2021

Accessory Nucleases Provide Robust Antiparasite Immunity for Type III CRISPR-Cas Systems

Jakob Træland Rostøl

Accessory nucleases provide robust anti-parasite immunity for type III CRISPR-Cas systems

A Thesis Presented to the Faculty of
The Rockefeller University
in Partial Fulfilment of the Requirements for
the degree of Doctor of Philosophy

By
Jakob Træland Rostøl
June 2021

Accessory nucleases provide robust anti-parasite immunity for type III CRISPR-Cas systems

Jakob Træland Rostøl, Ph.D.
The Rockefeller University 2021

To protect against parasites like bacteriophages and plasmids, bacteria employ diverse and sophisticated defence systems. Clustered, regularly interspaced short palindromic repeats (CRISPR)-Cas systems are adaptive immune systems that can integrate short "spacers" from a parasite into its CRISPR locus as a form of immunological memory. Upon reinfection, short RNAs transcribed from the CRISPR locus can guide Cas proteins to the viral genome through complementary base pairing. Cas nucleases then destroy the invader's genome. To date, six major types and multiple subtypes of CRISPR systems exist, each with their own signature genes and mechanisms of action.

Type III CRISPR systems are uniquely able to destroy both the parasite's DNA and RNA. Type III loci contain Cas10 and Csm2-5, which make up the main Cas10-Csm targeting complex. In addition, loci typically contain an ancillary RNase, *csm6* or csx1. Upon target transcription, the Cas10-Csm complex recognises a viral transcript containing a target, which activates DNase activity of Cas10, leading to the destruction of the invader. In addition, it was recently discovered that the Palm domain of Cas10 can synthesise cyclic oligoadenylate second messengers (cA). cA can activate Csm6 by binding to the latter's CARF domain.

In this work, I first elucidate and illuminate the role and mechanism of action of Csm6 during anti-plasmid immunity in staphylococci. I show that Csm6 is required for efficient immunity against a weakly transcribed target but is dispensable against a well-transcribed target. Moreover, *in vivo*, Csm6 is a non-specific RNase, targeting both host and invader transcripts. This induces a transient growth arrest in the host cell, which is relieved upon target clearance. This growth arrest "buys time" for the Cas10-Csm complex to eliminate the plasmid, which is required for clearance against weakly transcribed targets.

Further, I expand and characterise broader arsenal of cA-activated CARF genes that type III systems use during immunity. I identify Card1, a nuclease that can degrade both ssDNA and ssRNA *in vitro*. These activities required divalent cations, and were activated by cA_4. In *Staphylococcus aureus*, Card1 induces a growth arrest upon activation, and enhance anti-phage immunity. The protection is most likely primarily

through the ssDNase activity, since no RNA degradation was detected *in vivo*. Together with collaborators, we were also able to solve the crystal structure of apo-, cA_4-, and cA_6-bound Card1 structures, revealing the conformational changes allowing catalysis upon ligand binding.

I also identify TM-1, a transmembrane helix-CARF gene that also causes a growth arrest in *S. aureus* when stimulated by cA production. The mechanism of TM-1 remains to be elucidated, but likely represents the first CRISPR protection mechanism not mediated by degrading nucleic acid.

Altogether, my work both deepens and broadens our understanding of the ligand-mediated immune response of type III CRISPR systems. Robust immunity is obtained by coupling specific invader destruction (Cas10 DNase activity) with non-specific host and parasite growth arrest (Csm6/Card1/TM-1). This serves as a broader paradigm of how bacteria can use different catalytic activities and different systems to resist their parasites.

Til mams og paps

Acknowledgements

The Roman statesman and philosopher Seneca said “Luck is what happens when preparation meets opportunity”. In my case, my “luck” during my graduate career is what happens when a moderately prepared college graduate meets an inexhaustible level of opportunity. And by opportunity, I mean all the support, encouragement, and resources provided to me in the Marraffini laboratory at the Rockefeller University. The scientific, social, and administrative framework provided to me has been essential to where I am today, and to whatever scientific endeavours I embark on next.

I must start, as is right, with my PhD supervisor, Dr. Luciano Marraffini. The Greek historian and biographer Plutarch said “The mind is not a vessel to be filled, but a fire to be kindled”. Luciano has indeed started the metaphorical fire for my literal scientific career. Luciano has been an unyielding supporter of my ideas, with the door always open to discuss experiments and projects. I had an organic progression from a project suggestion assigned to me when I joined, to self-started projects I was emboldened to pursue once I had the necessary knowledge and ideas. Whatever I did, I had an ally and a sober voice advising on what was best for my short-term success, my long term learning, and my sanity. Luciano is extremely talented at seeing the context of a result in the larger picture, writing it in a convincing way, and wrapping it all up in a nice bow. Also, beyond the strictly scientific, Luciano’s top priority is to facilitate the success of his lab members, caring about the short- and long-term well-being of his employees. I will undoubtedly take these lessons with me going forward, and I hope some day to be as inspirational to young scientists as Luciano was to me.

I must also express my deepest gratitude to my committee, Drs. Shixin Liu, Seth Darst, and Charles Rice, all of whom I am honoured to have interacted with during my time at Rockefeller. Most students at Rockefeller pick two committee members, but I was only content to involve three, seeking out the expertise each member could uniquely provide. Being in a laboratory that already knows their way around bacterial genetics, I sought out complementary viewpoints and skills. Shixin can think in a manner unfamiliar to me, where molecules effect the biological functions individually, not in bulk. This is also true for Seth, whose proficiency in transcription (highly relevant to type III CRISPR) and biochemistry was of great help. Charlie, while lacking specialisation in bacteria and phages, can clearly see similar principles from his work in eukaryotic viruses and their hosts. Through my annual faculty committee meetings, I was reassured I was on the right track, while also thoroughly enjoying the meetings themselves.

Further, I want to thank Dr. Joseph Mougous for agreeing to be my external examiner. I sought out a microbiologist not yet corrupted by the CRISPR field, and certainly value his input on my work. His work is a stark reminder that it is not all a bacteria vs. phage world, when in fact bacteria also are each others’ worst enemies. I am saddened that a certain global pandemic prevents Joseph from being here in person, but I am glad that he can join via Zoom to hear what I have spent my last 5.5 years on. I will be keeping a close eye on what beautiful biology will come from his lab next.

Of course I must acknowledge the wonderful administrative and moral support originating from the Dean's office during my time at Rockefeller. Dr. Sid Strickland, Dr. Emily Harms, Dr. Andrea Morris, Cristian Rosario, Stephanie Hernandez, Marta Delgado, and Kristen Cullen have all provided a seamless and wonderful graduate experience. It was always a joy to visit their offices, and I knew I had allies if anything went awry in my lab (though it luckily never came to that!). A considerable regret of my last year is that a certain pandemic prevents me from playing a round of celebratory pool with Cris in Faculty Club upon graduating.

Another crucial aspect of my Rockefeller tenure is the Faculty and Students Club, an on-campus bar and lounge where I killed many an incubation period and sought out an experimental breather. The Greek philosopher Aristotle said "Virtue is the golden mean between two vices, the one of excess and the other of deficiency". In my case, this means that Faculty Club was an invaluable counterweight to the rigours of lab life. Faculty Club provided a refuge during the highs and lows of a PhD, and I made many a good friend, had many a good scientific conversation, and won or lost many a good pool game within the four walls of this sanctuary. I sincerely hope this oasis will continue to provide a refuge to bewildered Rockefeller scientists.

I must also acknowledge the terrific collaborators I have engaged with. Firstly, the laboratory of Dr. Martin Jinek, with Ole Niewoehner and Dr. Carmela Garcia-Doval, allowed me to partake in the exciting discovery of a second messenger-mediated type III CRISPR response. Later, I worked with the laboratory of Dr. Dinshaw Patel, and his talented post-doc Dr. Wei Xie, to delve deep into the workings of one of my favourite proteins, Card1. These multi-disciplinary approaches to more deeply understand our problems was of great use and excellent enjoyment, and I will definitely do my best to foster collaborations with scientists of different expertises in the future.

This section would be thoroughly incomplete without a mention of the fantastic colleagues I have encountered in the Marraffini lab. The Roman statesman and philosopher Cicero said "Friendship improves happiness and abates misery, by the doubling of our joy and the dividing of our grief". If there is anything you need during a PhD, it is the reassurance that a failed experiment is not your fault (abating misery), and being happy for someone else when something succeeds. From the moment I stepped my un-sure-footed Norwegian foot into the lab, I have experienced nothing but encouragement, but with a healthy dose of not taking anything too seriously. The Marraffini lab has been an amazing place to do science, with great discussions, a collaborative atmosphere, and a welcoming environment. I wish everyone the best of luck with their future endeavours.

My greatest scientific influences in lab are Dr. Alexander Meeske and Dr. Andrew Varble. I always enjoyed sparring with ideas with them, and much of my work is inspired from interactions with them, and I am excited to see what they do next. Dr. Charlie Mo

has also provided invaluable assistance with biochemistry and interesting type III conversations. Same goes for Dr. Poulami Samai, an early type III guru who helped me significantly, and whose work I built on. Dr. Joshua Modell and Dr. Naama Aviram have also been fine colleagues and mentors, contributing joy and jokes as well as hard scientific guidance.

The Marraffini lab has also enjoyed an unbroken string of talented and kind students that I have had the pleasure of working alongside. Dr. Wenyan Jiang and Dr. Gregory Goldberg blazed the type III CRISPR trail before I arrived, and helped me settle in. Dr. Jon McGinn supervised my rotation, and was of great help then and since. Dr. Philip Nussenzweig has offered fun scientific discussions mixed with clever/silly jokes, and we performed many a custom birthday song for lucky lab members. Amer Hossain provided friendship in lab and beyond, teaching me the intricacies of lambda gam and football, being the only one who could beat me at the penalty shootout competition. Dalton Bahn indulged my interests in vaporwave and PICIs, two topics where his knowledge is superior to mine, but that he was happy to educate me on. Claire Kenney is an unmitigated ray of sunshine in the lab environment, spreading her contagious positivity. Amanda Shilton is a type III loyalist who also happens to spread joy with her baking and glitter decorations. Finally, new students (and postdocs) Christian Baca, Gianna Stella, Hyejin Kim, and Dr. Maj Brodmann have recently enriched our lab, making me optimistic about the lab's future for after I'm gone. I also extend my gratitude to three rotation students I had the privilege of supervising, Kevin Kao, Ruby Froom, and Lin Mei, all of whom contributed to our projects and helped me progress on my projects.

I must also acknowledge the contributions by our lab managers and helpers, Rahul Bholse, Allison Richards, and Phone Ko, for helping the lab run smoothly, facilitating my reagent orders, media and competent cells consumption, and lab lunch orders. Four great lab technicians, Albina Kozlova, Jacob Mathai, Ashley Thornal, and Jessica Fyodorova, have also, in addition to working on their own projects, helped lab operations, as well as making the lab a younger and hipper.

Socrates said "Be slow to fall into friendship; but when thou art in, continue firm and constant". In this context, I must thank Pascal Maguin, Dr. Nora Pyenson, and Dr. Robert Heler, who have been great colleagues as well as close friends during my whole PhD career. They are amazing people who have always been there for me in and outside of lab, and my last six years would have been very different (and much worse) without them.

I must also thank Rockefeller friends outside the lab for making my experience at Rockefeller what it was, which include Eduardo Aguilar, Dr. Brandon Razooky, Dr. Emily Lorenzen, and my roommate Elias Scheer. They have provided much needed comfort, support, and distractions that helped make the time during my PhD and in New York an unforgettable experience.

Moreover, I was very privileged to receive support from the Boehringer Ingelheim Fonds during my PhD. This is run by talented and dedicated people who provided terrific fellowship meetings with other fellows, where I enhanced my communications skills and was able to meet other European PhD students, forging a network I will be able to use for my future research career.

Alexander the Great said “I am indebted to my father for living, but to my teacher for living well”. I owe an unpayable debt to all my teachers and mentors who have helped me so far. Of these, two in particular stand out. Liina Luukkonen was my high school International Baccalaureate coordinator. She was herself an Oxford graduate, and saw something in me that I could not yet see in myself, and I might not have studied outside of Norway, much less in Oxford, had it not be for her. She casts an inspirational shadow on whatever I have done since, making me believe in myself and try to make it as far as I can. Dr. Maureen Taylor, biochemistry tutor at Exeter College in Oxford, interviewed me, accepted me, and reared me to pursue science further. Her unyielding support guided me through my undergraduate degree, and made me recognise that science is not fun for the facts you drill into your head, but for the interesting connections you make with said facts, and the elegant experiments that led to these insights. Maureen’s invisible hand also guided me to Rockefeller through teaching science using seminal papers from Rockefeller scientists, including the aforementioned Dr. Seth Darst. And of course, her not-so-invisible hand wrote me a recommendation letter that got me to Rockefeller. Needless to say, I will never be able to uncouple my PhD careers and the two mentors that got me to my PhD. I wish upon anyone to benefit from as fruitful mentorship as I have.

Finally, I have to thank my family. My sisters Hanna and Vilde, and my parents Liv and Georg, have been nothing but supportive of my journey. My parents never commanded me to do anything academic, making me discover my own interests and strengths organically, perhaps explaining why I enjoy what I do to the extent that I do. My sisters always tolerated me being a nerd, but helped me understand there was more to life than academic pursuit, keeping my feet planted on the ground. My parents endured me going to foreign lands far away for my passion, even if that meant they only saw my face through a laptop camera most of the time. If it meant me getting to know my niece through a computer screen. Only through their humanity, patience, and love could I have done what I have done, academically and beyond. They are the shoulders of giants upon which I stand today. Further, extended family support from tante Turi, Torbjørg, Rune, and Bjørnar has been invaluable. I also want to dedicate this work to my late grandparents, Anna Træland and Sverre Rostøl, both of whom I would have loved to share this achievement with.

Table of contents

List of figures

List of tables

List of abbreviations

Δspc	Delta spacer, i.e. a CRISPR system lacking a targeting spacer
Abi	Abortive infection
aTc	Anhydrous tetracycline
BHI	Brain heart infusion media
bp	Base pairs
c-di-AMP	cyclic-di-adenosine monophosphate
cA	Cyclic oligoadenylate
cA_4	Cyclic tetraadenylates
cA_6	Cyclic hexaadenylates
Card1	Cyclic oligoadenylate-acticated RNase and DNase 1
CARF	CRISPR-Associated Rossman fold
Cas	CRISPR-associated
$Cas10^{HD}$	Cas10 HD domain mutant (H14A+D15A)
$Cas10^{Palm}$	Cas10 Palm domain mutant (D586A+D587A)
CBASS	cyclic-oligonucleotide-based anti-phage signalling systems
cfu	colony forming units
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
crRNA	CRISPR RNA
dCsm6	catalytically dead Csm6 (R364A+H369A)
dCard1	catalytically dead Card1 (E308A+K310A)
dsDNA	double-stranded DNA
dsDNase	double stranded DNase
HMC	hydroxymethylcytosine
HEPN	higher eukaryotes and prokaryotes nucleotide-binding
MOI	multiplicity of infection
NEB	New England Biolabs
nt	nucleotide
OD_{600}	optical density at wavelength of 600 nanometers
OMV	outer membrane vesicle
pAgo	prokaryotic Argonaute
PAM	protospacer-adjacent motif
pCRISPR	plasmid carrying a CRISPR system
PFU	plaque forming units
Phage	bacteriophage
PICI	Phage-Inducible Chromosomal Island
PLE	PICI-like elements
pTarget	plasmid carrying an aTc-inducible protospacer target

REase	restriction endonuclease
RM	restriction modification
RPM	reads per million
RT	Reverse Transcriptase
SaPI	*S. aureus* Pathogenicity Islands
s.e.m.	standard error of the mean
ssDNA	single-stranded DNA
ssDNase	single-stranded DNase
ssRNA	single-stranded RNA
ssRNase	single-stranded RNase
Sie	superinfection exclusion
TA	toxin-antitoxin
TM-1	Transmembrane 1, a CARF-TM gene from the Nitrosococcus halophilus type III-A CRISPR-Cas system
TSB	tryptic soy broth medium
WT	wild-type
E. coli	*Escherichia coli*
L. lactis	*Lactobacillus lactis*
S. epidermidis	*Staphylococcus epidermidis*
S aureus	*Staphylococcus aureus*
T. succinifaciens	*Treponema succinifaciens*
V. cholerae	*Vibrio cholerae*
P. aeruginosa	*Pseudomonas aeruginosa*

Chapter I. The bacterium-parasite arms race

Throughout evolution, bacteria have been preyed upon by parasitic bacteriophages (phages). Everywhere bacteria are found, they coexist with their respective phages, undergoing continuous cycles of infection. As a consequence, in order to survive and thrive, bacteria have developed an arsenal of anti-phage mechanisms which they deploy to prevent phage infection. Due to the immense evolutionary pressure imposed by phages, the diversity and sophistication of bacteria's anti-phage mechanisms are astounding, and we are only now beginning to appreciate the complexity of the interactions between bacteria and their parasites.

Since their discovery approximately 100 years ago, phages have been central players in the development of modern biology. As well as providing insights into bacterial ecology and evolution, phages have acted as probes to ask basic molecular biology questions. In addition, the study of anti-phage mechanisms has resulted in invaluable tools, like restriction enzymes and, more recently, revolutionary CRISPR-based gene editing and diagnostic techniques. With the advent of new technologies, genomics, and renewed interest in phage therapy due to the rise of antibiotic resistance, the study of anti-phage mechanisms has recently experienced a revival. This chapter aims to summarise the means by which bacteria use to resist their phages, with an emphasis of novel developments in the field.

1.1 Phages and bacteriophage-host interactions

Phages are obligate intracellular parasites of bacteria. They bind to their hosts through specific receptors, and then hijack the cellular machinery to replicate. Phages are ubiquitous, estimated to outnumber their hosts in most contexts [1], and play crucial roles in bacterial evolution and ecology. They are extremely diverse [2] and have profound effects in the different ecological niches they occupy, including the human microbiota [3]. Finally, temperate phages that can undergo lysogeny, staying dormant and replicating passively with their host, can alter host physiology, potentially providing novel genes and beneficial traits [4].

The red queen hypothesis states that an organism must constantly evolve to maintain their relative fitness in the face of a predator [5]. In the context of the bacteria-phage relationship, this means that bacteria continuously evolve and update anti-phage mechanisms, while phages adapt to overcome these mechanisms (Fig. 1.1). Competitive bacteria-phage coevolution, often referred to as an "evolutionary arms race", has produced a multitude of bacterial defence mechanisms that act to inhibit every stage of

the phage life cycle (Fig. 1.2). Although not discussed extensively in this review, phages have developed as many means to circumvent these defence strategies. As a result, in nature, bacteria and phages seem under many circumstances to coevolve and exist in stable equilibria without dramatic fluctuations or extinction events [6].

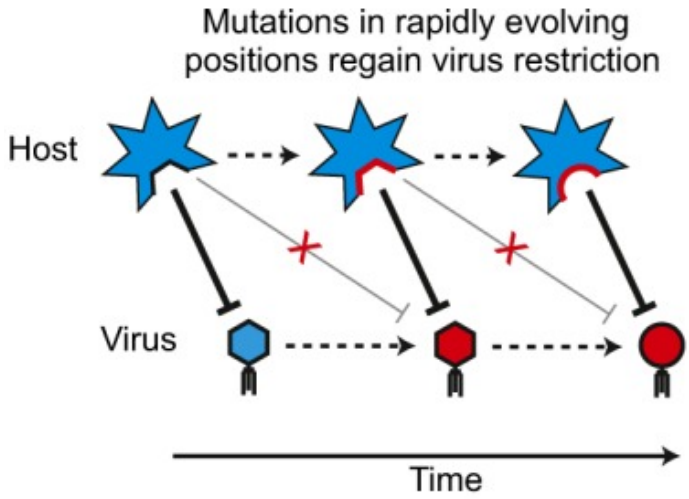

Figure 1.1. The red queen hypothesis of host-parasite co-evolution. *The virus, e.g. a phage, can mutate to no longer be recognized by the immune component of the host. In order to keep up, the host factor must also mutate. With time, successive rounds of virus evasion and host response drives the evolution of both players. Adapted from* [5]

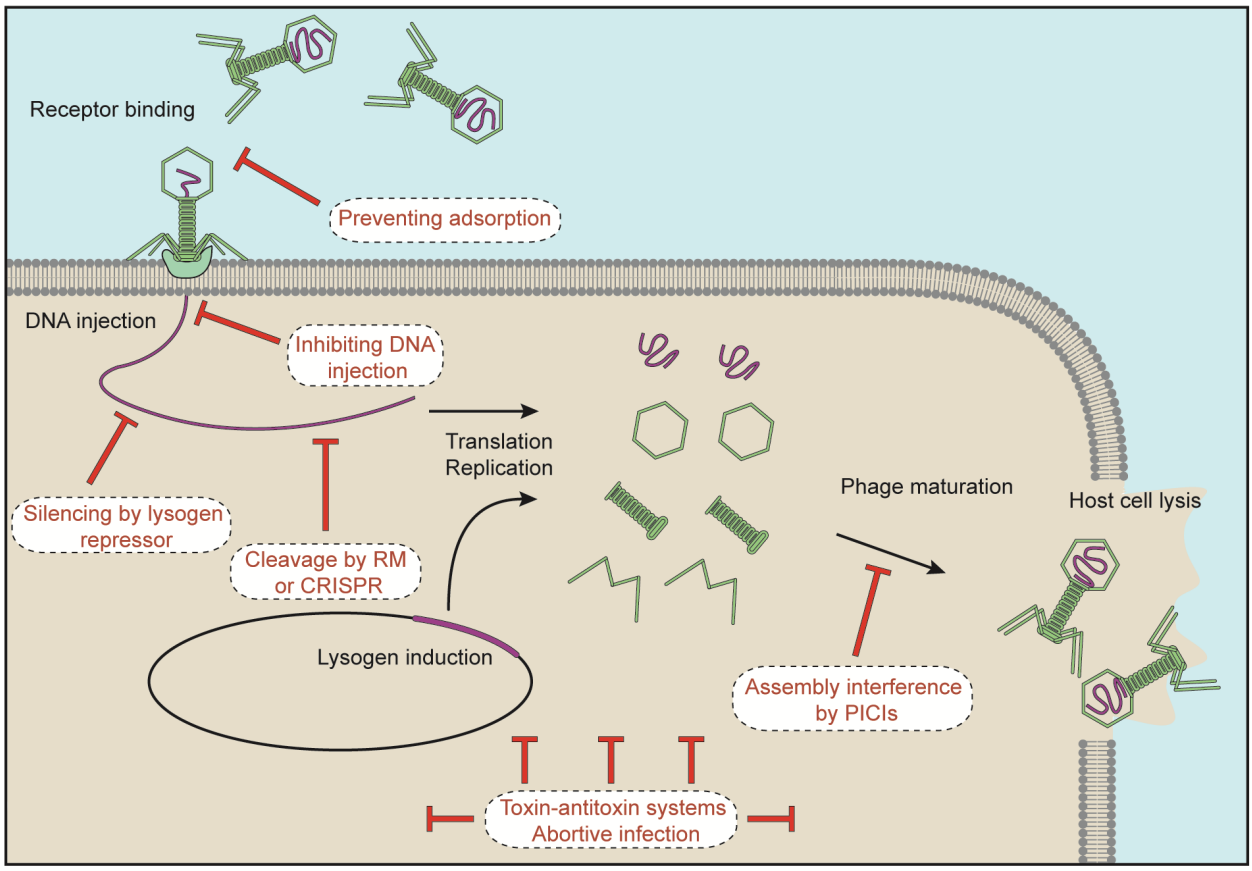

Figure 1.2. An overview of the phage life cycle, indicating at which stages various anti-phage mechanisms act. *Upon recognising a surface receptor, a phage injects its genome into the host cell. Alternatively, an incoming phage can enter lysogeny (not shown). The infecting phage (or excised lysogen) expresses phage genes, making proteins for new phage particles, as well as replicating its genome. New phages are then assembled, and the host cell is lysed to release phage particles. The host cell can interfere with all of these steps, as indicated by the red inhibitory arrows.*

Although more defence mechanisms should intuitively provide more robust protection, bacteria only tend to have a subset of the available diversity of anti-phage mechanisms. For example, while almost all bacteria have restriction-modification (RM) systems, only about 50% possess CRISPR systems [7] (and substantially less than 50% in some environments [8]). This, and other observations hints that there are fitness costs associated with carrying immune systems. Indeed, there are energetic burdens associated with expressing each system, and many systems are prone to low but significant levels of autoimmunity. Inhibiting incoming parasites also prevents uptake of potentially useful DNA by horizontal gene transfer, both from phages and plasmids. Additionally, altering or losing a phage receptor (e.g. bacterial pili) can, while preventing phage infection, also

result in loss of receptor function and lowered fitness overall. Bacterial defence systems are prone to horizontal gene transfer, are often found on plasmids, and are relatively easily gained or lost [9]. Bacteria must therefore tune the fitness trade-off between the cost of carrying anti-phage systems and the benefit of resisting phage infection [10]. In addition, immune systems can be regulated to minimize this cost, e.g. only turned on when phage infection is likely.

1.2 Preventing phage entry

For the phage to begin an infection cycle, it must bind to a susceptible host (adsorb) and inject its genome. Phages require specific exposed cellular proteins or cell wall components, and binding to the right receptor initiates genome transfer from phage to host cell. Bacteria use a range of mechanisms to prevent phage adsorption (Fig. 1.3) and injection, both broad and phage-specific, to avoid phage access to the bacterial cytoplasm.

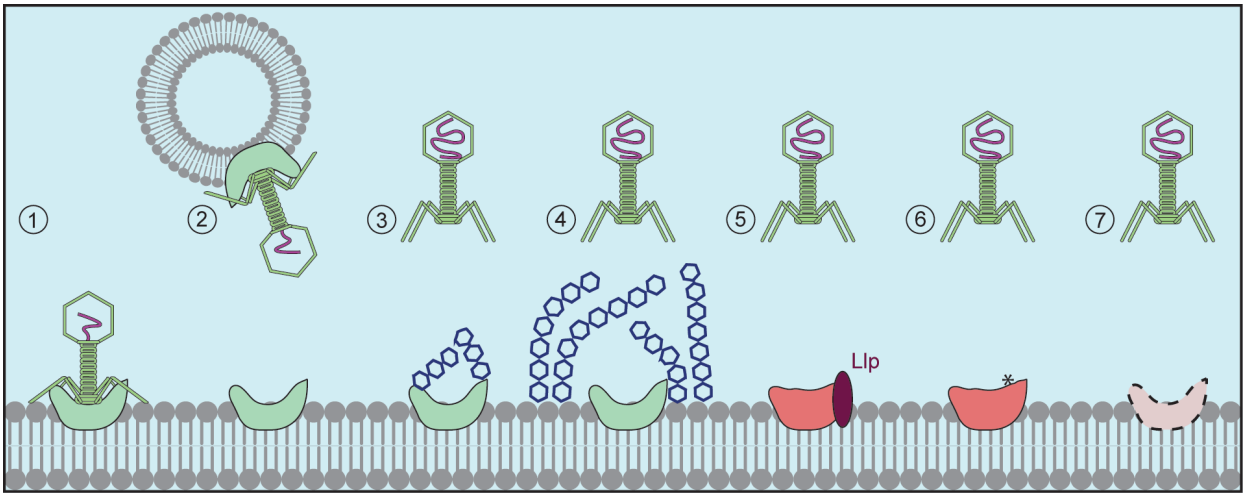

Figure 1.3. Examples of mechanisms bacteria employ to prevent phage adsorption
1. A successful binding of the phage to its receptor (green). 2. A phage particle is sequestered by OMVs containing the phage receptor (e.g. E. coli; V. cholerae). 3. The receptor is directly glycosylated to prevent phage binding (e.g. type IV pilus glycosylation in P. aeruginosa). 4. The bacterium hides its surface inside a peptidoglycan capsule (e.g. E. coli using a K1 capsule made from polysialic acid to prevent T7 infection). 5. Another protein changes the conformation of a phage receptor, which can no longer bind phage (e.g. phage T5-expressed Llp binding FhuA to prevent superinfection of other T5 virions). 6. Mutations in the receptor to abolish phage binding (e.g. LamB mutating to no longer be recognised by phage lambda). 7. The receptor is no longer expressed on the surface (e.g. Bvg⁻ phase Bordtella not expressing pertactin receptor for phage BPP-1; lysogen-encoded Tip protein preventing pilus exposure).

1.2.1 Preventing phage adsorption

Many bacteria spend much of their life cycle embedded in biofilms, an extracellular matrix made up of polymers where bacteria live in close proximity, often on surfaces. Biofilms protect bacteria in various ways, but how these structures affect phage-bacteria interactions remains incompletely understood. Computational modelling suggests that biofilms can conditionally survive and grow in the presence of phage [11]. This was also shown experimentally in *E. coli* with a virulent phage, where, depending on nutrient availability and phage infectivity, the colony size reaches equilibrium [12]. In this scenario, cells inside the colony divide, and are shielded by peripheral cells that get infected. Another study showed that while early biofilms were quickly eradicated, mature *E. coli* biofilms prevented cell killing by T7 phage [13]. Fluorescent labelling of cells and phages revealed that the biofilm structure prevents phage access to the biofilm interior. This depended on the presence of bacterial curli, through the formation of a proteinaceous matrix which promotes denser cell packing.

Many bacteria, especially gram-negatives, can secrete outer membrane vesicles (OMVs), spherical structures made up of outer membrane components and periplasmic cargo which pinch off the cell [14] . Since they contain exposed outer membrane proteins which can act as phage receptors, OMVs can act as decoys, sequestering extracellular phage. One report showed that pre-incubation with OMVs reduced T4 infectivity in *E. coli,* and phage-OMV complexes could be visualized with electron microscopy [15]. The same was done more recently with *Vibrio cholerae* OMVs interacting with three separate phages, and this neutralisation was dependent on the presence of phage receptor in the OMVs [16]. OMV-mediated phage protection is likely to be a widespread phenomenon, and future studies will further elucidate their significance to phage protection.

On the cell surface, phage receptors can be hidden or masked to be made inaccessible to incoming phage. Many phages bind structures like pili or flagella to initiate infection. *P. aeruginosa* counters this interaction through glycosylation of type IV pili, a bulky modification that prevents binding by the phage [17]. Receptors can also be blocked by polysaccharide capsules, which shield the whole bacterial surface. For example. the *E. coli* K1 capsule, made of polysialic acid, prevents T7 attachment to its receptor LPS, thereby reducing infectivity [18]. In response, phages can have enzymes in their tails that degrade various capsules, giving rise to an evolutionary arms race that results in the extreme diversification of capsule synthesis and hydrolysing enzyme genes of the host and phage, respectively.

Surface proteins can also hide phage receptors. *E. coli* lytic phage T5 uses the outer membrane iron uptake protein FhuA as its receptor. Early during infection T5 expresses the lipoprotein Llp, which binds FhuA and neutralizes its receptor properties. This prevents additional T5 particles, as well as other phages that use FhuA as receptor, such as T1 and phi80, from entering and disturbing T5's infection cycle [19]. This has the added benefit that upon lysis, newly made T5 does not bind to FhuA in the cell debris, something which would inactivate the phage particles for further infection [20]. This phenomenon is an example of a superinfection exclusion (Sie) mechanism, a process where a lysogen or intracellular phage blocks the infection of the same (homotypic) or a different (heterotypic) phage.

Finally, mutations within receptor genes that affecting the protein or its expression can prevent phage adsorption. The interaction between a phage and its receptor is highly specific, and altering the structure of the receptor can break this specific interaction. For example, the receptor for phage lambda is the mannose permease LamB, and mutating LamB can give rise to lambda-resistant strains [21]. In response to this, J tail protein mutations in lambda can restore infectivity [22], and receptor-phage interactions can co-evolve through various cycles of mutations. However, receptor mutations can incur fitness costs. For example, mutating LamB can abrogate *E. coli*'s ability to grow on mannose. For this reason, it might be more efficient to control the expression of the receptor, rendering the cell resistant to phage when the receptor is absent. This can be done through phase variation, which is a non-random genetic and heritable on/off switch of one or more genes that cells use to adapt to different environments [23]. *Bortdella* can change between a Bvg^+ state, expressing genes involved in virulence, and the Bvg^- state, which increases fitness when growing outside a host. Phage BPP-1 uses pertactin as its receptor, and this adhesion protein is only expressed in the Bvg^+ phase [24]. As a result, Bvg^- cells are 10^6-fold more resistant to infection by BPP-1. Receptor expression can also be modulated by lysogenic phages via Sie. The *P. aeruginosa* phage D3112 expresses the protein Tip, which inhibits type IV pilus polymerization [25]. Tip interacts with PilB, an ATPase essential for pilus extension, and this leads to the loss of exposed type IV pili on the cell surface. Phages that use pili as receptors, including D3112, are therefore unable to infect cells harboring this lysogen. Indeed, a systematic screen of *P. aeruginosa* Sie mechanisms identified many lysogens interfering with either type IV pilus function, or with the O-antigen, another typical *P. aeruginosa* phage entry receptor [26].

1.2.2 Preventing DNA injection

Blocking the entry of phage DNA into the cytoplasm is another mechanism of preventing phage infections, acting after of phage adhesion to the host and therefore

titrating phages out from the surroundings. The lambda-like *E. coli* phage HK97 can confer Sie against itself as well as resistance to the closely related phage HK75 [27]. gp15 is expressed during lysogeny, and is an inner membrane transmembrane protein that blocks DNA transfer across the inner membrane. Normally, the HK97 tape-measuring protein forms a pore across the inner membrane for DNA transfer from the viral capsid into the cytoplasm [28]. This process requires the help of the periplasmic chaperone FkpA, and an interaction with the glucose transporter PtsG. gp15 interferes with this process by interacting with PtsG, probably by driving the PtsG-tape measure protein complex into a non-productive conformation. Fruitloop, a mycobacteriophage that belongs to the F cluster, offers an example of heterotypic Sie working through the prevention of DNA injection of the unrelated B2 cluster phages. During the lytic cycle of Fruitloop, its protein gp52 interacts with and inactivates Wag31, an essential protein involved in cell wall synthesis at the poles [29]. This is thought to prevent DNA injection by B2 cluster phages, like Hedgerow and Rosebush, that rely on Wag31.

1.3 Targeting bacteriophage nucleic acid

Once the phage genome is injected into the host cell, it will begin a genetic programme to achieve viral propagation, or in the case of temperate phages, to integrate in the host chromosome as a prophage (lysogenic state). Prophages use a common homotypic Sie mechanism in which the repressor that prevents viral propagation prior to integration binds to incoming phage DNA, repressing its lytic programme [30]. An even more direct way to stop a phage's genetic programme is through the use of nucleases able to specifically degrade the injected genome. To achieve this, bacteria use restriction-modification (RM) systems, CRISPR-Cas immunity, and prokaryotic Argonautes.

1.3.1 Restriction-modification systems

RM systems (see [31] for a more detailed review) are ubiquitous, with many bacterial species harbouring multiple systems, and extremely diverse. RM systems are normally made up of two activities; the restriction endonuclease and the methyltransferase (the modification component). The restriction endonuclease recognizes small DNA motifs, usually 4-8 base-pairs long, and cuts the DNA, either within or away from the recognition site. These motifs exist in both host and invader, but the host protects its genome by using the methyltransferase to modify its own DNA and avoid recognition by the restriction enzyme. An invading phage is usually not methylated, and will therefore be cut upon cell entry. Occasionally, however, the methyltransferase modifies a phage restriction sites before the restriction endonuclease can cleave, resulting in immunity for the phage.

RM systems are classified into four major types, based their mechanism of action and subunit composition [31]. Both type I and III systems translocate along DNA and cleave away from the recognition sites. Type II, known for their use in molecular cloning, cleave within or near the recognition site. Type IV systems lack a methylase and only contain a restriction component, where the endonuclease specifically cleaves modified DNA. A phage might modify its DNA to avoid cleavage by types I-III RM systems, but that makes them susceptible to type IV cleavage instead. For example, T4 contains hydroxymethylcytosine (HMC) instead of cytosine in its DNA, inhibiting all RM systems that recognise sites containing cytosine. To counter this, *E. coli* uses McrBC, a type IV system specific for HMC-containing DNA [32]. In response, T4 can glycosylate its DNA, which impairs McrBC activity. Against this modification, an additional type IV system, the GmrS-GmrD system, is active [33].

Examples of "inverted" RM systems have also been found. The phage ϕC31 can infect *Streptomyces coelicolor* A2(3) harbouring the *pgl* locus but only mount one cycle of infection. The released phages are unable to reinfect Pgl^+ hosts [34], presumably due to the action of the methyltransferase *pglX* [35], which modifies the phage DNA to make it susceptible for restriction in the next Pgl+ host by an unknown mechanism. Interestingly, one of the recently described anti-phage BREX systems [36], type 2, shares two genes (*pglX* and *pglZ*) with the Pgl system. Type 2 BREX methylates DNA at a six nucleotide repeat, likely responsible for self/non-self discrimination of this system. Phage DNA is inactivated after injection prior to DNA replication by an unknown mechanism that does not involve cleavage.

1.3 2 Introduction to CRISPR-Cas systems

One of the most significant scientific advances in the last decades was the discovery of CRISPR-Cas bacterial immune systems (for a recent review, see[37]). These systems are present in approximately 50% of sequenced bacteria and 90% of sequenced archaea [7], and provide resistance against invading phages [38] and plasmids [39]. Uniquely, CRISPR systems are adaptive rather than innate immune systems, where exposure to a previous infection can be remembered. The molecular basis for immunological memory are short (30-40 base pairs) "spacer" sequences acquired form invader genomes flanked by similarly short semi-palindromic repeats. Alternating spacers and repeats form the CRISPR locus, which is transcribed and processed into small CRISPR RNA (crRNA) guides. CRISPR associated (*cas*) genes are found adjacent to the CRISPR locus, and encode the protein machinery required for the acquisition of new spacer sequences upon infection (the immunization phase) and for the sequence-specific elimination of the invader (the targeting phase). During the latter, RNA-guided Cas nucleases use the

crRNAs to find the target (also known as “protospacer”) with the invader’s genome via complementary base pairing. Based on the composition of *cas* genes, CRISPR systems can be classified into two classes, six types, and multiple subtypes [40], with diverse mechanisms of action.

