8-2021

Determination of the Structure, Function, and Mechanism of Type IV CRISPR-Cas Prokaryotic Defense Systems

Hannah Nicole Taylor
Utah State University

DETERMINATION OF THE STRUCTURE, FUNCTION, AND MECHANISM OF TYPE IV CRISPR-CAS PROKARYOTIC DEFENSE SYSTEMS

by

Hannah Nicole Taylor

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

DOCTOR OF PHILOSOPHY

in

Biochemistry

Approved:

_____	_____
Ryan Jackson, Ph.D. Major Professor	Lance Seefeldt, Ph.D. Committee Member
_____	_____
Sean Johnson, Ph.D. Committee Member	Joan Hevel, Ph.D. Committee Member
_____	_____
Gregory Podgorski, Ph.D. Committee Member	D. Richard Cutler, Ph.D. Interim Vice Provost of Graduate Studies

UTAH STATE UNIVERSITY

Logan, Utah

2021

ABSTRACT

Determination of the Structure, Function, and Mechanism of Type IV CRISPR-Cas Prokaryotic Defense Systems

by

Hannah Nicole Taylor, Doctor of Philosophy

Utah State University, 2021

Major Professor: Dr. Ryan N. Jackson
Department: Chemistry and Biochemistry

Mobile genetic elements (MGEs), including plasmids and bacteriophage, are a constant threat to the prokaryotes they invade. CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) – Cas (CRISPR-Associated) systems are one of the many systems prokaryotes use to defend against MGEs. CRISPR-Cas systems are incredibly diverse. There are two classes of CRISPR-Cas systems which are further divided into six types and over 33 subtypes. Each CRISPR-Cas subtype provides defense via a unique set of protein and RNA components. Type IV CRISPR-Cas systems are the least studied of the six CRISPR-Cas types. In fact, in some type IV systems, only the genetic architecture is known and their overall function is unclear. This dissertation contains many of the first structural and biochemical studies done on type IV systems. Reported here is the first structure of a type IV ribonucleoprotein complex, indicating that a certain subtype of type IV systems, IV-B, may have an exclusively non-defense function. Also reported is a biochemical characterization of a IV-A Cas endoribonuclease, the first evidence of defense against MGEs by a type IV system, and an initial exploration into potential

mechanisms of defense. These studies indicate that another type IV subtype, IV-A, is a defense system, consistent with its classification as a CRISPR-Cas system. Beyond expanding the realm of prokaryotic defense systems, these studies answer many questions surrounding type IV systems and provide a clear direction to further our understanding of these enigmatic CRISPR-Cas systems.

(287 pages)

PUBLIC ABSTRACT

Determination of the Structure, Function, and Mechanism of Type IV CRISPR-Cas Prokaryotic Defense Systems

Hannah Nicole Taylor

Bacteria are under constant threat of invasion by bacteriophage (viruses which infect bacteria). To prevent bacteriophage from entering and overtaking the bacteria, bacteria utilize defense systems to identify and destroy foreign elements. One method of defense is called CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repeats – CRISPR-Associated). Many different bacteria and most archaea use CRISPR-Cas systems. There are many diverse types of CRISPR-Cas systems, each of which provides defense in a slightly different way. One such CRISPR-Cas type is called type IV. The type IV CRISPR-Cas system is poorly understood and there are very few studies published on type IV systems. This dissertation details some of the first studies done on type IV systems, showing that some type IV systems are indeed defense systems, while others may have evolved a non-defense function. Several biochemical studies were performed to better understand the underlying mechanisms of type IV systems. Historically, the study of CRISPR-Cas mechanisms has led to innovations in gene editing, cancer research, diagnostics, therapeutics, and much more. The work described here significantly furthers the CRISPR-Cas field and may lead to the discovery of new, impactful mechanisms and biological tools.

ACKNOWLEDGMENTS

I thank Ryan Jackson for his mentorship and all the members of the Jackson lab for their teamwork and willingness to put up with me for the last several years.

I thank the members of my committee for pushing me to do better and set my own standards.

I thank my father for teaching me hard work and dedication. I thank my mother for teaching me the value of education and personal growth.

Most of all, I thank my husband and son for their unwavering support.

