Pseudomonas aeruginosa. *Journal of Bacteriology* **192**, 1921–1928. doi:10.1128/JB.01459-09.

Krüger, D. H. and Bickle, T. A. (1983). Bacteriophage survival: multiple mechanisms for avoiding the deoxyribonucleic acid restriction systems of their hosts. *Microbiological reviews* **47**, 345–360.

Krüger, D. H., Barcak, G. J., Reuter, M. and Smith, H. O. (1988). EcoRII can be activated to cleave refractory DNA recognition sites. *Nucleic Acids Research* **16**, 3997–4008.

Labrie, S. J., Samson, J. E. and Moineau, S. (2010). Bacteriophage resistance mechanisms.

Nature Reviews Microbiology **8**, 317–327. doi:10.1038/nrmicro2315.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. and Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics (Oxford, England)* **23**, 2947–2948. doi:10.1093/bioinformatics/btm404.

Laskowski, R. A., Rullmann, J. and MacArthur, M. W. (1996). AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR - Springer. *Journal of Biomolecular NMR* **8**, 477-486.

Lee, W., Westler, W. M., Bahrami, A., Eghbalnia, H. R. and Markley, J. L. (2009). PINE-SPARKY: graphical interface for evaluating automated probabilistic peak assignments in protein NMR spectroscopy. *Bioinformatics (Oxford, England)* **25**, 2085–2087. doi:10.1093/bioinformatics/btp345.

Levin, B. R. (2010). Nasty viruses, costly plasmids, population dynamics, and the conditions for establishing and maintaining CRISPR-mediated adaptive immunity in bacteria. *PLoS Genetics* **6**, e1001171. doi:10.1371/journal.pgen.1001171.

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

Mah, T.-F., Pitts, B., Pellock, B., Walker, G. C., Stewart, P. S. and O'Toole, G. A. (2003). A genetic basis for Pseudomonas aeruginosa biofilm antibiotic resistance. *Nature* **426**, 306–310. doi:10.1038/nature02122.

Mahenthiralingam, E., Campbell, M. E. and Speert, D. P. (1994). Nonmotility and phagocytic resistance of Pseudomonas aeruginosa isolates from chronically colonized patients with cystic fibrosis. *Infection and Immunity* **62**, 596–605.

Mahony, J., McGrath, S., Fitzgerald, G. F. and van Sinderen, D. (2008). Identification and characterization of lactococcal-prophage-carried superinfection exclusion genes. *Applied and Environmental Microbiology* **74**, 6206–6215. doi:10.1128/AEM.01053-08.

Maillou, J. and Dreiseikelmann, B. (1990). The sim gene of Escherichia coli phage P1: nucleotide sequence and purification of the processed protein. *Virology* **175**, 500–507.

Makarova, K. S., Aravind, L., Wolf, Y. I. and Koonin, E. V. (2011a). Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. *Biology Direct* **6**, 38. doi:10.1186/1745-6150-6-38.

Makarova, K. S., Grishin, N. V., Shabalina, S. A., Wolf, Y. I. and Koonin, E. V. (2006). A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biology Direct* **1**, 7. doi:10.1186/1745-6150-1-7.

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., van der Oost, J. and Koonin,

E. V. (2011b). Evolution and classification of the CRISPR-Cas systems. *Nature Reviews Microbiology* **9**, 467–477. doi:10.1038/nrmicro2577.

Manica, A., Zebec, Z., Teichmann, D. and Schleper, C. (2011). In vivo activity of CRISPR-mediated virus defence in a hyperthermophilic archaeon. *Molecular Microbiology* **80**, 481–491. doi:10.1111/j.1365-2958.2011.07586.x.

Manna, D., Wang, X. and Higgins, N. P. (2001). Mu and IS1 transpositions exhibit strong orientation bias at the Escherichia coli bgl locus. *Journal of Bacteriology* **183**, 3328–3335. doi:10.1128/JB.183.11.3328-3335.2001.

Marcó, M. B., Moineau, S. and Quiberoni, A. (2012). Bacteriophages and dairy fermentations. *Bacteriophage* **2**, 149–158. doi:10.4161/bact.21868.

Marraffini, L. A. and Sontheimer, E. J. (2008). CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322**, 1843–1845. doi:10.1126/science.1165771.