1.3.3 CRISPR targeting – destroying the invader

During targeting, crRNA-guided Cas nucleases specifically identify the target and effects the destruction of the invader nucleic acid, either destroying DNA, RNA, or both, depending on the CRISPR type (Fig. 1.4). The interference complexes of types I, II, and V bind the DNA of the invader, starting by recognising the protospacer adjacent motif (PAM), followed by directional unwinding of the DNA double helix and R loop formation. If the protospacer matches the crRNA spacer, DNA cleavage proceeds. For these types, the PAM sequence is required for target binding, and viral mutations in the PAM is frequently a means for the phage to escape type II immunity [41]. In addition, base pairing in the seed sequence (approximately the first eight nucleotides next to the PAM) is essential for type I and II [42,43], while single mismatches elsewhere in the crRNA-DNA interface can be tolerated and still allow for interference.

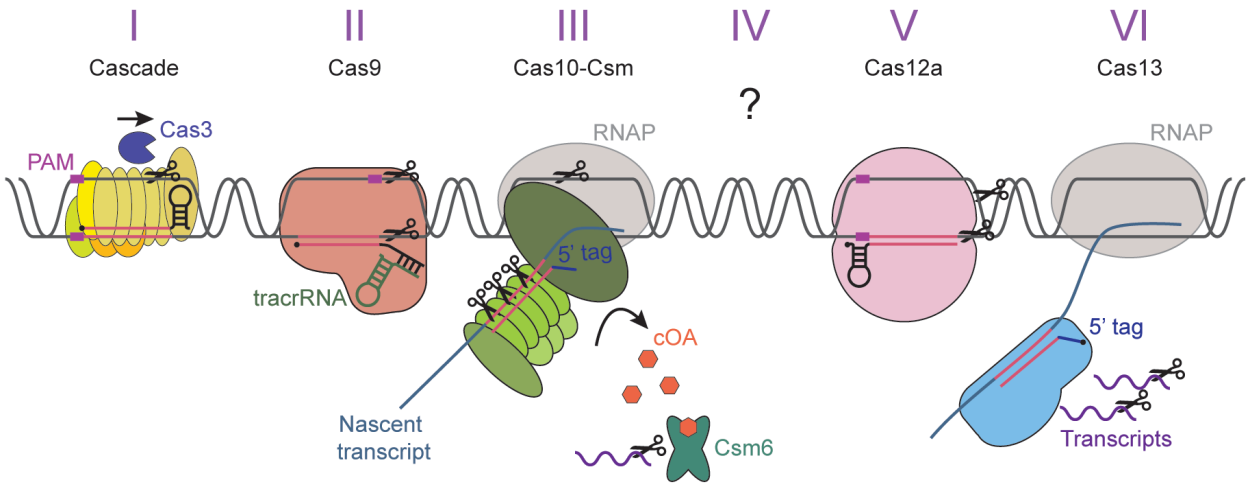

Figure 1.4. An overview of the CRISPR targeting mechanisms of the six CRISPR types. *For details, see text. PAMs are purple boxes, 5’ ends of crRNAs are black circles, spacers/protospacers are pink, and 5’ crRNA tags inhibiting type III/VI autoimmunity are blue. For types III and VI, RNA polymerase is actively transcribing, with the targeting complex recognising the nascent transcript rather than the DNA.*

In type I, the crRNA-guided complex that binds DNA is the large seahorse-shaped Cascade, which is responsible for finding the target but does not possess any catalytic activity. Conformational changes in Cascade upon DNA binding [44] allows Cas3 to be

recruited. Cas3 is a nuclease and ATP-dependent helicase, and will nick the displaced DNA strand, followed by unidirectional ATP-dependent DNA degradation [45] [46], with subsequent elimination of the phage. For type II CRISPR systems, Cas9 is the only protein required for interference. Once bound to a matching target, the HNH and RevC nuclease domains of Cas9 cleave the target and non-target strands, respectively, giving a blunt-end cut three nucleotides upstream of the PAM [43,47,48]. Type V, like type II, has a single subunit effector complex, but type V cleaves both DNA strands with the same RuvC domain [49], generating a staggered cut [50].

Type III CRISPR systems are composed of the main effector complex (Csm or Cmr for III-A or III-B, respectively), and of an accessory RNase (Csm6 or Csx1 for III-A or III-B, respectively). Transcription across the target is an absolute requirement for immunity [51]. This is because rather than recognizing the protospacer within the target DNA, the effector complex binds the protospacer within the nascent RNA through complementary RNA-RNA base-pairing. Target binding unleashes the non-specific single-stranded DNase activity in the HD domain of Cas10 (the main subunit of both the Csm and Cmr complexes), which cuts the non-template strand of the viral DNA associated with RNA polymerase [52] . In addition, target recognition results in the synthesis by the Palm domain of Cas10 of cyclic oligo-adenylates (cA), a ligand that activates the Csm6/Csx1 non-specific RNase [53,54]. This accessory RNase is required for targets that are poorly recognized by the crRNA guide and thus provide inefficient activation of the Cas10 ssDNase, such as weakly transcribed [55] or mutated targets [56]. Finally, the Csm3/Cmr4 subunit of the effector complex cleaves the protospacer RNA at six nucleotide intervals [57,58]. This cleavage event neutralizes both activities of the Cas10 subunit to restrict spatiotemporally the otherwise toxic non-specific DNase and RNase activities triggered by the Type III immune response. Type III-A targeting has no PAM or seed requirements [59], and therefore can tolerate more mismatches while still effecting immunity. As a consequence of this, Type III-A CRISPR systems offer more robust defense against a rapidly mutating phage invader than a type II [59] and Type I [60] systems, where single PAM or seed mutations can abolish targeting.

Type VI CRISPR systems are characterised by the effector protein Cas13. Uniquely, type VI only targets RNA, using an active site formed by two HEPN domains [61]. Similar to type III CRISPR systems, transcription across the target is required for interference and there is no PAM sequence requirement [62]. Binding to the target induces conformational changes that bring the two HEPN domains together [63], resulting in the cleavage of both target and non-target transcripts, the latter being the probable cause of the growth delay observed in cells undergoing Type VI CRISPR immunity [62] [61]. The role

of Cas13 in anti-phage immunity remains unclear, however. *L. shahii cas13-a* was able to confer protection against the single-stranded RNA virus MS2 in an *E. coli* heterologous host. Still, it is believed that only a minority of prokaryotic viruses are RNA viruses [64] , so protecting against RNA phages may not be the primary function of type VI CRISPR systems. Instead, it has been suggested that Cas13 activity might lead to growth arrest upon infection by DNA phages, buying time for other bacterial immune systems to clear the phage. Alternatively, programmed cell death could be induced, preventing the release of new virions from an infected host.

1.3.4 CRISPR adaptation – remembering the invader

Adaptation is the process where a short sequence from an invader is incorporated into the CRISPR array as a new spacer, offering protection against future invaders containing the same protospacer sequence (for a recent review, see [65]). Although rare in laboratory conditions, metagenomic studies have shown that bacteria actively adapt against phages in their natural environments through the acquisition of multiple spacers [66] . Diversity in the spacer repertoire of a bacterial population is important to successfully suppress phages and prevent the emergence of escape phages with mutated target sequences [67]. While the different CRISPR types have different signature effector proteins that work on the targeting phase, most share the same core genes, *cas1* and *cas2*, for spacer [40]. Adaptation has two steps; the selection of functional protospacers and their insertion into the leader end of the CRISPR array.

For the first part of adaptation, the adaptation machinery is faced with a difficult task; distinguishing self (chromosomal) from non-self (phage) DNA. Frequent adaptation against self would incur a great fitness cost to the host via self-targeting and autoimmunity. A study in *E. coli* carrying the type I-E CRISPR-Cas system [68] showed that both replication and the RecBCD complex (involved in repair of dsDNA breaks) were required for adaptation, with adaptation hotspots in regions where dsDNA breaks are common, like replication forks. More recently, it was demonstrated that the adaptation machinery of the type II-A CRISPR-Cas system preferentially samples the free DNA end injected by the staphylococcal phage ϕ12 [69], also requiring AddAB, the Gram-positive RecBCD ortholog. Both of these studies suggest that free DNA ends, generated either through breaks during replication or during injection of the phage genome, are captured by the Cas machinery as “soon-to-be” spacers (also known pre-spacers). Resection by the RecBCD dsDNA repair pathway generates more fragments with free DNA ends that amplifies the adaptation process.

Another obstacle faced by the adaptation machinery is to select targets flanked by a proper PAM sequence to produce functional viral targets. In type II-A systems, mutations within Cas9 that abrogate PAM recognition result in the acquisition of random, non-functional spacer sequences; i.e. that do not match a protospacer followed by a PAM [70]. This suggests that Cas9, which forms a complex with Cas1, Cas2, and Csn2 (all required for spacer acquisition), is responsible for the selection PAM-adjacent spacers. Recently, three studies identified a similar mechanism for type I systems harbouring the *cas4* gene [71-73]. The Cas4 protein helps select pre-spacers with a PAM sequence compatible with the system's Cascade complex, while also ensuring the correct spacer length and orientation during spacer integration into the array.

Pre-spacers are integrated into the CRISPR array by the Cas1-Cas2 complex, via a reaction mechanism similar to that of retroviral integrases and transposases [74] . Cas1 facilitates two staggered nucleophilic attacks by the 3'-OH hydroxyl at the end of each pre-spacer DNA strand onto the phosphodiester backbone of the repeat DNA [75]. Spacers are always added to the leader end of the CRISPR array, and in type II systems this is achieved by the recognition of a 'leader-anchoring sequence' directly upstream of the first repeat by the Cas1-Cas2 complex [76,77]. Significantly, leader-proximal spacers provide better immunity, and integrating the spacers from the most recent infection in this polarized manner ensures protection against the most pressing viral threat [76].

A so far unique aspect of type I adaptation is the presence of priming, where pre-existing spacers against an invader enhances further adaptation against the same threat [78]. This allows the CRISPR system to keep up with rapidly mutating phages that might have altered their target sequence to escape CRISPR targeting, and to offer fast adaptation against phages that are related to pre-encountered phages. Primed adaptation requires the binding by Cascade to mutated protospacers, harbouring a non-functional PAM or seed sequence mismatch with the crRNA. The complex then adopts an "open" conformation [79], which recruits Cas1-Cas2 along with Cas3 [80] to capture pre-spacers from the target. Finally, interference-driven adaptation can occur in the presence of a perfect target cleaved by Cas3, which is recruited by Cascade adopting a "closed" confirmation [81].

Adaptation of type III systems has only been observed in the relatively rare subset of loci that encode reverse transcriptase-Cas1 (RT-Cas1) fusions [82]. In a type III-B system from *Marinomonas mediterranea,* adaptation was shown to integrate DNA spacers derived from cellular RNA via a reverse transcription reaction mediated by RT-Cas1 [83]. This is thus an elegant way for the transcription-dependent type III systems to sample

from well-transcribed regions and ensure functional spacers. In the same organism, the type III targeting machinery can use crRNAs derived from the type I CRISPR array [60]. Cross-talk between type I and type III systems could be the reason for their frequent co-occurence, and might help explain the lack of dedicated adaptation machinery in many type III loci (which would use spacers acquired by the type I system) [7].

1.3.5 Prokaryotic Argonautes

Prokaryotic Argonautes (pAgo) represent a recently discovered bacterial innate defence mechanism found in approximately 9% of bacterial genomes and 32% of archaeal genomes (for a recent review, see [84]). They are often encoded within defence islands, i.e. in regions enriched for phage resistance systems, and have undergone extensive horizontal gene transfer [85], two factors which suggest a defensive role. So far, several mechanisms have been demonstrated, including DNA-guided DNA silencing and RNA-guided DNA silencing. For the former mechanism, it was shown in two systems that the apo form of pAgo can first degrade a target DNA non-specifically. The degradation products from this DNA can then be incorporated and used as guide DNAs, which allows sequence-specific interference against the same target [86,87]. Some pAgos are also predicted to be catalytically inactive, but are encoded near other nuclease genes that might be guided by pAgo to the invader. So far, however, *in vivo* pAgo immunity has only been demonstrated against invasive plasmids, and further studies will determine what role this class of mechanistically diverse defence proteins plays against phages.

1.4 Abortive infection and toxin-antitoxin systems

Abortive infection (Abi) and toxin-antitoxin (TA) systems are highly versatile and widespread, albeit poorly understood, anti-phage and stress systems. They target both self and invader, and through autoimmunity are able to inhibit the phage life cycle, preventing virion release from the infected cell. This might doom the bacterium, but provides protection for the clonal population. An alternative outcome from Abi/TA activation might be reversible dormancy induction, which affords the infected cell extra time to eliminate the invader. Abi systems are phenotypically, rather than genetically, defined, and are always involved in disrupting phage infection. TAs, on the other hand, comprise of a well-defined TA gene pair, where the antitoxin inhibits the toxic catalytic activity of the toxin. Abi and TAs have some overlap; some TAs might work through Abi-like mechanisms, e.g. genetic analysis of the lactococcal Abi systems AbiQ (Samson et al., 2013) and AbiE (Dy et al., 2014) revealed that they were *bona fide* TA systems. A central outstanding question is to what extent TA systems, but also Abi systems, irreversibly lead to cell death as opposed to temporary growth arrest and dormancy.

1.4.1 Abortive infection

Abi is a process by which cells prevent release of functional phage virions at the expense of host cell survival/fitness (for a more detailed treatment, see [20]). As such, it is considered an altruistic action, "programmed cell death", where the cell protects the surrounding clonal population from phage spread at its own expense. This overall fitness advantage requires cells to be in a structured environment, where sacrificing oneself benefits nearby cells that are kin [88]. The mechanisms of Abis are highly diverse, and often poorly understood, but tend to target essential processes in ways that are detrimental to both host and phage, like translation, transcription, replication, or inducing membrane leakage. How phages are recognised and trigger Abi is often unknown, but specific phage proteins can be involved.

Although Abis are widespread, most characterised systems have been studied in *E. coli* and *Lactobacillus lactis*. The Lit and PrrC systems of *E. coli* both target translation. The Lit protease of *E. coli* K12 is activated by the 29 amino acid GoI peptide of the T4 major capsid protein, a gene that is transcribed late in the phage infection cycle [89]. Lit is a zinc metalloprotease which, when activated, specifically cleaves the ribosomal elongation factor EF-Tu, thereby arresting translation for both the phage and its host. PrrC of *E. coli* CT196 cleaves the $tRNA^{Lys}$ in the anticodon loop; this depletes the $tRNA^{Lys}$ pool and inhibits global translation [90].

In *L. lactis*, a bacterium important in the dairy industry which is susceptible to phage attack, more than 20 Abi systems have been identified. The single-gene *abiK* system is able to reduce phage infectivity 10^6-fold. AbiK possesses polymerase activity, synthesising long DNA molecules with random sequences *in vitro*, an activity that is essential for its Abi function *in vivo* [91]. Since phage mutants that escape AbiK have mutations in phage-encoded recombinases, it was hypothesized that AbiK-synthesised DNA interferes with phage recombination, preventing phage replication and maturation. How this activity is also detrimental to the host, and what triggers AbiK during infection, remains unclear. Another lactococcal Abi system, AbiZ, reduces the burst size of phage Φ31 a 100-fold. AbiZ seems to act cooperatively with the phage pore-forming protein holin to induce premature lysis and the release of immature, non-infectious phage particles [92].

Recently, a kinase-mediated Abi mechanism protecting against Siphoviridae phages was uncovered in *Staphylococcus epidermidis* [93] (Figure 1.5). The eukaryotic-like Serine/Threonine kinase Stk2, located next to CRISPR-Cas and RM systems, was found to be sufficient to result in host cell death upon phage infection without the release of viral progeny. Stk2 phosphorylates a range of targets in diverse cellular pathways, including

transcription, translation, replication, and metabolism, presumably inhibiting their activities to robustly effect host death. Phages able to escape Stk2 Abi defense carry mutations in the *pacK* gene, suggesting that PacK induces Stk2 phosphorylation; however phages lacking the *pacK* gene could also activate the pathway, indicating that other factors might also trigger Stk2.

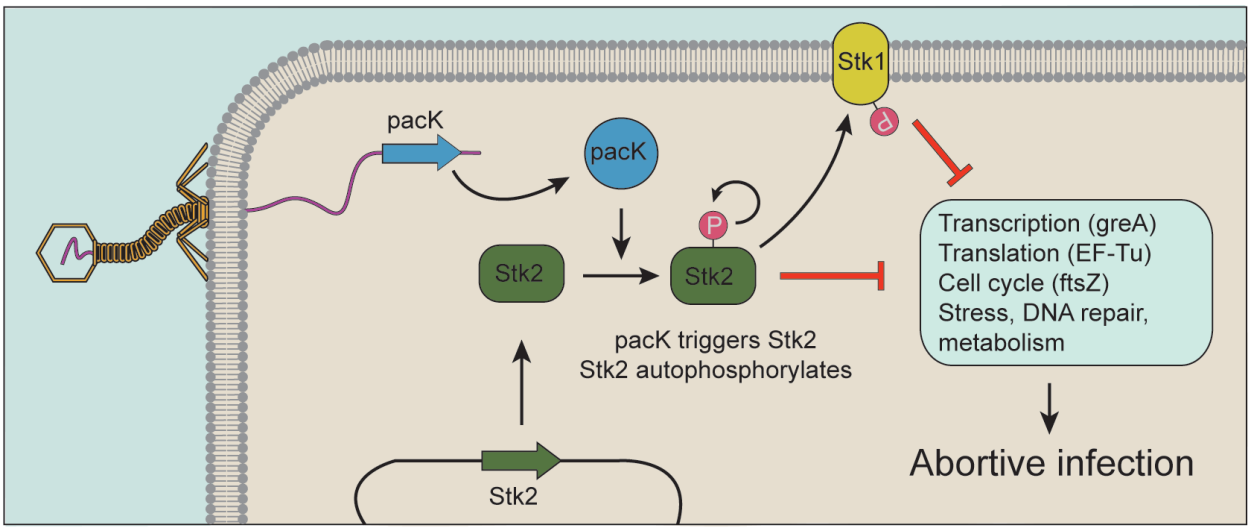

Figure 1.5. The staphylococcal Abi system Stk2. *The phage protein pacK triggers the S. epidermidis Stk2 kinase. Stk2 autophosphorylates, and goes on to phosphorylate Stk1 and miscellaneous cellular factors, eventually leading to abortive infection and cell death.*

1.4.2 Toxin-antitoxin systems

TA systems are widespread genetic elements consisting of a toxin-antitoxin pair, where the activity of a toxin is inhibited by its cognate antitoxin (for a recent review, see [94]). Typically, the antitoxin is labile and prone to degradation, and must be continuously expressed and replenished to remain at appropriate stoichiometric ratios with the toxin. How antitoxin degradation is triggered under physiological conditions remains mysterious in many cases. The toxin can possess various catalytic activities, including DNase and RNase, or can inhibit DNA replication, ATP synthesis, or cell division machinery. There are at least six TA types, categorised based on the nature of the toxin and antitoxin (protein or RNA), and the mechanism of toxin neutralisation, with many strains harbouring dozens of TA gene pairs (*E. coli* K-12 has more than 35 TA pairs for example [94]). This high genetic diversity is reflected in the many functions found for TA systems: In addition to phage defence, they have been implicated in stress responses, plasmid maintenance, and persister cell formation.

There are many lines of evidence linking TA pairs to anti-phage immunity. First, many TA systems are encoded near other bacterial defence systems, suggesting a

protective function. Second, phages have evolved mechanisms to overcome TA systems, indicating that TAs are part of the host-phage arms race. For example, Lon can degrade different labile antitoxin proteins upon infection, releasing the toxin to interfere with the phage life cycle. T7 encodes a Lon protease inhibitor that prevents antitoxin degradation and allows infection in the presence of the TA pair [95]. Phage T4 carries the ADP-ribosyltransferase Alt, which, upon infection, modifies the toxin MazF to inhibit its RNase activity [96]. Two other toxins are inhibited by the T4 *dmd* gene, which acts as an antitoxin and binds and sequesters the toxins directly [97]. Third, TAs can directly inhibit the phage life cycle. In *E. coli*, the type II TA system MazF/MazE can suppress the T4 infection [96]. Upon infection, T4 inhibition of host transcription leads to reduced MazE antitoxin expression, thereby releasing the RNase toxin MazF to inhibit the phage and host life cycles [98]. The plant pathogen *P. atrosepticum* possesses the type III ToxN/ToxI, where the endoribonuclease ToxN is sequestered by binding the noncoding RNA antitoxin ToxI [99]. Upon phage infection, the RNase activity of ToxN is unleashed to target both host and phage transcripts, arresting the infection. In spite of these examples, only a handful more TA systems have so far been shown to protect against phage infection.

Even if the activity of the toxin is recognized, in many cases how toxicity contributes to phage defence is unclear. TA systems are often described as inducing programmed cell death akin to Abi systems, where TA induction causes altruistic suicide of the cell while also aborting phage propagation. However, the activation of toxins *per se* can in some cases be reversible. With the MazF/MazE and ToxN/ToxI systems, cells where the toxin MazF or ToxN was expressed could be rescued and were viable upon induction of the expression of the antitoxin (MazE or ToxI, respectively) after a delay [99,100]. Recently, a comprehensive analysis of MazF cleavage sites revealed that most mRNAs, as well as rRNA precursors, are cleaved at multiple sites [101]. Presumably, upon toxin neutralization, the cells can replenish their RNA pool and resume growth. Given the extensive diversity of TA systems, it is probable that TAs can work both as dormancy induction and programmed cell death (Abi) systems, and the outcome will depend on a range of factors including the toxin mechanism of action, the duration of the toxin's activity, and the life cycle of the phage.

1.5 Bacteriophage assembly interference

Phage-Inducible Chromosomal Islands (PICIs) form a group of genetic elements that parasitise phage for replication and transmission (for a recent review, see [102]). PICIs are integrated into a bacterial chromosome and excise in the presence of a 'helper phage' (by infection or lysogen induction). Although the main role of PICIs seems to be the dissemination of the genetic material they harbour (in many cases virulence

determinants), they interfere with the phage life cycle, and therefore can be classified as an anti-phage mechanism.

PICI genomes are often small (~15 kb), encoding genes required for excision and integration, factors to manipulate helper phages to facilitate PICI packaging and dissemination and a repressor that prevents their expression in the absence of the helper phage. They are best characterized in *Staphylococcus aureus*, where they are named 'SaPIs' (*S. aureus* Pathogenicity Islands). SaPIs are induced when their repressor, Stl, is sequestered from the SaPI promoter by an anti-repressor expressed early during the helper phage lytic cycle [103] (Fig. 1.6). Derepression induces the expression of the SaPI genes and couples its propagation cycle with that of the helper phage. Specialized SaPI structural proteins modulate the assembly of the helper phage's capsid to produce a virion that can only be packaged with the SaPI's smaller genome. As a result, the host bacterium lyses, but primarily releases SaPI virions that infect neighbouring cells to disseminate the SaPI genomes and the virulence factors they encode. Recently, PICIs were found to also be widespread in gram-negative bacteria [104]. Instead of being regulated by a repressor, gram-negative PICIs are induced by a PICI-encoded activator whose expression requires the helper phage.

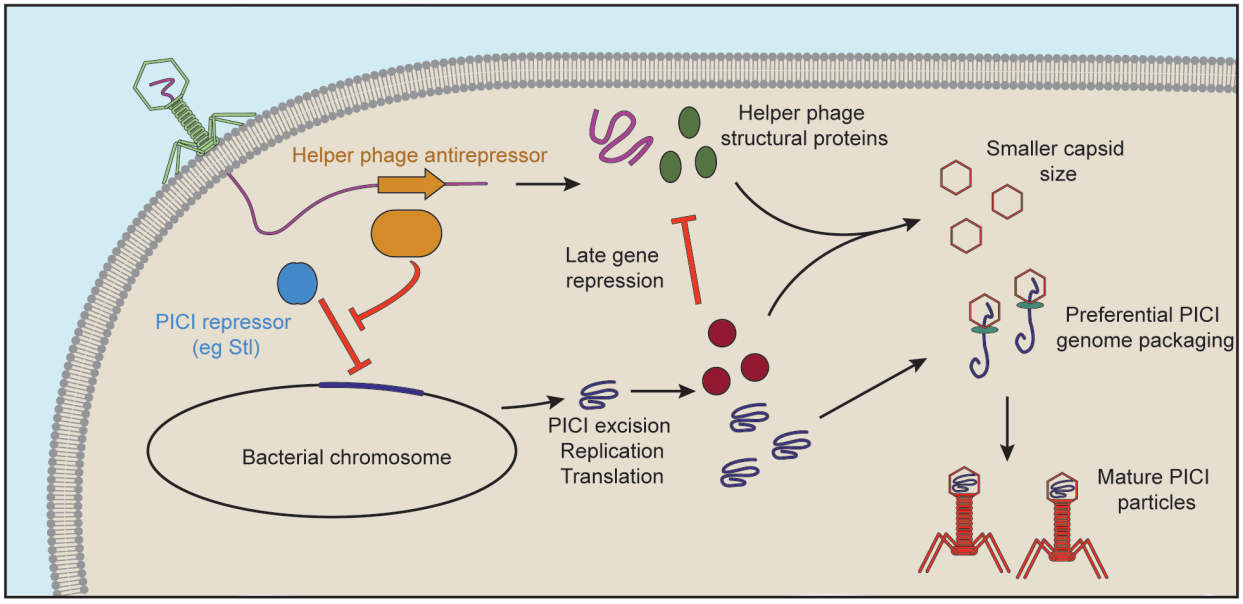

Figure 1.6. PICI-mediated interference of phage assembly. *A phage antirepressor inactivates the PICI repressor (or for gram-negative PICIs, an activator is expressed in the presence of the helper phage), leading to PICI excision from the host chromosome. The PICI genome replicates, and proteins modulating the behaviour of the helper phage are expressed. These proteins repress late phage genes, they alter the phage capsid size to be more appropriate for the PICI genome size, and they lead to preferential packing of PICI genomes. Mature PICIs are produced at the cost of helper phage propagation.*

To fight phage, *V. cholerae* encodes PICI-like elements (PLEs) [105], which are similar to other gram-negative PICIs except for their somewhat larger size and different gene content. Upon infection of the host bacterium by the phage ICP1, the PLE-encoded recombinase Int excises PLE from the *V. cholerae* chromosome [106]. PLE then replicates to high levels and inhibits ICP1 phage replication by yet unknown mechanisms. Interestingly, phage ICP1 encodes its own CRISPR system capable of targeting PLEs [105], allowing propagation in PLE+ *V. cholerae* strains.

1.6 Recently discovered anti-phage mechanisms

In spite of decades of intense study, much remains unknown about anti-phage mechanisms. For example, CRISPR and Argonautes were only recently described, and other systems probably await discovery. Interestingly, bacterial defence systems often cluster in defence islands [107]. This has allowed a 'guilt-by-association' approach to uncover new defence mechanisms, the BREX system [36] (see above) being one successful example of this method. Another recent example is the DISARM system [108], which was identified as a five-gene cassette frequently associated with genomic defence islands. This system provides broad anti-phage immunity through a novel RM mechanism that includes a methyltransferase modifying a five-nucleotide motif and a multi-component restriction element that probably cleaves unmodified phage DNA early in the phage life cycle, though the exact mechanism remains to be elucidated. Also using this genetic neighbourhood approach, a study identified 26 candidate systems harbouring genes predicted to interact with nucleic acids (including helicases and nucleases) that were ectopically expressed in *E. coli* or *Bacillus subtilis* and assayed for anti-phage activity [109]. Of these, eight provided robust protection against at least one type of phage (and one worked against plasmids), including the 'Zorya' system, found in 3% of sequenced bacterial genomes. Zorya most likely works as an Abi system that causes membrane depolarization of the host using a predicted proton channel similar to MotA/MotB.

The guilt-by-association approach was also used to identify genes regularly associated with CRISPR loci [110,111]. These studies identified a diverse range of CRISPR-associated accessory candidate genes, likely to complement or expand on the functionality of the core *cas* gene machinery. Finally, an extensive screen of mycobacterial phages revealed varied mechanisms of phage-encoded Sie, which included a (p)ppGpp synthetase and a single-subunit RM system, as well as classical Sie modes like promoter repression and inhibition of entry [112]. This suggests that phage-encoded anti-phage mechanisms are more widespread than previously thought.

1.7 Putting it all together

The phage resistance mechanisms discussed so far have mostly been studied individually in the lab, though this is rarely how a bacterium's arsenal is applied against phage. Bacteria usually employ several complementary lines of defence, none of which are mutually exclusive, and a phage has to overcome each system to allow successful infection. On the other hand, contrary to experimental settings, the environment typically contains a heterogeneous mix of phages. Whether multiple defense systems have an additive effect or work synergistically, and how exposure to multiple phages affects the development and effectiveness of these defenses, is not well understood.

Complicating the picture further are potential synergistic effects between anti-phage mechanisms, where the activity of one system might enhance that of another. RM and CRISPR systems often co-exist, and spacer acquisition by the CRISPR system is enhanced in the presence of an RM system [113]. *Marinomonas mediterranea* contains two CRISPR systems, and spacers incorporated from phage into the type I-B array can be used by the III-B machinery against phage [60]. The cross-talk between the type I and III systems makes protection more robust since it is harder for phages to escape type III targeting through protospacer mutations [59]. Although not experimentally demonstrated, it is tempting to speculate that there can also be synergy between dormancy-inducing components (TA systems) and effector components (i.e. CRISPR and RM systems). Indeed, their clustering in genomic defence islands suggests a functional coupling between systems [114]. The rationale is that TA or Abi systems could "buy time" for the cell, inducing short-term dormancy while CRISPR or RM mechanisms eliminate the phage. An example of this is seen in the type III-A CRISPR-Cas response, where the non-specific RNase Csm6 causes a transient growth arrest until the DNase activity of the Cas10-Csm complex has eliminated the plasmid invader[55]. Induction of dormancy could also afford time for the acquisition of new spacers during CRISPR adaptation. Alternatively, if the invader is not cleared and immunity fails, continuous activity of TA or Abi could lead to cell death, acting as a failsafe to prevent further spread of the phage.

1.8 Outlook and future directions

Recently, both novel technologies and experimental approaches, as well as renewed interest in the field, have dramatically boosted our knowledge of how bacteria resist their parasites. Scientists have probed both broader (new systems) and deeper (expanding repertoire of known systems). Future studies will surely continue this trend, which will most likely translate into clinical outcomes and technological innovations. Still, there are hurdles to overcome. The study of more natural models, or more integrative models where it is possible to appreciate how different immune mechanisms interact and

complement each other, as well as more ecological approaches, where the fluxes of multiple and different bacteria and phages can be observed, will undoubtedly expand our understanding of prokaryotic immunity. Last but not least, given the importance of phages as mediators of horizontal gene transfer, the study of defence mechanisms will help us understand prokaryotic evolution. Considering the millions of years bacteria and phages have had to coevolve, there surely remains a cornucopia of unknown unknowns for us to discover.

Chapter II. The discovery of cyclic oligoadenylate second messengers in type III CRISPR-Cas immunity

At the beginning of my research, little was known about how the main interference complex of type III CRISPR systems interact and activate the accessory RNase Csm6, or ancillary factors more generally. However, we initiated a collaboration with the laboratory of Prof. Martin Jinek (University of Zurich, Switzerland) that helped define cyclic oligoadenylate second messengers as central to type III activation of downstream effectors. This work was published in 2017[53], and will briefly be outlined in this chapter.

2.1 Previous knowledge on type III accessory proteins

The type III-A CRISPR-Cas locus of *S. epidermidis* RP62a type III-A is shown in Fig. 2.1. Previous work from our group demonstrated that the main interference complex for this CRISPR type is composed of Cas10 and Csm2-5 [115]. The RNase Csm6, on the other hand, did not interact with the Cas10-Csm complex, at least under non-targeting circumstances. One hypothesis was the existence of a transient interaction between the Cas10-Csm complex and Csm6, allowing recruitment and allosteric activation of Csm6 during interference. Another hypothesis was ligand binding of a nucleotide derivative to the CARF domain of Csm6, based on bioinformatic predictions [116] [117]. This was also supported by the similarity of the Palm domain of Cas10 to nucleotidyl transferases and DNA polymerases. Yet, many of these predictions were not supported by strong evidence, and required experimental testing and validation.

Figure 2.1. The *S. epidermidis* RP62a type III-A CRISPR-Cas locus. *The naturally occurring spacers 1 to 3 are shown as coloured boxes. The adaptation machinery (*cas1*,* cas2*) is in purple, genes encoding the Cas10-Csm complex are green, the accessory RNase* csm6 *is yellow, and the gene responsible for crRNA maturation (*cas6*) is brown.*

2.2 The discovery of cyclic oligoadenylates

To address this problem, our collaborators made two assumptions. Firstly, they assumed there was an activating ligand that was nucleotide-derived. Secondly, they assumed the ligand was symmetrical, based on the symmetrical proposed ligand binding site in the homodimeric Csm6[118]. In the concomitant assays, they tested various ligands' ability to activate the RNase activity of Csm6. One ligand, cyclic-di-AMP (c-di-AMP), robustly activated Csm6, and they initiated the collaboration with our laboratory to see if I could validate their findings *in vivo*. c-di-AMP is a bacterial second messenger involved

in various processes like cell growth, homeostasis and osmotic regulation[119]. Thus, I initiated work in *S. aureus* to try to establish a link between c-di-AMP production/degradation and Csm6 function during CRISPR immunity.

However, the further work by our collaborators showed that the c-di-AMP link was an artifact. Upon purchasing another commercial batch of c-di-AMP, the Csm6 activation was no longer observed. The previous (functional) batch was synthesised chemically, while the newer (non-functional) batch was made enzymatically. It turned out that trace impurities in the first batch were responsible for activating Csm6. Subsequent experiments confirmed that other oligoadenylates were responsible. Further biochemical experiments by our collaborators demonstrated that the Cas10-Csm complex could, once bound to a target RNA, synthesise oligoadenylates of various sizes from ATP. Of these, cyclic hexa-oligoadenylate (cA_6) activated the RNase activity of Csm6.

To corroborate these findings *in vivo*, I performed experiments in *S. aureus* during infection by phage ϕNM1γ6. During infection, against a protospacer transcribed late in the phage life cycle, Csm6 is required for immunity[56]. Since the Palm domain of Cas10 was shown biochemically to be required for cA_6 synthesis, the Palm domain should also be required for phage protection, i.e. it should phenocopy a loss of Csm6. Moreover, a double loss of the Palm domain of Cas10 and of Csm6 should phenocopy both single mutants. This is indeed what was observed (Fig. 2.2). Thus, the genetic link between the Palm domain and Csm6 activation was demonstrated *in vivo*.

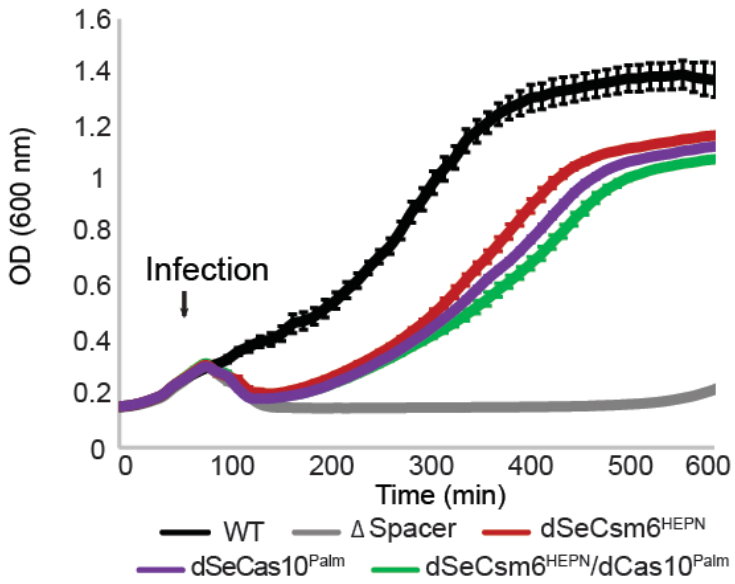

Figure 2.2. The Palm domain of Cas10 activates Csm6 *in vivo*. *Cells harbouring a type III-A CRISPR system in a wild-type setting (WT), with no targeting spacer (ΔSpacer), with catalytically dead Csm6 (dSeCsm6^{HEPN}), inactivated Cas10 Palm domain (dSeCas10^{Palm}), or a double mutant variant. Cells are infected at 60 minutes with a multiplicity of infection (MOI) of 30, each data point representing the mean of three biological replicates ±s.e.m.*

The results of these studies were published in 2017 [53]. Another research group independently published similar results at the same time [54]. The finding of ligand-mediated activation of accessory CRISPR proteins had important consequences for my later research. In particular, I demonstrated that the Cas10-Csm complex of one system (*S. epidermidis* III-A) can generate cA to activate an accessory gene from a different type III system (in this case, the Csm6 of *Enterococcus italicus*), highlighting the modularity of these factors.

Interestingly, another nucleotide second messenger-centred anti-phage immune system, cyclic-oligonucleotide-based anti-phage signalling systems (CBASS) has recently been identified[120-122]. These systems sense phage infection through an unknown mechanism, and initiate the synthesis of diverse di- and tri-nucleotides. These signalling molecules then activate downstream effectors. In spite of being genetically and mechanistically distinct, there are striking parallels between type III CRISPR systems and CBASS systems.