Hannah Nicole Taylor

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Cas	"CRISPR-associated"
	Usually refers to a gene or protein associated with a CRISPR system or operon.
CRISPR	"Clustered regularly interspaced short palindromic repeats"
	Describes the genomic locus that typifies CRISPR-Cas system and stores memories of past invasions.
crRNA	"CRISPR-derived RNA"
	The RNA component of CRISPR-Cas RNP complexes. It is derived from transcripts of the CRISPR locus.
Cryo-EM	"cryo-electron microscopy"
Csf	"CRISPR-Cas subtype as in *Acidithiobacillus ferrooxidans*"
	The prefix for type IV CRISPR-Cas genes and proteins, e.g. *csf1*, Csf2, Csf RNP complex.
CysH	"cysteine biosynthesis H"
	An enzyme involved in cysteine biosynthesis. CasCysH is a CysH-like protein that is associated with type IV CRISR-Cas systems.
dsDNA, ssDNA	"double-stranded DNA", "single-stranded DNA"
DinG	"damage inducible gene G"
	An enzyme involved in DNA repair. CasDinG is a DinG-like protein that is associated with type IV CRISR-Cas systems.
HD domain	"histidine (H) and aspartate (D) domain"
	Nuclease domain typified by histidine and aspartate residues and composed of several motifs.
MGE	"mobile genetic element"
	Bits of nucleic acid which encode for its own movement around or between genomes. MGEs include plasmids, transposons, and bacteriophages.

PAM, rPAM	"protospacer adjacent motif", "RNA protospacer adjacent motif"
	Describes a 2-5 nucleotide motif adjacent to protospacers (DNA/RNA complementary to crRNA) required for Cas RNP complexes to bind the protospacer.
PFS	"protospacer flanking sequence"
	A PAM-like motif adjacent to protospacers in the type VI CRISPR-Cas system.
RNP	"ribonucleoprotein"
	A complex composed of both RNA and protein components.

CHAPTER 1

INTRODUCTION

CRISPR-CAS SYSTEMS ARE A DIVERSE GROUP OF PROKARYOTIC DEFENSE SYSTEMS

There is an intricate evolutionary relationship between mobile genetic elements (MGEs) (bacteriophage, plasmids, etc.) and their prokaryotic hosts. For example, bacteriophage (phage), viruses which infect bacteria and archaea, reproduce by hijacking the cellular machinery of the host and forcing replication of the phage genome and translation of phage proteins (Campbell, 2003; Frost et al., 2005). This hijacking is deleterious to the prokaryotic host, often resulting in cell death. Defense systems exist in bacteria and archaea to prevent the effects of infection by bacteriophage (Labrie et al., 2010; Rostøl & Marraffini, 2019). In response to prokaryotic defense systems, bacteriophage encode anti-defense systems (Hampton et al., 2020). Most prokaryotes encode several types of defense systems, making it harder for a bacteriophage to bypass all defensive layers. Thus, the evolutionary history of prokaryotes and their parasitic bacteriophages is linked through an arms race (Gómez & Buckling, 2011; Hall et al., 2011; Gurney et al., 2019).

One type of prokaryotic defense system is the CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)-Cas (CRISPR-associated) system. CRISPR-Cas systems harbor nucleic acid memories of past MGE invaders by integrating short (~30bp) segments of invasive DNA into the bacterial/archaeal genome at the CRISPR locus (Barrangou et al., 2007; Nuñez et al., 2014; Rollie et al., 2015; Wang et al., 2015). The CRISPR-based genetic memory is then transcribed and processed to be used by Cas

proteins to identify and eliminate successive invasions by complementary nucleic acids (Brouns et al., 2008; Carte et al., 2008; C. Hale et al., 2008; C. R. Hale et al., 2009; Garneau et al., 2010) (**Figure 1-1**).

Like the prokaryotes they inhabit and the bacteriophage they defend against, CRISPR-Cas systems are incredibly diverse. There are two classes (class 1-2), six types (type I-VI), and over 33 subtypes (e.g. type I-A, I-B, I-C, etc.) of CRISPR-Cas systems (Makarova et al., 2015, 2020) (**Figure 1-2**). Class 1 CRISPR-Cas systems form a multi-subunit ribonucleoprotein (RNP) complex with a CRISPR-derived RNA (crRNA) to surveil the cell and identify invasive nucleic acids, while the RNP complexes of class 2 CRISPR-Cas systems are composed of a single protein subunit and crRNA (**Figure 1-3**). The CRISPR-Cas types are classified by the genes they encode and their operon architecture with each type encoding a type-specific, or signature, gene (Makarova et al., 2018, 2020). Variants within each type are classified into subtypes with some encoding subtype-specific genes. The basic functions of the six CRISPR-Cas types have all been characterized, except the type IV system (**Figure 1-2**). To further our understanding of CRISPR-Cas systems, it is critical that we understand the function of type IV CRISPR-Cas systems.