Maxwell, K. L. and Davidson, A. R. (1998). Mutagenesis of a buried polar interaction in an SH3 domain: sequence conservation provides the best prediction of stability effects. *Biochemistry* **37**, 16172–16182. doi:10.1021/bi981788p.

McGrath, S., Fitzgerald, G. F. and van Sinderen, D. (2002). Identification and characterization of phage-resistance genes in temperate lactococcal bacteriophages. *Molecular Microbiology* **43**, 509–520.

McGrath, S., Seegers, J.F.M.L., Fitzgerald, G. F. and van Sinderen, D. (1999). Molecular characterization of a phage-encoded resistance system in *Lactococcus lactis*. *Applied and Environmental Microbiology* **65**, 1891-1899.

Medina-Aparicio, L., Rebollar-Flores, J. E., Gallego-Hernández, A. L., Vázquez, A., Olvera, L., Gutiérrez-Ríos, R. M., Calva, E. and Hernández-Lucas, I. (2011). The CRISPR/Cas immune system is an operon regulated by LeuO, H-NS, and leucine-responsive regulatory protein in Salmonella enterica serovar Typhi. *Journal of Bacteriology* **193**, 2396–2407. doi:10.1128/JB.01480-10.

Millen, A. M., Horvath, P., Boyaval, P. and Romero, D. A. (2012). Mobile CRISPR/Cas-mediated bacteriophage resistance in Lactococcus lactis. *PloS one* **7**, e51663. doi:10.1371/journal.pone.0051663.

Minot, S., Bryson, A., Chehoud, C., Wu, G. D., Lewis, J. D. and Bushman, F. D. (2013). Rapid evolution of the human gut virome. *Proceedings of the National Academy of Sciences*. doi:10.1073/pnas.1300833110.

Minot, S., Sinha, R., Chen, J., Li, H., Keilbaugh, S. A., Wu, G. D., Lewis, J. D. and Bushman, F. D. (2011). The human gut virome: inter-individual variation and dynamic response to diet. *Genome Research* **21**, 1616–1625. doi:10.1101/gr.122705.111.

Mirold, S., Rabsch, W., Rohde, M., Stender, S., Tschäpe, H., Rüssmann, H., Igwe, E. and

Hardt, W. D. (1999). Isolation of a temperate bacteriophage encoding the type III effector protein SopE from an epidemic Salmonella typhimurium strain. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 9845–9850.

Mitchell, J., Siboo, I. R., Takamatsu, D., Chambers, H. F. and Sullam, P. M. (2007). Mechanism of cell surface expression of the Streptococcus mitis platelet binding proteins PblA and PblB. *Molecular Microbiology* **64**, 844–857. doi:10.1111/j.1365-2958.2007.05703.x.

Modi, S. R., Lee, H. H., Spina, C. S. and Collins, J. J. (2013). Antibiotic treatment expands the resistance reservoir and ecological network of the phage metagenome. *Nature* **499**, 219–222. doi:10.1038/nature12212.

Mojica, F. J. M., Diez-Villasenor, C., Garcia-Martinez, J. and Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**, 733–740. doi:10.1099/mic.0.023960-0.

Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. and Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of molecular evolution* **60**, 174–182. doi:10.1007/s00239-004-0046-3.

Morales, D. K., Jacobs, N. J., Rajamani, S., Krishnamurthy, M., Cubillos-Ruiz, J. R. and Hogan, D. A. (2010). Antifungal mechanisms by which a novel Pseudomonas aeruginosa phenazine toxin kills Candida albicans in biofilms. *Molecular Microbiology* **78**, 1379–1392. doi:10.1111/j.1365-2958.2010.07414.x.

Morgan, G. J., Hatfull, G. F., Casjens, S. and Hendrix, R. W. (2002). Bacteriophage Mu genome sequence: analysis and comparison with Mu-like prophages in Haemophilus, Neisseria and Deinococcus. *Journal of Molecular Biology* **317**, 337–359. doi:10.1006/jmbi.2002.5437.