Chapter III – The role of, and the mechanism of action of, the accessory type III RNase Csm6

My primary research project upon the start of my PhD was to investigate the role of the accessory RNase Csm6. Specifically, its role in anti-plasmid immunity was not understood, and its mechanism of action, i.e. what is being targeted upon activation, was unknown. The findings from chapter II, that Csm6 is activated by the ligand cA_6,aided in experimental design, but also raised questions of how Csm6 might specifically target invader RNA. At the time, many researchers believed that Csm6 was a cell death factor acting through abortive infection, irreversibly killing the host bacterium in order to prevent parasite spread. Through my work, however, I show that Csm6 induced a reversible dormant state, allowing the cell to resume growth upon clearance of the invader. In the context of anti-plasmid immunity, I showed that Csm6 is required specifically when transcription across the target is weak, necessitating a more robust immune response. This chapter is based on our publication from early 2019 [55].

3.1 Background on Csm6 and type III-A anti-plasmid immunity

To date, six major types of CRISPR systems have been identified, which vary in their *cas* gene composition and mechanism of action[40]. Type III systems are uniquely able to degrade both the DNA and RNA of the invader[123]. The Cas10-Csm (type III-A) and Cas10-Cmr (type III-B) complexes use crRNA guides to detect and anneal to transcripts harbouring a complementary sequence[57,58,124,125]. This base pair interaction unleashes the single-stranded DNase activity of Cas10[52,126,127], and the sequence-specific RNase activity of Csm3[58,124,128] or Cmr4[57,125]. DNA degradation by Cas10 is transient, and cleavage of the protospacer RNA by Csm3/Cmr4 results in Cas10 inactivation and the dissociation of the effector complex from its target[52,126,127].

In addition to the Cas10 complexes, type III systems also use an accessory RNase, Csm6 for type III-A and Csx1 for type III-B[7]. Csm6 is an endoribonuclease[118] whose activity is modulated by cyclic oligoadenylate (cA), a second messenger synthesised by the Palm domain of Cas10 upon target recognition by the crRNA[53,54,117]. The de-activation of Csm6 is the consequence of two events: the lack of synthesis of new cA molecules by the Palm domain after cleavage of the target transcript by Csm3/Cmr4[53,54,129], and the degradation of the existing cA, presumably by specific ring nucleases[130].

The function of Csm6 has been studied during the type III-A CRISPR-Cas immune response against lambda-like dsDNA phages, where Csm6 RNase activity is required only when the target is transcribed late in the phage life cycle[56]. Because late targeting

cannot prevent phage replication, it is hypothesized that Csm6 degradation of viral transcripts prevents the completion of the lytic cycle and allows Cas10 DNase activity to clear the phage genomes that accumulated before the transcription of the target. Csm6 has also been shown to be required for the prevention of plasmid conjugation and plasmid transformation by type III-A CRISPR-Cas systems[115,131]; however how Csm6 contributes to plasmid clearance is still not understood. Here I show that low levels of target transcription are sufficient to activate Csm6 and trigger non-specific degradation of both host and plasmid transcripts. This accelerates plasmid clearance by the Cas10-Csm complex, presumably through the depletion of transcripts required for efficient plasmid replication and maintenance. Simultaneously, the destruction of host transcripts produces a growth arrest, as was previously proposed[53,54,132]. Since plasmid DNA degradation leads to the disappearance of the targets that activate Csm6, the growth arrest is short lived and the cells resume normal growth following plasmid clearance. Our study furthers our understanding of the mechanisms by which type III-A CRISPR-Cas systems employ a two-pronged combination of DNase and RNase activities to provide robust immunity against foreign genetic parasites.

3.2 Csm6 RNase activity is required for immunity against poorly transcribed targets in pG0400

Spc1 in the CRISPR-*cas* locus of *S. epidermidis* RP62A (Fig. 3.1a) targets the nickase (*nes*) gene of the conjugative plasmid pG0400, preventing pG0400 transfer[39]. However, Northern analysis of the crRNAs derived from *spc1* indicated that the CRISPR locus is transcribed unidirectionally [133] and that the *spc1* crRNA has the same, not the complementary, sequence as the putative transcript of the *nes* gene. Given the requirement of direct binding between the crRNA guide and a target RNA for type III-A CRISPR-Cas immunity[51,134], I performed RNA-seq analysis of pG0400 transcripts to look for the presence of *nes* antisense transcripts that could elicit the *spc1*-mediated anti-plasmid response. I found very low levels of reads corresponding to transcripts derived from the *nes* non-template strand, which I hypothesised would result in infrequent activation of the DNase activity of the *spc1*-crRNA/Cas10-Csm complex. In turn, inefficient clearance of the pG0400 genome would lead to the Csm6 requirement for anti-plasmid immunity that our laboratory reported previously[115].

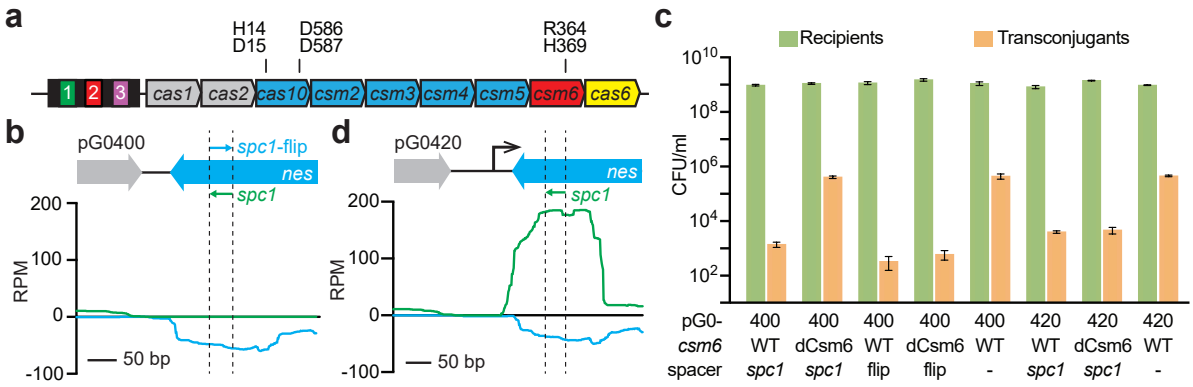

Figure 3.1. Csm6 is required for interference against pG0400 when the target is weakly transcribed. ***(a),*** *The S. epidermidis RP62A type III-A CRISPR, with the catalytic residues of Cas10 and Csm6 highlighted. The black boxes, representing repeats, flank the coloured boxes, symbolising spacers. Spc1 matches the* nes *antisense transcript of the conjugative plasmid pG0400.* ***(b)*** *The architecture around the* nes *protospacer in pG0400, with RNA-seq traces representing read depth. The annealing positions for spc1 and spc1-flip are indicated. RPM, reads per million.* ***(c)*** *Conjugation of pG0400 or pG0420 into S. aureus cells containing pCRISPRs with the indicated csm6 variant and spacer, after filter mating. Each bar represents the mean of three biological replicates ±s.e.m.* ***(d)*** *Like* ***(b)****, but for plasmid pG0420, with the inserted transcriptional promoter shown as a black arrow.*

To test this hypothesis, I generated a construct harbouring a reversed *spc1* (*spc1*-flip) that would produce a crRNA complementary to the highly abundant *nes* transcript (Fig. 3.1b). This was achieved by using *Staphylococcus aureus* RN4220[135] harbouring an *S. epidermidis* type III-A CRISPR-Cas locus on the chloramphenicol-resistant pC194 staphylococcal plasmid, pCRISPR[115], with modified spacer sequences. I found that, in contrast to s*pc1* where targeting requires *csm6*[115] (Fig. 3.1c), *spc1*-flip blocks pG0400 conjugation in recipients expressing a catalytically dead (R364A, H369A) *csm6* version (d*csm6*) [56,118] (Fig. 3.11c). I also constructed a pG0400 derivative (pG0420) in which I inserted a strong promoter that drives high transcription of the *nes* non-template strand (Fig. 3.1d). Next, I tested the ability of *spc1* to mediate immunity against pG0420 in the presence or absence of Csm6 RNase activity and found that, similarly to *spc1*-flip, the increased transcription of the target RNA eliminated the Csm6 requirement observed for pG0400 (Fig. 3.1c). Altogether, these results support our hypothesis that Csm6 RNA degradation is not required for plasmid clearance in conditions of high target transcription.

3.3. Csm6 RNase activity accelerates plasmid clearance in conditions of low target transcription

To determine how Csm6 affects anti-plasmid immunity in the presence of different levels of target transcription, I placed the *gp43* protospacer[56] (from phage ϕNM1γ6[51]) under the control of a tight and tuneable anhydrotetracycline (aTc)-inducible promoter[136], flanked by strong transcriptional terminators (pTarget, Fig. 3.2a). I performed Northern blots to validate these useful features of pTarget (Fig. 3.2b); I did not detect a target transcript in the absence of aTc, and found that target transcription levels correlate with the aTc concentration.

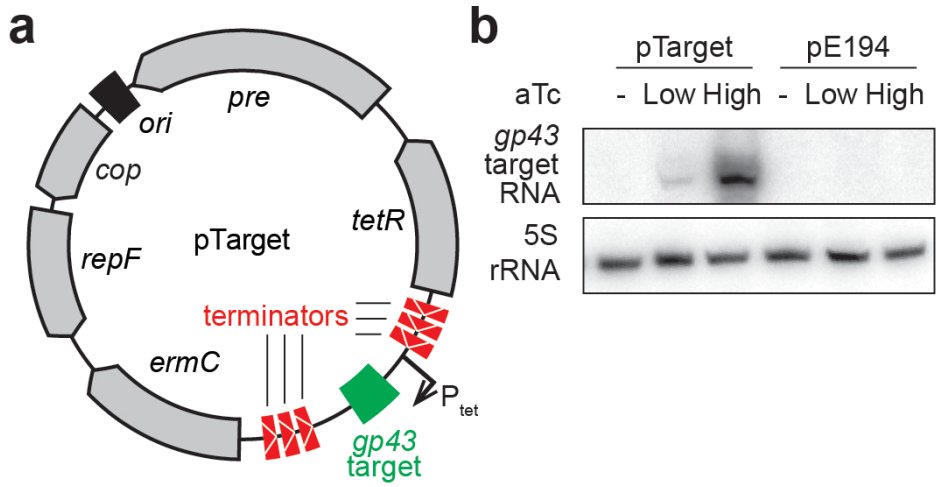

Figure 3.2. Properties of an aTc-inducible target plasmid, pTarget. ***(a)*** *Overview of the pTarget plasmid, with salient features labelled. Upon the addition of the inducer anhydrotetracycline (aTc), transcription is initiated from promoter* P_{tet}*, and the gp43 protospacer is transcribed. This transcription allows interference against pTarget by the III-A CRISPR system (pCRISPR). The transcriptional terminators prevent non-specific background transcription across the gp43 target. (**b**) Northern blot analysis of pTarget, detecting the gp43 target transcript prior to, or two minutes after, addition of either low or high levels of aTc. The empty plasmid pE194 is used as a control.*

I first investigated the effect of Csm6 on the fate of plasmid DNA. I generated *S. aureus* cultures harbouring both pTarget and pCRISPR, added different concentrations of aTc, purified the plasmids, and visualized them using agarose gel electrophoresis. I found that at high levels of aTc, i.e. high target transcription, plasmid loss was equally efficient in wild-type and d*csm6* backgrounds (Fig. 3.3a). This result corroborates our previous result that type III-A immunity is independent of Csm6 at high levels of target transcription, and that this ribonuclease is not necessary for DNA degradation *per se*. In contrast, at low levels of aTc, pTarget remained stable in hosts carrying the d*csm6* allele

(Fig. 3.3b), demonstrating that the RNase activity of Csm6, which is essential for immunity in these conditions, leads to an accelerated rate of plasmid DNA loss.

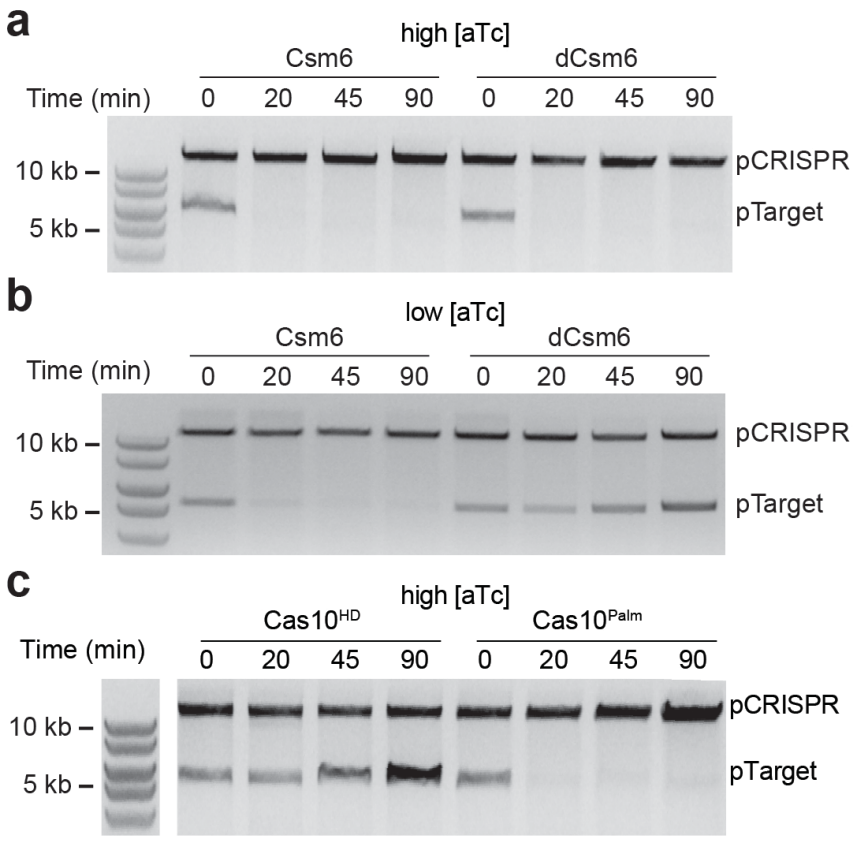

Figure 3.3. Csm6 accelerates plasmid clearance when interfering against a weakly transcribed protospacer. ***(a)*** *pTarget plasmid curing assay, where plasmid DNA is extracted from cells containing pTarget and a pCRISPR with the specified csm6 allele, before or at the specified time after adding high levels of aTc, and visualised by gel electrophoresis. Gel image is representative of three independent experiments. (**b**) Like **a**, but with induction of protospacer transcription by low levels of aTc. Gel image is representative of three independent experiments. (**c**) Like **a**, but with mutations in the Cas10 Palm or HD domains, both in a dcsm6 background. Gel image is representative of three independent experiments.*

3.4. Csm6 activation leads to non-specific degradation of host and plasmid transcripts

How can an RNase impact plasmid DNA stability? I investigated if Csm6 was responsible for the degradation of host and/or plasmid transcripts important for plasmid

replication/maintenance using RNA-seq. Because plasmid degradation carried out by the DNA cleavage activity of the type III-A system would lead to differences in transcript abundance due to different plasmid template levels (not only due to Csm6 RNase activity), I performed RNA-seq in a *cas10* mutant genetic background lacking DNase activity.

Previously, our laboratory reported that the Palm domain of the Cas10 subunit of the type III-A *S. epidermidis* complex is implicated in DNA cleavage[128]. However, recent work on other type III systems showed that Cas10 nuclease activity relies on its HD domain[52,126,127,137] and that the Palm domain is involved in the synthesis of Csm6's inducer, cA[53,54,129]. To clarify this discrepancy, I repeated the experiment of Fig. 3.3a in a genetic background harbouring HD or Palm domain mutations [H14A, D15A (Cas10^{HD}) or D586A, D587A (Cas10^{Palm})] (Fig. 3.1a). I tested conditions of high target transcription, in a dCsm6 background, to avoid the interference of RNA degradation in the interpretation of the results. I found that the HD, but not the Palm, domain of Cas10 is required for the clearance of pTarget (Fig. 3.3c), and therefore used a pCRISPR-*cas10*^{HD} mutant for our RNA-seq experiment.

For the RNA-seq experiment, wild-type or d*csm6* cells were harvested in duplicate either before or two minutes after induction with a low concentration of aTc. RNA was extracted, sequenced, and the average transcript abundance of every gene at 0 or 2 minutes after induction was plotted. Host transcripts exhibited a marked decrease in global transcript abundance in the presence of Csm6, i.e. most genes falling below the identity line (Fig. 4.1a). In the absence of the RNase activity of Csm6, however, this reduction was not detected. pTarget transcripts showed a similar difference in abundance (Fig. 4.1b). To corroborate the RNA-seq results, Northern blot analysis was performed to detect a subset of transcripts (Fig. 4.1c). The transcripts corresponding to the pTarget protospacer, the plasmid replication *repF* gene, and the chromosomal gene *def* (peptide deformylase) were all rapidly degraded in the presence of Csm6 RNase activity. Together, these experiments demonstrate that Csm6 activation during the type III-A CRISPR-Cas immune response results in significant depletion of both host and plasmid transcripts.

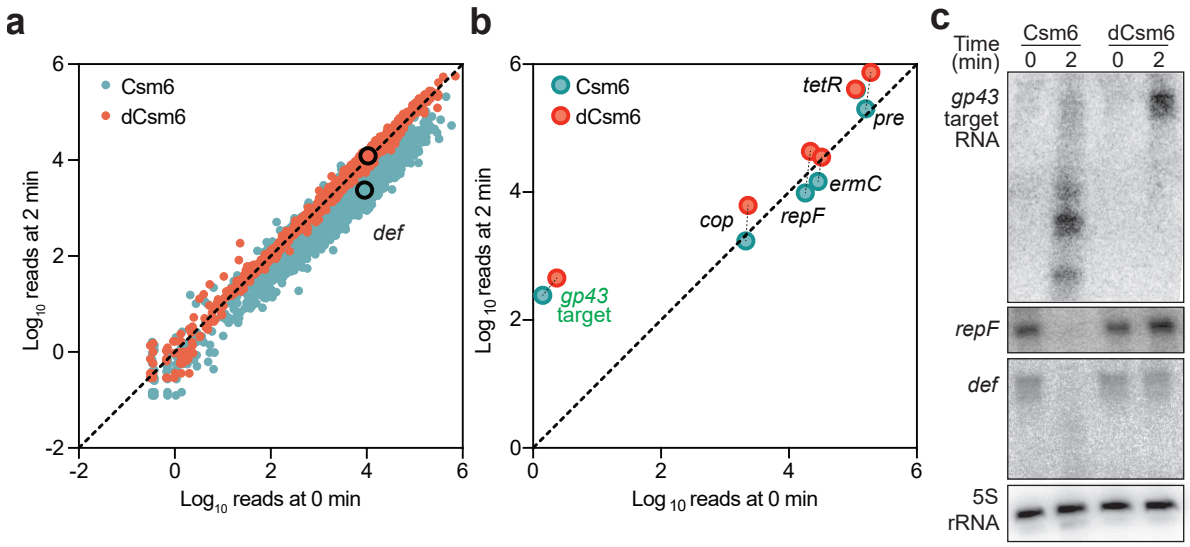

Figure 3.4. Csm6 activation results in non-specific degradation of host and plasmid transcripts. (a) *RNA-seq was performed with S. aureus cells harbouring a pCRISPR with wild-type (red) or mutated csm6 (blue), extracting RNA before or after inducing target transcription by adding low levels of aTc. L. seeligeri spike-in RNA was added to allow absolute normalisation of read depth. Chromosomal genes falling on the identity line are unchanged from 0 to 2 minutes, while genes falling below the line are depleted after 2 minutes. Both are done in a Cas10HD background. The def gene (peptide deformylase) is highlighted. The average of two replicates for each condition is shown.* ***(b)*** *Like a, but showing pTarget genes.* ***(c)*** *Northern blot analysis of RNA from cells with pTarget and pCRISPRs containing either csm6 or dcsm6, before or after inducing pTarget transcription with low levels of aTc (performed once). Probes detecting either the gp43 target transcript, the repF transcript of pTarget, or the chromosomal transcript def are used.*

3.5 Degradation of host transcripts induces a growth arrest

The results described above suggest two hypotheses: (i) that the general destruction of host transcripts should result in host cell toxicity during type III-A CRISPR-Cas immunity, and (ii) that the degradation of transcripts important for replication and/or maintenance should prevent the replenishment of plasmid DNA and therefore facilitate plasmid clearance by the DNase activity of the Cas10-Csm complex. I tested the first hypothesis by looking at the effect of CRISPR immunity against pTarget on host growth. To determine if Csm6 activation results in collateral RNA degradation I measured the effect of type III-A targeting on culture growth in the absence of erythromycin, i.e. without plasmid selection. Under low target transcription conditions that require Csm6 for rapid plasmid loss (Fig. 3.3b), I detected a significant growth arrest that was dependent on the

presence of Csm6's RNase activity (Fig. 3.5a). To measure plasmid destruction, I enumerated the cells that still contained pTarget at the end of the growth experiment by counting the erythromycin-resistant colony forming units (cfu) that remain after induction (Fig. 3.5b). I found a *csm6*-dependent decrease in cfu that reflects extensive plasmid clearance in the culture. These results show that Csm6 activation, which is required for plasmid loss under low target transcription conditions, is also responsible for generating a growth defect of the host. Interestingly, this growth defect is triggered by Csm6's enzymatic activity even in conditions of high transcription, when this RNase is not required for plasmid clearance (Fig. 3.5c-d).

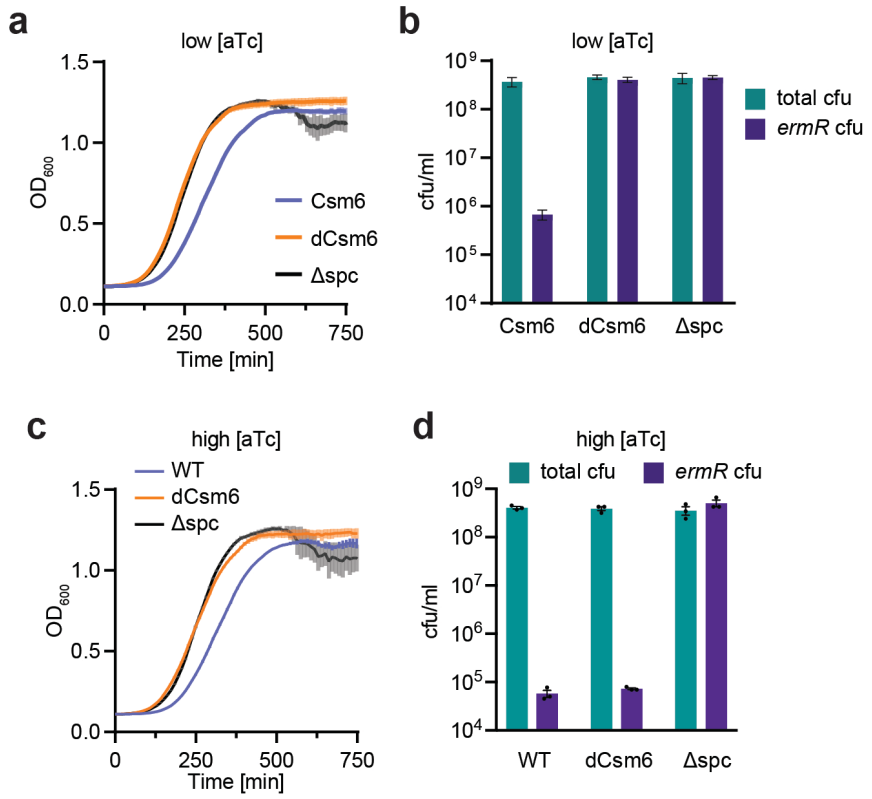

Figure 3.5. Prevention of expression of genes important for plasmid replication accelerates plasmid clearance. (*a*) *Growth of staphylococci containing pTarget and pCRISPRs with either csm6 or dcsm6, upon inducing transcription across the target with low levels of aTc, in the absence of erythromycin. Δspc is a no spacer control. OD600 measurements are taken every 10 minutes, each data point representing the mean of three biological replicates ±s.e.m.* ***(b)*** *Cells at the end of the experiment in a are spotted onto TSB plates in the presence or absence of erythromycin, selecting for the presence of pTarget, and enumerated. Each bar represents the mean of three biological replicates ±s.e.m.* ***(c)****, Same as panel* ***b****, but without erythromycin. For the Δspc control curve, the same data as from Fig. 3.7a is shown.* ***(d)****, Cells at the end of the experiment in panel* ***c***

are spotted onto TSB plates in the presence or absence of erythromycin, selecting for the presence of pTarget, and enumerated. Each bar represents the mean of three replicates ±s.e.m.

3.6 Non-specific degradation of host and plasmid transcripts facilitates plasmid clearance

Next, I investigated the second hypothesis; whether the prevention of expression of genes important for plasmid replication could accelerate plasmid clearance. pTarget genes such as *repF* and *cop* are important for plasmid replication and maintenance, and our RNA-seq analysis determined that these transcripts are targeted by Csm6. In addition, Csm6-mediated depletion of host transcripts required for plasmid replication, such as DNA polymerase and single-stranded DNA binding protein[138], could also impair plasmid stability and facilitate clearance by the type III-A CRISPR-Cas system. To test if a global reduction of gene expression, similar to that caused by Csm6 activation, can accelerate plasmid loss during the type III-A CRISPR-Cas immune response, I used neomycin, an inhibitor of the 30S ribosomal subunit, and measured pTarget clearance after induction with low levels of aTc. In contrast to Fig. 3.3b results, repeating this experiment in the presence of neomycin led to rapid pTarget clearance also in the dCsm6 background (Fig. 3.6), comparable to that of cells expressing wild-type Csm6. Importantly, neomycin did not affect pTarget stability in non-targeting hosts (Fig. 3.6, Δ*spc*). These results support the idea that reduction in expression of both plasmid and host genes, either by the non-specific mRNA degradation by Csm6 or by neomycin-induced inhibition of translation, can facilitate the type III-A CRISPR-Cas immunity against plasmids with low rates of protospacer transcription.

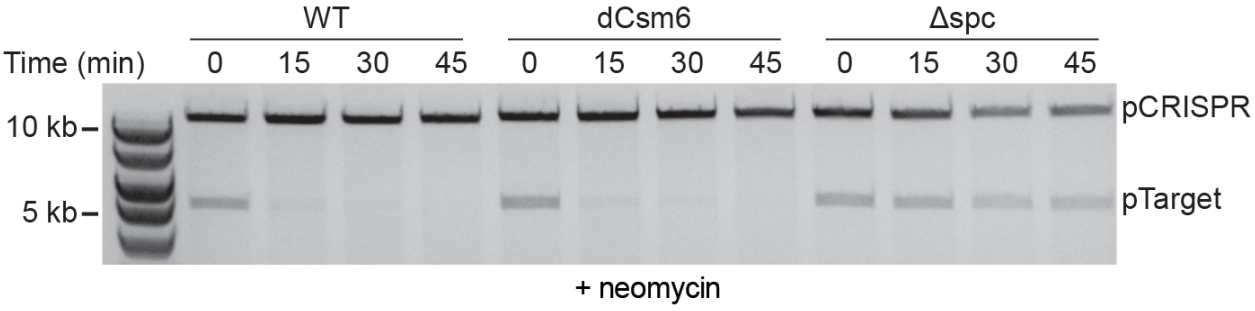

Figure 3.6. A general suppression of gene expression is sufficient to promote plasmid clearance. *Plasmid curing assay, where plasmid DNA is extracted from cells harbouring the specified pCRISPR and pTarget, before or at different times after induction of protospacer transcription by the addition of low levels of aTc, and visualised by gel electrophoresis. The cells are grown in the presence of the translational inhibitor neomycin. Gel image is representative of three independent experiments.*

3.7 Complete plasmid clearance requires DNA cleavage by Cas10

Finally, I examined whether transcript degradation by Csm6 alone is sufficient for anti-plasmid type III-A CRISPR-Cas immunity. I measured the growth of cells carrying the *cas10*^{HD} mutation after induction of target transcription in the absence of erythromycin selection. I found a growth arrest that was much more severe than detected with wild-type Cas10 (Fig. 3.5a), likely due to the absence of DNA cleavage and the persistence of pTarget, leading to continuous activation of Csm6 (Fig. 3.7a). This growth arrest was dependent on the ability of the Cas10 Palm domain to make cA, and on the RNase activity of Csm6. In addition, all the cells that recovered at the end of the experiment were able to grow on erythromycin and therefore retained pTarget (Fig. 3.7b), escaping arrest through the accumulation of mutations that abrogate type III-A immunity and the activation of Csm6 (Figs. 3.7c-e). Altogether these results demonstrate that the Csm6-mediated non-specific degradation of host and plasmid transcripts that affect plasmid replication is not sufficient to mount an efficient type III-A CRISPR-Cas immune response, which also requires the destruction of plasmid DNA by the Cas10 nuclease.

I also investigated the effect of Csm6-mediated growth arrest on the type III-A immune response against pG0400, the natural plasmid target of the *S. epidermidis* CRISPR system. Previously, our laboratory showed that mutations in *csm6* and *cas10* Palm domain, but not in its HD domain, led to an increase in the number of pG0400 transconjugants, i.e. lack of type III-A immunity[115]. Similar results were recently obtained by analyzing this system in an *Escherichia coli* heterologous host[131]. Given my findings with pTarget, I thought that previous work might have failed to see pG0400 transconjugants in the *cas10*^{HD} background if the Csm6-mediated growth arrest affected colony growth. Therefore, I performed conjugation assays and incubated the plates seeded with transconjugants for a longer time (24 hours, as opposed to the 16 hours used in previous assays). I found that indeed *cas10*^{HD} hosts do produce a large number of transconjugants, indicating a loss of CRISPR immunity (Figs. 3.8a-b). However, transconjugant colonies are markedly small (Fig. 3.8b). As I hypothesized, Csm6 is responsible for this phenotype, which is eliminated by the addition of the d*csm6* mutation (Cas10^{HD} vs Cas10^{HD}/dCsm6, Fig. 3.8b). Altogether, these results with both pTarget and pG0400 suggest that Csm6-mediated growth arrest is only temporary and not essential for type III-A immunity against plasmids. Cas10-mediated DNA cleavage, on the other hand, is absolutely necessary to fully eliminate the plasmid target and turn off Csm6's non-specific RNase activity.

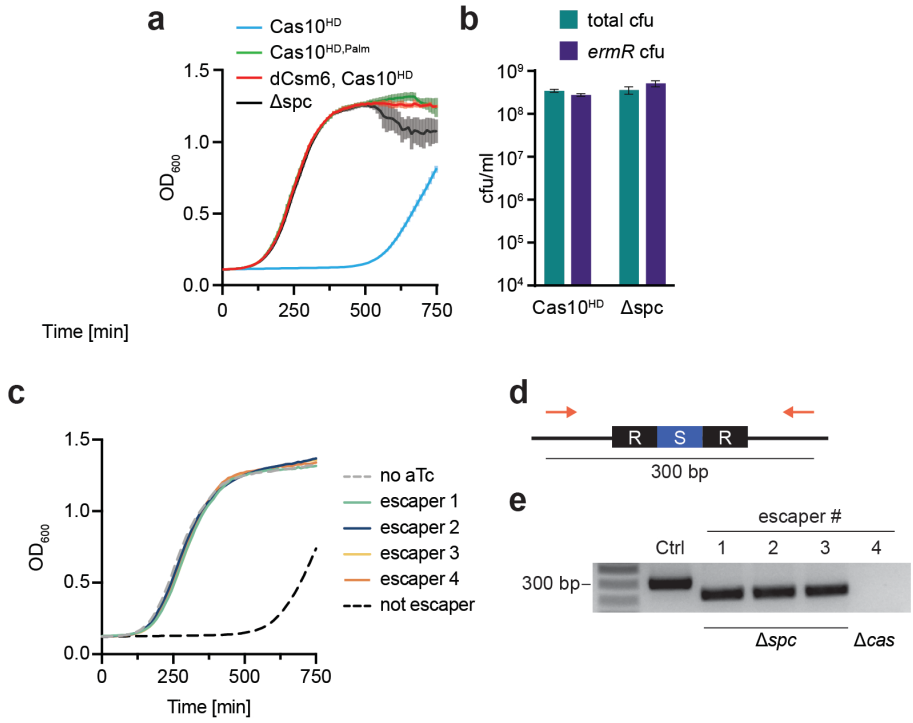

Figure 3.7. The Cas10 HD domain is required for efficient plasmid clearance during type III-A pTarget immunity. ***(a)*** *The OD_{600} of staphylococci harbouring pTarget and the specified pCRISPR was measured every 10 min after induction of protospacer transcription by the addition of high levels of aTc, in the absence of erythromycin. Each data point represents the mean of three biological replicates ± s.e.m.* ***(b)*** *Cells at the end of the experiment in a were spotted onto TSB plates in the presence or absence of erythromycin, selecting for the presence of pTarget, and enumerated. Each bar represents the mean of three biological replicates ± s.e.m.* ***(c)*** *Cells that recovered at the end of the experiment in in Fig. 3.7a (expressing the $Cas10^{HD}$ mutant) were streaked out to obtain single colonies. Four colonies were picked and seeded into a new microwell plate where growth was tracked by measuring the OD_{600} every 10 minutes in the presence of high levels of aTc. The four isolates displayed unrestricted growth, similar to a control culture lacking aTc induction of CRISPR immunity. A culture that had previously not grown in the presence of aTc (non-escaper) is shown as a targeting control.* ***(d)*** *All four escapers were analysed by PCR using primers that include the CRISPR locus. and Sanger sequencing to determine the presence of escaper mutations. A 300 bp PCR amplicon spanning the pTarget-specific spacer (S) and its two flanking repeats (R) is expected from wild-type loci.* ***(e)*** *Agarose gel electrophoresis of the PCR products from the amplification illustrated in panel* ***b.*** *A wild-type locus was used a control (ctrl). Escapers 1 through 3 contained a reduced CRISPR array. Sanger sequencing revealed that they carry a deletion of the spacer and a one of the flanking repeats (Δspc). Escaper 4 did not produce a PCR product. Sanger sequencing of plasmid DNA isolated from this escaper showed that it harboured a complete deletion (~9kb) of the cas operon (Δcas). All these deletions render pCRISPR unable to target pTarget, and therefore allow to escape the Csm6-mediated growth arrest.*

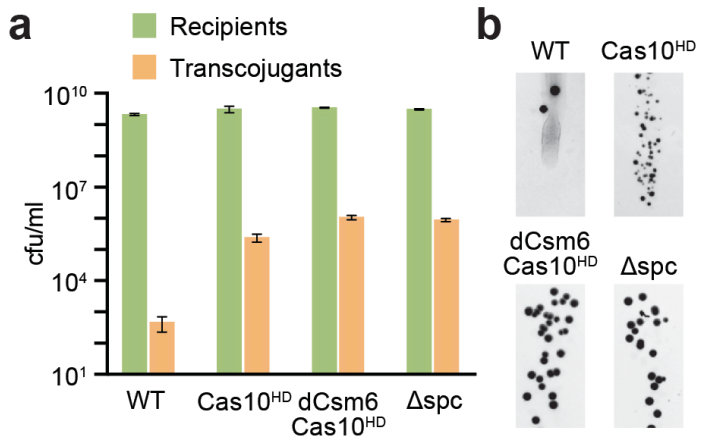

Figure 3.8. The Cas10 HD domain is required for pG0400 immunity. ***(a)*** *Conjugation of pG0400 into S. aureus cells carrying the specified pCRISPRs, after filter mating. Each bar represents the mean of three biological replicates ± s.e.m.* ***(b)****, Representative plate images from the conjugation experiment in panel c, performed once. Scale bars, 5 mm.*

3.8 Discussion

Here, I show that Csm6 RNase activity is only required for *S. epidermidis* type III-A CRISPR-Cas immunity under conditions of low transcription of the plasmid target sequence. I found that activated Csm6 results in the non-specific degradation of host and plasmid transcripts and promotes the fast clearance of the target plasmid by reducing the expression of genes important for its replication and/or maintenance. In addition, the global depletion of host transcripts leads to a growth arrest. Csm6 RNase activity is not sufficient for plasmid clearance and the slow degradation of plasmid DNA by the Cas10 nuclease is also required. Upon complete plasmid loss promoted by both activities, the target transcript is eliminated and cA production ceases (and is eventually degraded), resulting to Csm6 deactivation and the return of normal growth. Although our studies only addressed type III-A CRISPR-Cas systems, type III-B systems have a similar targeting mechanism, and it is therefore likely that their accessory RNase, Csx1[139], has a similar function to that described here for Csm6.

I believe that when target transcription is low there are few recruitment events of the Cas10-Csm complex to the invading plasmid, and DNA cleavage is inefficient. Under

such conditions a “two-hit punch” is required for immunity, where Csm6 prevents the plasmid from replicating while the Cas10-Csm complex slowly degrades the plasmid. This is in line with studies of the type III-A anti-phage response[56], where it was found that Csm6 is required when the target is expressed late in the phage lytic cycle. In this case, the accumulation of many phage genomes before the activation of type III-A immunity would overwhelm the DNA cleavage capacity of the Cas10-Csm complex. Csm6-dependent degradation of both host and phage transcripts would hinder phage replication and allow time for the gradual degradation of phage genomes by Cas10. Also, Csm6 was determined to be essential when the phage target contains multiple mismatches with the crRNA guide. It is conceivable that such mismatches will specifically inhibit activation of the HD domain but not the Palm domain, leading to low DNA cleavage but normal Csm6-mediated RNA degradation. Therefore Csm6 appears to be an accessory RNase that rescues type III-A immunity when DNA targets are difficult to detect or eliminate.