TYPE IV SYSTEMS ARE MINIMAL, MOBILE CRISPR-CAS SYSTEMS

Some functional aspects of type IV CRISPR-Cas systems can be inferred by comparing their operons to the operons of other CRISPR-Cas types. Firstly, type IV systems lack the *cas1* gene (Makarova et al., 2015). This is especially striking as nearly all other CRISPR-Cas types encode cas1, regardless of class or type (Wiedenheft et al.,

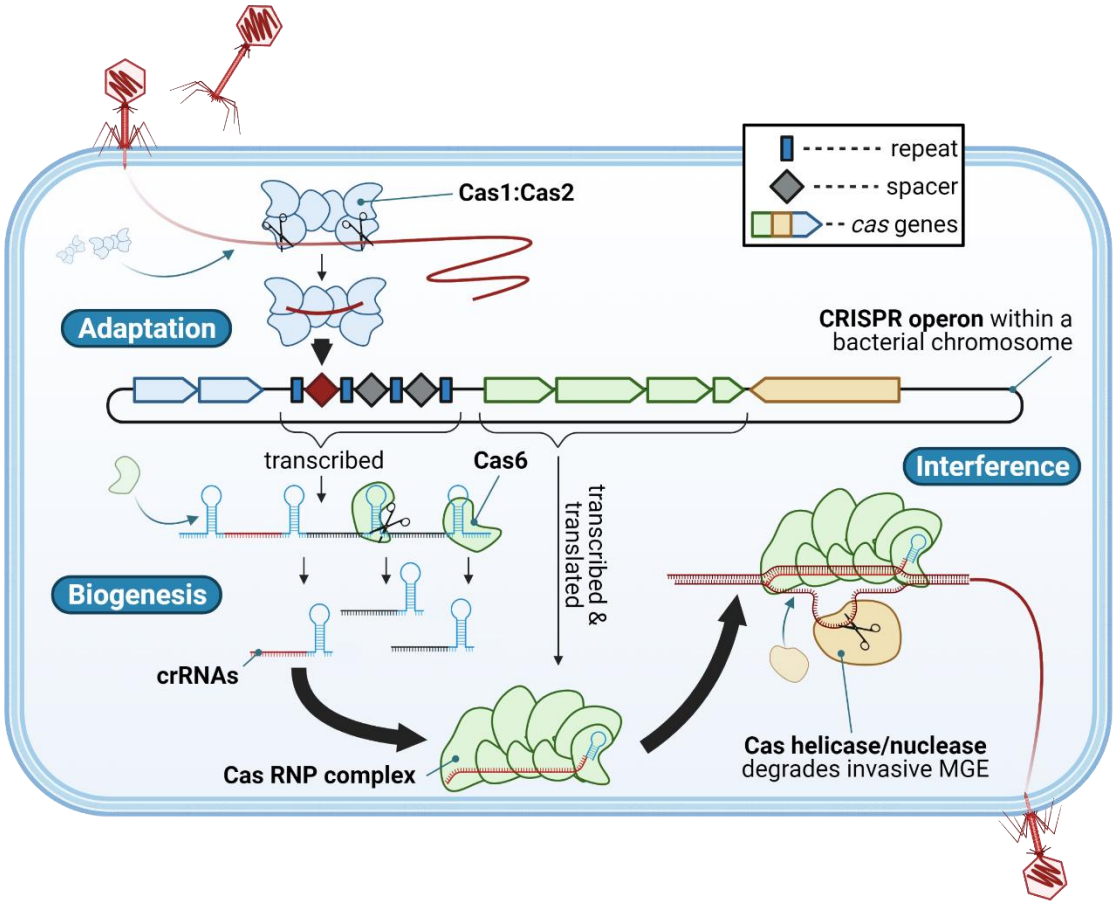

Figure 1-1. The three stages of CRISPR-Cas defense. During adaptation Cas1 and Cas2 integrate small chucks of foreign DNA into the CRISPR. During biogenesis the CRISPR is transcribed and Cas6 endoribonucleases cleave the transcript to produce a library of crRNAs. The crRNAs are bound by Cas proteins to form Cas ribonucleoprotein (RNP) complexes. The RNP complex binds nucleic acids complementary to the crRNA, inducing destruction of the invasive nucleic acid. Created with BioRender.com.