Müller, M. G., Ing, J. Y., Cheng, M. K.-W., Flitter, B. A. and Moe, G. R. (2013). Identification of a phage-encoded Ig-binding protein from invasive Neisseria meningitidis. *Journal of immunology (Baltimore, Md : 1950)* **191**, 3287–3296. doi:10.4049/jimmunol.1301153.

Nakayama, K., Kanaya, S., Ohnishi, M., Terawaki, Y. and Hayashi, T. (1999). The complete nucleotide sequence of phi CTX, a cytotoxin-converting phage of Pseudomonas aeruginosa: implications for phage evolution and horizontal gene transfer via bacteriophages. *Molecular Microbiology* **31**, 399–419.

Navarre, W. W., Porwollik, S., Wang, Y., McClelland, M., Rosen, H., Libby, S. J. and Fang, F. C. (2006). Selective silencing of foreign DNA with low GC content by the H-NS protein in Salmonella. *Science* **313**, 236–238. doi:10.1126/science.1128794.

Nesper, J., Blass, J., Fountoulakis, M. and Reidl, J. (1999). Characterization of the major control region of Vibrio cholerae bacteriophage K139: immunity, exclusion, and integration. *Journal of Bacteriology* **181**, 2902–2913.

Newton, G. J., Daniels, C., Burrows, L. L., Kropinski, A. M., Clarke, A. J. and Lam, J. S. (2001). Three-component-mediated serotype conversion in Pseudomonas aeruginosa by bacteriophage D3. *Molecular Microbiology* **39**, 1237–1247.

Nozawa, T., Furukawa, N., Aikawa, C., Watanabe, T., Haobam, B., Kurokawa, K., Maruyama, F. and Nakagawa, I. (2011). CRISPR inhibition of prophage acquisition in Streptococcus pyogenes. *PloS one* **6**, e19543. doi:10.1371/journal.pone.0019543.

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. *Nature structural & molecular biology*. doi:10.1038/nsmb.2820.

O'Brien, A. D., Newland, J. W., Miller, S. F., Holmes, R. K., Smith, H. W. and Formal, S. B. (1984). Shiga-like toxin-converting phages from Escherichia coli strains that cause hemorrhagic colitis or infantile diarrhea. *Science* **226**, 694–696.

O'Flaherty, S., Coffey, A., Edwards, R., Meaney, W., Fitzgerald, G. F. and Ross, R. P. (2004). Genome of staphylococcal phage K: a new lineage of Myoviridae infecting gram-positive bacteria with a low G+C content. *Journal of Bacteriology* **186**, 2862–2871.

Ochman, H., Gerber, A. S. and Hartl, D. L. (1988). Genetic applications of an inverse polymerase chain reaction. *Genetics* **120**, 621–623.

Ojeniyi, B. (1988). Bacteriophages in sputum of cystic fibrosis patients as a possible cause of in vivo changes in serotypes of Pseudomonas aeruginosa. *APMIS : acta pathologica, microbiologica, et immunologica Scandinavica* **96**, 294–298.

Oliver, K. M., Degnan, P. H., Hunter, M. S. and Moran, N. A. (2009). Bacteriophages encode factors required for protection in a symbiotic mutualism. *Science* **325**, 992–994. doi:10.1126/science.1174463.

Palmer, K. L. and Gilmore, M. S. (2010). Multidrug-resistant enterococci lack CRISPR-cas. *mBio* **1**. doi:10.1128/mBio.00227-10.

Pawluk, A., Bondy-Denomy, J., Cheung, V. H. W., Maxwell, K. L. and Davidson, A. R. (2014). A new group of phage anti-CRISPR genes inhibits the type I-E CRISPR-Cas system of Pseudomonas aeruginosa. *mBio* **5**, e00896. doi:10.1128/mBio.00896-14.

Pedulla, M. L., Ford, M. E., Houtz, J. M., Karthikeyan, T., Wadsworth, C., Lewis, J. A., Jacobs-Sera, D., Falbo, J., Gross, J., Pannunzio, N. R., Brucker, W., Kumar, V., Kandasamy, J., Keenan, L., Bardarov, S., Kriakov, J., Lawrence, J. G., Jacobs, W. R., Hendrix, R. W. and Hatfull, G. F. (2003). Origins of highly mosaic mycobacteriophage genomes. *Cell* **113**, 171–182.