One fundamental difference in the Csm6-mediated response against phages and plasmids is the observation of a host growth defect only in the latter. I speculate that as result of both the high concentration of phage genomes and the preferential transcription of these genomes during infection, the host cell contains a much higher proportion of phage transcripts[56]. Therefore, assuming that the cA degradation enzymes prevent accumulation of the inducer in most of the bacterial cytoplasm except around its site of production where there is a local concentration of active Csm6 molecules, during the anti-viral type III-A CRISPR response this volume will be occupied mostly by phage RNA, and the impact on host transcripts (and thus growth) is less noticeable.

Interestingly, the indiscriminate RNA degradation of Csm6 has similarities to ribonuclease toxins. Indeed, HEPN domains are found in a wide range of predicted toxin-antitoxin (TA) modules and abortive infection systems in all three domains of life. In contrast to the function sometimes ascribed to these systems, our data suggests that the role of Csm6 in immunity is not through the induction of host cell death, but instead leads to a temporary growth arrest. Although not found in our experimental system, there are possible scenarios where Csm6 activation could trigger cell death. For example, at very high phage concentrations the Cas10-Csm complex might be unable to clear the virus, prolonging Csm6 activity to degrade total cellular RNA. This could kill the host, but would also prevent the release of functional viral particles, similar to programmed cell death pathways of eukaryotic cells[140] and abortive infection defense systems of prokaryotes[20]. Further, there might be circumstances where the DNase activity of Cas10 is completely compromised, such as base modifications in phage DNA that affect cleavage, or anti-CRISPR proteins that could specifically inhibit the HD domain of Cas10. Future studies

will determine if Csm6 global degradation of host transcripts can mediate abortive infection-like immunity.

Chapter IV – Discovering novel ligand-activated type III CRISPR effectors

In this short chapter, I outline the approach used to identify further novel cA-activated accessory genes in diverse type III CRISPR systems. This work begun while finishing the work on Csm6 described in the previous chapter, and generated two projects I pursued further (in chapters V and VI).

4.1 Rationale behind type III-associated CARF genes

Type III CRISPR systems are capable of producing the ligand cA upon target recognition to activate downstream effectors. The most common accessory gene in type III loci is *csm6* (and the related gene *csx1*), encoding a non-specific RNase which provides another layer of immunity, required against weakly transcribed targets[55] or against phage targets expressed late in the infection cycle[56]. The CARF domain is responsible for cA binding in Csm6, with two CARF domains from the Csm6 homodimer forming a composite ligand binding site. Each CARF domain is fused to a HEPN domain, and ligand binding activates the RNase activity of the HEPN domain.

Yet, bioinformatic studies have revealed a range of different CARF domain-containing genes beyond just *csm6/csx1* [110,111,116,141]. These genes often contain a predicted effector domain fused to a CARF domain. For *csm6*, the effector domain is the HEPN ribonuclease, but other effector domains include other RNase domains like RelE or PIN domains, predicted restriction endonuclease-like (REase) domains, transmembrane domains, proteases, and transcription factors. It is therefore likely that these are diverse accessory factors activated by Cas10 and cA ligands within their respective type III CRISPR systems. These effectors might then induce toxicity in the host cell, either leading to dormancy (inhibiting the replication of the invader) or cell death in a sequence non-specific manner, analogous to Csm6.

4.2 Identifying and testing new CARF genes for functionality in S. aureus

To investigate the hypothesis of CARF genes non-specifically inducing growth arrest in the presence of cA production, I first identified promising candidates, choosing from previously published lists of predicted CARF genes [116]. Since they would have to be active in a heterologous host (*S. aureus*), I prioritised genes found in mesophilic gram-positive bacteria wherever possible. I selected 14 genes of three categories, being REase-containing genes (six chosen), RNase-containing genes (four RelE and two PIN genes chosen), and transmembrane domain-containing genes (two chosen). Candidates are summarized in table 4.1.

These CARF candidates were ordered to be synthesised as DNA fragments, all of them codon optimised for *S. aureus*. Each gene was then cloned onto a pTarget plasmid, which in addition to expressing the candidate constitutively, has an aTc-inducible promoter followed by a protospacer (see chapter III). A second plasmid, pCRISPR, contains the matching spacer, as well as the full type III system. However, this plasmid has a mutated Cas10 HD domain and mutated Csm6. Thus, this locus has no direct interference mechanism, but is able to synthesise cA upon activation. These two plasmids were transformed into *S. aureus*. For the experiment itself, aTc was added to cells habouring both plasmids to initiate pTarget protospacer transcription, and subsequent cA synthesis, and OD_{600} was monitored over time. An inactive CARF gene candidate will show no difference in growth, whereas an active CARF gene candidate, with non-specific catalytic activity, will result in a growth delay or growth arrest. The results are summarised in table 1.

Table 1. Overview of CARF genes tested for autotoxicity.

Bacterial origin	CARF gene class (predicted function)	Activity observed?
Syntrophomonas wolfei subsp. wolfei str. Goettingen	REase (DNase)	No
Prevotella nigrescens ATCC 33563	REase (DNase)	Yes
Bergeyella zoohelcum CCUG 30536	REase (DNase)	Yes
Treponema succinifaciens DSM 2489	REase (DNase)	Yes
Cardiobacterium hominis ATCC 15826	REase (DNase)	No
Ignavibacterium album JCM 16511	REase (DNase)	Yes
Thermococcus CL1	PIN (RNase)	Maybe (very minor)
Methanotorris igneus Kol 5	PIN (RNase)	No
Desulfococcus oleovorans Hxd3	RelE (RNase)	Maybe (very minor)
Desulfobacca acetoxidans DSM 11109	RelE (RNase)	Yes (minor)
Rhodomicrobium vannielii ATCC 17100	RelE (RNase)	Yes (minor)
Candidatus Cloacamonas acidaminovorans Evry	RelE (RNase)	No
Nitrosococcus halophilus Nc 4	Transmembrane helix (pore?)	Yes
Methylosarcina lacus LW14	Transmembrane helix (pore?)	No

The gene classes of the highest interest, CARF-REase and CARF-TM, had 4/6 and 1/2 genes providing robust toxicity, respectively (not shown). For further work, the CARF-REase from Treponema succinifaciens DSM 2489 was chosen and renamed Card1 (chapter V). The CARF-TM gene from Nitrosococcus halophilus Nc 4 was also selected, and renamed TM-1. Of the remaining genes, some RNases displayed modest toxicity, but were also generally quite toxic to cells in the absence of stimulation. Further characterisation of these could have been attempted with expression of each gene on a lower copy plasmid, but this was not prioritised. Yet, these results, combined with additional CRISPR-associated CARF gene candidates recently identified

bioinformatically [141] highlight the diverse and fascinating mechanisms of type III CRISPR immunity that remain to be understood.

Chapter V - Characterising the *in vitro* and *in vivo* functions of the new type III effector Card1

In this chapter, my work on the CARF accessory gene Card1 will be described. Card1 was identified to be active in *S. aureus* in the previous chapter, being able to induce cell toxicity upon cA production. Based on its REase domain, often being found in restriction enzymes and other enzymes that interact with and cut DNA, Card1 was predicted to also be a DNase. However, *in vitro* experiments revealed that Card1 can in fact cut both ssDNA and ssRNA. Still, *in vivo*, I was unable to detect any RNase activity, suggesting that the ssDNase activity is responsible for the anti-phage and anti-plasmid immunity observed, at least a heterologous host. This project resulted in a fruitful collaboration with the laboratory of Prof. Dinshaw Patel, of Memorial Sloan Kettering Cancer Center (NYC, USA). His group, through the excellent experimental work of Dr. Wei Xie, was able to solve the atomic structure of Card1, and contributed invaluable biochemical input. This chapter largely includes work performed by me, but some data generated from our collaborators is included, which is indicated appropriately throughout. A paper based on the work in this chapter is currently under peer review.

5.1 Introduction

Depending on their *cas* gene content, CRISPR-Cas systems can be classified into six different types [141]. Of these, type III systems display the most complex targeting mechanism. The crRNA in the type III Cas10 effector complex recognizes complementary invader's transcripts [52,57], resulting in the activation of two catalytic domains within Cas10. The HD domain initiates single-stranded DNA (ssDNA) cleavage near the target transcription site [52,128]; i.e. within the genome of the invader. At the same time the Palm domain converts ATP into 3'-5' cyclic oligoadenylate (cA) of various sizes, commonly cA_4 and cA_6[53,54]. These molecules function as secondary messengers that bind the CRISPR-Cas Associated Rossmann Fold (CARF) domain of Csm6 [142] or Csx1 [143], accessory RNases most commonly found in type III-A or III-B loci, respectively. Binding of cA to the CARF domain activates an RNase HEPN domain, through which Csm6 degrades host and invader transcripts non-specifically [55], inducing a growth arrest essential for the type III-A CRISPR-Cas immune response against targets that are transcribed either at low levels [55,131] or late in the viral infection cycle [56].

Recent bioinformatics studies revealed the existence of a great diversity of genes associated with type III CRISPR-*cas* loci [110,111]. Many of them contain CARF domains fused to different effector domains with predicted catalytic or regulatory functions [116]. One of the most abundant (found in 929/6665 CARF-containing proteins) is the PD-D/ExK

domain [141]. Biochemical and structural analysis determined that one such protein, *Thermus thermophilus* Can1, is activated by cA_4 binding to introduce nicks only in supercoiled DNA at 60 °C [144]. However, whether and how these type III-associated, CARF:PD-D/ExK proteins, can be activated to provide immunity to prokaryotes remains to be demonstrated. To investigate this, we characterized Tresu_2185, found in the type III-A CRISPR-*cas* locus of the mesophilic gram-negative spirochete *Treponema succinifaciens* [116] (Fig. 5.1). Tresu_2185 contains 373 amino acids (43.9 kDa), is a member of the Pfam family pfam09002 (domain of unknown function 1887, DUF1887), and is composed of an N-terminal CARF domain and a C-terminal restriction endonuclease-like (PD-D/ExK) domain, typically found in type II restriction endonucleases or Holliday junction resolvases involved in DNA restriction, recombination, replication and repair [145], where the two acidic residues coordinate a divalent cation important for catalysis.

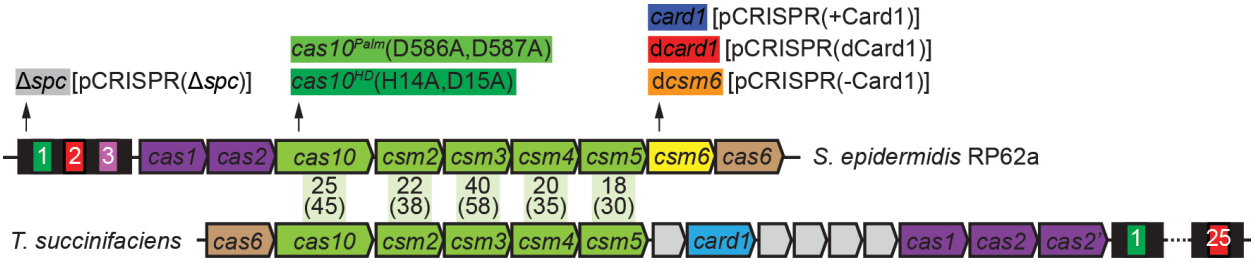

Figure 5.1. Context of Card1 in the *Treponema succinifaciens* CRISPR-Cas locus. *Schematic of the S. epidermidis type III-A locus showing the replacement of* csm6 *by* card1 *and the different mutations used later in this chapter. A comparison with the T. succinifaciens type III-A locus is provided. Numbers indicate the % identity between the genes that encode the Cas10 complex (with % similarity in parenthesis), as determined by EMBOSS Needle pairwise sequence alignment*[146].

5.2 ssDNase activity of Card1 in vitro

To evaluate the biochemical activity of Tresu_2185, I expressed and purified it from *Escherichia coli*, and incubated with different nucleic acids and cAs. I found that the addition of cA_4, but not cA_6, resulted in the degradation of circular ssDNA ΦX174 (Fig. 5.1a) and M13 ssDNA (5.1b), but cA_4 did not promote cleavage of supercoiled or linearised ΦX174 dsDNA (Figs. 5.1c-d), at least at 37 °C. ssDNA degradation required the addition of manganese, but not magnesium, calcium, or zinc, divalent cation (Fig. 5.1e) and resulted in a smear of products, suggesting that ssDNA cleavage is non-specific. These results demonstrate that Tresu_2185 is a cA_4-activated, non-specific ssDNA nuclease.

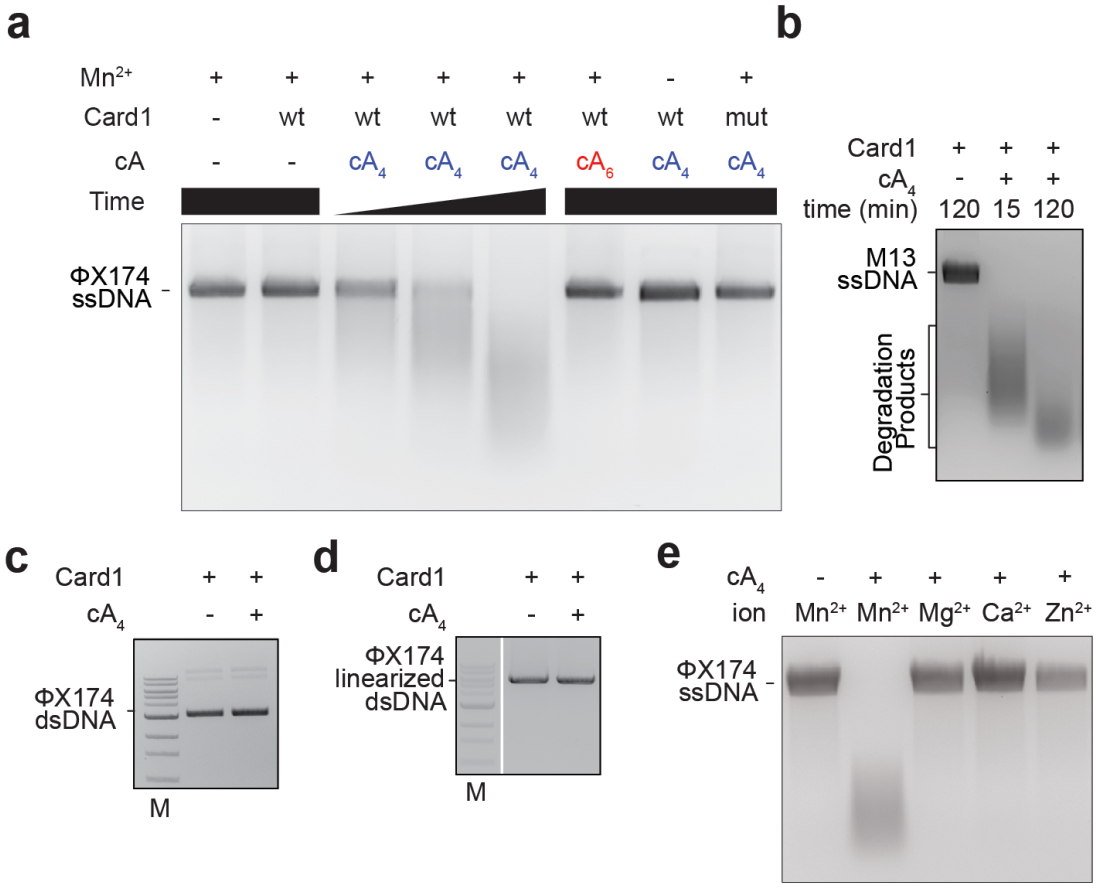

Figure 5.2. Card1 is a non-specific ssDNase. ***(a)*** *Card1 cleavage of ΦX174 circular ssDNA for 15 minutes (2, 5, 15 minutes in the increasing time annotation) in the presence of cA_4 and cA_6. Wild-type (wt) or catalytically dead (K308A+E310A) (mut) Tresu_2185 (Card1) is used, and Mn^{2+} is present or absent.* ***(b-d)*** *Cleavage of (b) M13 ssDNA, (c) uncut, supercoiled ΦX174 dsDNA, or (d) ΦX174 linearised dsDNA, for the time indicated* ***(b)*** *or for 30 minutes* ***(c-d)****, all visualised by agarose gel electrophoresis.* ***(e)*** *Card1 digestion of ΦX174 ssDNA (60 minutes) in the presence of cA_4 and different divalent cations, visualized by agarose gel electrophoresis*

To further understand the ssDNA cleavage requirements of Card1, cleavage products resulting from 2 hours of digestion of either ΦX174 ssDNA or M13 ssDNA were analysed by next-generation sequencing (Fig. 5.3). Since each DNA 5' end represents a cut site, the cut site preference of Card1 could be determined. Weblogos from ΦX174 digestion (Fig. 5.3a) and M13 digestion (Fig. 5.3d) showed a similar cleavage preference upstream of T(A/G) sites, with cleavage occurring across both genomes (Figs. 5.3b,e), and DNA fragments being on average 150-165 nucleotides in length (Figs. 5.3c,f).

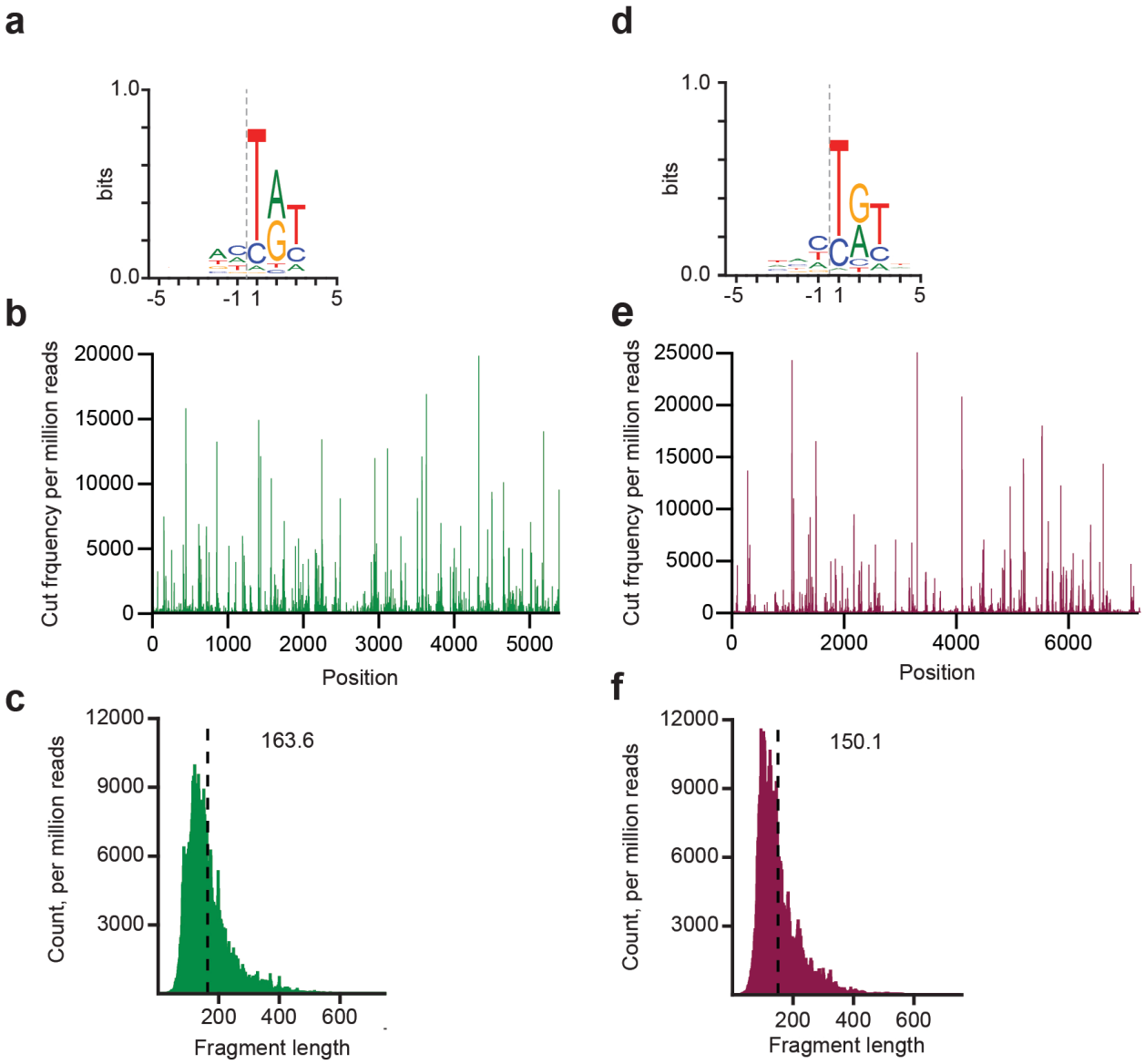

Figure 5.3. Next-generation sequencing analysis of Card1 ssDNase specificity. ***(a)***. *Cleavage preference of Card1, represented as a WebLogo, determined after next generation sequencing of ΦX174 degradation products (2 hours digestion). Five nucleotide positions upstream (-5 to -1) and downstream (1 to 5) of the detected cleavage sites are shown.* ***(b)*** *Overview of Card1 cleavage sites across the ΦX174 genome based on the 5' end mapping of DNA degradation products obtained after 2 hours of digestion, per 1 million reads. There appears to be preferential cleavage sites that may reflect lack of Card1 access to secondary structures formed within the ssDNA molecule. 26.7% of cuts occur at the 25 most frequent positions.* ***(c)*** *Fragment size distribution of the ΦX174 degradation products after 2 hours of Card1 digestion. The average fragment length (163.6 nucleotides) is marked by the dotted line.* ***(d-f)*** *Like* ***a-c****, but from digestion of M13 ssDNA by Card1. For* ***e****, 31.1% of cuts occur at the 25 most frequent positions.*

5.3 ssRNase activity of Card1 in vitro

Even if the REase catalytic domain found in Card1 is predicted to cut DNA, and is found in type II restriction enzymes and other enzymes interacting with DNA, I also tested Card1's activity against RNA substrates. Surprisingly, Card1 robustly degraded ssRNA oligonucleotides (Fig. 5.4a), but not dsRNA (Fig 5.4b). Interestingly, both Mn and Mg cations supported the ssRNase activity (Fig. 5.4c). I explored the sequence specificity using polyA, polyC and polyU oligonucleotides harboring fluorescent-quencher pairs, and found that all were equally degraded (Fig 5.4d), similarly to the non-specific RNase I control. Neither Card1 nor RNase I cleaved polyG, most likely due to the formation of higher order structures by G quartets[147]. Therefore, given the cleavage of both ssRNA and ssDNA, we renamed Tresu_2185 **cA**-activated ss**R**Nase and ss**D**Nase 1, or Card1.

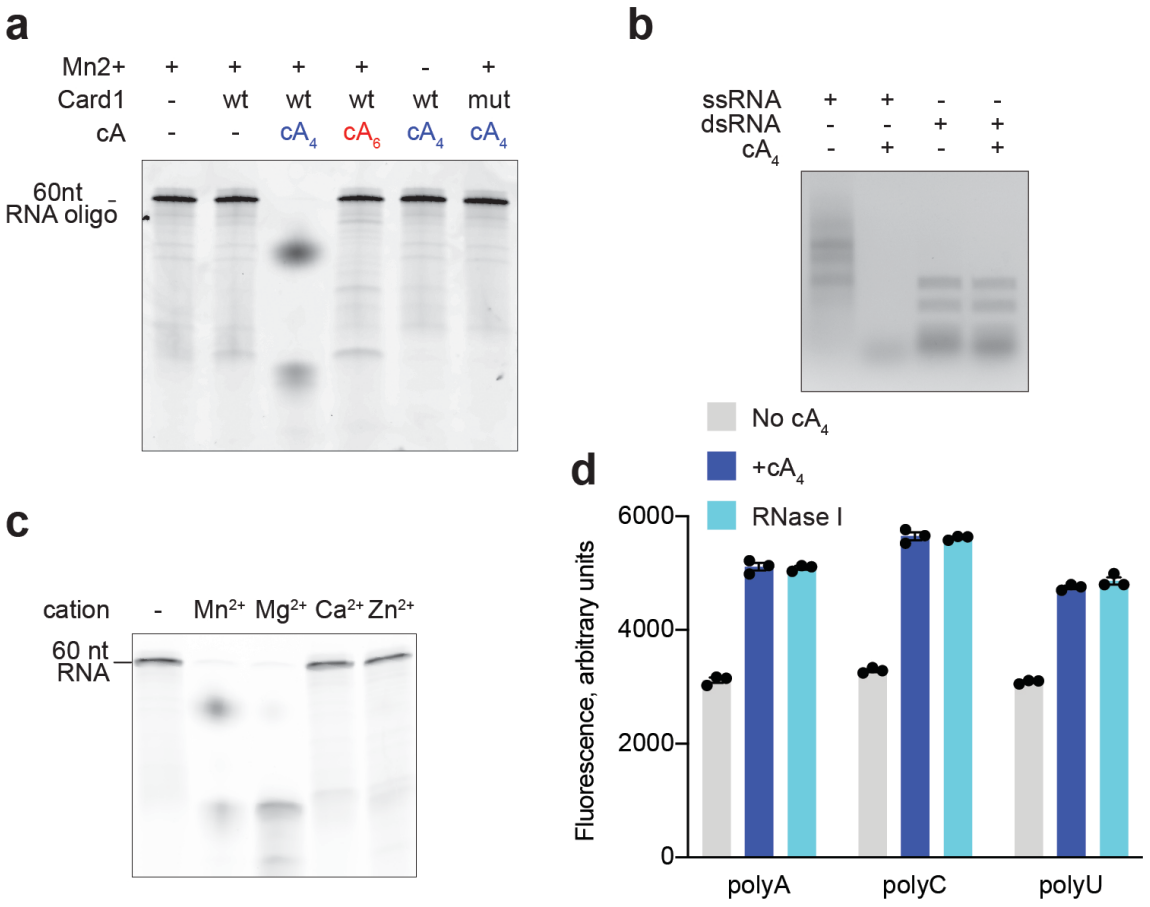

Figure 5.4. cAn-mediated cleavage of ssRNA by Card1 at 37 ºC. ***(a)*** *Card1 cleavage of a 60-nt ssRNA oligonucleotide for 15 minutes in the presence of cA_4 and cA_6, witho or without Mn^{2+}, or with a catalytically dead Card1 (K308A+E310A) (mut).* ***(b)*** *Card1 digestion of a ssRNA or a dsRNA molecular weight ladder. Card1 rapidly degrades the*

ssRNA, but not the dsRNA, ladder. ***(c)*** *Digestion of a 60-nucleotide (nt) RNA species for 15 minutes in buffers containing either no divalent cation, or either Mn^{2+}, Mg^{2+}, Ca^{2+}, or Zn^{2+}.* ***(d)*** *Cleavage of RNA oligonucleotides containing a fluorophore-quencher pair, measured as the increase in fluorescence, by Card1 with or without cA_4, or with the non-specific RNase I as a positive control. The RNA oligonucleotides are either poly-A_{15}, poly-C_{15}, or poly-U_{15}. Cleavage of poly-G_{15} could not be tested due to its resistance to cleavage by RNases. Each bar represents the mean of three replicates ± s.e.m., given as relative fluorescent units*

In order to compare the relative efficiencies of the DNase and RNase activities, I labeled 30-nt ssDNA and 50-nt ssRNA oligonucleotides with Cy3, as well as the 30-nt ssRNA and 50-nt ssDNA oligonucleotides with Cy5, incubated the 30-nt species together, or the 50-nt species together, with increasing concentrations of Card1/cA_4 in the presence of Mn (Fig. 5.5). I observed that both ssRNA oligos were cleaved at lower nuclease concentrations than those required to observe ssDNA degradation.

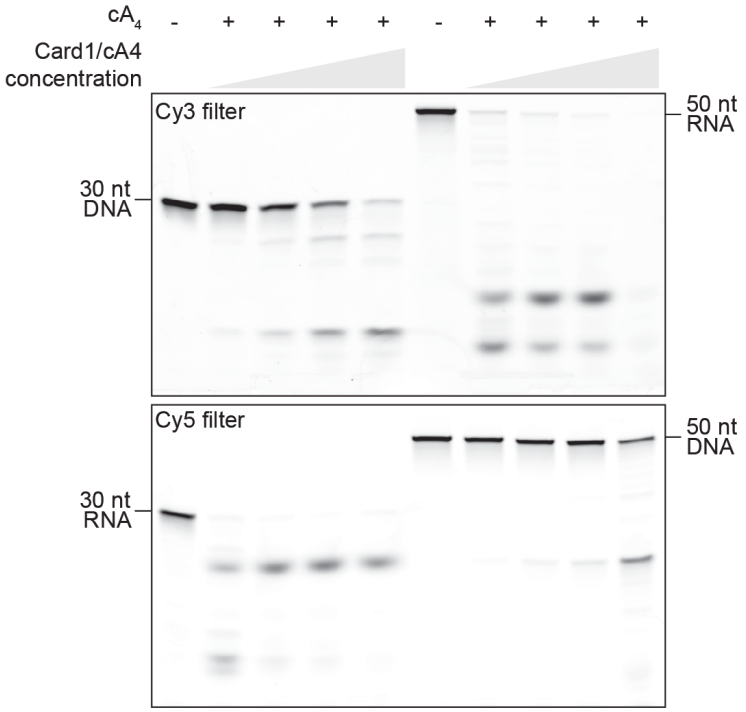

Figure 5.5. A comparison of the relative rates of the DNase and RNase activities of Card1. *Simultaneous Card1 digestion of a pair of 30-nt DNA and RNA oligonucleotides, or of a pair of 50-nt DNA or RNA oligonucleotides, with increasing concentration of Card1 and cA_4. This results in direct competition between the DNase and RNase activities of Card1 in each reaction. For each pair, one oligonucleotide is labelled with Cy3 and the other with Cy5 fluorescent groups, and the two panels display the same gel imaged through different filters. All reactions were quenched after 15 minutes.*

5.4 Structural studies of Card1

To further understand Card1 at a molecular level, our collaborators solved the structure of Card1 in the apo-, cA_4-, and cA_6-bound forms. The 2.3 Å crystal structure of apo-Card1 showed a symmetric dimeric topology (space group *C*2; 1460 $Å^2$ interface) with a sizeable central hole between the CARF and REase domains. The crystal structure of the co-crystallized cA_4-Card1 complex at 3.0 Å, which retained the dimer topology (space group $P2_1$, 1608 $Å^2$ interface), shows the second messenger bound in the central cavity, with a Mn cation bound to the individual catalytic pocket of each monomer. Pronounced conformational changes are observed following superposition of the apo- and cA_4-bound dimeric structures of Card1 with an r.m.s.d. = 5.2 Å (over 709 residues Fig. 5.6c). However, superposition of individual monomers of apo- and cA_4-bound Card1 yields a much smaller r.m.s.d. of 1.0 Å over 301 residues, suggesting that the transition to the active conformation involves a rotation of the monomers relative to each other after cA_4 binding. Molecular modelling confirmed that one single-stranded DNA can sterically fit in each REase subunit.

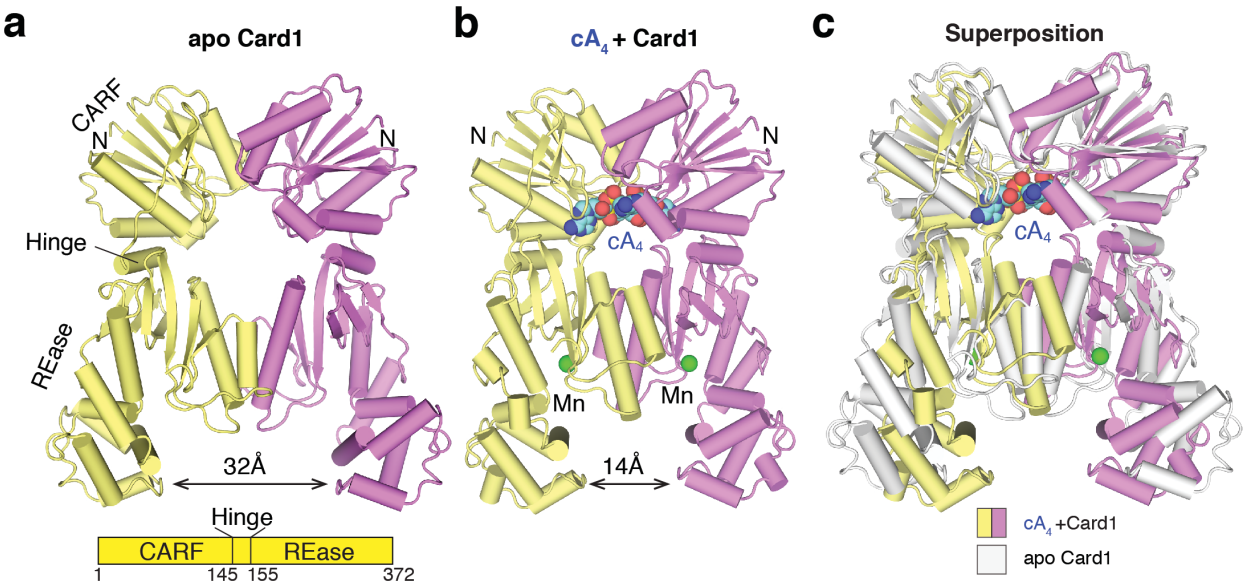

Figure 5.6. The crystal structure of apo- and cA4-bound Card1. (a) *Crystal structure of dimeric apo-Card1 at 2.3 Å resolution. The monomers are colored yellow and magenta in the symmetrical dimer, with labeling of the CARF, hinge and REase domains, and highlighting the central hole between domains.* ***(b)*** *Crystal structure of co-crystallized cA_4-bound to dimeric Card1 at 3.0 Å resolution, with one cA_4 (in space-filling representation) bound per dimer and positioned within the periphery of the central hole. In addition, one Mn (in green) per monomer is bound in each REase catalytic pocket.* ***(c)*** *Conformational transitions following superposition of apo- (in silver) and cA_4-bound (in color) states of dimeric Card1. Data courtesy of Dr. Wei Xie and Prof. Dinshaw Patel.*

Isothermal calorimetry showed that cA_4 and cA_6 bind Card1 with similar affinities (15 nM and 78 nM respectively) (not shown). Indeed, the structure of cA_6-bound Card1 showed cA_6 bind in the same pocket as cA_4. However, binding by cA_6 shows no significant conformational change relative to the apo structure, helping to explain why Card1 is activated by cA_4 and not cA_6.

The structural studies also revealed the exact binding mode of cA_4, including important residues, and conformational changes likely responsible for allowing catalysis. This will not be discussed in further detail here.

5.5 Effect of Card1 activation in staphylococci

Next, I investigated the function of Card1 during the type III-A CRISPR-Cas immune response. I hypothesized that the non-specific ssDNA cleavage could introduce DNA lesions, for example on R-loops generated during transcription or ssDNA intermediates that result from DNA replication. On the other hand, the ssRNase activity could lead to the degradation of host transcripts. Since deleterious activity against DNA would in principle also affect cellular RNAs, I first looked at the effect of Card1 in the transcriptome. To do this, I constructed pCRISPR(+Card1) by cloning into the staphylococcal plasmid pC194 [148]the *Staphylococcus epidermidis* RP62 type III-A locus, which Cas10 complex genes display a high similarity (30-58 %) to those of the *T. succinifaciens* type III-A locus, carrying the *card1* open reading frame instead of that of the cA-activated accessory protein of staphylococci, Csm6 (Fig. 5.1). As a control I introduced mutations that inactivate Card1 (E308A, K310A) in pCRISPR(dCard1), that inactivate Csm6 in pCRISPR(-Card1) or that lack a targeting spacer in pCRISPR(Δ*spc*). Each of these plasmids were transformed into *Staphylococcus aureus* RN4220 [135] cells containing pTarget (described in section 3.3), a second plasmid producing a target transcript, i.e. complementary to the crRNA expressed by the pCRISPR plasmids, that is under the control of an anhydrotetracycline(aTc)-inducible promoter[55]. In this experimental system, the addition of the inducer triggers the type III-A response, with the production of cA by the Palm domain of Cas10 and the subsequent activation of Card1. RNA was extracted from +Card1 or dCard1 cultures 3 minutes after addition of aTc for RNA-seq analysis. The standard normalization procedure, against the total number of reads in the sample, cannot be used if Card1 causes global RNA destruction. Therefore, an equal amount of *Listeria seeligeri* RNA was added to each sample prior to RNA extraction for normalization of the *S. aureus* RNA reads. From two replicates, the average number of reads mapping to *S. aureus*, pCRISPR and pTarget genes was calculated and the transcript abundance of every gene at 0 or 3 minutes

after induction was plotted. I was not able to detect significant differences between the presence or absence of Card1 nucleases activities (Fig. 5.7a-c). This is in contrast to previous results obtained for the type III-associated RNase Csm6 (chapter 3.4), whose activation leads to significant RNA degradation (Fig. 3.4a-c). To corroborate the absence of Card1 effects on cellular transcripts, I performed Northern blot analysis of the target RNA and *repF* transcript in pTarget, as well as the *def*, and *msaB* mRNAs of the host, either before or three minutes after aTc induction of the two cultures (Fig. 5.7d). I also included a wild-type pCRISPR control, expressing the non-specific RNase Csm6. While the selected transcripts were degraded by Csm6, no difference was detected in the presence of Card1 activity. Altogether these results suggest that Card1 activation, either through attacking host or plasmid ssDNA or ssRNA, does not substantially affect the transcriptome of *S. aureus*.