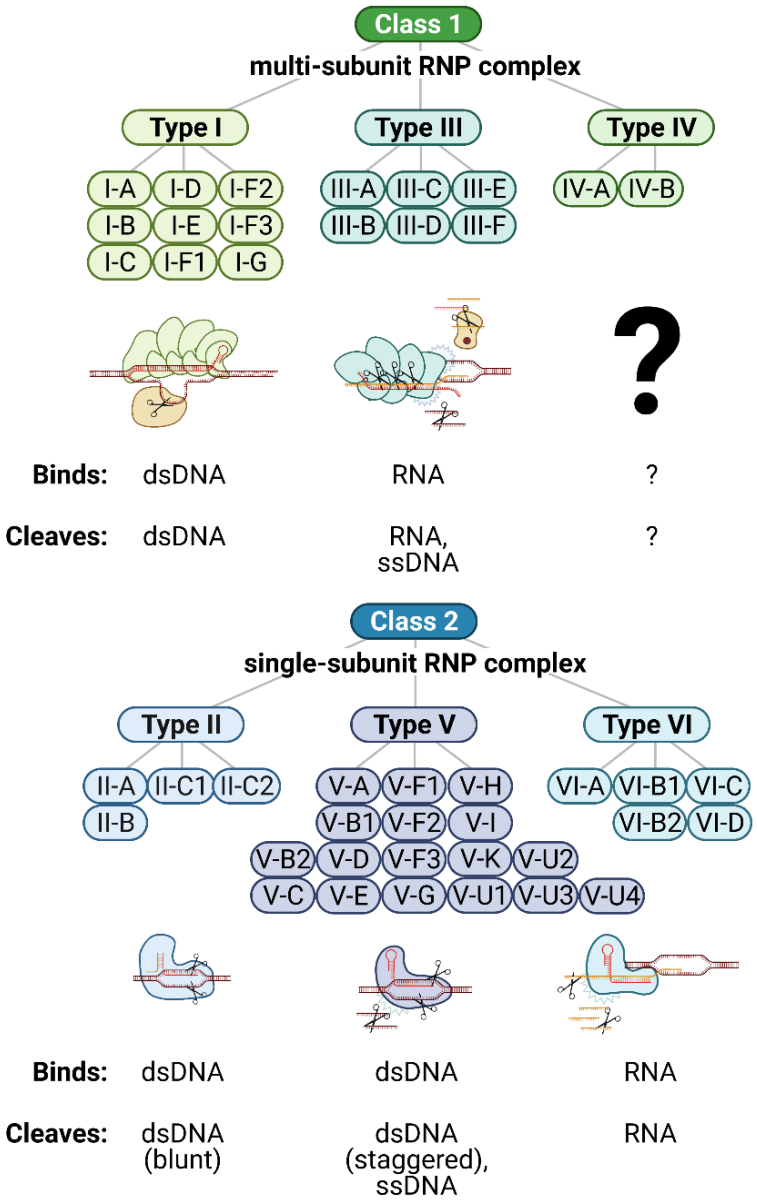

Figure 1-2. Classification of CRISPR-Cas systems into two classes, six types, and many subtypes. Diagrams depicting the Cas RNP complex for each type are shown, including descriptions of how nucleic acids are cleaved. RNP complex formation and possible nucleic acid binding and/or cleavage is unknown for type IV CRISPR-Cas systems. See also **Figure 1-3**. Created with BioRender.com.