Pérez-Martínez, I. and Haas, D. (2011). Azithromycin inhibits expression of the GacA-dependent small RNAs RsmY and RsmZ in Pseudomonas aeruginosa. *Antimicrobial Agents and Chemotherapy* **55**, 3399–3405. doi:10.1128/AAC.01801-10.

Pitt, T. L. (1986). Biology of Pseudomonas aeruginosa in relation to pulmonary infection in

cystic fibrosis. *Journal of the Royal Society of Medicine* **79 Suppl 12**, 13–18.

Pourcel, C., Salvignol, G. and Vergnaud, G. (2005). CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology (Reading, England)* **151**, 653–663. doi:10.1099/mic.0.27437-0.

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. *Molecular Microbiology* **75**, 1495–1512. doi:10.1111/j.1365-2958.2010.07073.x.

Qiu, D., Damron, F. H., Mima, T., Schweizer, H. P. and Yu, H. D. (2008). PBAD-Based Shuttle Vectors for Functional Analysis of Toxic and Highly Regulated Genes in Pseudomonas and Burkholderia spp. and Other Bacteria. *Applied and Environmental Microbiology* **74**, 7422–7426. doi:10.1128/AEM.01369-08.

Ran, H., Hassett, D. J. and Lau, G. W. (2003). Human targets of Pseudomonas aeruginosa pyocyanin. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 14315–14320. doi:10.1073/pnas.2332354100.

Reeks, J., Naismith, J. H. and White, M. F. (2013). CRISPR interference: a structural perspective. *The Biochemical journal* **453**, 155–166. doi:10.1042/BJ20130316.

Reeve, J. N. and Shaw, J. E. (1979). Lambda encodes an outer membrane protein: the lom gene. *Molecular & general genetics : MGG* **172**, 243–248.

Reidl, J. and Mekalanos, J. J. (1995). Characterization of Vibrio cholerae bacteriophage K139 and use of a novel mini-transposon to identify a phage-encoded virulence factor. *Molecular Microbiology* **18**, 685–701.

Reyes, A., Haynes, M., Hanson, N., Angly, F. E., Heath, A. C., Rohwer, F. and Gordon, J. I. (2010). Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature* **466**, 334–338. doi:10.1038/nature09199.

Reyes, A., Wu, M., McNulty, N. P., Rohwer, F. L. and Gordon, J. I. (2013). Gnotobiotic mouse model of phage-bacterial host dynamics in the human gut. *Proceedings of the National Academy of Sciences* **110**, 20236–20241. doi:10.1073/pnas.1319470110.

Rho, M., Wu, Y.-W., Tang, H., Doak, T. G. and Ye, Y. (2012). Diverse CRISPRs evolving in human microbiomes. *PLoS genetics* **8**, e1002441. doi:10.1371/journal.pgen.1002441.

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. doi:10.1371/journal.pone.0049549.

Rohwer, F. and Thurber, R. V. (2009). Viruses manipulate the marine environment. *Nature* **459**, 207–212. doi:10.1038/nature08060.

Samai, P., Smith, P. and Shuman, S. (2010). Structure of a CRISPR-associated protein Cas2

from Desulfovibrio vulgaris. *Acta crystallographica. Section F, Structural biology and crystallization communications* **66**, 1552–1556. doi:10.1107/S1744309110039801.

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. doi:10.1038/nature12048.

Samson, J. E., Magadán, A. H., Sabri, M. and Moineau, S. (2013). Revenge of the phages: defeating bacterial defences. *Nature Reviews Microbiology* **11**, 675–687. doi:10.1038/nrmicro3096.

Sashital, D. G., Wiedenheft, B. and Doudna, J. A. (2012). Mechanism of foreign DNA selection in a bacterial adaptive immune system. *Molecular Cell* **46**, 606–615. doi:10.1016/j.molcel.2012.03.020.

Scholz, I., Lange, S. J., Hein, S., Hess, W. R. and Backofen, R. (2013). CRISPR-Cas systems in the cyanobacterium Synechocystis sp. PCC6803 exhibit distinct processing pathways involving at least two Cas6 and a Cmr2 protein. *PloS one* **8**, e56470. doi:10.1371/journal.pone.0056470.