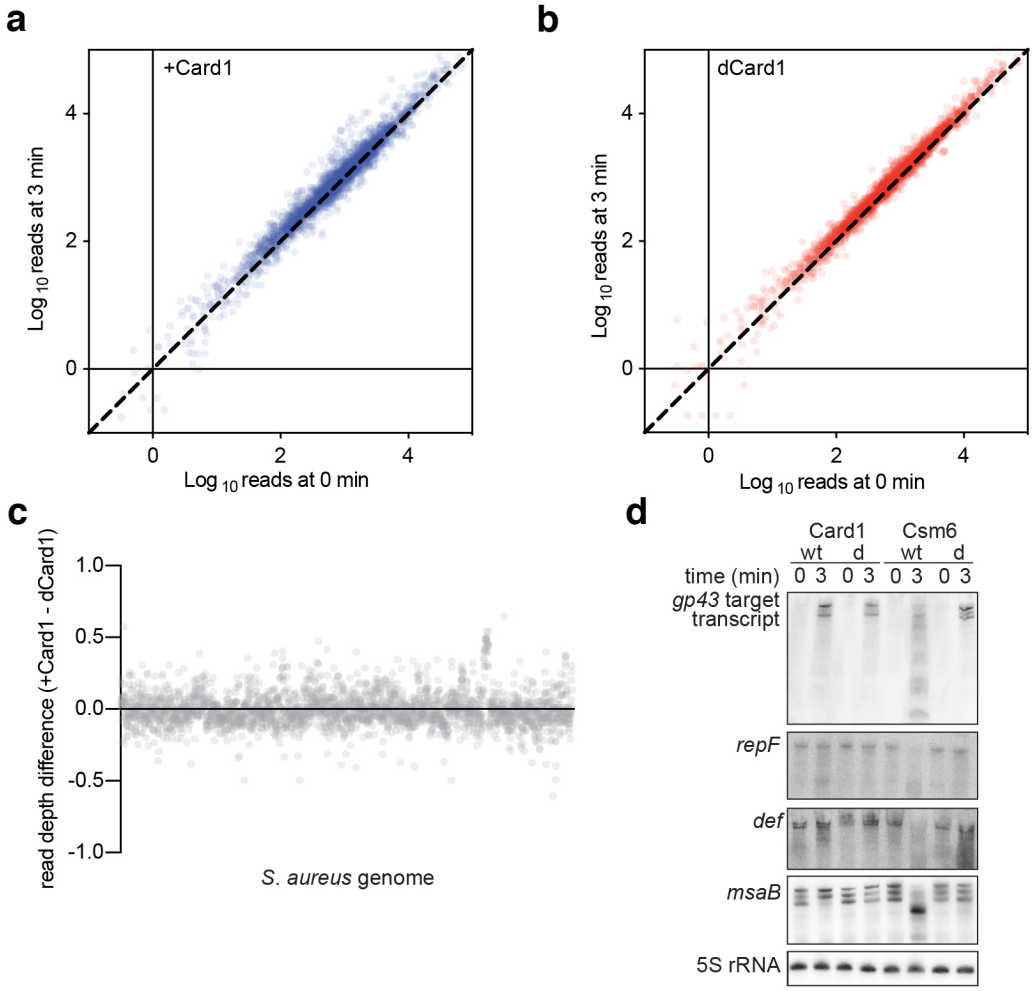

Figure 5.7. The RNase activity of Card1 is not detected in vivo. ***(a)*** *RNA-seq of staphylococci harbouring pTarget and pCRISPR-Cas10^{HD}, and +Card1. At 0 minutes, targeting is induced by the addition of aTc, and cells are harvested after 3 minutes. An equal amount of RNA from Listeria seeligeri was added to all samples prior to RNA purification to allow absolute comparison between timepoints. Each dot represents a gene, and is the average of two biological replicates. Genes that fall on or near the identity line are unchanged by 3 minutes of Card1 activity.* ***(b)*** *Like (**a**), but in cells carrying a catalytically dead Card1 (dCard1).* ***(c)*** *A comparison between the log_{10} read depth for all individual chromosomal genes between +Card1 cells and dCard1 cells, at 3 minutes. A value of 0 means that a gene showed no difference between +Card1 and dCard1 cells. Overall, there is no clear trend for depletion (or enrichment) in +Card1 cells relative to dCard1 cells.* ***(d)*** *Northern blot analysis of cells carrying pTarget and pCRISPR-Cas10^{HD}, with either +Card1, dCard1, +Csm6, or dCsm6. Targeting was induced at time 0 with the addition of aTc, and RNA was analysed with probes specific to the protospacer target transcript (in pTarget), the plasmid replication gene repF (in pTarget), the* def *gene (peptide deformylase, in the S. aureus chromosome), or the* msaB *gene (in the msaABCR operon, in the S. aureus chromosome). 5S rRNA is used as a loading control. Card1 activation showed no detectable RNA degradation, in contrast to robust RNA depletion following Csm6 activation. OD_{600} measurements confirmed that the +Card1 and +Csm6 cells both experienced growth arrest.*

Next, I explored whether Card1 attacks the host DNA. However, the quantification of non-specific nicks across chromosomal or plasmid DNA using next-generation sequencing is not possible. Instead, I hypothesized that chromosomal DNA lesions caused by the ssDNA activity of Card1 would result in cell toxicity. To test this, I monitored the growth of the different cultures after the addition of aTc, in the absence of antibiotic selection of pTarget (Fig. 5.8a). dCard1 cultures continued exponential growth, similarly to Δ*spc* cells that cannot trigger the type III-A CRISPR-Cas immune response. In contrast, +Card1 cultures displayed a severe growth defect after the induction of target transcription, which was completely eliminated by the introduction of mutations in the Palm domain of Cas10 (D586A, D587A; Cas10^{Palm}) that prevent the production of cA (Fig. 5.8a). I observed an increase in OD_{600} at around 10 hours after addition of aTc, which was a result of the propagation of “escaper” cells carrying re-arranged, non-functional pCRISPR or pTarget plasmids within +Card1 cultures (Figs. 5.8b,c).

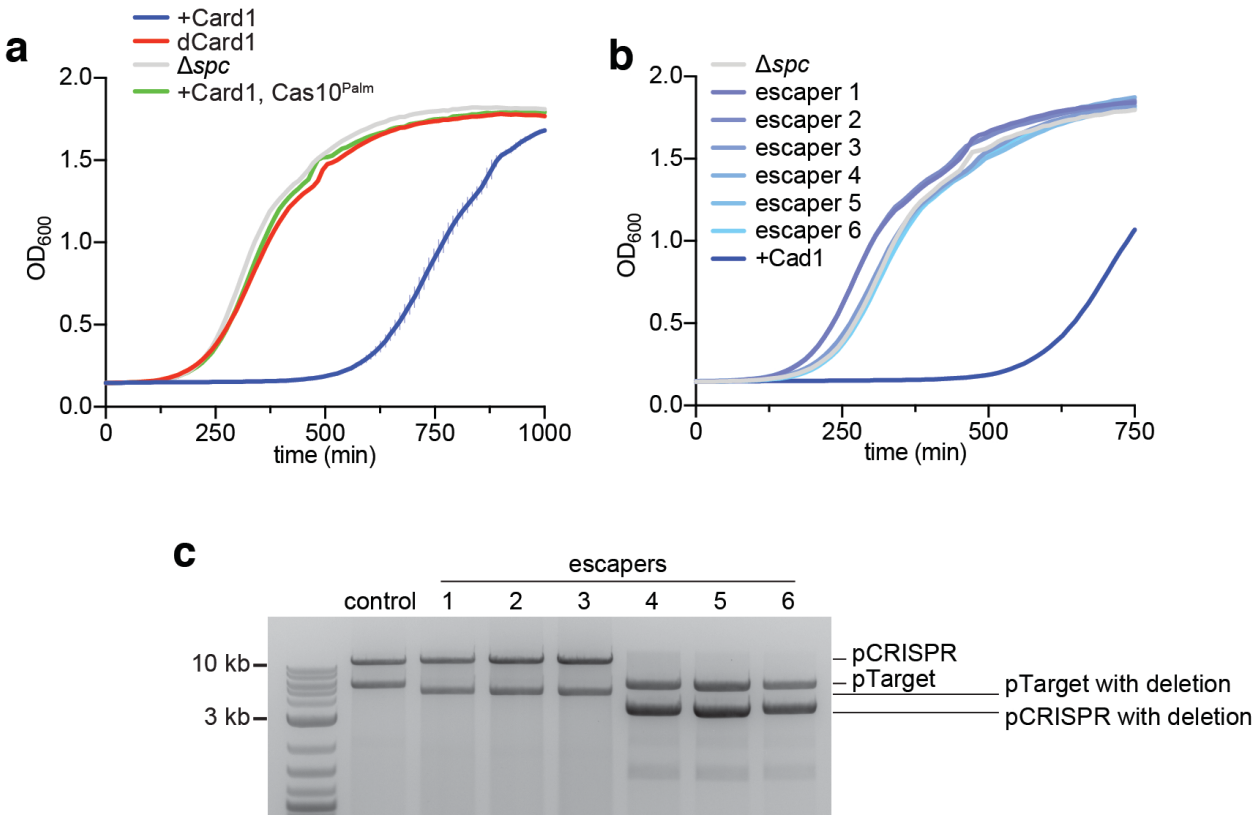

Figure 5.8. Effect of Card1 activation on growth of staphylococci. ***(a)*** *Growth of staphylococci carrying pTarget and different pCRISPR variants, measured as OD_{600} after the addition of aTc to induce the production of cA_4 by the Cas10 complex, in the absence of antibiotic selection for pTarget. Mean of three biological triplicates ± s.e.m. are reported.* ***(b)*** *Growth of staphylococci carrying different pCRISPR(+Card1) taken from six escaper colonies obtained from streaking out surviving cells from +Card1 in* ***(a)****, measured as OD_{600} after the addition of aTc to induce the production of cA_4 by the Cas10 complex. Mean of three biological triplicates ± s.e.m. are reported.* ***(c)*** *Agarose gel electrophoresis of plasmid DNA was extracted from escaper cells grown in* ***(b)****, showing deletions in pTarget or pCRISPR. Sanger sequencing determined the same promoter deletion in pTarget escapers 1-3, and similar pCRISPR deletions in escapers 4-6, all comprising the whole CRISPR-cas locus.*

The lack of growth induced by Card1 could be due to either the arrest or the death of individual cells within the culture. To distinguish between these possibilities, I enumerated viable staphylococci after Card1 induction, plating culture aliquots taken at different times after addition of aTc on solid media lacking the inducer (Fig. 5.9). This procedure removes the inducer and turns off Card1, allowing the formation of colonies from cells that were arrested in the liquid culture, but not from those that died after activation of Card1. Over a course of three hours of target transcription and Card1

activation, I observed that while dCard1 cultures displayed a steady growth and an increase in colony formation, +Card1 cultures showed an initial decrease in colony counts of an order of magnitude that slowed down after one hour, demonstrating the presence of a population of viable cells that cannot grow, but do not quickly die, upon activation of the nuclease. To determine what fraction of these colonies are escapers, I also enumerated colonies resistant to aTc induction (Fig. 5.9). I found that escaper numbers rise over time and comprise approximately half of the cells in the culture at the end of the experiment; the other half are *bona fide* dormant cells.

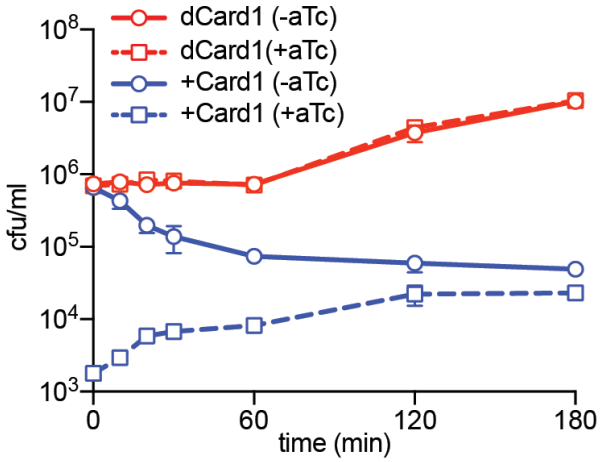

Figure 5.9. Enumeration of surviving cells following continuous Card1 activity. *Enumeration of colony-forming units (cfu) from staphylococcal cultures carrying different pCRISPR variants where cA_4 production was activated by the addition of aTc to start target transcription. At the indicated times after induction aliquots were removed and plated on solid media with or without aTc to count the remaining viable cells. Mean of three biological replicates ± s.e.m are reported.*

Next, I looked at the effects of Card1 activation on pTarget by agarose gel electrophoresis of plasmid DNA extracted at different time points after aTc addition (Fig. 5.10a). In dCard1 cultures the plasmid remained intact for 120 minutes after the activation of cA production, similarly to the Δ*spc* control. In contrast, +Card1 cells cleared pTarget, but not pCRISPR, 20 minutes after addition of the inducer. However, in cells lacking the ssDNase activity of Cas10 (H14A, D15A; $Cas10^{HD}$), but that are still able to produce cA, pTarget remained intact (Fig. 5.10b) and the Card1-mediated growth arrest was maintained (Fig. 5.10c), a result that highlights the importance of Cas10 for target DNA destruction.

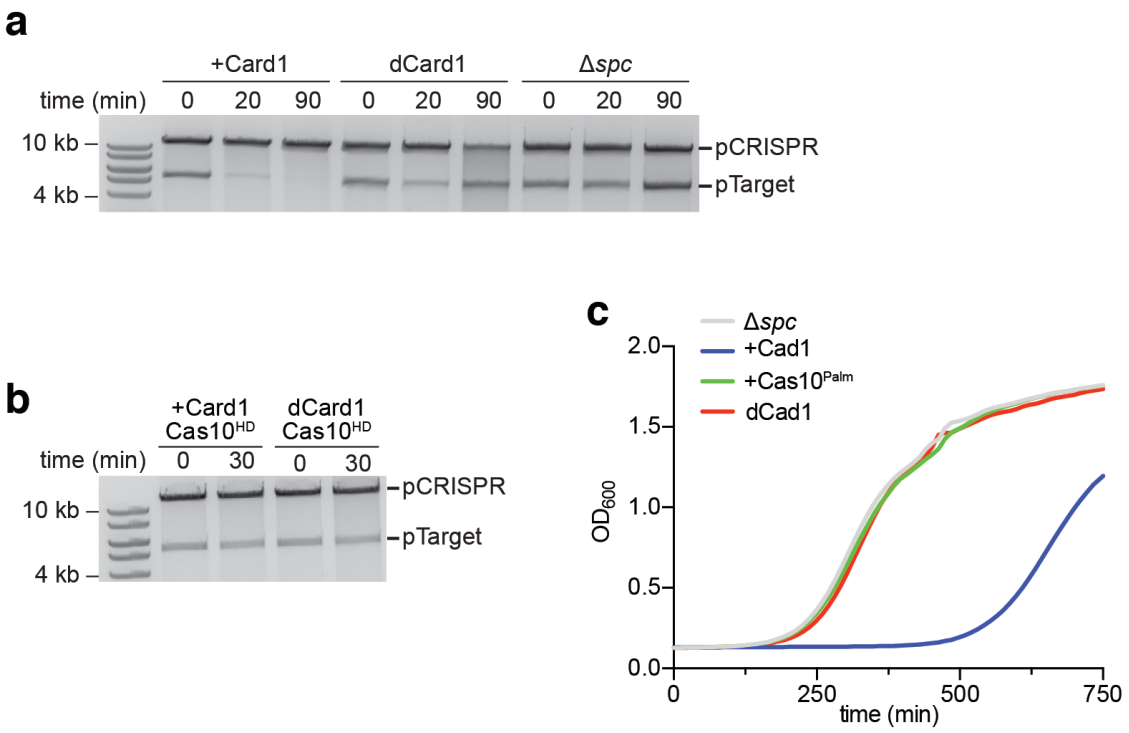

Figure 5.10. Effect of Card1 on pTarget stability. ***(a-b)*** *pTarget plasmid curing assay, where plasmid DNA was extracted from cells containing pTarget and different pCRISPR plasmids after induction with aTc. Plasmids were linearized and visualized by gel electrophoresis. Gel images are representative of three independent experiments.* ***(c)*** *Growth of staphylococci carrying different pCRISPR variants expressing Cas10^{HD}, measured as OD_{600} after the addition of aTc to induce the production of cA_4 by the Cas10^{HD} complex. Mean of three biological triplicates ± s.e.m. are reported.*

Together with the data showing the absence of detectable transcript degradation, these results show that the ssDNase activity of Card1, can produce a growth arrest of the host cell, presumably by introducing DNA lesions in the host chromosome, and act synergistically with Cas10 to specifically eliminate the target DNA (but not pCRISPR, which has the same replication mechanism as pTarget).

5.6 The role of Card1 in anti-phage immunity in staphylococci

I also tested the importance of Card1 during immunity against phage infection. Due to the dependence on target transcription to activate the HD domain of Cas10, type III-A immunity results in the rapid elimination of the phage DNA from the host when the target is expressed early during infection and the viral genome has not yet replicated to increase its copy number[56]. In contrast, when the viral target is located in a late-expressed transcript, the Cas10 complex can only initiate its attack on the invading DNA after the

phage has replicated and accumulated in the host. In this situation the complete degradation of the viral genomes within the infected cells is much slower[56] and, during the type III-A response in staphylococci, efficient immunity requires the activity of the Csm6 RNase [56]. I programmed the different pCRISPR plasmids with spacers targeting either early- or late-expressed viral genes, infected the cultures with staphylococcal virulent phages at a multiplicity of infection (MOI) of 2-8 and followed their growth to determine the effectiveness of the type III-A immune response in the presence or absence of Card1. As expected from previous results[56], when the early *ORF9* transcript of the lambda-like, dsDNA phage Φ12γ3 [69] was targeted, the presence of Card1 activity was not required for immunity (Figs. 5.11a, b). In contrast, when immunity was activated by the late *ORF27* transcript, +Card1 but not dCard1 cultures were able to survive infection (Fig. 5.11c). Moreover, *in vitro* data from our collaborators revealed that a series of mutated Card1 variants (S11A, Q13A, M42A, Y122A, I125A) were unable to bind cA_4 and perform catalysis (data not shown). During phage infection, survival of staphylococci also required the ability to bind cA_4, as the same mutations in the nucleotide binding pocket abrogated Card1-mediated immunity (Fig. 5.11d).

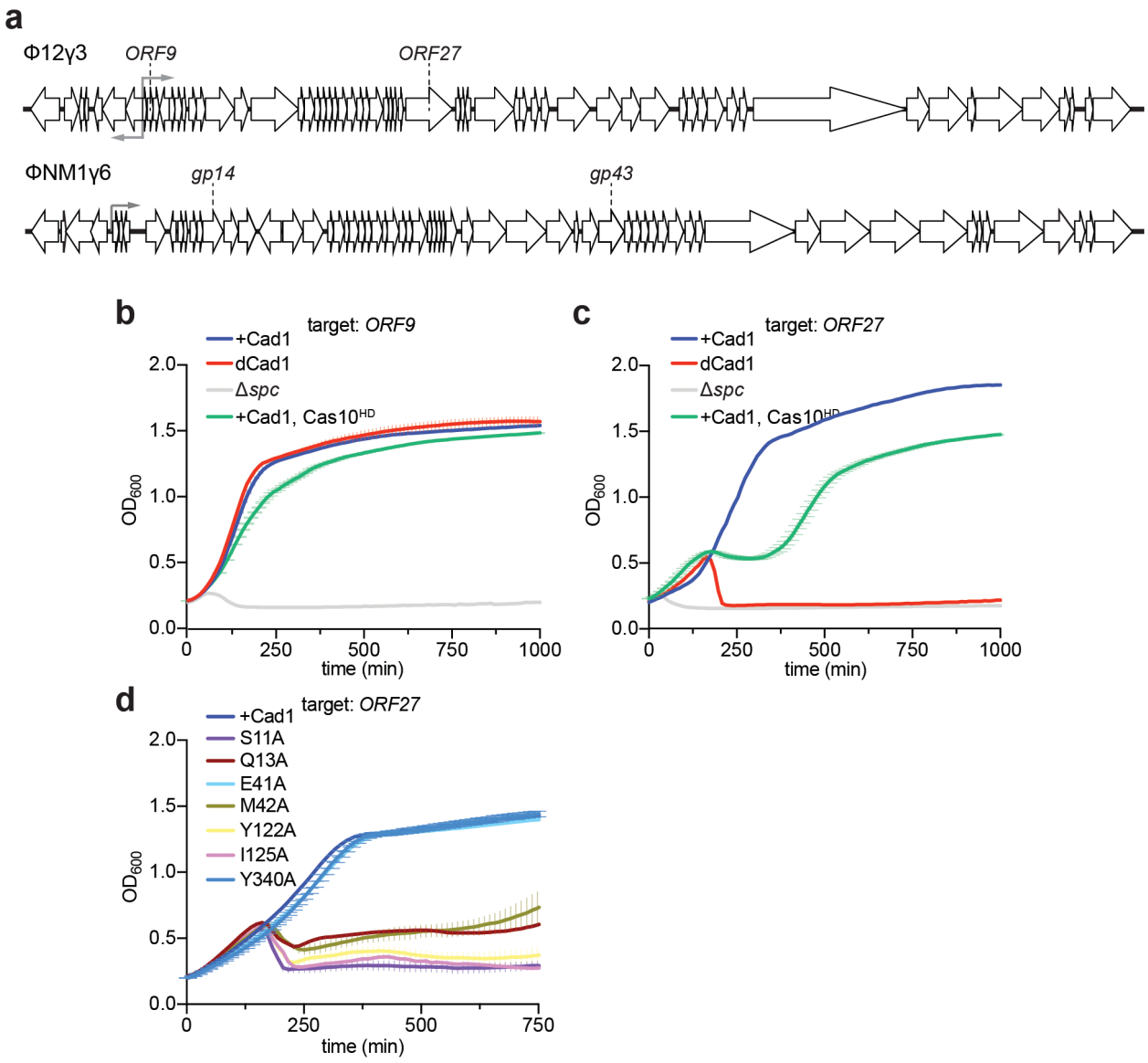

Figure 5.11. Card1 provides anti-phage immunity against phage Φ12γ3. ***(a)*** *Schematic of the genomes of the staphylococcal phages used in this study, Φ12γ3 and ΦNM1γ6, showing the location of the transcripts targeted by the type III-A CRISPR-Cas system. Grey arrows indicate promoters.* ***(b)*** *Growth of staphylococci carrying different pCRISPR variants programmed to target the ORF9 transcript of Φ12γ3, measured as* OD_{600} *at different times after infection at a multiplicity of infection (MOI) between 2 and 8. Mean of three biological triplicates ± s.e.m. are reported.* ***(c)*** *Same as in* ***(b)*** *but targeting the ORF27 transcript at an MOI ~8. Mean of three biological triplicates ± s.e.m. are reported.* ***(d)*** *Same as in* ***(c)*** *but following cultures carrying different mutations in the* cA_4 *binding pocket of Card1, at an MOI ~15.*

Similar results were obtained when the pCRISPR plasmids were programmed with spacers that target an early or late transcript of the staphylococcal phage ΦNM1γ6 [51] (Figs. 5.12a-b).

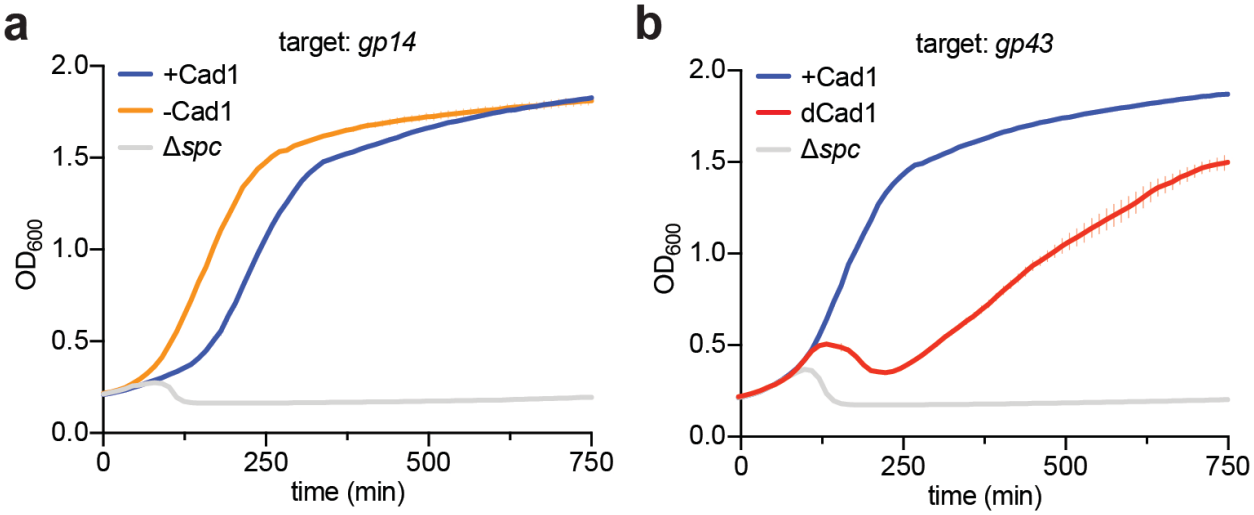

Figure 5.12. Card1 provides anti-phage immunity against phage ΦNM1γ6. (a). *Growth of staphylococci carrying different pCRISPR variants programmed to target the* gp14 ***(a)*** *or* gp43 ***(b)*** *transcript of ΦNM1γ6, measured as* OD_{600} *at different times after infection at a multiplicity of infection (MOI) of 15. Mean of three biological triplicates ± s.e.m. are reported.*

I also measured phage propagation in cells programmed to target the ORF27 transcript of Φ12γ3 by counting plaque forming units (pfu) at different times after infection (Fig. 5.13a). I found that while the phage propagated to high titers in both -Card1 and Δ*spc* cultures, +Card1 cells effectively suppressed Φ12γ3 from the culture.

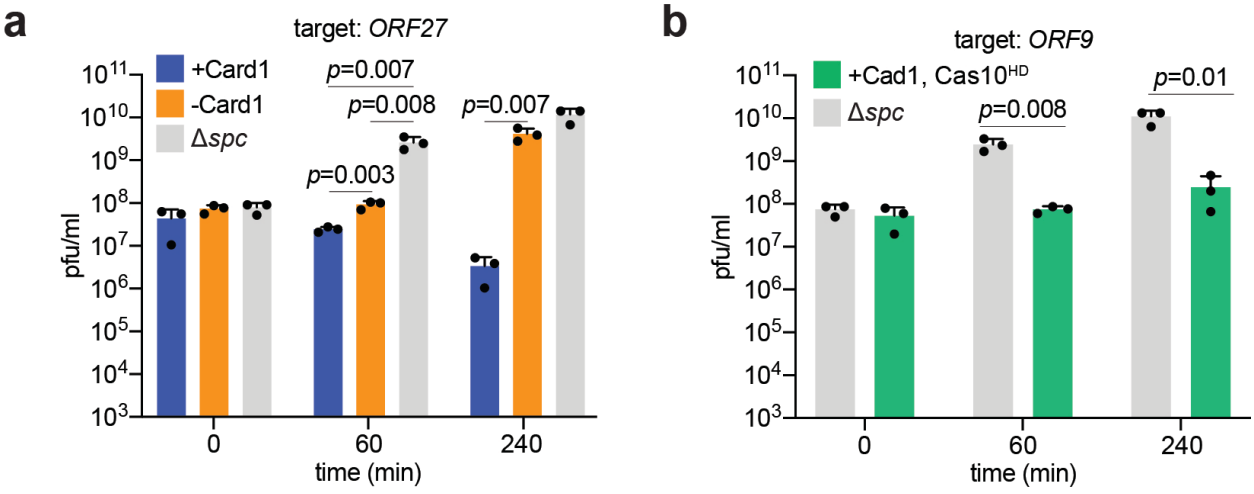

Figure 5.13. Card1 impedes phage propagation in staphylococci. ***(a)*** *Enumeration of plaque-forming units (pfu) within staphylococcal cultures carrying different pCRISPR variants after infection with Φ12γ3 at an MOI ~10. At the indicated times after infection aliquots were removed and plated on top agar media seeded with a susceptible strain. Mean of three biological replicates ± s.e.m are reported. Significant p values (p<0.05), obtained with two-sided t-test, are shown.* ***(b)*** *Enumeration of plaque-forming units (pfu) within staphylococcal cultures carrying different pCRISPR variants after infection with Φ12γ3 at an MOI ~10. At the indicated times after infection aliquots were removed and plated on top agar media seeded with a susceptible strain. Mean of three biological replicates ± s.e.m are reported. Significant p values (p<0.05), obtained with two-sided t-test, are shown.*

I also investigated whether Card1 activity was sufficient to provide immunity by infecting cells that express Cas10^{HD} in the presence or absence of Card1. Remarkably, both when the *ORF9* (Fig. 5.11b) and *ORF27* (Fig. 5.11c) transcripts were targeted by the Cas10 complex, activated Card1 alone was able to provide immunity to growing cells as well as reduce the phage titer in the cultures (Fig. 5.13b).

To further examine the immunity provided by Card1 alone, I performed infection assays at a higher MOI (~25), when the great majority of cells are infected. Both when targeting *ORF9* (Fig. 5.14a) and *ORF27* (Fig. 5.14b) Card1 failed to provide immunity, suggesting that in the absence of Cas10 nuclease activity and low MOIs, defense is achieved through the growth of the cells that are not infected and not arrested by Card1 activity.

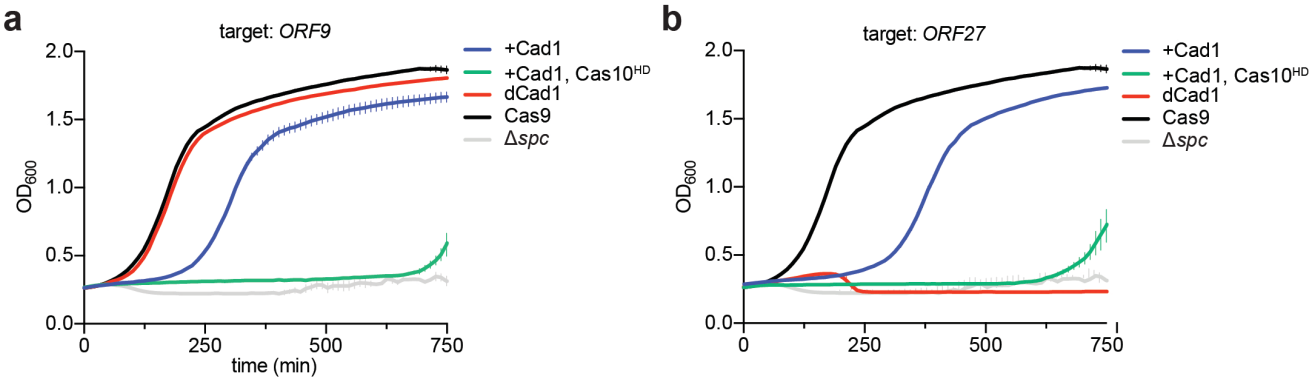

Figure 5.14. Anti-phage protection of Card1 at high MOIs. ***(a)*** *Growth of staphylococci carrying different pCRISPR variants programmed to target the ORF9 transcript of Φ12γ3, measured as OD_{600} at different times after infection at an MOI ~25. The immunity provided by the Cas9 nuclease, which directly recognizes and cleaves the phage genome shortly after its injection and therefore allows the survival of the infected cells, is used as a control to show that the observed growth delays are not due to an excessive amount of phage*

added in the experiment. Mean of three biological triplicates ± s.e.m. are reported. ***(b)*** *Same as in* ***(a)*** *but targeting the ORF27 transcript. In both* ***(a)*** *and* ***(b)****, Cas10^{HD} cells with +Card1 do not lyse from infection (as it is the case for Δspc cells), indicating an incomplete phage life cycle. Mean of three biological triplicates ± s.e.m. are reported. In* ***(a)*** *and* ***(b)****, the data representing Cas9 and Δspc is from the same experiment.*

Altogether, these results demonstrate that Card1 activity is sufficient to provide anti-phage defense in staphylococci at low MOIs and also required for an efficient type III-A CRISPR-Cas immune response when the target is expressed late after infection.

5.7 Discussion

To date, two prokaryotic defense systems have been described that use cyclic oligonucleotide second messengers to activate auxiliary proteins needed for immunity, type III CRISPR-Cas[53,54] and CBASS (cyclic-oligonucleotide-based anti-phage signaling system) [120-122,149,150]. While the mechanisms of invader recognition and cyclic oligonucleotide synthesis are very different, both pathways lead to the activation of non-specific nucleases that affect host viability[55,150]. For type III CRISPR, the most common effector proteins are the non-specific ssRNases Csm6 and Csx1[110,111]. In contrast, for CBASS, the cyclic oligonucleotides have so far mainly been shown to activate a majority of non-specific dsDNases such as NucC [149] and Cap4 [150]. Here I described that Card1, a ssDNase and ssRNase, provides immunity to the host in the context of type III-A CRISPR-Cas immunity. During plasmid targeting, Card1 activation did not affect the transcriptome of the host cell and resulted in cell toxicity. This produced a growth arrest followed by the death of a substantial fraction of the host population. In addition, both Cas10 and Card1 nuclease activities were required for efficient clearance of a target plasmid. During phage infection, Card1 was necessary for defense when the target transcript recognized by the crRNA in the Cas10 complex is expressed late in the viral lytic cycle, but it was also sufficient to allow survival of a host population lacking the ssDNase activity of Cas10, both when activated by cA_4 production early and late during infection.

Based on these data I propose that, at least in staphylococci, Card1 protective function is achieved by two separate but overlapping mechanisms. First, Card1 toxicity can provide an abortive infection mechanism of defense in which compromised cells stop growing and prevent the exponential replication of the phage. This activity is similar to the function of Csm6 during type III-A immunity against plasmid-borne, weakly transcribed targets[55,131] and viral threats recognized late in the infection cycle[56]. One difference I observed, which could be the result of the more toxic effects caused by degradation of the DNA, as opposed to the RNA, of the host, is that while Csm6 induction results in a stable population of dormant cells[55], Card1 activation leads to a gradual but continuous

decrease in cell viability (Fig. 5.9). This toxicity is believed to not only constrain viral propagation and allow the growth of the non-infected cells, but also to facilitate the clearance of the foreign DNA within infected, non-growing cells by Cas10. Interestingly, Card1 orthologs are present in type III-D systems[116], where Cas10 naturally lacks a functional HD domain and is predicted to be unable to destroy the invader's DNA. Therefore, our results suggest that these systems might protect the host population via a crRNA-guided abortive infection mechanism, similar to the defense provided by type VI systems, which also elicit a growth arrest upon infection, but through the direct recognition of the target transcript (i.e.; without the need of a second messenger) by the RNA-guided, non-specific RNase Cas13 [151] (although organisms that carry type III and VI CRISPR systems usually also possess *bona fide* abortive infection systems as well). Second, in contrast to Csm6, Card1 could directly destroy the phage genome. Many phages and plasmids copy their DNA through rolling-circle replication, which involves the formation of ssDNA intermediates[152], likely making them sensitive to Card1 digestion. Moreover, since Cas10 also cuts ssDNA, possibly at the transcription fork of the target[128], it could generate more ssDNA intermediates that are sensitive to Card1 cutting, a hypothesis that explains our result showing synergy between both nucleases to specifically destroy pTarget (Figs. 5.10a,b).

Finally, I can also compare Card1's *in vivo* effects to the function the CBASS effector dsDNases NucC and Cap4. Likely due to the introduction of more severe lesions to the host genome in the form of dsDNA cuts, these nucleases cause the irreversible death of the cell[149,150] to provide an abortive infection mechanism of defense where the only survivors are the non-infected bacteria. In contrast, our results show that Card1 is less immediately toxic, possibly allowing the time needed to clear the invader before cell death occurs and enabling the rescue of the infected host.

Card1, ssRNA-specific Csm6 [142] Csx1 [143], and supercoiled dsDNA-specific nickase Can1[144] all bind cA_4, while dsDNA-specific Cap4 binds cA_3 [150], in each case within a dimeric pocket formed by a pair of CARF domains. There is little sequence conservation amongst the CARF domains of Card1 (classified as CARF4 family[141]), Csm6 (CARF1), Can1 (CARF4) and Cap4 (SAVED3), and Card1 binds cA_4 in a different manner to these other proteins (collaborator's analysis, not shown here). In addition, Csm6 quickly converts cA_4 to ApA>p [142] to auto-regulate its RNase activity; in contrast, degradation of cA_4 was not observed in the Card1 structure (collaborator's data, not shown here), raising the possibility that its ssDNase activity could be controlled by *trans*-acting, CARF domain-containing ring nucleases[130]. Also, as opposed to the substantial conformational changes

detected across the REase domain of Card1 upon cA_4 binding (Fig. 5.6), a cA_4-dependent conformational change has yet to be identified for either Csm6 [142] or Csx1 [143], limiting our understanding of how the second messenger activates their RNase domains.

Card1 can also be compared with Can1, a type III-associated, CARF-containing nickase activated by cA_4 whose structure was recently solved, but its role during type III CRISPR-Cas immunity is unknown[144]. Can1 contains a pair of CARF domains interspaced between nuclease and nuclease-like domains and binds cA_4 as a monomer. Both CARF domains are involved in cA_4 recognition with the nuclease and nuclease-like domains brought together to form a postulated nucleic acid binding site. It is unclear whether these structural differences account for the distinct substrate specificity of each nuclease. Because Can1 nuclease assays were performed at 60°C [144], it is possible that at this high temperature the supercoiled DNA used as substrate could partially melt, unwind and expose short ssDNA regions for Can1 cutting, and therefore its activity could be similar to the ssDNase of Card1. Finally, Card1 has many structural differences with CBASS effectors. For example, NucC adopts a trimeric scaffold, with bound cA_3 ligands promoting the formation of a dimer of trimers with endonuclease activity[149]. Cap4 [150], which is composed of an endonuclease followed by a SAVED domain, recognizes diverse cA_3 with mixed 2',5' and 3',5 linkages through a pair of tandem CARF domains within the SAVED module. Cap4 proteins are activated through ligand-dependent oligomerization, with this higher order state proposed to mediate cleavage of target dsDNA.

Due to the impossibility to culture and genetically manipulate *Treponema succinifaciens* bacteria, I decided to study Card1 function as an accessory protein of the type III-A CRISPR-Cas system of *S. epidermidis*, in staphylococci. A caveat of this approach is the possibility that Card1 function could be different in its native host. In particular, I wonder whether the ssRNase activity, which I failed to detect in staphylococci, is more relevant in *T. succinifaciens*. Card1-mediated RNA degradation is robust *in vitro*, and there is the possibility that the *S. aureus* cellular environment is inhibitory of this activity. On the other hand, a recent neighborhood analysis[141] showed an enrichment of the *csx1* RNase (cd09741, cd09732, and cd09747) near *card1* (pfam09002). The absence of significant *in vivo* RNase degradation for Card1 would assign these two genes with complementary, rather than redundant, functions, and perhaps influence the observed co-localization. Finally, because CRISPR-Cas loci are able to transfer horizontally between different species to provide defense without the need of host factors (with only a few exceptions), I believe that our findings for Card1 would reflect its function in the native host. Supporting this idea, I previously found that the Palm domain of *S.*

epidermidis Cas10 produces cA_6 to heterologously activate *Enterococcus italicus* Csm6 in staphylococcal hosts (chapter II) [53]. Interestingly, the results showing that Card1 is activated by cA_4 but not cA_6 indicate that the *S. epidermidis* Cas10 complex is able to produce cA rings of different sizes to activate a wide range of CARF-containing proteins, offering the possibility for the functional genetic exchange of type III accessory proteins. My study highlights the variety of defense systems and mechanisms that prokaryotic organisms have evolved to counteract the diversity and rapid evolution of their genetic parasites.

Chapter VI – Preliminary studies on TM-1, a membrane-associated type III accessory protein

In this chapter, my research on TM-1, an effector gene identified in chapter V, is described. TM-1 showed robust toxicity upon cA production *in vivo*, and behaves in a similar manner to Csm6 and Card1 during immunity. With merely a single transmembrane helix and a cytosolic CARF domain, the mechanism of toxicity and immunity of this gene is less obvious than that of the predicted nuclease domains of Csm6 and Card1. One possibility is the oligomerisation of individual TM-1 monomers, resulting in a membrane pore and the loss of protonmotive force or osmotic pressure. The project is currently unfinished and not ready for publication, though future experiments will be outlined at the end of the chapter, which will be performed by another lab member. In addition, structural studies of TM-1 have been initiated by Dr. Wei Xie and Prof. Dinshaw Patel (the same collaborators as chapter V), though this effort has so far not yielded any results.

6.1 Introduction

During type III CRISPR-Cas immunity, the Cas10-Csm complex binds a complementary target transcript through base pairing. This initiates the single-stranded DNase activity of the HD domain of Cas10 which helps destroy the invader. In addition, the Palm domain of Cas10 starts synthesising cA of various sizes. cA can bind the CARF domain of Csm6/Csx1 and Card1 (chapter V), activating the non-specific catalytic activities of these accessory genes. The type III CRISPR immune response is therefore a multi-pronged attack, combining specific DNA degradation of the invader with non-specific, growth arrest-inducing activities mediated by ancillary genes.

Bioinformatic studies have revealed a variety of diverse accessory, CARF-containing genes in bacterial and archaeal genomes, often associated with type III CRISPR-Cas loci. Beyond the already characterised Csm6 (RNase), Csx1 (RNase), Can1 (DNase?), and Card1 (DNase/RNase), other genes have other predicted RNase domains, protease domains, transcription factor-like motifs, and transmembrane helices. These are likely to also be activated by cA produced by the Cas10 domain, to either cause growth arrest through diverse mechanisms, or to coordinate the immune response in other ways.

6.2 TM-1 from Nitrosococcus halophilus

In chapter IV, I performed a screen for cA4-mediated toxicity by various CARF-containing genes when activated in *S. aureus*. One hit, which produced robust toxicity in growth assays, was a gene from the type III-A CRISPR-Cas system from Nitrosococcus

halophilus Nc 4 (GenBank accession number CP001798), a halophilic Gram-negative bacterium isolated from saline ponds. This gene was listed in the bioinformatic study [111], in cluster 11023, in the family called "CARF+2TM". The gene encodes a relatively small protein of 206 amino acids, with a domain architecture shown in Fig. 6.1. Even if the general protein family from [111] states 2TM, and most genes of this family do have two predicted transmembrane helices, TM-1 only has one predicted transmembrane helix. Beyond this helix (approximately amino acids 20-40) and the CARF domain (approximately amino acids 66-187), there are no other domains, limiting the potential mechanisms of action of TM-1. One possibility is that TM-1 is a membrane-tethered toxin with catalytic activity in the cytosolic CARF domain. However, CARF domains themselves have not previously been shown to possess catalytic activity (except for slow cleavage of bound cA[142]). Another possibility is the formation of a membrane pore by oligomerisation of TM-1. In this scenario, the affinity of CARF domains to bind each other might be increased by the presence of cA.

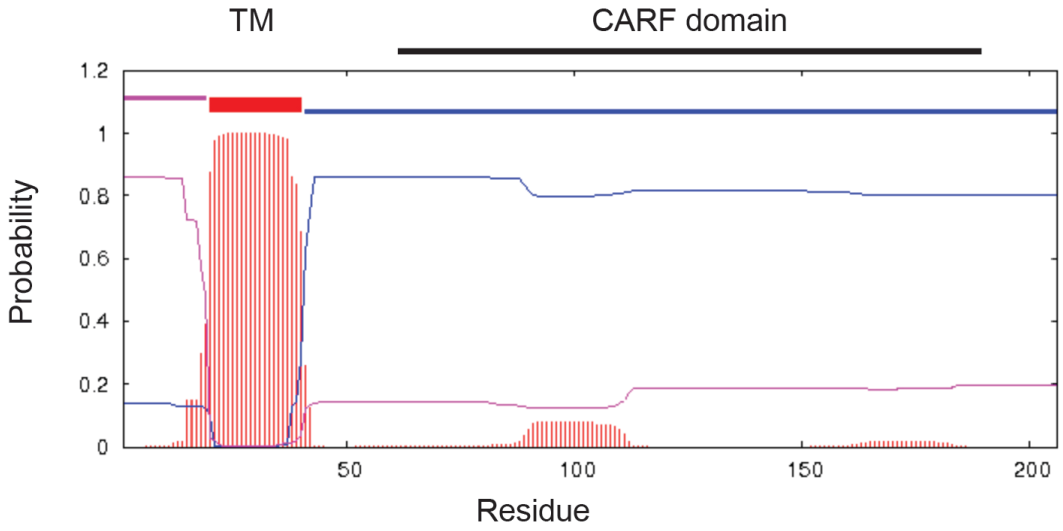

Figure 6.1. The architecture of TM-1. *The transmembrane (TM) domain is shown as a red horizontal bar, and the CARF domain is shown as a black horizontal line. The probability of there being a transmembrane helix is shown on the y-axis. The N-terminus (pink) is predicted to be periplasmic, while the blue portion is predicted to be cytosolic. The output is generated by TMHMM* [153].

6.3 Toxicity of TM-1 in staphylococci

In chapter IV, TM-1 was shown to be toxic in the context of cA production, with TM-1 on a separate plasmid. To confirm these findings, TM-1 was cloned into pCRISPR, being expressed after the last *S. epidermidis* type III-A cas gene, *cas6*. Cells harbouring the aTc-inducible pTarget (Fig. 3.2) were then transformed with different versions of pCRISPR (containing TM-1), and their growth was tracked in 96-well plate with a

microplate reader upon the addition of aTc (induction of targeting). All pCRISPRs harboured a mutation in the HD domain of Cas10 to prevent pTarget elimination, and lacked functional Csm6. As expected, TM-1 induced a growth defect upon the addition of aTc relative to a no targeting spacer (TM1, Δspc) (Fig. 6.2a). The cells that recover after 500 minutes in the TM-1 condition are targeting escapers, which were no longer able to induce toxicity (data not shown). This toxicity depended on cA production by a functional Cas10 Palm domain (TM-1, Cas10^{Palm}). I also tested whether both main domains of TM-1 were important to toxicity, so I generated a plasmid with a TM-1 lacking the transmembrane domain (deleting residues 2-40) and without the CARF domain (deleting residues 51-206) (referred to as TM-1$^{no\ TM}$ and TM-1$^{no\ CARF}$ respectively). Both these mutations failed to induce toxicity (Fig. 6.2b), showing that TM-1 is not a membrane-tethered cytosolic toxin (since TM-1$^{no\ TM}$ should still be active), and that the transmembrane portion alone is not sufficient.

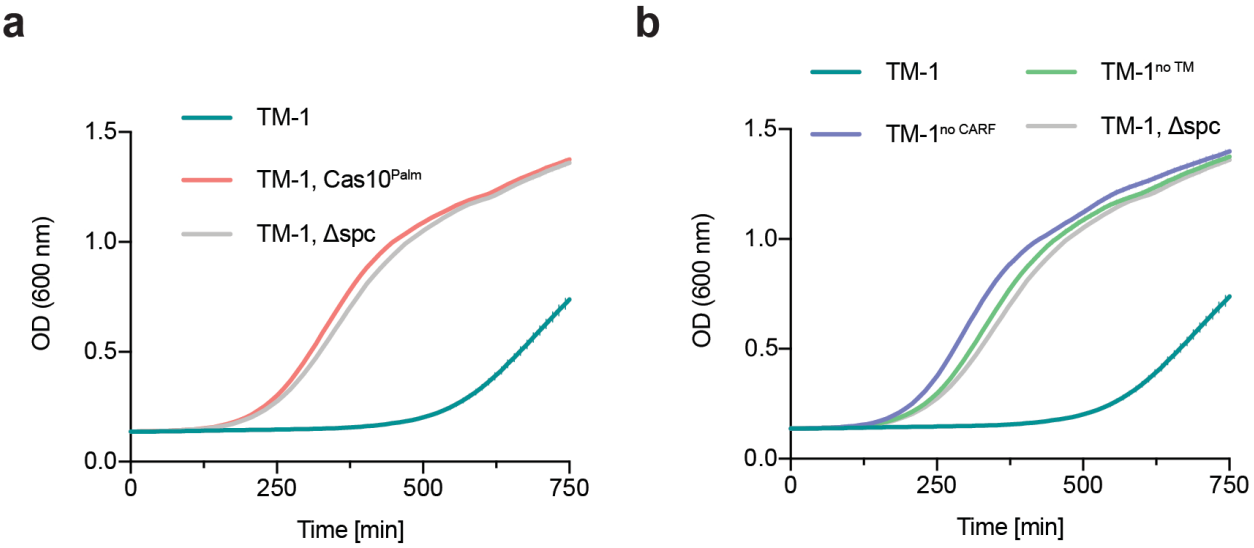

Figure 6.2. cA-mediated toxicity by TM-1 in staphylococci. (a-b) *Growth of cells habouring pTarget and the pCRISPR specified, upon pTarget transcription and cA production inducuction by aTc. The TM-1 and TM1, Δspc data are the same in both curves. The average of three biological replicates ±s.e.m. is shown.*

6.4 Anti-phage protection of TM-1

Next, I tested whether TM-1 could provide anti-phage protection in *S. aureus*. With TM-1 on pCRISPR, in a wild-type Cas10 HD context, against ORF27 (late target) of Φ12γ3 (Fig. 5.11a), protection was robust at MOIs of both 0.5 and 10 (Fig. 6.3a). (Note, however, that there is no minus TM-1 control in this experiment.) Moreover, in the absence of functional Cas10 DNase activity (Cas10^{HD}), protection against Φ12γ3 allowed survival of the culture at MOIs of 0.5 and 10 when targeting ORF9 (early target) (Fig. 5.11a), in an abortive infection-like mechanism (Fig. 6.3b).

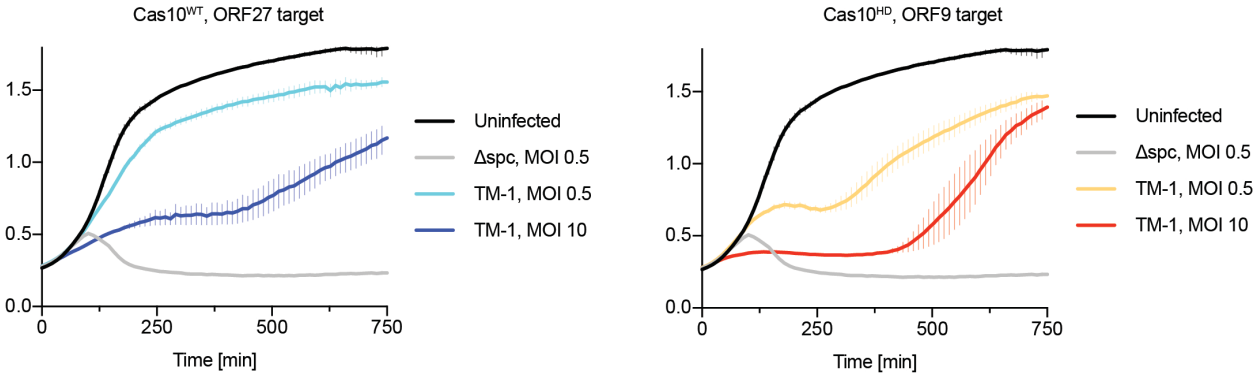

Figure 6.3. Anti-phage protection mediated by TM-1. *S. aureus cells carrying the pCRISPRs indicated are infected by Φ12γ3 at time 0, and their growth is measured over time. (a) cells carry a wild-type Cas10 DNase, while (b) harbours a Cas10 HD mutation. The data for Uninfected and Δspc, MOI 0.5 are the same in both graphs. Each data point represents the average of three biological replicates ±s.e.m.*

TM-1 therefore seems to offer protection in a manner similar to Csm6 and Card1, i.e. allowing extra protection against a late expressed phage target in the context of a functional Cas10 DNase, and mediating abortive infection protection in the absence of Cas10 DNase activity. However, the protection offered by TM-1 is a bit less robust, being weaker at higher MOIs. This could be due to TM-1 being in a heterologous host, made more incompatible by *S. aureus* being a Gram-positive bacterium (c.f. Gram-negative natural host), and not being halophilic, as different salt concentrations might affect a proposed membrane pore.

6.5 Proposed mechanism of TM-1 and remaining experiments

The mechanism of action of TM-1 remains unknown, and the experiments performed so far do not provide much insight. However, the cytosolic part of TM-1 alone is not sufficient for toxicity (Fig. 6.2b), suggesting that the cytosolic domain (the CARF domain) is not alone toxic, and that TM-1 is not just a membrane-tethered toxin. Instead, the transmembrane helix-CARF domain combination is essential for function. One possible mechanism is the formation of an oligomeric inner membrane pore that assembles in the presence of cA. Membrane pores that either collapse the protonmotive force or form larger membrane holes are known to be involved in growth arrest and cell death, for example with phage holins (phage pores that help phage escape infected cells at the end of the life cycle) [154] or as a pore-forming effectors used together with bacteria-bacteria type VI secretion systems [155]. Moreover, it has been suggested that some

CBASS effectors, also activated by nucleotide second messengers, are pore-forming proteins that induce growth arrest or cell death upon phage infection.

In this model, in the absence of cA, TM-1 (and other genes of this family) would exist as monomers in the inner membrane of cells. Upon cA binding, individual CARF domains might oligomerise, bringing individual transmembrane helices in close proximity which allows pore formation. Even in the absence of cA, TM-1 is mildly toxic, with *S. aureus* cells displaying a small colony phenotype when carrying TM-1. Moreover, overexpressing TM-1 in *E. coli*, where higher copy number plasmids and stronger promoters allow higher production, lead to long-lasting growth defect (data not shown). It is possible that at high concentrations, TM-1 monomers can oligomerise even in the absence of cA. Indeed, both Csm6 and Card1 exist as homodimers, so TM-1 might already be a dimer which can undergo conformational changes with cA binding, increasing the affinity for TM-1 homodimers for each other.

In an attempt to test this hypothesis, I stained cells with activated TM-1 with the dye $DiOC_2(3)$ (from the ThermoFisher BacLight Bacterial Membrane Potential Kit). *S. aureus* cells with pTarget and a pCRISPR carrying TM-1 were incubated with aTc for 20 minutes to provide enough time for TM-1 to be sufficiently activated, with possible depolarisation. The cells were then stained with $DiOC_2(3)$, and analysed using flow cytometry. $DiOC_2(3)$ stains all cells with green fluorescence equally, but higher membrane potentials give more red fluorescent staining. Thus, depolarised cells should have a lower red/green ratio. Some cells were stained with CCCP as a positive control for depolarisation. The data shown in Fig. 6.4 shows that there appears to be depolarisation with aTc for TM-1, but this difference is not statistically significant, and the experiment was only performed once. It therefore cannot yet be concluded that TM-1 induces dormancy by neutralising the protonmotive force.

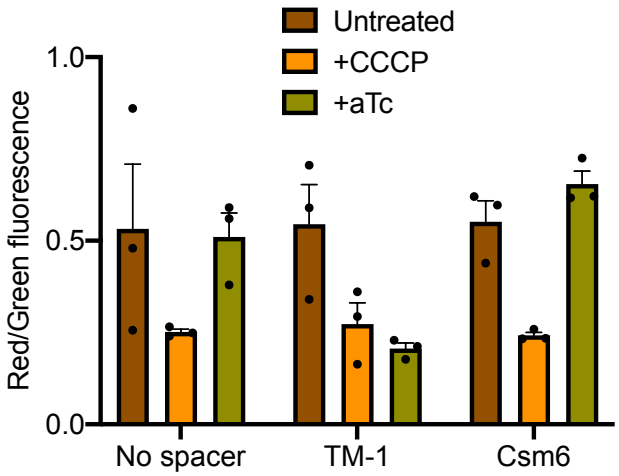

Figure 6.4. Attempting to observe a loss in membrane potential upon activation of TM-1. *The relative fluorescent signals from red or green per cell is measured from each sample. Untreated cells were not depolarised in any way, +CCCP cells were depolarised by the uncoupler agent CCCP, and aTc was added to induce cA synthesis and TM-1 activation. Csm6 was used as a control for cells that undergo growth arrest.*

Finally, I also attempted to observe membrane localisation of TM-1 in *S. aureus*. I cloned a TM-1-mCherry fluorescent protein and tried to observe enrichment of mCherry fluorescence in cell edges. However, this was difficult to clone into the pC194 backbone I normally use, and I failed to observe fluorescence in cells with a lower copy number plasmid. This could possibly be due high levels of TM-1 stabilised by mCherry, giving a higher protein concentration and spontaneous pore formation. Instead, I have tagged TM-1 with a C-terminal His tag (which still allows TM-1 function, data not shown). For future work, bacteria expressing TM-1 will be fractionated, and the cytosolic and membrane fractions will separately be stained with a His tag-specific antibody to detect the subcellular localisation of TM-1.

6.5 TM-1 perspectives

Of the many CARF-containing accessory type III CRISPR genes, the transmembrane helix-containing CARF genes are among the most interesting, possibly effecting a cA-mediated host cell growth defect without cleaving nucleic acid. My hypothesis of TM-1 forming oligomeric membrane pores that arrests the cell by destroying the membrane potential/protonmotive force remains to be proven. Yet, I think it is likely due to i) The small colony phenotype in *S. aureus*, and the complete growth arrest upon overexpression in *E. coli*, both consistent with spontaneous oligomerisation at high concentrations in the absence of cA, ii) The small size of TM-1, too small to contain additional catalytic domains that can help explain the growth defect in the presence of cA,

iii) The insufficient but promising depolarisation data from Fig. 6.4, and iv) pore-forming channels being present or being suggested in other host-parasite interactions. Future experiments, to be performed by other members in the lab, will shed light on the mechanism of TM-1. Additionally, our structural biology collaborators might successfully purify and solve the structure of TM-1, giving invaluable insight into molecular details that mediate the potential oligomerisation of TM-1. I also note that the anti-phage protection offered by TM-1 is less robust than that of Csm6 and Card1, possibly suggesting that TM-1 naturally works in conjunction with other proteins from its native locus, or that expression or function in a heterologous host is compromised.

Chapter VII – Perspectives and reflections

In the last decade, CRISPR has seen a meteoric rise to the front stage of science and biotechnology, and has also enjoyed widespread fame with the general public, lauded for its promise to revolutionise medicine. CRISPR-Cas deserves all this and more, Nobel prize included, but one should not forget its humble roots in basic, curiosity-driven microbiology research. Genetic conflict, the struggle between entities that compete for limited resources, lies at the heart of evolution, and if evolution indeed is a tinkerer and not an engineer [156], nowhere can be greener pastures for tinkering than in the interactions between bacteria and their viruses, bacteriophages. Astronomical numbers of hosts and parasites, quick life spans, and high rates of mutation and gene transfer all set the stage for the emergence of elegant and ingenious mechanisms of offence and defence. Bacteria-phage interactions have, in addition to CRISPR, given rise to other innovations like restriction enzymes, phage display, phage therapy, lambda red recombination systems, DNA/RNA ligation, to mention a few. With the advent of CRISPR, a phage research renaissance has occurred, with exciting recent discoveries and more surely to come.

In CRISPR-Cas immunity, two common themes hold true for all six types and multiple subtypes. The first is that Cas1-Cas2 participates in incorporating immunological memories into the CRISPR array; the property that makes CRISPR-Cas immunity adaptive. The second is that short CRISPR RNAs (crRNAs) guide a main protein targeting complex to the target upon exposure to a parasite; the property that makes CRISPR highly specific. However, recent research in the field, including my work, has shown that the specificity trademark of CRISPR is not absolute.

As is often the case with new discoveries, bioinformatic studies foreshadowed and predicted what was later shown experimentally. The ability of type III CRISPR systems to generate nucleotide-based ligands (as well as CBASS immunity mediated by nucleotide-based signalling molecules) was suggested by earlier reports [116,117]. However, it was difficult to square the circle of how a diffusible ligand could provide the specificity so characteristic of other CRISPR systems. Also, bioinformatics can be wrong, like when it was postulated that CRISPR loci encoded DNA repair systems [157]. In parallel, the mechanism of action of Csm6, the most common accessory RNase found in type III systems, was unknown. Work from our laboratory had established that it was required during immunity against the conjugative plasmid pG0400 [115], and against phage when the target was a late transcribed late gene [56]. Still, what Csm6 targeted during immunity, and how this achieved protection, was unclear, and many reports suggested a strict

abortive infection mechanism, where the cell dies upon exposure to the invader, was frequently proposed [54].

Yet, the nucleotide-based messenger hypothesis was confirmed in 2017 with two reports [53] [54], one of which I was involved in (chapter II), describing cyclic oligoadenylate production by the Cas10 Palm domain upon Cas10-Csm recognition of a target RNA. Suddenly, a downstream immune signalling pathway existed in type III CRISPR immunity, which was involved in Csm6 activation during immunity [53], and which likely could coordinate other CARF-containing (cA-binding) genes in other type III systems. Yet, how exactly this contributed to immunity remained unclear.

To date, the work described herein (chapter III) is the most thorough and mechanistic research done on Csm6, and indeed on any CARF gene. My starting point was with the conjugative plasmid pG0400, a natural target of the *S. epidermidis* type III-A CRISPR system. One advantage of a plasmid target is that the effect seen on growth of the host cell is largely from the CRISPR system directly. In contrast, when working with phage, it is difficult to distinguish direct effects of collateral CRISPR targeting and toxicity from the phage infection. I was firstly able to solve why Csm6 was required against pG0400, which was previously counter-intuitive, since pG0400 is a DNA target which should not require an RNase for clearance. The natural crRNA of the *S. epidermidis* system targeted the “wrong” strand of pG0400, i.e. the non-transcribed strand, which RNA-seq confirmed was very poorly transcribed. Therefore, the CRISPR system was infrequently activated and inefficient, and immunity was compromised. Indeed, when targeting the well-transcribed strand of pG0400, Csm6 was dispensable, with the DNase activity of the Cas10-Csm complex being sufficient for immunity. Moreover, using pTarget, an inducible target plasmid, revealed that Csm6 targets host and plasmid transcripts non-specifically, resulting in a temporary growth arrest.

From this emerges a theme of type III CRISPR immunity, namely the coupling of a specific targeting activity and a non-specific targeting activity. Once activated, Csm6 can induce a host cell growth defect, preventing the replication of both the host and the plasmid. This “buys” time for the CRISPR system to eliminate the plasmid, even if targeting is inefficient. Upon target clearance, Cas10 is inactivated with no further cA synthesis, and cA is degraded, turning off Csm6. Against “difficult” invaders, like with a weakly transcribed plasmid target, or a phage with a late transcribed target, the Cas10-Csm complex alone is insufficient for adequate immunity, and Csm6 arrests the cell for as long as it takes to clear the invader. However, in the absence of target clearance, like

in a Cas10 HD domain mutation, the arrest is permanent, showing that Csm6 can be a *bona fide* abortive infection module.

Bioinformatic studies have revealed a range of other CRISPR-associated genes, many containing CARF domains [110,111,116,141]. Knowing that cA is a diffusible ligand that bind and activate CARF domains, and that many CARF genes are fused to other domains of predicted catalytic functions, it is likely that these new CARF genes also work through inducing arrest in a non-specific manner. In particular, genes containing the PD-D/ExK motif (restriction endonuclease) represented a promising class of genes which might work like Csm6, but targeting host and invader DNA, not RNA. Through my screen in chapter IV I found multiple genes that could be toxic during cA production by the *S. epidermidis* Cas10-Csm complex, and I pursued one such gene, named Card1 (Cyclic oligoadenylate-activated RNase and DNase 1).

Biochemical characterisation revealed that Card1 is a ssDNase *in vitro*, activated by cA_4 but not by cA_6. However, to my surprise, Card1 could also degrade ssRNA, in fact more efficiently than ssDNA (Fig. 5.5). All previous bioinformatics suggested Card1 to be a DNase, and a recent report of a similar cA-activated protein with a similar catalytic motif [144] (albeit a dsDNase targeting supercoiled DNA). Some known nucleases, like the commercially available mung bean nuclease and S1 nuclease, as well as some toxins [158], can target both DNA and RNA. It can be imagined that targeting both DNA and RNA might more rapidly and robustly induce a growth arrest in the case of Card1.

In vivo, however, I was unable to detect any Card1 RNase activity through RNA-seq or Northern blotting, which previously showed strong RNase activity for Csm6. Yet, Card1 induces growth arrest in a Cas10 Palm-dependent manner, and can protect against phages in a similar way to Csm6. Detecting ssDNase activity *in vivo* was challenging and ultimately unsuccessful, but given the absence of RNase activity, which should be readily detected if present (like with Csm6), it is likely that the ssDNase activity is primarily responsible for the toxicity and immunity observed. In addition to being able to induce a general growth arrest, inhibiting the parasite as well as the host, a ssDNase might directly target ssDNA intermediates from phages and plasmids. Both of these parasites often employ rolling circle replication with significant ssDNA stages, and might be particularly vulnerable to the ssDNase activity of Card1.

The project investigating the final CARF gene candidate, TM-1, was not completed (chapter VI). Like Csm6 and Card1, *in vivo* activation of TM-1 by cA induces cell dormancy (Fig. 6.2) and offers anti-phage protection (Fig. 6.4), suggesting the theme of non-specific

immunity is true also here. Although it remains to be proven, I suspect it works by oligomerisation and pore formation, and if true, would be the first example of a CRISPR-Cas effector that works without targeting nucleic acid.

Fig. 1.2 shows an overview of immunity mechanisms employed by bacterial cells upon phage infection. CRISPR-Cas systems typically act directly against invading DNA (represented by the red inhibitory), recognising the viral DNA directly and promptly destroying it. This DNA recognition step is highly specific, which is what allows e.g. Cas9 to be used as a precise genome engineering tool. Before my work, it was known that even if the Cas10-Csm complex recognises transcribed RNA, it still destroys the DNA the viral RNA is transcribed from [128].

Also in Fig. 1.2, in the middle at the bottom, are multiple inhibitory arrows from "Toxin-antitoxin systems/Abortive infection". These represent cellular components and pathways that respond to viral infection by shutting down the cell. This can be a transient growth arrest or a programmed cell death [159], though the latter outcome is more likely in the absence of a component that can directly destroy the viral genome. Even if the host cell dies, phage propagation is hampered and surrounding cells can survive.

With the accessory cA-activated CARF genes found in CRISPR loci, I propose that type III CRISPR systems combine the specificity of the DNA-targeting CRISPR system with the non-specific, collateral, and dormancy-inducing catalytic activities of the ligand-activated CARF proteins. Whether this activity is an RNase, DNase, or potentially pore forming, the outcome remains host cell growth arrest, buying time for the DNA-targeting specific arm of the type III CRISPR immune system to destroy the parasite. It is possible that the dormancy sometimes results in cell death, for example with high MOIs or against phages that can protect themselves against the direct DNA targeting of CRISPR[160,161], and future research will address this. Interestingly, type VI CRISPR systems possess only non-specific RNase activity upon viral recognition, and is a *bona fide* abortive infection system[151]. However, it is likely that type VI CRISPR systems can combine with a DNA-targeting component, e.g. a restriction-modification system, to result in reversible dormancy and viral clearance.

Much work remains to obtain a fuller understanding of the role of cA-activated ancillary type III genes in diverse bacteria. Firstly, the mechanism of action of TM-1, as well as other predicted RNases (chapter IV), need to be elucidated. These might possess surprising functions. Likewise, other CARF genes more recently identified [141] include more complex transmembrane helix proteins and proteases, which might have novel roles

in immunity. There are also reports that CBASS-associated effectors might be activated by cA_3[149], which might be produced both by CBASS systems and type III CRISPR systems[161]. This raises the possibility that there is cross-talk between effectors of these two separate nucleotide-based bacterial immune systems.

More importantly, however, is the study of these CARF genes in their native loci, and if possible, in their native bacterial hosts. The universality of the cA ligand allows activation of proteins with heterologous Cas10-Csm complexes (like I did with Card1 and TM-1), but much interesting biology might get lost. Type III CRISPR loci are often more complex than the *S. epidermidis* locus (Fig. 2.1), with multiple genes of known function that are not part of the Cas10-Csm complex. These might have regulatory roles, dampening or amplifying the activity of a given CARF gene. Also, more than one CARF gene can occur in a type III locus, possibly resulting in multiple non-specific activities to provide a more robust dormant state. For example, the CARF-RNase *csx1* is sometimes enriched in *card1* neighbourhoods. This would result two effectors with combined ssDNase and ssRNase activities. The study of full systems from different organisms is difficult to make work (I tried a few, not presented here), and it is even more difficult to work in a non-model bacterium. Still, the fascinating biology left to discover will surely make it worth pursuing.

Chapter VII - Methods

7.1 Bacterial strains and growth conditions

S. aureus RN4220[135] was grown in tryptic soy broth (TSB) medium or Brain Heart Infusion (BHI) medium at 37°C, supplemented with 10 μg/ml of chloramphenicol or erythromycin for maintenance of plasmids pC194[148] or pE194[162] respectively. 5 μM $CaCl_2$ was supplemented in phage experiments.

7.2 Molecular cloning

The sequences of the oligonucleotides used in this study are provided in Table 1. The plasmids used in this study are shown in Table 2. The plasmid cloning strategies are showed in Table 3. All pCRISPR plasmids contain one spacer flanked by two repeats. For obtaining the coding sequence of Card1, the amino acid sequence of Tresu_2185 (NCBI Reference Sequence WP_013702306.1) from *Treponema succinifaciens* DSM 2489 (Genbank accession number CP002631) was codon optimised for expression in *S. aureus* and synthesised by Genewiz (NJ, USA).

7.3 Conjugation

Conjugation was performed using filter mating as previously described[115] into recipients containing the specified pCRISPRs (pGG25, pGG-BsaI-R, pJTR111, pJTR135, pJTR138, pJTR175, or pJTR177). Pictures of colonies for Fig. 3.8b were obtained by imaging the plates with the Axygen Scientific GD1000 gel documentation system, 24 hours after plating.

7.4 Plasmid curing assay

For chapter III plasmid curing assays, overnight RN4220 cultures carrying pTarget (pJTR162) and the specified pCRISPRs (pWJ191, pWJ241, pJTR125, pJTR147, pGG-BsaI-R) were diluted to an OD of exactly 0.15 in TSB containing 10 μg/ml of chloramphenicol. Where relevant, 200 μg/ml of neomycin was added. aTc was added to a final concentration of either 7.5 ng/ml (“low aTc condition”) or 50 ng/ml (“high aTc condition”), and cells were isolated and plasmid extracted at the specified timepoints. 300 ng of total plasmid was then linearised with the common single cutter BamHI-HF (NEB), and imaged by gel electrophoresis.

For chapter V, plasmid curing experiments were largely performed as above. Overnight cultures of *S. aureus* cells harbouring pTarget and a pCRISPR containing either Card1, dCard1, or no spacer (*Δspc*) were diluted to exactly OD 0.15 in tryptic soy

broth with 10 ug/ml chloramphenicol. After removing a cell aliquot for the 0 timepoint, aTc was added to a concentration of 9.3 ng/ml (Fig. 5.10a) or 125 ng/ml (Fig. 5.10b)

7.5 Growth curves

For chapter III, triplicate overnight RN4220 cultures carrying pTarget and the specified pCRISPRs (pWJ191, pWJ241, pGG-BsaI-R, pJTR109, pJTR121, pJTR125) were diluted 1:100 in fresh TSB with 10 μg/ml of chloramphenicol, and 10 μg/ml of erythromycin where specified, and outgrown for an hour. Cells were then diluted and normalised for OD, moved to a 96-well plate in triplicate, and aTc was added to a final concentration of either 2.5 ng/ml ("low aTc condition") or 12.5 ng/ml ("high aTc condition"). OD_{600} readings were then taken every 10 minutes by a microplate reader (TECAN Infinite 200 PRO). For measuring colony forming units from each well, plateau phase cells from the end of the experiment were resuspended, serial diluted, and spotted on TSB agar plates. Plates contained 10 μg/ml of chloramphenicol when selecting for pCRISPR only, or 10 μg/ml of chloramphenicol and 10 μg/ml of erythromycin when also selecting for pTarget. For the Cas10 dHD targeting escaper growth curves, cells that recovered from the end of the experiment in 5a ($Cas10^{HD}$ cells grown in high aTc conditions) were streaked out, and single colonies were picked for a new growth experiment, in the presence of high levels of aTc. To analyse these escapers, DNA was isolated, subjected by PCR using primers JTR390 and W1022, with the products being visualised by gel electrophoresis. PCR products (escapers 1-3) or isolated plasmid (escaper 4) were sent for Sanger sequencing to confirm the observed deletions.

For chapter V, with *in vivo* Card1 toxicity induction, triplicate RN4220 overnight cultures harbouring pTarget and a pCRISPR are diluted 1:100, outgrown for about an hour, and normalised for OD. Cells are then seeded in a 96 well plate. To induce targeting, 6.25-12.5 ng/ml of anhydrotetracycline (aTc) is added to the appropriate wells. Absorbance at 600 nm is then measured every 10 minutes by a microplate reader (TECAN Infinite 200 PRO). To analyse targeting escapers, cells from the end of the experiment (either cells from wells without aTc, i.e. naïve cells, or cells from wells that recovered later in the time course due to Card1 toxicity) are re-streaked on BHI agar plates, and individual colonies were launched in liquid culture, diluted the next day, and used for a new time course experiment. From these overnight cultures, plasmid DNA was isolated (QIAgen Spin Miniprep Kit), digested by BamHI-HF (single-cutter for both pTarget and pCRISPR) (New England Biolabs), and visualised by gel electrophoresis. The deletion of important features in pTarget (making it unable to be targeted by pCRISPR) or pCRISPR was confirmed by Sanger sequencing.

For *in vivo* anti-phage immunity, cells harbouring various pCRISPRs were launched in triplicate overnight, diluted 1:100, outgrown for about an hour, and normalised for OD. Cells were seeded into a 96 well plate. Phage Φ12γ3 [69] ΦNM1γ6 [51] or was added at the specified multiplicity of infection, and OD measurements were taken every 10 minutes.

7.6 Liquid anti-phage infection for counting pfu (chapter V)

To obtain CFU and PFU counts over time from cultures infected with phage, RN4220 cultures harbouring various pCRISPRs were launched overnight, diluted 1:100, and outgrown for about one hour. Cells were then infected with phage Φ12γ3 at an MOI of 10, and an aliquot was taken shortly after to obtain PFUs at time 0. The cultures were then incubated further, with aliquots taken at 1 and 4 hours.

7.7 RNA purification

For isolating RNA for pG0400/pG0420 RNA-seq, 5ml of *S. aureus* RN4220 cells at OD 0.6 containing the relevant plasmid were spun down. For isolating RNA for Csm6 targeting RNA-seq, 20 ml of *S. aureus* RN4220 cells at OD 0.15 were spun down at 0 min or 2 min after aTc addition. For both RNA-seq runs, cells were lysed in PBS with treatment with 1 mg/ml lysostaphin and 2 mg/ml lysozyme for five minutes, followed by addition of 1% sarcosyl. For the Csm6 targeting RNA-seq, 2.5 μg of *Listeria seeligeri* RNA was added at this stage. RNA was then purified using Quick-RNA Miniprep Plus Kit (Zymo research). For Northern blot analysis, 80 ml of OD 0.15 *S. aureus* RN4220 cells were spun down and lysed as above. The RNA was then isolated by resuspending the lysed cells in Trizol (Thermo Fisher Scientific), and following the Trizol manufacturer's protocol.