Schuch, R. and Fischetti, V. A. (2009). The secret life of the anthrax agent Bacillus anthracis: bacteriophage-mediated ecological adaptations. *PloS one* **4**, e6532. doi:10.1371/journal.pone.0006532.

Schuch, R., Nelson, D. and Fischetti, V. A. (2002). A bacteriolytic agent that detects and kills Bacillus anthracis. *Nature* **418**, 884–889. doi:10.1038/nature01026.

Schunder, E., Rydzewski, K., Grunow, R. and Heuner, K. (2013). First indication for a functional CRISPR/Cas system in Francisella tularensis. *International journal of medical microbiology : IJMM* **303**, 51–60. doi:10.1016/j.ijmm.2012.11.004.

Sebaihia, M., Wren, B. W., Mullany, P., Fairweather, N. F., Minton, N., Stabler, R., Thomson, N. R., Roberts, A. P., Cerdeño-Tárraga, A. M., Wang, H., Holden, M. T. G., Wright, A., Churcher, C., Quail, M. A., Baker, S., Bason, N., Brooks, K., Chillingworth, T., Cronin, A., Davis, P., Dowd, L., Fraser, A., Feltwell, T., Hance, Z., Holroyd, S., Jagels, K., Moule, S., Mungall, K., Price, C., Rabbinowitsch, E., Sharp, S., Simmonds, M., Stevens, K., Unwin, L., Whithead, S., Dupuy, B., Dougan, G., Barrell, B. and Parkhill, J. (2006). The multidrug-resistant human pathogen Clostridium difficile has a highly mobile, mosaic genome. *Nature genetics* **38**, 779–786. doi:10.1038/ng1830.

Seed, K. D., Lazinski, D. W., Calderwood, S. B. and Camilli, A. (2013). A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. *Nature* **494**, 489–491. doi:10.1038/nature11927.

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. *Proceedings of the National Academy of Sciences* **108**, 10098–10103. doi:10.1073/pnas.1104144108.

Seo, H. S., Xiong, Y. Q., Mitchell, J., Seepersaud, R., Bayer, A. S. and Sullam, P. M. (2010). Bacteriophage Lysin Mediates the Binding of Streptococcus mitis to Human Platelets through Interaction with Fibrinogen. *PLoS pathogens* **6**, e1001047. doi:10.1371/journal.ppat.1001047.t001.

Shanks, R. M. Q., Caiazza, N. C., Hinsa, S. M., Toutain, C. M. and O'Toole, G. A. (2006). Saccharomyces cerevisiae-based molecular tool kit for manipulation of genes from gram-negative bacteria. *Applied and Environmental Microbiology* **72**, 5027–5036. doi:10.1128/AEM.00682-06.

Shinedling, S., Parma, D. and Gold, L. (1987). Wild-type bacteriophage T4 is restricted by the lambda rex genes. *Journal of Virology* **61**, 3790–3794.

Slavcev, R. A. and Hayes, S. (2003). Stationary phase-like properties of the bacteriophage lambda Rex exclusion phenotype. *Molecular genetics and genomics : MGG* **269**, 40–48. doi:10.1007/s00438-002-0787-x.

Snyder, L. (1995). Phage-exclusion enzymes: a bonanza of biochemical and cell biology reagents? *Molecular Microbiology* **15**, 415–420.

Soutourina, O. A., Monot, M., Boudry, P., Saujet, L., Pichon, C., Sismeiro, O., Semenova, E., Severinov, K., Le Bouguenec, C., Coppee, J.-Y., Dupuy, B. and Martin-Verstraete, I. (2013). Genome-wide identification of regulatory RNAs in the human pathogen Clostridium difficile. *PLoS genetics* **9**, e1003493. doi:10.1371/journal.pgen.1003493.

Stern, A., Keren, L., Wurtzel, O., Amitai, G. and Sorek, R. (2010). Self-targeting by CRISPR: gene regulation or autoimmunity? *Trends in genetics : TIG* **26**, 335–340. doi:10.1016/j.tig.2010.05.008.

Stern, A., Mick, E., Tirosh, I., Sagy, O. and Sorek, R. (2012). CRISPR targeting reveals a reservoir of common phages associated with the human gut microbiome. *Genome Research* **22**, 1985–1994. doi:10.1101/gr.138297.112.

Sternberg, S. H., Haurwitz, R. E. and Doudna, J. A. (2012). Mechanism of substrate selection by a highly specific CRISPR endoribonuclease. *RNA (New York, N.Y.)* **18**, 661–672. doi:10.1261/rna.030882.111.

Stummeyer, K., Schwarzer, D., Claus, H., Vogel, U., Gerardy-Schahn, R. and Mühlenhoff, M. (2006). Evolution of bacteriophages infecting encapsulated bacteria: lessons from Escherichia coli K1-specific phages. *Molecular Microbiology* **60**, 1123–1135. doi:10.1111/j.1365-2958.2006.05173.x.

Sun, X., Göhler, A., Heller, K. J. and Neve, H. (2006). The ltp gene of temperate Streptococcus thermophilus phage TP-J34 confers superinfection exclusion to Streptococcus thermophilus and Lactococcus lactis. *Virology* **350**, 146–157. doi:10.1016/j.virol.2006.03.001.

Sutherland, E., Coe, L. and Raleigh, E. A. (1992). McrBC: a multisubunit GTP-dependent restriction endonuclease. *Journal of Molecular Biology* **225**, 327–348.

Suttle, C. A. (2005). Viruses in the sea. *Nature* **437**, 356–361. doi:10.1038/nature04160.

Suttle, C. A. (2007). Marine viruses--major players in the global ecosystem. *Nature Reviews Microbiology* **5**, 801–812. doi:10.1038/nrmicro1750.

Tan, M. W., Mahajan-Miklos, S. and Ausubel, F. M. (1999a). Killing of Caenorhabditis elegans by Pseudomonas aeruginosa used to model mammalian bacterial pathogenesis. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 715–720.

Tan, M. W., Rahme, L. G., Sternberg, J. A., Tompkins, R. G. and Ausubel, F. M. (1999b). Pseudomonas aeruginosa killing of Caenorhabditis elegans used to identify P. aeruginosa virulence factors. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 2408–2413.

Tock, M. R. and Dryden, D. T. F. (2005). The biology of restriction and anti-restriction. *Current opinion in microbiology* **8**, 466–472. doi:10.1016/j.mib.2005.06.003.

Touchon, M. and Rocha, E. P. C. (2010). The small, slow and specialized CRISPR and anti-CRISPR of Escherichia and Salmonella. *PloS one* **5**, e11126. doi:10.1371/journal.pone.0011126.

Touchon, M., Charpentier, S., Clermont, O., Rocha, E. P. C., Denamur, E. and Branger, C. (2011). CRISPR distribution within the Escherichia coli species is not suggestive of immunity-associated diversifying selection. *Journal of Bacteriology* **193**, 2460–2467. doi:10.1128/JB.01307-10.

Touchon, M., Charpentier, S., Pognard, D., Picard, B., Arlet, G., Rocha, E. P. C., Denamur, E. and Branger, C. (2012). Antibiotic resistance plasmids spread among natural isolates of Escherichia coli in spite of CRISPR elements. *Microbiology* **158**, 2997–3004. doi:10.1099/mic.0.060814-0.

Tyler, J. S., Mills, M. J. and Friedman, D. I. (2004). The operator and early promoter region of the Shiga toxin type 2-encoding bacteriophage 933W and control of toxin expression. *Journal of Bacteriology* **186**, 7670–7679. doi:10.1128/JB.186.22.7670-7679.2004.

Uc-Mass, A., Loeza, E. J., la Garza, de, M., Guarneros, G., Hernández-Sánchez, J. and Kameyama, L. (2004). An orthologue of the cor gene is involved in the exclusion of temperate lambdoid phages. Evidence that Cor inactivates FhuA receptor functions. *Virology* **329**, 425–433. doi:10.1016/j.virol.2004.09.005.