7.8 RNA-seq of pG0400/pG0420

RNA was isolated from cells harbouring pG0400/pG0420 as described above, DNase treated (Invitrogen TURBO DNA-free kit according to the manufacturer's protocol), and rRNase depleted (Illumina Ribo-Zero rRNA Removal Kit (Bacteria) according to the manufacturer's protocol). Library preparation was done using TruSeq Stranded mRNA kit (Illumina), and the sequencing was performed by Illumina MiSeq. Reads were aligned using STAR[163] version 2.5 to either pG0400[164] (Genebank reference KT780705) or *S. aureus* NCTC 8325 (Genebank reference CP000253), without normalisation.

7.9 RNA-seq of pTarget using a spike RNA (chapter III)

In duplicate, overnight cultures of *S. aureus* RN4220 carrying pTarget and pCRISPRs containing Cas10^{HD} and either Csm6 or dCsm6 (pJTR109 or pJTR125) were

diluted to an OD_{600} of exactly 0.15 in TSB. Cells were then harvested for the 0 minute timepoint. Then, aTc was added to a final concentration of 7.5 ng/ml (similar to "low" concentration for plasmid curing), and cells were harvested after 2 minutes, quenching the reaction with cold TSB. RNA was then isolated as described above. At the cell lysis stage (after adding the TRI reagent of the Quick-RNA Miniprep Plus Kit (Zymo research)), an equal amount of purified *Listeria seeligeri* RNA was added to each sample, to be carried through the purification to allow absolute comparisons between the samples. Library preparation was then carried out like for pG0400/pG0420 RNA-seq, and the samples were submitted to NextSeq (Illumina) sequencing at the Rockefeller University Genomics core (New York, USA). For the analysis, inspired by[165], the normalisation protocol relied on the number of spike reads mapping to the *Listeria seeligeri* genome (NC_013891.1). In the first round of read mapping, the reads for each sample were mapped to *S. aureus* using STAR aligner, with standard parameters except for allowing maximum one mismatch. The unmapped reads were then aligned to the *L. seeligeri* genome. Since each sample initially had the same number of *L. seeligeri* spike reads at the lysis stage, a scaling factor was calculated to make number of *L. seeligeri* reads identical between the samples. This scaling factor is later used to normalise the reads mapping to the *S. aureus* chromosome or pTarget between samples. Then, for the second round of mapping, all reads are first aligned to *L. seeligeri* (to eliminate spike reads), and the remaining reads were mapped to either *S. aureus* or pTarget. The total assigned reads per gene was determined using featureCounts[166] with largestOverlap set to TRUE. The number of assigned reads to each gene was then normalised by multiplying with the previously calculated scaling factor, thus allowing absolute comparison between the number of assigned reads to a gene between different samples.

7.10 RNA-seq of +Card1 and dCard1 cells (chapter V)

To investigate the RNase activity of Card1 *in vivo*, RNA-seq using spike-in normalisation was performed similarly to described before[55]. In duplicate, cells containing pTarget and pCRISPR with $Cas10^{HD}$ and either +Card1 or dCard1 were grown overnight and normalised to an OD of exactly 0.15. Cells were harvested (quenching growth with the addition of ice-cold media) for the 0 minute timepoint, and aTc was added to 125 ng/ml to initiate pTarget transcription and Card1 activation. At 3 minutes, the cell growth was quenched. The cells were lysed in PBS with 1 mg/ml of lysostaphin and 2 mg/ml of lysozyme at 37°C for 5 minutes before adding TRI reagent. At this step, 3 μg of Listeria seeligeri was added to each sample to serve as an internal normalisation control. RNA was then purified with the Quick-RNA Miniprep Plus kit according to the manufacturer's protocol (Zymo Research). The RNA was DNase treated (Invitrogen TURBO DNA-free kit) and rRNA depleted (Ribo-Zero Plus rRNA Depletion Kit (Illumina)), before undergoing

library preparation with a TruSeq Stranded mRNA kit (Illumina). The samples were sequenced using an Illumina NextSeq at the Rockefeller University Genomics Core. Mapping was performed by STAR v2.7.5 allowing one mismatch. To determine the scaling factor to apply to the reads (to normalise the read depth for spike reads), the reads were first mapped to the genome of *S. aureus* NCTC 8325 (GenBank reference CP000253) to remove non-spike reads, and the remaining reads were mapped to the *L. seeligeri* genome (NC_013891.1). This process was then reversed, first mapping all the original reads to *L. seeligeri* (to remove spike reads), and the remaining reads were mapped to *S. aureus.* Reads were assigned to each gene with featureCounts[166] with largestOVERLAP set to TRUE. The number of reads per gene for each sample was then scaled according to the previously calculated scaling factor, allowing direct comparison of read depth per gene between all samples.

7.11 Northern blot

Overnight RN4220 cultures carrying pTarget and the specified pCRISPRs (pWJ191 or pWJ241) were diluted to an OD of exactly 0.15 in TSB containing 10 *μ*g/ml of chloramphenicol. 80 ml of cells were harvested for the 0 minute timepoint. Then aTc was added to a final concentration of 7.5 ng/ml, and 80 ml of cells were again harvested after 2 minutes. RNA was then isolated described as above. The RNA was separated on a 6% PAGE gel by electrophoresis, and blotted onto nylon filters (Invitrogen BrightStar Plus) using a semi-dry blotting apparatus. The oligonucleotide probes were radiolabelled with γ-^{32}P-labelled ATP using PNK (NEB), and incubated with the nylon membranes overnight at 42 °C in the presence of 0.1 mg/ml salmon sperm DNA. The membranes were then visualised by phosphorimaging (Typhoon FLA 7000, GE Life Sciences). For assaying pTarget induction at “low” or “high” aTc concentrations, cells containing either pJTR162 or pE194 were isolated before or 2 minutes after adding 7.5 ng/ml or 50 ng/ml, respectively. The rest of the protocol was done as described above.

The Northern blot in chapter V was performed as above, but in BHI, with three minutes incubation time, and with 125 ng/ml of aTc

7.12 Protein expression and purification

The corresponding sequence of full-length Card1 (1-372) was cloned to plasmid pJTR330 with a C-terminal hexahistidine (His6)-tag. The protein was overexpressed in *E. coli* strain BL21-CodonPlus(DE3)-RIL (Stratagene). Bacteria were grown at 37 °C to OD600 of 0.8 and induced by 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at 18 °C overnight. Bacteria cells were lysed by sonication in buffer A (20 mM Tris-HCl, 500 mM NaCl, pH 8.0) supplemented with 20 mM imidazole and 1 mM phenylmethylsulfonyl

fluoride (PMSF). Cell lysates were centrifuged, and the supernatant was loaded onto a 5 mL HisTrap FF column (GE Healthcare) with extensive washing by buffer A supplemented with 50 mM imidazole. The target protein was eluted with buffer A supplemented with 300 mM imidazole. The eluate was further purified on 5 mL HiTrap Heparin column (GE Healthcare) by a linear gradient from 100 mM to 1 M NaCl, and then on Superdex 200 16/60 column pre-equilibrated in buffer B (20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1mM DTT). The high purity eluting fractions were detected by SDS-PAGE and collected. The protein was flash-frozen in liquid nitrogen and stored at –80 °C.

7.13 Crystallization and structure determination (performed by collaborators)

Crystallization conditions were determined with crystal screens (Qiagen) by sitting-drop vapor diffusion. As for apo Card1, the protein was prepared at 15 mg/mL in buffer B. As for cA_4-Card1 or cA_6-Card1 complex, cA_4 or cA_6 at the final concentration of 1 mM was added to the 15 mg/mL Card1 in buffer B and incubated on ice for 1 hr before crystallization. Apo Card1 crystals were grown from drops with 1.5 µL protein solution and 1.5 µL reservoir solution (0.2 M K_2HPO_4, and 20% PEG3350 (w/v)). cA_4-Card1 crystals were grown from drops with 1.5 µL protein solution and 1.5 µL reservoir solution (0.1 M citric acid, 10% PEG6000 (w/v), final pH 5.0). cA_6-Card1 crystals were grown from drops with 1.5 µL protein solution and 1.5 µL reservoir solution (0.2 M NaCl, 0.1 M Na/K phosphate, pH 6.2, and 20% PEG1000 (w/v)). cA_4-Card1(D294N) crystals were grown from drops with 1.5 µL protein solution and 1.5 µL reservoir solution (0.1 M HEPES, pH 7, and 30% Jeffamine M-600 (v/v)). Crystals were cryoprotected by mother liquor containing 25% glycerol and flash-frozen in liquid nitrogen.

All the diffraction data sets were collected on the 24-ID beamline at the Advanced Photon Source (APS) at the Argonne National Laboratory, and auto-processed by XDS package [167] in the NE-CAT RAPD online server. The structure of apo- Card1 was solved by single wavelength anomalous diffraction method with the AutoSol and AutoBuild programs and followed by molecular replacement with PHASER program [168] in the PHENIX package [169]. The structure of cA_4-Card1 complex was solved by molecular replacement using one monomer of apo Card1 structure as the search model. The structure of cA_6-Card1 complex was solved by molecular replacement using the dimeric apo-Card1 structure as the search model. The structure of cA_4-Card1(D294N) complex was solved by molecular replacement using the dimeric cA_4-Card1 structure as the search model. Iterative manual model building was performed using the program COOT[170], and refinement with phenix.refine[171] to produce the final models. Figures were generated using PyMOL (http://www.pymol.org).

7.14 In vitro DNA/RNA cleavage assays for Card1

All reactions were performed at 37°C in a reaction buffer containing 20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1mM DTT, and 5 mM $MnCl_2$, unless otherwise stated. For the reactions in Figs. 5.2, the reactions were done as above, but with 250 nM Card1 and 2.5 uM of cA4, with 2 µg M13 ssDNA (NEB), 500 ng of non-linearized and linearized ΦX174 dsDNA (NEB), or 2 µg ΦX174 ssDNA (NEB), for 15 minutes, before being quenched by 25 mM EDTA. For testing the metal dependence of Card1, the reaction was performed as above, but with 5 mM of $MgCl_2$, $CaCl_2$, or $ZnSO_4$, in the presence of 1 mM EDTA.

For the RNA oligo cleavage assay, 250 nM of a Cy3-labelled RNA oligo was added to the reaction. The reaction products were run on Mini-PROTEAN TBE-Urea precast gel with 15% acrylamide (Bio-Rad), and visualised on an Amersham Typhoon 5 Biomolecular Imager. For degradation of ssRNA or dsRNA ladders, 1 ug of either ssRNA ladder (NEB) or dsRNA ladder (NEB) was digested. After 15 minutes, the reactions were stopped by the addition of 25 mM of EDTA. The reaction products were visualised by agarose gel electrophoresis.

For determining the nucleotide cleavage preference of the RNase activity of Card1, the reaction was performed as above, with 1 μM of each RNA oligo (IDT). The RNA oligos had a 5' end fluorophore (FAM) and a 3' end quencher (Iowa black), generating a fluorescent signal upon cleavage of the linker RNA. Fluorescent measurements were taken in a microplate reader (TECAN Infinite 200 PRO), using values from when the reaction was complete. 0.5 ul of RNase I (Thermo Fisher Scientific), which cuts next to all four RNA nucleotides, was used as a positive control. Guanine polynucleotides could not be tested due to their propensity to form degradation-resistant secondary structures.

For comparing relative catalytic rates of the DNase and RNase activities of Card1, Card1 was incubated with the reaction buffer described before, in the presence of one pair of DNA/RNA fluorescently labelled oligonucleotide of the same sequence (IDT). The increasing Card1 concentrations are 50 nM, 250 nM, 1.25 μM, and 6.25 μM respectively, and the cA4 concentrations are 250 nM, 1.25 μM, 6.25 μM, and 31.25 μM, respectively. The 30 nucleotide oligonucleotides (with the sequence of the top ΦX174 cut site as determined by NGS) had the DNA oligonucleotide labelled with Cy3 and the RNA oligonucleotide labelled with Cy5, and for the 50 nucleotide oligonucleotides (with the sequence of the top M13 cut site as determined by NGS) the DNA species was labelled with Cy5, and the RNA species labelled with Cy3. This allowed the RNA and DNA

oligonucleotides to be visualised on the same gel using different fluorescent filters. Each fluorophore was at 250 nM, and both reactions were done for 15 minutes before being quenched by 25 mM EDTA. The products were separated on a 15% acrylamide TBE-urea gels, and visualised on an Amersham Typhoon 5 Biomolecular Imager.

7.15 Card1 toxicity assay

To measure the effect of Card1 activity on *S. aureus* viability over time, colonies of *S. aureus* harbouring pTarget and the specified pCRISPR were launched in liquid culture overnight in triplicate. The next day, cells were diluted 1:100 and grown out for about an hour, and normalised for OD. One aliquot was taken from each culture, and then aTc was added to induce CRISPR targeting and Card1 activity to a concentration of 3 ng/ml. At each timepoint, cell aliquots were removed, centrifuged, resuspended in media lacking aTc, and serial dilutions were plated on solid BHI agar plates with or without aTc. All viable cells should grow on the solid agar plates, but only targeting escapers (cells that recover due to mutations in pTarget or pCRISPR) should form CFUs on plates with aTc.

7.16 Next-generation sequencing of ssDNA degradation products

To assess the ssDNA cleavage patterns of Card1, 2 ug of ΦX174 virion DNA (NEB) or M13 ssDNA (M13mp18) (NEB) was first digested by 250 nM Card1 with 2.5 uM of cA4. At the specified time points, the reaction was quenched by adding 25 mM of EDTA. Half the reaction was visualised by agarose gel electrophoresis. The remaining digestion products from the 2 hour timepoint were purified by phenol chloroform extraction.

Without further fragmentation, the purified digested DNA was subjected to the Accel-NGS 1S Plus DNA Library Kit (Swift Biosciences), proceeding according to the manufacturer's protocol, using a 1.5 x ratio of magnetic beads (AMPure XP beads by Beckman Coulter) to also include small DNA fragments. One of the library preparation steps involves the addition of on average 8 nucleotides to the 3' end of the DNA. The 5' end of the input DNA molecules remains unchanged. Paired-end sequencing was performed on an Illumina MiSeq. The 5' end of each read R1 represents the start of a DNA molecule, and thus a Card1 cut site. Using a custom python script, the location of 7,020,067 ΦX174 reads (mapping to Genbank reference NC_001422) and 7,670,616 M13 reads (mapping to Genbank reference X02513) was determined. To account for reads mapping at the circular junctions, 65 nucleotides of the first 5' end of the maps were copied and added at the 3' end of the maps. The DNA sequence 20 nucleotides upstream

and downstream of the cut sites was extracted using a custom Python script, and the Card1 cleavage motifs for ΦX174 and the M13 were determined separately using Weblogo3[172], with basal nucleotide compositions determined by the base compositions in each map (ΦX174 with A:24.0, C:21.4, G:23.3, T:31.3, and M13 with A:24.4, C:21.1, G:21.1, T:33.4). For the fragment size analysis, 8 nucleotides were removed from all the reads from the 3' end pair mate by the "Trim Ends" option in the Geneious Bioinformatics Software platform[173]. Using the STAR aligner (version 2.7.3) [163], 7,505,136 reads were successfully mapped to ΦX174, and 8,179,356 reads were successfully mapped to M13, using default arguments with the addition of --alignIntronMax 1 --alignMatesGapMax 6000 --peOverlapNbasesMin 5 --alignEndsProtrude 10 ConcordantPair.

7.17 Membrane depolarisation flow cytometry experiment for TM-1

This experiment was performed using the BacLight Bacterial Membrane Potential Kit according to the manufacturer's protocol (Thermo Fisher Scientific). Cells containing pTarget and pCRISPR with TM-1 or Csm6, and one with no targeting spacer, were diluted to an OD_{600} 0.035 in 1 ml BHI (equals 1E7 cells), each in biological triplicate. To +aTc conditions, 125 ng/ml of aTc was added, and all wells were incubated for 12 minutes at 37 ºC. Cells were then spun down and resuspended in PBS. To +CCCP samples, CCCP was added up to 5 μM, incubated for 5 minutes. DiOC2(3) was then added to 30 μM, incubated for 30 minutes at room temperature in the dark, and put on ice. The cells were then analysed with flow cytometry, with channels for Alexa 488 and Texas Red.

Table 2. Oligonucleotides used in this work.

Name	Sequence	Purpose
AV552	tcgagtcagaaaaatatacctgtatct	Cloning
AV553	cctagaccatgggtatggacagatc	Cloning
GG424	CATATTGCCTGATGAAGTGAATAG	Cloning
GG425	CTATTCACTTCATCAGGCAATATG	Cloning
JTR405	gaacctttgtactgatgatttatatacttcggcatacgt	Cloning
JTR406	gatcacgtatgccgaagtatataaatcatcagtacaaag	Cloning
JTR476	GAACAGTGTCTAACAACTGCAATTCACTAAATGCTGTAA	Cloning
JTR477	GATCTTACAGCATTTAGTGAATTGCAGTTGTTAGACACT	Cloning
JTR592	TTTTGTGTGTTGCGGCTCCTATTCTCCCGACTTTGGTACC	Protospacer target transcript Northern probe
JTR595	GCGGGAACCAATCATCAAATTTAAACTTCATTGCATAATC	*repF* Northern probe
JTR600	CTTTAGTTAATGGTAATTCTAACTCAGCTGCTTTTTGACG	*def* Northern probe
JTR606	GTGTTCGGCATGGGAACAGGTGTGACCTCC	5S rRNA Northern probe
JTR632	ATGATAAATAAAATTACAGTAGAGTTAGACTTGC	Cloning
JTR633	TATAGCACCTCATTATTTAACTCTTGAAAAC	Cloning
JTR638	CAAGAGTTAAATAATGAGGTGCTATAATGAAAGAGACTATTTTGGTTAACTTGG	Cloning
JTR639	CTAACTCTACTGTAATTTTATTTATCATATGTATATCTCCTTCTCATAATTGTGTACCGTCTTTAATGTC	Cloning
JTR678	GTTACACGTGATAGCATGTGCTAGTTTCGTCGATGGAAACG	Cloning
JTR679	GACGAAACTAGCACATGCTATCACGTGTAACTTATTGTCTTTGTC	Cloning
JTR859	GATCCGGCTGCTAACAAAGC	Cloning
JTR860	ATGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTCTAG	Cloning
JTR861	GTTTAACTTTAAGAAGGAGATATACATATGAAAGAGACTATTTTGGTTAACTTGG	Cloning

JTR862	CGGGCTTTGTTAGCAGCCGGATCTTAGTGATGGTGATGGTGATGTCCTG	Cloning	
JTR951	CAACAAAAGTATTTAAACTTTCTTTTAAGACTC	Cloning	
JTR952	GAAAGTTTAAATACTTTTGTTGCCAATGTCATTCACCTACTTAATTTTAAGATTTG	Cloning	
JTR972	AACAATTCCTAATGTACAATTTATCAAGTGG	Cloning	
JTR973	GATAAATTGTACATTAGGAATTGTTGCCTCACTAACCAAGTTAACCAAAATAGTC	Cloning	
JTR974	GAGCAAACAATTCCTAATGTACAATTTATC	Cloning	
JTR975	GTACATTAGGAATTGTTTGCTCTGCAACCAAGTTAACCAAAATAGTCTCTTTC	Cloning	
JTR976	ATGGAACAGAAAGAAAAATCATTGTTC	Cloning	
JTR977	CAATGATTTTTCTTTCTGTTCCATTGCCTTAGTACTCACTAACAATATCTTCATTGG	Cloning	
JTR978	GAACAGAAAGAAAAATCATTGTTCATCAAG	Cloning	
JTR979	GAACAATGATTTTTCTTTCTGTTCTGCCTCCTTAGTACTCACTAACAATATCTTCATTGG	Cloning	
JTR980	AGGTAAAGAATTGCAGGAGTTATACC	Cloning	
JTR981	GTATAACTCCTGCAATTCTTTACCTGCAGGTTGGTAGAATATCTCTGTGTTAGG	Cloning	
JTR982	GTAAAGCAACACTTGTATACGAAAAG	Cloning	
JTR983	CTTTTCGTATACAAGTGTTGCTTTACTGCTAATCCGAACTTACTCTTGATGATCG	Cloning	
JTR984	ATGTAAGAGTTTCGTCGATGGAAACG	Cloning	
JTR985	CCATCGACGAAACTCTTACATGCTATCACGTGTAACTTATTGTCTTTGTCC	Cloning	
JTR986	AATGTAAGAGTTTCGTCGATGG	Cloning	
JTR987	CCATCGACGAAACTCTTACATTGTATCACGTGTAACTTATTGTCTTTGTCC	Cloning	
JTR993	CAACCTATAGGTAAAGAATTGCAGG	Cloning	
JTR994	CAATTCTTTACCTATAGGTTGTGCGAATATCTCTGTGTTAGGTTTGTTATTAAAGAAATC	Cloning	
JTR995	GTTATCTACTTGGACAAAGACAATAAGTTAC	Cloning	
JTR996	GTCTTTGTCCAAGTAGATAACTGCCAACTCGTTTTTGTCATTTCCC	Cloning	
JTR1052	TAATACGACTCACTATAGGGTATAGTTTAACAACGTTTGCAGCTTGTGGAC	*msaB* probe	Northern

W614	GGTTATACTAAAAGTCGTTTGTTGG	Cloning
PM797	AAACCATCCTCAAGACTTATTAAGTCAATTAGTTG	Cloning
PM798	AAAACAACTAATTGACTTAATAAGTCTTGAGGATG	Cloning
poly-rU	/56-FAM/rUrUrUrUrUrUrUrUrUrUrUrUrU/3IABkFQ/	Card1 ssRNase specificity fluorescence assay
poly-rA	/56-FAM/rArArArArArArArArArArArArArA/3IABkFQ/	Card1 ssRNase specificity fluorescence assay
poly-rC	/56-FAM/rCrCrCrCrCrCrCrCrCrCrCrCrC/3IABkFQ/	Card1 ssRNase specificity fluorescence assay
W852	CCAACAAACGACTTTTAGTATAACC	Cloning
W1169	CCGATTAAAAATAAAGCTGCACCGCCTGAATATATAGCAGTAATTTG	Cloning
W1170	TATTCAGGCGGTGCAGCTTTATTTTTAATCGGTGCATGGGATG	Cloning
soJTR1	/5Cy3/TGATATTAATAACACTATAGACCACCGCCC	Cy3 30-nt DNA for RNase reaction. Used with soJTR2
soJTR2	/5Cy5/rUrGrArUrArUrUrArArUrArArCrArCrUrArUrArGrArCrCrArCrCrGrCrCrC	Cy5 30-nt RNA for RNase reaction. Used with soJTR2
soJTR3	/5Cy5/GCGTTGGTAAGATTCAGGATAAAATTGTAGCTGGGTGCAAAATAGCAACT	Cy5 50-nt DNA for RNase reaction. Used with soJTR4
soJTR4	/5Cy3/rGrCrGrUrUrGrGrUrArArGrArUrUrCrArGrGrArUrArArArArUrUrGrUrArGrCrUrGrGrGrUrGrCrArArArArUrArGrCrArArCrU	Cy3 50-nt RNA for RNase reaction. Used with soJTR3

60-nt RNA Cy3	/5Cy3/rGrGrCrArCrArCrCrCrGrCrArGrGrGrArGrGrArGrCrCrArArArGrCrArCrGrUrCrCrArUrCrArUrUrCrCrGrUrUrGrCrCrArCrArGrCrArGrArArGrCrCrC	Cy-3 labelled 60nt RNA for use in Card1 RNase assays

Table 3. List of plasmids used in this work

Plasmid name	Plasmid contents
pE194	pTarget backbone
pGG-BsaI-R	Type III with no spacer
pGG25	pCRISPR with nes spacer
pG0420	pG0400 with promoter
pJTR109	pCRISPR with gp43 spacer, dHD
pJTR111	pCRISPR with nes spacer, dCsm6
pJTR121	pCRISPR with gp43 spacer, dHD, dPalm
pJTR125	pCRISPR with gp43 spacer, dHD, dCsm6
pJTR135	pCRISPR with flipped nes spacer
pJTR138	pCRISPR with flipped nes spacer, dCsm6
pJTR170	Type III, Φ12γ3 ORF27 spacer, Csm6
pJTR172	Type III, Φ12γ3 ORF27 spacer, dCsm6
pJTR175	pCRISPR with nes spacer, dHD
pJTR177	pCRISPR with nes spacer, dHD, dCsm6
pJTR193	Genome editing plasmid, with spacer targeting homology arm of pG0400, and homology arms with a promoter inside
pJTR330	Cad1-His_6 on pET23 overexpression
pJTR378	dCad1-His_6 on pET23 overexpression
pJTR393	Type III, no targeting spacer, Cad1
pJTR394	Type III, anti-pTarget spacer, $Cas10^{HD}$, $Cas10^{Palm}$, Cad1
pJTR395	Type III, anti-pTarget spacer, Cad1
pJTR396	Type III, anti-pTarget spacer, $Cas10^{HD}$, Cad1
pJTR400	Type III, Φ12γ3 ORF9 spacer, Cad1
pJTR401	Type III, Φ12γ3 ORF9 spacer, $Cas10^{HD}$, Cad1
pJTR402	Type III, Φ12γ3 ORF27 spacer, Cad1
pJTR403	Type III, Φ12γ3 ORF27 spacer, $Cas10^{HD}$, Cad1
pJTR405	Type III, anti-pTarget spacer, Cad1
pJTR406	Type III, anti-pTarget spacer, $Cas10^{HD}$, dCad1
pJTR424	Type III, Φ12γ3 ORF27 spacer, $Cad1^{Q13A}$
pJTR426	Type III, Φ12γ3 ORF27 spacer, $Cad1^{E41A}$
pJTR427	Type III, Φ12γ3 ORF27 spacer, $Cad1^{M42A}$
pJTR428	Type III, Φ12γ3 ORF27 spacer, $Cad1^{I125A}$

pJTR429	Type III, Φ12γ3 ORF27 spacer, Cad1^{Y340A}
pJTR431	Type III, Φ12γ3 ORF27 spacer, Cad1^{E308Q}
pJTR434	Type III, Φ12γ3 ORF27 spacer, Cad1^{S11A}
PJTR435	Type III, Φ12γ3 ORF27 spacer, Cad1^{E308A}
pJTR436	Type III, Φ12γ3 ORF27 spacer, Cad1^{Y122}
pJTR437	Type III, Φ12γ3 ORF27 spacer, Cad1^{D294A}
pJTR439	Type III, Φ12γ3 ORF9 spacer, dCad1
pJTR441	Type III, Φ12γ3 ORF27 spacer, dCad1
pJTR443	Type III, ΦNM1γ6 gp14 spacer, Cad1
pJTR444	Type III, anti-pTarget spacer, dCad1
pJTR446	Type III, anti-pTarget spacer, Cas10^{Palm}, dCad1
pPM169	Cas9 targeting Φ12γ3
pTarget	aTc-inducible promoter in front of a protospacer (also known as pJTR162)
pWJ191	pCRISPR with gp43 spacer
pWJ241	pCRISPR with gp43 spacer, dCsm6
pWJ246	Type III, ΦNM1γ6 gp14 spacer, dCsm6 (-Cad1)

Table 4. Cloning strategies used for plasmids used in this work

Name	Cloning strategy
pG0420	Homology-directed allelic exchange with Cas9 elimination was performed as in [174]. Briefly, a plasmid (pJTR193) with the promoter to be inserted between 1 kb homology arms was transformed into pG0400-containing S. aureus RN4220 cells. This plasmid also contained Cas9 with a spacer targeting the unedited pG0400 template. Subsequently, pJTR193 was cured by growing the cells at the non-permissive temperature (37C)
pJTR109	PCR amplification of pWJ191 with W852/GG425, and pJTR119 with W614/GG424, following a Gibson assembly of the two products
pJTR111	PCR amplification of pGG25 with W852/GG425, and pWJ241 with W614/GG424, following a Gibson assembly of the two products
pJTR121	PCR amplification of pWJ191 with W852/GG425, and pJTR120 with W614/GG424, following a Gibson assembly of the two products
pJTR125	PCR amplification of pWJ191 with W852/GG425, and pJTR124 with W614/GG424, following a Gibson assembly of the two products
pJTR135	BsaI-HF (NEB) digestion of plasmid pGG-BsaI-R, followed by ligation with annealed oligos JTR405 and JTR406 with compatible overhangs
pJTR138	PCR amplification of pJTR135 with W852/GG425, and pWJ241 with W614/GG424, following a Gibson assembly of the two products
pJTR170	Cleavage of plasmid pGG-BsaI-R with BsaI-HF, then ligation of the linearized plasmid with the annealed oligo pair JTR476 and JTR477
pJTR172	PCR amplification of pJTR170 with W852 and GG425, and of pWJ241 with GG424 and W614, followed by Gibson assembly of the two PCR products
pJTR175	PCR amplification of pGG25 with W852/GG425, and pJTR109 with W614/GG424, following a Gibson assembly of the two products
pJTR177	PCR amplification of pGG25 with W852/GG425, and pJTR25 with W614/GG424, following a Gison assembly of the two products
pJTR193	PCR amplification of pJTR173 with AV552/W614 and AV553/W852, and pJTR190 with JTR530/JTR531, following a Gibson assembly of the three products
pJTR330	PCR amplification of pPS3 with JTR859 and JTR860, and with pJTR325 by JTR861 and JTR862, followed by Gibson assembly of the two PCR products

pJTR378	PCR amplification of pJTR330 by JTR861 and JTR679, and with JTR860 and JTR678
pJTR393	PCR amplification of pGG-BsaI-R with W852 and GG425, and of pRF7 with GG424 and W614, followed by Gibson assembly of the two PCR products
pJTR394	PCR amplification of pJTR121 with W852 and GG425, and of pRF7 with GG424 and W614, followed by Gibson assembly of the two PCR products
pJTR395	PCR amplification of pWJ191 by JTR633 and JTR632, and of pJTR224 by JTR638 and JTR639, followed by Gibson assembly of the two PCR products
pJTR396	PCR amplification of pJTR125 with JTR632 and JTR633, and of pJTR224 with JTR638 and JTR639, followed by Gibson assembly of the two PCR products
pJTR400	PCR amplification of pJTR169 with W852 and GG424, and of pJTR395 with W614 and GG424, followed by Gibson assembly of the two PCR products
pJTR401	PCR amplification of pJTR169 with W852 and GG424, and of pJTR396 with W614 and GG424, followed by Gibson assembly of the two PCR products
pJTR402	PCR amplification of pJTR170 with W852 and GG425, and of pJTR395 with GG424 and W614, followed by Gibson assembly of the two PCR products
pJTR403	PCR amplification of pJTR170 with W852 and GG425, and of pJTR396 with GG424 and W614, followed by Gibson assembly of the two PCR products
pJTR405	PCR amplification of pJTR395 with JTR678 and W614, and with JTR679 and W852, followed by Gibson assembly of the two PCR products
pJTR406	PCR amplification of pJTR396 with JTR678 and W614, and with JTR679 and W852, followed by Gibson assembly of the two PCR products
pJTR424	PCR amplification of pJTR402 with JTR972 and W614, and with JTR973 and W852, followed by Gibson assembly of the two PCR products
pJTR426	PCR amplification of pJTR402 with JTR976 and W614, and with JTR977 and W852, followed by Gibson assembly of the two PCR products

pJTR427	PCR amplification of pJTR402 with JTR978 and W614, and with JTR979 and W852, followed by Gibson assembly of the two PCR products
pJTR428	PCR amplification of pJTR402 with JTR980 and W614, and with JTR981 and W852, followed by Gibson assembly of the two PCR products
pJTR429	PCR amplification of pJTR402 with JTR982 and W614, and with JTR983 and W852, followed by Gibson assembly of the two PCR products
pJTR431	PCR amplification of pJTR402 with JTR986 and W614, and with JTR987 and W852, followed by Gibson assembly of the two PCR products
pJTR434	PCR amplification of pJTR402 with JTR974 and W614, and with JTR975 and W852, followed by Gibson assembly of the two PCR products
pJTR435	PCR amplification of pJTR402 with JTR984 and W614, and with JTR985 and W852, followed by Gibson assembly of the two PCR products
pJTR436	PCR amplification of pJTR402 with JTR993 and W614, and with JTR994 and W852, followed by Gibson assembly of the two PCR products
pJTR437	PCR amplification of pJTR402 with JTR995 and W614, and with JTR996 and W852, followed by Gibson assembly of the two PCR products
pJTR439	PCR amplification of pJTR400 with W852 and JTR679, and with W614 and JTR678, followed by Gibson assembly of the two PCR products
pJTR441	PCR amplification of pJTR402 with W852 and JTR679, and with W614 and JTR678, followed by Gibson assembly of the two PCR products
pJTR443	PCR amplification of pJTR288 with W852 and JTR952, and with W614 and JTR951, followed by Gibson assembly of the two PCR products
pJTR444	PCR amplification of pJTR109 with W852 and GG425, and of pJTR441 with W614 and GG424, followed by Gibson assembly of the two PCR products
pJTR446	PCR amplification of pJTR395 with W852 and W1169, and with W614 and W1170, followed by Gibson assembly of the two PCR products

Bibliography

1. Parikka, K. J., Le Romancer, M., Wauters, N. & Jacquet, S. Deciphering the virus-to-prokaryote ratio (VPR): insights into virus-host relationships in a variety of ecosystems. *Biol Rev Camb Philos Soc* **92,** 1081–1100 (2017).
2. Paez-Espino, D., Eloe-Fadrosh, E. A., Pavlopoulos, G. A., Thomas, A. D., Huntemann, M., Mikhailova, N., Rubin, E., Ivanova, N. N. & Kyrpides, N. C. Uncovering Earth's virome. *Nature* **536,** 425–430 (2016).
3. Mirzaei, M. K. & Maurice, C. F. Ménage à trois in the human gut: interactions between host, bacteria and phages. *Nature Publishing Group* **15,** 397–408 (2017).
4. Howard-Varona, C., Hargreaves, K. R., Abedon, S. T. & Sullivan, M. B. Lysogeny in nature: mechanisms, impact and ecology of temperate phages. *The ISME Journal* **11,** 1511–1520 (2017).
5. McLaughlin, R. N. & Malik, H. S. Genetic conflicts: the usual suspects and beyond. *J Exp Biol* **220,** 6–17 (2017).
6. Fernández, L., Rodríguez, A. & García, P. Phage or foe: an insight into the impact of viral predation on microbial communities. *The ISME Journal* **12,** 1171–1179 (2018).
7. Makarova, K. S., Wolf, Y. I., Alkhnbashi, O. S., Costa, F., Shah, S. A., Saunders, S. J., Barrangou, R., Brouns, S. J. J., Charpentier, E., Haft, D. H., Horvath, P., Moineau, S., Mojica, F. J. M., Terns, R. M., Terns, M. P., White, M. F., Yakunin, A. F., Garrett, R. A., van der Oost, J., Backofen, R. & Koonin, E. V. An updated evolutionary classification of CRISPR–Cas systems. *Nat Rev Micro* **13,** 722–736 (2015).
8. Burstein, D., Sun, C. L., Brown, C. T., Sharon, I., Anantharaman, K., Probst, A. J., Thomas, B. C. & Banfield, J. F. Major bacterial lineages are essentially devoid of CRISPR-Cas viral defence systems. *Nature Communications* **7,** 1–8 (2016).
9. Stern, A. & Sorek, R. The phage-host arms race: shaping the evolution of microbes. *BioEssays* **33,** 43–51 (2011).
10. van Houte, S., Buckling, A. & Westra, E. R. Evolutionary Ecology of Prokaryotic Immune Mechanisms. *Microbiol. Mol. Biol. Rev.* **80,** 745–763 (2016).
11. Simmons, M., Drescher, K., Nadell, C. D. & Bucci, V. Phage mobility is a core determinant of phage-bacteria coexistence in biofilms. *The ISME Journal* **12,** 531–543 (2018).
12. Eriksen, R. S., Svenningsen, S. L., Sneppen, K. & Mitarai, N. A growing microcolony can survive and support persistent propagation of virulent phages. *Proc. Natl. Acad. Sci. U.S.A.* **115,** 337–342 (2018).
13. Vidakovic, L., Singh, P. K., Hartmann, R., Nadell, C. D. & Drescher, K. Dynamic biofilm architecture confers individual and collective mechanisms of viral protection. *Nature Microbiology* **3,** 26–31 (2018).