Vaca-Pacheco, S., Paniagua Contreras, G. L., García González, O. and la Garza, de, M. (1999). The clinically isolated FIZ15 bacteriophage causes lysogenic conversion in Pseudomonas aeruginosa PAO1. *Current microbiology* **38**, 239–243.

van Duijn, E., Barbu, I. M., Barendregt, A., Jore, M. M., Wiedenheft, B., Lundgren, M., Westra, E. R., Brouns, S. J. J., Doudna, J. A., van der Oost, J. and Heck, A. J. R. (2012). Native tandem and ion mobility mass spectrometry highlight structural and modular similarities in clustered-regularly-interspaced shot-palindromic-repeats (CRISPR)-associated

protein complexes from Escherichia coli and Pseudomonas aeruginosa. *Molecular & cellular proteomics : MCP* **11**, 1430–1441. doi:10.1074/mcp.M112.020263.

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 genetics* **9**, e1003454. doi:10.1371/journal.pgen.1003454.

Vica Pacheco, S., García González, O. and Paniagua Contreras, G. L. (1997). The lom gene of bacteriophage lambda is involved in Escherichia coli K12 adhesion to human buccal epithelial cells. *FEMS Microbiology Letters* **156**, 129–132.

Vostrov, A. A., Vostrukhina, O. A., Svarchevsky, A. N. and Rybchin, V. N. (1996). Proteins responsible for lysogenic conversion caused by coliphages N15 and phi80 are highly homologous. *Journal of Bacteriology* **178**, 1484–1486.

Waldor, M. K. and Mekalanos, J. J. (1996). Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science* **272**, 1910–1914.

Wang, P. W., Chu, L. and Guttman, D. S. (2004). Complete sequence and evolutionary genomic analysis of the Pseudomonas aeruginosa transposable bacteriophage D3112. *Journal of Bacteriology* **186**, 400–410.

Wang, X., Kim, Y., Ma, Q., Hong, S. H., Pokusaeva, K., Sturino, J. M. and Wood, T. K. (2010). Cryptic prophages help bacteria cope with adverse environments. *Nature Communications* **1**, 147. doi:10.1038/ncomms1146.

Weinberger, A. D. and Gilmore, M. S. (2012). CRISPR-Cas: to take up DNA or not-that is the question. *Cell host & microbe* **12**, 125–126. doi:10.1016/j.chom.2012.07.007.

Weldon, S. R., Strand, M. R. and Oliver, K. M. (2013). Phage loss and the breakdown of a defensive symbiosis in aphids. *Proceedings Biological sciences / The Royal Society* **280**, 20122103. doi:10.1098/rspb.2012.2103.

Westra, E. R., Pul, U., Heidrich, N., Jore, M. M., Lundgren, M., Stratmann, T., Wurm, R., Raine, A., Mescher, M., Van Heereveld, L., Mastop, M., Wagner, E. G. H., Schnetz, K., van der Oost, J., Wagner, R. and Brouns, S. J. J. (2010). H-NS-mediated repression of CRISPR-based immunity in Escherichia coli K12 can be relieved by the transcription activator LeuO. *Molecular Microbiology* **77**, 1380–1393. doi:10.1111/j.1365-2958.2010.07315.x.

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 genetics* **9**, e1003742. doi:10.1371/journal.pgen.1003742.

Westra, E. R., Swarts, D. C., Staals, R. H. J., Jore, M. M., Brouns, S. J. J. and van der Oost, J. (2012a). The CRISPRs, they are a-changin': how prokaryotes generate adaptive immunity. *Annual Review of Genetics* **46**, 311–339. doi:10.1146/annurev-genet-110711-

155447.

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., de Vos, W. M., Dame, R. T., de Vries, R., Brouns, S. J. J. and van der Oost, J. (2012b). CRISPR Immunity Relies on the Consecutive Binding and Degradation of Negatively Supercoiled Invader DNA by Cascade and Cas3. *Molecular Cell* **46**, 595–605. doi:10.1016/j.molcel.2012.03.018.

Wiedenheft, B., Lander, G. C., Zhou, K., Jore, M. M., Brouns, S. J. J., van der Oost, J., Doudna, J. A. and Nogales, E. (2011a). Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature* **477**, 486–489. doi:10.1038/nature10402.

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., Heck, A., Boekema, E. J., Boekema, E., Dickman, M. J., Dickman, M. and Doudna, J. A. (2011b). RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proceedings of the National Academy of Sciences* **108**, 10092–10097. doi:10.1073/pnas.1102716108/-/DCSupplemental.