14. Schwechheimer, C. & Kuehn, M. J. Outer-membrane vesicles from Gram-negative bacteria: biogenesis and functions. *Nat Rev Micro* **13,** 605–619 (2015).
15. Manning, A. J. & Kuehn, M. J. Contribution of bacterial outer membrane vesicles to innate bacterial defense. *BMC Microbiol.* **11,** 258 (2011).
16. Reyes-Robles, T., Dillard, R. S., Cairns, L. S., Silva-Valenzuela, C. A., Housman, M., Ali, A., Wright, E. R. & Camilli, A. Vibrio cholerae outer membrane vesicles inhibit bacteriophage infection. *J. Bacteriol.* **200,** 1539 (2018).
17. Harvey, H., Bondy-Denomy, J., Marquis, H., Sztanko, K. M., Davidson, A. R. & Burrows, L. L. Pseudomonas aeruginosa defends against phages through type IV pilus glycosylation. *Nature Microbiology* **3,** 47–52 (2018).
18. Scholl, D., Adhya, S. & Merril, C. Escherichia coli K1's capsule is a barrier to bacteriophage T7. *Applied and Environmental Microbiology* **71,** 4872–4874 (2005).
19. Pedruzzi, I., Rosenbusch, J. P. & Locher, K. P. Inactivation in vitro of the Escherichia coli outer membrane protein FhuA by a phage T5-encoded lipoprotein. *FEMS Microbiol. Lett.* **168,** 119–125 (1998).
20. Labrie, S. J., Samson, J. E. & Moineau, S. Bacteriophage resistance mechanisms. *Nat Rev Micro* **8,** 317–327 (2010).
21. Clément, J. M., Lepouce, E., Marchal, C. & Hofnung, M. Genetic study of a membrane protein: DNA sequence alterations due to 17 lamB point mutations affecting adsorption of phage lambda. *The EMBO Journal* **2,** 77–80 (1983).
22. Werts, C., Michel, V., Hofnung, M. & Charbit, A. Adsorption of bacteriophage lambda on the LamB protein of Escherichia coli K-12: point mutations in gene J of lambda responsible for extended host range. *J. Bacteriol.* **176,** 941–947 (1994).
23. Henderson, I. R., Owen, P. & Nataro, J. P. Molecular switches - the ON and OFF of bacterial phase variation. *Mol Microbiol* **33,** 919–932 (1999).
24. Liu, M., Deora, R., Doulatov, S. R., Gingery, M., Eiserling, F. A., Preston, A., Maskell, D. J., Simons, R. W., Cotter, P. A., Parkhill, J. & Miller, J. F. Reverse transcriptase-mediated tropism switching in Bordetella bacteriophage. *Science* **295,** 2091–2094 (2002).
25. Chung, I.-Y., Jang, H.-J., Bae, H.-W. & Cho, Y.-H. A phage protein that inhibits the bacterial ATPase required for type IV pilus assembly. *Proc. Natl. Acad. Sci. U.S.A.* **111,** 11503–11508 (2014).
26. Bondy-Denomy, J., Qian, J., Westra, E. R., Buckling, A., Guttman, D. S., Davidson, A. R. & Maxwell, K. L. Prophages mediate defense against phage infection through diverse mechanisms. *The ISME Journal* **10,** 2854–2866 (2016).
27. Cumby, N., Edwards, A. M., Davidson, A. R. & Maxwell, K. L. The bacteriophage HK97 gp15 moron element encodes a novel superinfection exclusion protein. *J. Bacteriol.* **194,** 5012–5019 (2012).
28. Cumby, N., Reimer, K., Mengin-Lecreulx, D., Davidson, A. R. & Maxwell, K. L. The phage tail tape measure protein, an inner membrane protein and a periplasmic chaperone play connected roles in the genome injection process of E. coliphage HK97. *Mol Microbiol* **96,** 437–447 (2015).

29. Ko, C.-C. & Hatfull, G. F. Mycobacteriophage Fruitloop gp52 inactivates Wag31 (DivIVA) to prevent heterotypic superinfection. *Mol Microbiol* **108,** 443–460 (2018).
30. Johnson, A. D., Poteete, A. R., Lauer, G., Sauer, R. T., Ackers, G. K. & Ptashne, M. lambda Repressor and cro--components of an efficient molecular switch. *Nature* **294,** 217–223 (1981).
31. Tock, M. R. & Dryden, D. T. F. The biology of restriction and anti-restriction. *Current Opinion in Microbiology* **8,** 466–472 (2005).
32. Raleigh, E. A. & Wilson, G. Escherichia coli K-12 restricts DNA containing 5-methylcytosine. *Proceedings of the National Academy of Sciences* **83,** 9070–9074 (1986).
33. Bair, C. L. & Black, L. W. A type IV modification dependent restriction nuclease that targets glucosylated hydroxymethyl cytosine modified DNAs. *Journal of Molecular Biology* **366,** 768–778 (2007).
34. Chinenova, T. A., Mkrtumian, N. M. & Lomovskaia, N. D. [Genetic characteristics of a new phage resistance trait in Streptomyces coelicolor A3(2)]. *Genetika* **18,** 1945–1952 (1982).
35. Sumby, P. & Smith, M. C. M. Genetics of the phage growth limitation (Pgl) system of Streptomyces coelicolor A3(2). *Mol Microbiol* **44,** 489–500 (2002).
36. Goldfarb, T., Sberro, H., Weinstock, E., Cohen, O., Doron, S., Charpak-Amikam, Y., Afik, S., Ofir, G. & Sorek, R. BREX is a novel phage resistance system widespread in microbial genomes. *The EMBO Journal* **34,** 169–183 (2015).
37. Nussenzweig, P. M. & Marraffini, L. A. Molecular Mechanisms of CRISPR-Cas Immunity in Bacteria. *Annu. Rev. Genet.* **54,** annurev–genet–022120–112523 (2020).
38. Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A. & Horvath, P. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315,** 1709–1712 (2007).
39. Marraffini, L. A. & Sontheimer, E. J. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322,** 1843–1845 (2008).
40. Koonin, E. V., Makarova, K. S. & Zhang, F. Diversity, classification and evolution of CRISPR-Cas systems. *Current Opinion in Microbiology* **37,** 67–78 (2017).
41. Deveau, H., Barrangou, R., Garneau, J. E., Labonte, J., Fremaux, C., Boyaval, P., Romero, D. A., Horvath, P. & Moineau, S. Phage Response to CRISPR-Encoded Resistance in Streptococcus thermophilus. *J. Bacteriol.* **190,** 1390–1400 (2008).
42. Semenova, E., Jore, M. M., Datsenko, K. A., Semenova, A., Westra, E. R., Wanner, B., van der Oost, J., Brouns, S. J. J. & Severinov, K. Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl. Acad. Sci. U.S.A.* **108,** 10098–10103 (2011).
43. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A. & Charpentier, E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337,** 816–821 (2012).

44. Hayes, R. P., Xiao, Y., Ding, F., van Erp, P. B. G., Rajashankar, K., Bailey, S., Wiedenheft, B. & Ke, A. Crystal structure of E. coli Cascade bound to a PAM-containing dsDNA target at 2.45 angstrom resolution. (2016). doi:10.2210/pdb5h9f/pdb
45. Sinkunas, T., Gasiunas, G., Waghmare, S. P., Dickman, M. J., Barrangou, R., Horvath, P. & Siksnys, V. In vitro reconstitution of Cascade-mediated CRISPR immunity in Streptococcus thermophilus. *The EMBO Journal* **32,** 385–394 (2013).
46. Hochstrasser, M. L., Taylor, D. W., Bhat, P., Guegler, C. K., Sternberg, S. H., Nogales, E. & Doudna, J. A. CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. *Proc. Natl. Acad. Sci. U.S.A.* **111,** 6618–6623 (2014).
47. Garneau, J. E., Dupuis, M.-È., Villion, M., Romero, D. A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadán, A. H. & Moineau, S. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468,** 67–71 (2010).
48. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. U.S.A.* **109,** E2579–86 (2012).
49. Swarts, D. C., van der Oost, J. & Jinek, M. Structural Basis for Guide RNA Processing and Seed-Dependent DNA Targeting by CRISPR-Cas12a. *Molecular Cell* **66,** 221–233.e4 (2017).
50. Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., Volz, S. E., Joung, J., van der Oost, J., Regev, A., Koonin, E. V. & Zhang, F. Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell* 1–14 (2015). doi:10.1016/j.cell.2015.09.038
51. Goldberg, G. W., Jiang, W., Bikard, D. & Marraffini, L. A. Conditional tolerance of temperate phages via transcription-dependent CRISPR-Cas targeting. *Nature* **514,** 633–637 (2014).
52. Kazlauskiene, M., Tamulaitis, G., Kostiuk, G., Venclovas, Č. & Siksnys, V. Spatiotemporal Control of Type III-A CRISPR-Cas Immunity: Coupling DNA Degradation with the Target RNA Recognition. *Molecular Cell* **62,** 295–306 (2016).
53. Niewoehner, O., Garcia-Doval, C., Rostøl, J. T., Berk, C., Schwede, F., Bigler, L., Hall, J., Marraffini, L. A. & Jinek, M. Type III CRISPR–Cas systems produce cyclic oligoadenylate second messengers. *Nature* **548,** 543–548 (2017).
54. Kazlauskiene, M., Kostiuk, G., Venclovas, Č., Tamulaitis, G. & Siksnys, V. A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems. *Science* **357,** 605–609 (2017).
55. Rostøl, J. T. & Marraffini, L. A. Non-specific degradation of transcripts promotes plasmid clearance during type III-A CRISPR-Cas immunity. *Nature Microbiology* **4,** 656–662 (2019).
56. Jiang, W., Samai, P. & Marraffini, L. A. Degradation of Phage Transcripts by CRISPR- Associated RNases Enables Type III CRISPR-Cas Immunity. *Cell* **164,** 710–721 (2016).

57. Hale, C. R., Zhao, P., Olson, S., Duff, M. O., Graveley, B. R., Wells, L., Terns, R. M. & Terns, M. P. RNA-Guided RNA Cleavage by a CRISPR RNA-Cas Protein Complex. *Cell* **139,** 945–956 (2009).
58. Tamulaitis, G., Kazlauskiene, M., Manakova, E., Venclovas, Č., Nwokeoji, A. O., Dickman, M. J., Horvath, P. & Siksnys, V. Programmable RNA Shredding by the Type III-A CRISPR-Cas System of Streptococcus thermophilus. *Molecular Cell* **56,** 506–517 (2014).
59. Pyenson, N. C., Gayvert, K., Varble, A., Elemento, O. & Marraffini, L. A. Broad Targeting Specificity during Bacterial Type III CRISPR-Cas Immunity Constrains Viral Escape. *Cell Host and Microbe* **22,** 343–353.e3 (2017).
60. Silas, S., Lucas-Elio, P., Jackson, S. A., Aroca-Crevillén, A., Hansen, L. L., Fineran, P. C., Fire, A. Z. & Sánchez-Amat, A. Type III CRISPR-Cas systems can provide redundancy to counteract viral escape from type I systems. *Elife* **6,** aaf5573 (2017).
61. Abudayyeh, O. O., Gootenberg, J. S., Konermann, S., Joung, J., Slaymaker, I. M., Cox, D. B. T., Shmakov, S., Makarova, K. S., Semenova, E., Minakhin, L., Severinov, K., Regev, A., Lander, E. S., Koonin, E. V. & Zhang, F. C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* **353,** aaf5573–17 (2016).
62. Meeske, A. J. & Marraffini, L. A. RNA Guide Complementarity Prevents Self-Targeting in Type VI CRISPR Systems. *Molecular Cell* **71,** 791–801.e3 (2018).
63. Liu, L., Li, X., Ma, J., Li, Z., You, L., Wang, J., Wang, M., Zhang, X. & Wang, Y. The Molecular Architecture for RNA-Guided RNA Cleavage by Cas13a. *Cell* **170,** 714–726.e10 (2017).
64. Koonin, E. V., Dolja, V. V. & Krupovic, M. Origins and evolution of viruses of eukaryotes: The ultimate modularity. *Virology* **479-480,** 2–25 (2015).
65. McGinn, J. & Marraffini, L. A. Molecular mechanisms of CRISPR–Cas spacer acquisition. *Nat Rev Micro* **3,** 711 (2018).
66. Andersson, A. F. & Banfield, J. F. Virus population dynamics and acquired virus resistance in natural microbial communities. *Science* **320,** 1047–1050 (2008).
67. Ekroth, A. K. E., Broniewski, J. M., Chabas, H., Ben Ashby, Bondy-Denomy, J., Gandon, S., Boots, M., Paterson, S., van Houte, S., Buckling, A. & Westra, E. R. The diversity-generating benefits of a prokaryotic adaptive immune system. *Nature* 1–11 (2016). doi:10.1038/nature17436
68. Levy, A., Goren, M. G., Yosef, I., Auster, O., Manor, M., Amitai, G., Edgar, R., Qimron, U. & Sorek, R. CRISPR adaptation biases explain preference for acquisition of foreign DNA. *Nature* **520,** 505–510 (2015).
69. Modell, J. W., Jiang, W. & Marraffini, L. A. CRISPR–Cas systems exploit viral DNA injection to establish and maintain adaptive immunity. *Nature* 1–18 (2017). doi:10.1038/nature21719
70. Heler, R., Samai, P., Modell, J. W., Weiner, C., Goldberg, G. W., Bikard, D. & Marraffini, L. A. Cas9 specifies functional viral targets during CRISPR-Cas adaptation. *Nature* 1–16 (2015). doi:10.1038/nature14245

71. Lee, H., Zhou, Y., Taylor, D. W. & Sashital, D. G. Cas4-Dependent Prespacer Processing Ensures High-Fidelity Programming of CRISPR Arrays. *Molecular Cell* **70,** 48–59.e5 (2018).
72. Kieper, S. N., Almendros, C., Behler, J., McKenzie, R. E., Nobrega, F. L., Haagsma, A. C., Vink, J. N. A., Hess, W. R. & Brouns, S. J. J. Cas4 Facilitates PAM-Compatible Spacer Selection during CRISPR Adaptation. *Cell Rep* **22,** 3377–3384 (2018).
73. Shiimori, M., Garrett, S. C., Graveley, B. R. & Terns, M. P. Cas4 Nucleases Define the PAM, Length, and Orientation of DNA Fragments Integrated at CRISPR Loci. *Molecular Cell* **70,** 814–824.e6 (2018).
74. Nuñez, J. K., Lee, A. S. Y., Engelman, A. & Doudna, J. A. Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. *Nature* **519,** 193–198 (2015).
75. Wright, A. V., Liu, J.-J., Knott, G. J., Doxzen, K. W., Nogales, E. & Doudna, J. A. Structures of the CRISPR genome integration complex. *Science* **357,** 1113–1118 (2017).
76. McGinn, J. & Marraffini, L. A. CRISPR-Cas Systems Optimize Their Immune Response by Specifying the Site of Spacer Integration. *Molecular Cell* 1–20 (2016). doi:10.1016/j.molcel.2016.08.038
77. Wright, A. V. & Doudna, J. A. Protecting genome integrity during CRISPR immune adaptation. *Nature Publishing Group* 1–9 (2016). doi:10.1038/nsmb.3289
78. Datsenko, K. A., Pougach, K., Tikhonov, A., Wanner, B. L., Severinov, K. & Semenova, E. Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nature Communications* **3,** 945 (2012).
79. Xue, C., Whitis, N. R. & Sashital, D. G. Conformational Control of Cascade Interference and Priming Activities in CRISPR Immunity. *Molecular Cell* **64,** 826–834 (2016).
80. Redding, S., Sternberg, S. H., Marshall, M., Gibb, B., Bhat, P., Guegler, C. K., Wiedenheft, B., Doudna, J. A. & Greene, E. C. Surveillance and Processing of Foreign DNA by the Escherichia coli CRISPR-Cas System. *Cell* **163,** 854–865 (2015).
81. Künne, T., Kieper, S. N., Bannenberg, J. W., Vogel, A. I. M., Miellet, W. R., Klein, M., Depken, M., Suarez-Diez, M. & Brouns, S. J. J. Cas3-Derived Target DNA Degradation Fragments Fuel Primed CRISPR Adaptation. *Molecular Cell* **63,** 852–864 (2016).
82. Silas, S., Makarova, K. S., Shmakov, S., Paez-Espino, D., Mohr, G., Liu, Y., Davison, M., Roux, S., Krishnamurthy, S. R., Fu, B. X. H., Hansen, L. L., Wang, D., Sullivan, M. B., Millard, A., Clokie, M. R., Bhaya, D., Lambowitz, A. M., Kyrpides, N. C., Koonin, E. V. & Fire, A. Z. On the Origin of Reverse Transcriptase-Using CRISPR-Cas Systems and Their Hyperdiverse, Enigmatic Spacer Repertoires. *mBio* **8,** (2017).
83. Silas, S., Mohr, G., Sidote, D. J., Markham, L. M., Sanchez-Amat, A., Bhaya, D., Lambowitz, A. M. & Fire, A. Z. Direct CRISPR spacer acquisition from RNA by a

natural reverse transcriptase-Cas1 fusion protein. *Science* **351,** aad4234–aad4234 (2016).
84. Hegge, J. W., Swarts, D. C. & van der Oost, J. Prokaryotic Argonaute proteins: novel genome-editing tools? *Nat Rev Micro* **16,** 5–11 (2017).
85. Makarova, K. S., Wolf, Y. I., van der Oost, J. & Koonin, E. V. Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements. *Biol. Direct* **4,** 29 (2009).
86. Zander, A., Willkomm, S., Ofer, S., van Wolferen, M., Egert, L., Buchmeier, S., Stöckl, S., Tinnefeld, P., Schneider, S., Klingl, A., Albers, S.-V., Werner, F. & Grohmann, D. Guide-independent DNA cleavage by archaeal Argonaute from Methanocaldococcus jannaschii. *Nature Microbiology* **2,** 17034 (2017).
87. Swarts, D. C., Szczepaniak, M., Sheng, G., Chandradoss, S. D., Zhu, Y., Timmers, E. M., Zhang, Y., Zhao, H., Lou, J., Wang, Y., Joo, C. & van der Oost, J. Autonomous Generation and Loading of DNA Guides by Bacterial Argonaute. *Molecular Cell* **65,** 985–998.e6 (2017).
88. Fukuyo, M., Sasaki, A. & Kobayashi, I. Success of a suicidal defense strategy against infection in a structured habitat. *Scientific Reports* **2,** 1828 (2012).
89. Bingham, R., Ekunwe, S. I., Falk, S., Snyder, L. & Kleanthous, C. The major head protein of bacteriophage T4 binds specifically to elongation factor Tu. *J. Biol. Chem.* **275,** 23219–23226 (2000).
90. Kaufmann, G. Anticodon nucleases. *Trends Biochem. Sci.* **25,** 70–74 (2000).
91. Wang, C., Villion, M., Semper, C., Coros, C., Moineau, S. & Zimmerly, S. A reverse transcriptase-related protein mediates phage resistance and polymerizes untemplated DNA in vitro. *Nucleic Acids Res* **39,** 7620–7629 (2011).
92. Durmaz, E. & Klaenhammer, T. R. Abortive phage resistance mechanism AbiZ speeds the lysis clock to cause premature lysis of phage-infected Lactococcus lactis. *J. Bacteriol.* **189,** 1417–1425 (2007).
93. Depardieu, F., Didier, J.-P., Bernheim, A., Sherlock, A., Molina, H., Duclos, B. & Bikard, D. A Eukaryotic-like Serine/Threonine Kinase Protects Staphylococci against Phages. *Cell Host and Microbe* 1–12 (2016). doi:10.1016/j.chom.2016.08.010
94. Harms, A., Brodersen, D. E., Mitarai, N. & Gerdes, K. Toxins, Targets, and Triggers: An Overview of Toxin-Antitoxin Biology. *Molecular Cell* **70,** 768–784 (2018).
95. Sberro, H., Leavitt, A., Kiro, R., Koh, E., Peleg, Y., Qimron, U. & Sorek, R. Discovery of functional toxin/antitoxin systems in bacteria by shotgun cloning. *Molecular Cell* **50,** 136–148 (2013).
96. Alawneh, A. M., Qi, D., Yonesaki, T. & Otsuka, Y. An ADP-ribosyltransferase Alt of bacteriophage T4 negatively regulates the Escherichia coli MazF toxin of a toxin-antitoxin module. *Mol Microbiol* **99,** 188–198 (2016).
97. Otsuka, Y. & Yonesaki, T. Dmd of bacteriophage T4 functions as an antitoxin against Escherichia coli LsoA and RnlA toxins. *Mol Microbiol* **83,** 669–681 (2012).
98. Pecota, D. C. & Wood, T. K. Exclusion of T4 phage by the hok/sok killer locus from plasmid R1. *J. Bacteriol.* **178,** 2044–2050 (1996).

99. Fineran, P. C., Blower, T. R., Foulds, I. J., Humphreys, D. P., Lilley, K. S. & Salmond, G. P. C. The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair. *Proc. Natl. Acad. Sci. U.S.A.* **106,** 894–899 (2009).
100. Pedersen, K., Christensen, S. K. & Gerdes, K. Rapid induction and reversal of a bacteriostatic condition by controlled expression of toxins and antitoxins. *Mol Microbiol* **45,** 501–510 (2002).
101. Culviner, P. H. & Laub, M. T. Global Analysis of the E. coli Toxin MazF Reveals Widespread Cleavage of mRNA and the Inhibition of rRNA Maturation and Ribosome Biogenesis. *Molecular Cell* **70,** 868–880.e10 (2018).
102. Penadés, J. R. & Christie, G. E. The Phage-Inducible Chromosomal Islands: A Family of Highly Evolved Molecular Parasites. *Annu Rev Virol* **2,** 181–201 (2015).
103. Tormo-Más, M. A., Mir, I., Shrestha, A., Tallent, S. M., Campoy, S., Lasa, I., Barbé, J., Novick, R. P., Christie, G. E. & Penadés, J. R. Moonlighting bacteriophage proteins derepress staphylococcal pathogenicity islands. *Nature* **465,** 779–782 (2010).
104. Fillol-Salom, A., Martínez-Rubio, R., Abdulrahman, R. F., Chen, J., Davies, R. & Penadés, J. R. Phage-inducible chromosomal islands are ubiquitous within the bacterial universe. *The ISME Journal* **12,** 2114–2128 (2018).
105. Seed, K. D., Lazinski, D. W., Calderwood, S. B. & Camilli, A. A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. *Nature* **494,** 489–491 (2013).
106. McKitterick, A. C. & Seed, K. D. Anti-phage islands force their target phage to directly mediate island excision and spread. *Nature Communications* **9,** 2348 (2018).
107. Makarova, K. S., Wolf, Y. I., Snir, S. & Koonin, E. V. Defense islands in bacterial and archaeal genomes and prediction of novel defense systems. *J. Bacteriol.* **193,** 6039–6056 (2011).
108. Ofir, G., Melamed, S., Sberro, H., Mukamel, Z., Silverman, S., Yaakov, G., Doron, S. & Sorek, R. DISARM is a widespread bacterial defence system with broad anti-phage activities. *Nature Microbiology* **3,** 90–98 (2018).
109. Doron, S., Melamed, S., Ofir, G., Leavitt, A., Lopatina, A., Keren, M., Amitai, G. & Sorek, R. Systematic discovery of antiphage defense systems in the microbial pangenome. *Science* **359,** eaar4120 (2018).
110. Shah, S. A., Alkhnbashi, O. S., Behler, J., Han, W., She, Q., Hess, W. R., Garrett, R. A. & Backofen, R. Comprehensive search for accessory proteins encoded with archaeal and bacterial type III CRISPR-cas gene cassettes reveals 39 new cas gene families. *RNA Biology* **45,** 1–13 (2018).
111. Shmakov, S. A., Makarova, K. S., Wolf, Y. I., Severinov, K. V. & Koonin, E. V. Systematic prediction of genes functionally linked to CRISPR-Cas systems by gene neighborhood analysis. *Proc. Natl. Acad. Sci. U.S.A.* **115,** E5307–E5316 (2018).
112. Dedrick, R. M., Jacobs-Sera, D., Bustamante, C. A. G., Garlena, R. A., Mavrich, T. N., Pope, W. H., Reyes, J. C. C., Russell, D. A., Adair, T., Alvey, R., Bonilla, J.

A., Bricker, J. S., Brown, B. R., Byrnes, D., Cresawn, S. G., Davis, W. B., Dickson, L. A., Edgington, N. P., Findley, A. M., Golebiewska, U., Grose, J. H., Hayes, C. F., Hughes, L. E., Hutchison, K. W., Isern, S., Johnson, A. A., Kenna, M. A., Klyczek, K. K., Mageeney, C. M., Michael, S. F., Molloy, S. D., Montgomery, M. T., Neitzel, J., Page, S. T., Pizzorno, M. C., Poxleitner, M. K., Rinehart, C. A., Robinson, C. J., Rubin, M. R., Teyim, J. N., Vazquez, E., Ware, V. C., Washington, J. & Hatfull, G. F. Prophage-mediated defence against viral attack and viral counter-defence. *Nature Microbiology* 1–13 (2017). doi:10.1038/nmicrobiol.2016.251

113. Hynes, A. P., Villion, M. & Moineau, S. Adaptation in bacterial CRISPR-Cas immunity can be driven by defective phages. *Nature Communications* **5,** 4399 (2014).
114. Makarova, K. S., Wolf, Y. I. & Koonin, E. V. Comparative genomics of defense systems in archaea and bacteria. *Nucleic Acids Res* **41,** 4360–4377 (2013).
115. Hatoum-Aslan, A., Maniv, I., Samai, P. & Marraffini, L. A. Genetic Characterization of Antiplasmid Immunity through a Type III-A CRISPR-Cas System. *J. Bacteriol.* **196,** 310–317 (2013).
116. Makarova, K. S., Anantharaman, V., Grishin, N. V., Koonin, E. V. & Aravind, L. CARF and WYL domains: ligand-binding regulators of prokaryotic defense systems. *Front Genet* **5,** 102 (2014).
117. Burroughs, A. M., Zhang, D., Schäffer, D. E., Iyer, L. M. & Aravind, L. Comparative genomic analyses reveal a vast, novel network of nucleotide-centric systems in biological conflicts, immunity and signaling. *Nucleic Acids Res* **43,** 10633–10654 (2015).
118. Niewoehner, O. & Jinek, M. Structural basis for the endoribonuclease activity of the type III-A CRISPR-associated protein Csm6. *RNA* **22,** 318–329 (2016).
119. Corrigan, R. M. & Gründling, A. Cyclic di-AMP: another second messenger enters the fray. *Nature Publishing Group* **11,** 513–524 (2013).
120. Whiteley, A. T., Eaglesham, J. B., de Oliveira Mann, C. C., Morehouse, B. R., Lowey, B., Nieminen, E. A., Danilchanka, O., King, D. S., Lee, A. S. Y., Mekalanos, J. J. & Kranzusch, P. J. Bacterial cGAS-like enzymes synthesize diverse nucleotide signals. *Nature* **567,** 194–199 (2019).
121. Cohen, D., Melamed, S., Millman, A., Shulman, G., Oppenheimer-Shaanan, Y., Kacen, A., Doron, S., Amitai, G. & Sorek, R. Cyclic GMP-AMP signalling protects bacteria against viral infection. *Nature* **574,** 691–695 (2019).
122. Millman, A., Melamed, S., Amitai, G. & Sorek, R. Diversity and classification of cyclic-oligonucleotide-based anti-phage signalling systems. *Nature Microbiology* **18,** 113–8 (2020).
123. Pyenson, N. C. & Marraffini, L. A. Type III CRISPR-Cas systems: when DNA cleavage just isn't enough. *Current Opinion in Microbiology* **37,** 150–154 (2017).
124. Staals, R. H. J., Zhu, Y., Taylor, D. W., Kornfeld, J. E., Sharma, K., Barendregt, A., Koehorst, J. J., Vlot, M., Neupane, N., Varossieau, K., Sakamoto, K., Suzuki, T., Dohmae, N., Yokoyama, S., Schaap, P. J., Urlaub, H., Heck, A. J. R., Nogales, E., Doudna, J. A., Shinkai, A. & van der Oost, J. RNA Targeting by the Type III-A

CRISPR-Cas Csm Complex of Thermus thermophilus. *Molecular Cell* **56,** 518–530 (2014).
125. Zhang, J., Rouillon, C., Kerou, M., Reeks, J., Brugger, K., Graham, S., Reimann, J., Cannone, G., Liu, H., Albers, S.-V., Naismith, J. H., Spagnolo, L. & White, M. F. Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity. *Molecular Cell* **45,** 303–313 (2012).
126. Elmore, J. R., Sheppard, N. F., Ramia, N., Deighan, T., Li, H., Terns, R. M. & Terns, M. P. Bipartite recognition of target RNAs activates DNA cleavage by the Type III-B CRISPR–Cas system. *Genes Dev.* **30,** 447–459 (2016).
127. Estrella, M. A., Kuo, F.-T. & Bailey, S. RNA-activated DNA cleavage by the Type III-B CRISPR–Cas effector complex. *Genes Dev.* **30,** 460–470 (2016).
128. Samai, P., Pyenson, N., Jiang, W., Goldberg, G. W., Hatoum-Aslan, A. & Marraffini, L. A. Co-transcriptional DNA and RNA Cleavage during Type III CRISPR-Cas Immunity. *Cell* **161,** 1164–1174 (2015).
129. Rouillon, C., Athukoralage, J. S., Graham, S., Grüschow, S. & White, M. F. Control of cyclic oligoadenylate synthesis in a type III CRISPR system. *Elife* **7,** e02565–14 (2018).
130. Athukoralage, J. S., Rouillon, C., Graham, S., Grüschow, S. & White, M. F. Ring nucleases deactivate type III CRISPR ribonucleases by degrading cyclic oligoadenylate. *Nature* **562,** 277–280 (2018).
131. Foster, K., Kalter, J., Woodside, W., Terns, R. M. & Terns, M. P. The ribonuclease activity of Csm6 is required for anti-plasmid immunity by Type III-A CRISPR-Cas systems. *RNA Biology* **16,** 449–460 (2019).
132. Anantharaman, V., Makarova, K. S., Burroughs, A. M., Koonin, E. V. & Aravind, L. Comprehensive analysis of the HEPN superfamily: identification of novel roles in intra-genomic conflicts, defense, pathogenesis and RNA processing. *Biol. Direct* **8,** 15 (2013).
133. Marraffini, L. A. & Sontheimer, E. J. Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature* **463,** 568–571 (2010).
134. Deng, L., Garrett, R. A., Shah, S. A., Peng, X. & She, Q. A novel interference mechanism by a type IIIB CRISPR-Cmr module in Sulfolobus. *Mol Microbiol* **87,** 1088–1099 (2013).
135. Kreiswirth, B. N., Löfdahl, S., Betley, M. J., O'Reilly, M., Schlievert, P. M., Bergdoll, M. S. & Novick, R. P. The toxic shock syndrome exotoxin structural gene is not detectably transmitted by a prophage. *Nature* **305,** 709–712 (1983).
136. Helle, L., Kull, M., Mayer, S., Marincola, G., Zelder, M. E., Goerke, C., Wolz, C. & Bertram, R. Vectors for improved Tet repressor-dependent gradual gene induction or silencing in Staphylococcus aureus. *Microbiology* **157,** 3314–3323 (2011).
137. Liu, T. Y., Iavarone, A. T. & Doudna, J. A. RNA and DNA Targeting by a Reconstituted Thermus thermophilus Type III-A CRISPR-Cas System. *PLoS ONE* **12,** e0170552–20 (2017).

138. MachóN, C., Espinosa, M., Ruiz-Masó, J. A., Del Solar, G., Bordanaba-Ruiseco, L. & Coll, M. Plasmid Rolling-Circle Replication. *Microbiology Spectrum* **3,** 1–23 (2015).
139. Sheppard, N. F., Glover, C. V. C., III, Terns, R. M. & Terns, M. P. The CRISPR-associated Csx1 protein of Pyrococcus furiosusis an adenosine-specific endoribonuclease. *RNA* **22,** 216–224 (2016).
140. Upton, J. W. & Chan, F. K.-M. Staying Alive: Cell Death in Antiviral Immunity. *Molecular Cell* **54,** 273–280 (2014).
141. Makarova, K. S., Timinskas, A., Wolf, Y. I., Gussow, A. B., Siksnys, V., Venclovas, Č. & Koonin, E. V. Evolutionary and functional classification of the CARF domain superfamily, key sensors in prokaryotic antivirus defense. *Nucleic Acids Res* **48,** 8828–8847 (2020).
142. Jia, N., Jones, R., Yang, G., Ouerfelli, O. & Patel, D. J. CRISPR-Cas III-A Csm6 CARF Domain Is a Ring Nuclease Triggering Stepwise cA4 Cleavage with ApA>p Formation Terminating RNase Activity. *Molecular Cell* **75,** 944–956.e6 (2019).
143. Molina, R., Stella, S., Feng, M., Sofos, N., Jauniskis, V., Pozdnyakova, I., López-Méndez, B., She, Q. & Montoya, G. Structure of Csx1-cOA4 complex reveals the basis of RNA decay in Type III-B CRISPR-Cas. *Nature Communications* **10,** 4302–14 (2019).
144. McMahon, S. A., Zhu, W., Graham, S., Rambo, R., White, M. F. & Gloster, T. M. Structure and mechanism of a Type III CRISPR defence DNA nuclease activated by cyclic oligoadenylate. *Nature Communications* **11,** 500–11 (2020).
145. Kosinski, J., Feder, M. & Bujnicki, J. M. The PD-(D/E)XK superfamily revisited: identification of new members among proteins involved in DNA metabolism and functional predictions for domains of (hitherto) unknown function. *BMC Bioinformatics* **6,** 172–13 (2005).
146. Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A. R. N., Potter, S. C., Finn, R. D. & Lopez, R. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res* **47,** W636–W641 (2019).
147. Balaratnam, S. & Basu, S. Divalent cation-aided identification of physico-chemical properties of metal ions that stabilize RNA G-quadruplexes. *Biopolymers* **103,** 376–386 (2015).
148. Horinouchi, S. & Weisblum, B. Nucleotide sequence and functional map of pC194, a plasmid that specifies inducible chloramphenicol resistance. *J. Bacteriol.* **150,** 815–825 (1982).
149. Lau, R. K., Ye, Q., Birkholz, E. A., Berg, K. R., Patel, L., Mathews, I. T., Watrous, J. D., Ego, K., Whiteley, A. T., Lowey, B., Mekalanos, J. J., Kranzusch, P. J., Jain, M., Pogliano, J. & Corbett, K. D. Structure and Mechanism of a Cyclic Trinucleotide-Activated Bacterial Endonuclease Mediating Bacteriophage Immunity. *Molecular Cell* **77,** 723–733.e6 (2020).
150. Lowey, B., Whiteley, A. T., Keszei, A. F. A., Morehouse, B. R., Mathews, I. T., Antine, S. P., Cabrera, V. J., Kashin, D., Niemann, P., Jain, M., Schwede, F., Mekalanos, J. J., Shao, S., Lee, A. S. Y. & Kranzusch, P. J. CBASS Immunity

Uses CARF-Related Effectors to Sense 3'-5'- and 2'-5'-Linked Cyclic Oligonucleotide Signals and Protect Bacteria from Phage Infection. *Cell* **182,** 38–49.e17 (2020).
151. Meeske, A. J., Nakandakari-Higa, S. & Marraffini, L. A. Cas13-induced cellular dormancy prevents the rise of CRISPR-resistant bacteriophage. *Nature* (2019).
152. Wawrzyniak, P., Płucienniczak, G. & Bartosik, D. The Different Faces of Rolling-Circle Replication and Its Multifunctional Initiator Proteins. *Front. Microbiol.* **8,** 2353 (2017).
153. Krogh, A., Larsson, B., Heijne, von, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology* **305,** 567–580 (2001).
154. Cahill, J. & Young, R. Phage Lysis: Multiple Genes for Multiple Barriers. *Adv Virus Res* **103,** 33–70 (2019).
155. Mariano, G., Trunk, K., Williams, D. J., Monlezun, L., Strahl, H., Pitt, S. J. & Coulthurst, S. J. A family of Type VI secretion system effector proteins that form ion-selective pores. *Nature Communications* **10,** 5484–15 (2019).
156. Jacob, F. Evolution and tinkering. *Science* **196,** 1161–1166 (1977).
157. Makarova, K. S., Aravind, L., Grishin, N. V., Rogozin, I. B. & Koonin, E. V. A DNA repair system specific for thermophilic Archaea and bacteria predicted by genomic context analysis. *Nucleic Acids Res* **30,** 482–496 (2002).
158. Boampong, K., Smith, S. L. & Delahay, R. M. Rapid growth inhibitory activity of a YafQ-family endonuclease toxin of the Helicobacter pylori tfs4 integrative and conjugative element. *Scientific Reports* **10,** 18171–14 (2020).
159. Rostøl, J. T. & Marraffini, L. (Ph)ighting Phages: How Bacteria Resist Their Parasites. *Cell Host and Microbe* **25,** 184–194 (2019).
160. Mendoza, S. D., Nieweglowska, E. S., Govindarajan, S., Leon, L. M., Berry, J. D., Tiwari, A., Chaikeeratisak, V., Pogliano, J., Agard, D. A. & Bondy-Denomy, J. A bacteriophage nucleus-like compartment shields DNA from CRISPR nucleases. *Nature* **577,** 244–248 (2020).
161. Malone, L. M., Warring, S. L., Jackson, S. A., Warnecke, C., Gardner, P. P., Gumy, L. F. & Fineran, P. C. A jumbo phage that forms a nucleus-like structure evades CRISPR-Cas DNA targeting but is vulnerable to type III RNA-based immunity. *Nature Microbiology* **5,** 48–55 (2020).
162. Horinouchi, S. & Weisblum, B. Nucleotide sequence and functional map of pE194, a plasmid that specifies inducible resistance to macrolide, lincosamide, and streptogramin type B antibodies. *J. Bacteriol.* **150,** 804–814 (1982).
163. Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. & Gingeras, T. R. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15–21 (2013).
164. Ray, M. D., Boundy, S. & Archer, G. L. Transfer of the methicillin resistance genomic island among staphylococci by conjugation. *Mol Microbiol* **100,** 675–685 (2016).
165. Lamberte, L. E., Baniulyte, G., Singh, S. S., Stringer, A. M., Bonocora, R. P., Stracy, M., Kapanidis, A. N., Wade, J. T. & Grainger, D. C. Horizontally acquired

AT-rich genes in Escherichia coli cause toxicity by sequestering RNA polymerase. *Nature Microbiology* **2,** 16249–16 (2017).
166. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30,** 923–930 (2014).
167. Kabsch, W. XDS. *Acta Crystallogr D Biol Crystallogr* **66,** 125–132 (2010).
168. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. Phaser crystallographic software. *J Appl Crystallogr* **40,** 658–674 (2007).
169. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Zwart, P. H. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66,** 213–221 (2010).
170. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66,** 486–501 (2010).
171. Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* **68,** 352–367 (2012).
172. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res* **14,** 1188–1190 (2004).
173. Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P. & Drummond, A. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28,** 1647–1649 (2012).
174. Chen, W., Zhang, Y., Yeo, W.-S., Bae, T. & Ji, Q. Rapid and Efficient Genome Editing in Staphylococcus aureusby Using an Engineered CRISPR/Cas9 System. *J. Am. Chem. Soc.* **139,** 3790–3795 (2017).