Wiedenheft, B., Zhou, K., Jinek, M., Coyle, S. M., Ma, W. and Doudna, J. A. (2009). Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense. *Structure (London, England : 1993)* **17**, 904–912. doi:10.1016/j.str.2009.03.019.

Winstanley, C., Langille, M. G. I., Fothergill, J. L., Kukavica-Ibrulj, I., Paradis-Bleau, C., Sanschagrin, F., Thomson, N. R., Winsor, G. L., Quail, M. A., Lennard, N., Bignell, A., Clarke, L., Seeger, K., Saunders, D., Harris, D., Parkhill, J., Hancock, R. E. W., Brinkman, F. S. L. and Levesque, R. C. (2009). Newly introduced genomic prophage islands are critical determinants of in vivo competitiveness in the Liverpool Epidemic Strain of Pseudomonas aeruginosa. *Genome Research* **19**, 12–23. doi:10.1101/gr.086082.108.

Wommack, K. E. and Colwell, R. R. (2000). Virioplankton: viruses in aquatic ecosystems. *Microbiology and molecular biology reviews : MMBR* **64**, 69–114.

Wong, C. S., Jelacic, S., Habeeb, R. L., Watkins, S. L. and Tarr, P. I. (2000). The risk of the hemolytic-uremic syndrome after antibiotic treatment of Escherichia coli O157:H7 infections. *The New England journal of medicine* **342**, 1930–1936. doi:10.1056/NEJM200006293422601.

Yamamoto, T., Obana, N., Yee, L. M., Asai, K., Nomura, N. and Nakamura, K. (2014). SP10 Infectivity Is Aborted after Bacteriophage SP10 Infection Induces nonA Transcription on the Prophage SPβ Region of the Bacillus subtilis Genome. *Journal of Bacteriology* **196**, 693–706. doi:10.1128/JB.01240-13.

Yang, Y., Kurokawa, T., Takahama, Y., Nindita, Y., Mochizuki, S., Arakawa, K., Endo, S. and Kinashi, H. (2011). pSLA2-M of Streptomyces rochei is a composite linear plasmid characterized by self-defense genes and homology with pSLA2-L. *Bioscience, biotechnology, and biochemistry* **75**, 1147–1153.

Yasmin, A., Kenny, J. G., Shankar, J., Darby, A. C., Hall, N., Edwards, C. and Horsburgh, M. J. (2010). Comparative genomics and transduction potential of Enterococcus faecalis temperate bacteriophages. *Journal of Bacteriology* **192**, 1122–1130. doi:10.1128/JB.01293-09.

Yosef, I., Goren, M. G. and Qimron, U. (2012). Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. *Nucleic Acids Research* **40**, 5569–5576. doi:10.1093/nar/gks216.

Young, J. C., Dill, B. D., Pan, C., Hettich, R. L., Banfield, J. F., Shah, M., Fremaux, C., Horvath, P., Barrangou, R. and Verberkmoes, N. C. (2012). Phage-induced expression of CRISPR-associated proteins is revealed by shotgun proteomics in Streptococcus thermophilus. *PloS one* **7**, e38077. doi:10.1371/journal.pone.0038077.

Zegans, M. E., Wagner, J. C., Cady, K. C., Murphy, D. M., Hammond, J. H. and O'Toole, G. A. (2009). Interaction between bacteriophage DMS3 and host CRISPR region inhibits group behaviors of Pseudomonas aeruginosa. *Journal of Bacteriology* **191**, 210–219. doi:10.1128/JB.00797-08.

Zhang, Y., Heidrich, N., Ampattu, B. J., Gunderson, C. W., Seifert, H. S., Schoen, C., Vogel, J. and Sontheimer, E. J. (2013). Processing-Independent CRISPR RNAs Limit Natural Transformation in Neisseria meningitidis. *Molecular Cell* **50**, 488–503. doi:10.1016/j.molcel.2013.05.001.

Zhao, Y., Temperton, B., Thrash, J. C., Schwalbach, M. S., Vergin, K. L., Landry, Z. C., Ellisman, M., Deerinck, T., Sullivan, M. B. and Giovannoni, S. J. (2013). Abundant SAR11 viruses in the ocean. *Nature* **494**, 357–360. doi:10.1038/nature11921.