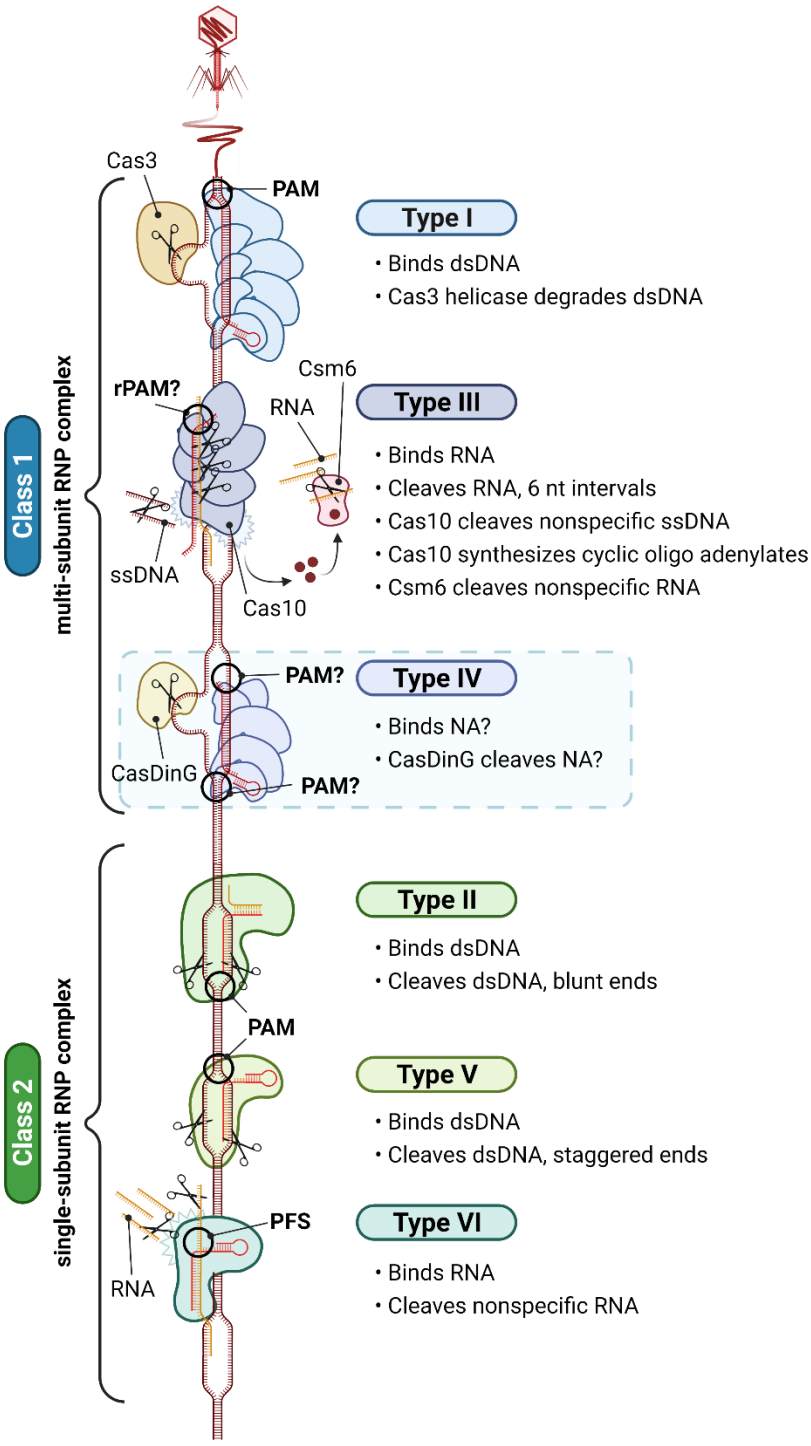

Figure 1-3. Classification of CRISPR-Cas systems indicating binding and cleavage patterns. Binding and cleavage by Type IV systems is only hypothesized. Indicated protospacer adjacent motifs (PAMs) and protospacer flanking sequences (PFSs) are required for binding and/or activation of the system. See also **Figure 1-2**. Created with BioRender.com.

2009). Cas1 and its counterparts, Cas2 and Cas4, are referred to as adaptation or maintenance machinery, as they are responsible for integrating new nucleic acid sequences, called protospacers, into the CRISPR locus (Nuñez et al., 2014; Rollie et al., 2015; Wang et al., 2015; Kieper et al., 2018; Lee et al., 2018). The CRISPR locus is composed of short, repeated sequences, called repeats, interspaced with short, unique sequences derived from invasive nucleic acids, called spacers (Bolotin et al., 2005; Mojica et al., 2005; Pourcel et al., 2005) (**Figure 1-1**). At the 5' end of the CRISPR is a leader sequence. Cas1 and Cas2 form an oligomeric complex which integrates protospacers into the CRISPR, always immediately after the leader sequence (Nuñez et al., 2014; Xiao et al., 2017). Despite their lack of a *cas1* gene, many type IV systems do encode a CRISPR. It has been hypothesized that type IV system may borrow adaptation machinery from other CRISPR-Cas types, as has been seen with some type III CRISPR-Cas systems (Staals et al., 2013, 2014; Elmore et al., 2015; Bernheim et al., 2020). Supporting this hypothesis, type IV systems are often present in organisms which encode two or more CRISPR-Cas systems of various types and some crosstalk between these systems has been implicated (Pinilla-Redondo et al., 2019). Type IV systems are considered minimal systems, as they lack components (Cas1/Cas2/Cas4) typically required for CRISPR-Cas immunity.

Secondly, type IV systems encode several *cas* genes which are putative subunits of an RNP complex, thus indicating that type IV systems are class 1 CRISPR-Cas systems (Makarova et al., 2011, 2015). Recent work has shown that type IV systems form multi-subunit RNP complexes, confirming their identity as class 1 systems (Özcan et al., 2018; Zhou et al., 2021). The Type IV cas genes *csf2* and *csf3* have predicted gene

products that are, respectively, Cas7-like and Cas5-like. Multiple Cas7 proteins form the backbone of Cas RNP complexes and directly bind the crRNA (Jore et al., 2011; Lintner et al., 2011; Staals et al., 2013; Jackson et al., 2014; Mulepati et al., 2014; Osawa et al., 2015) (**Figure 1-4**). The type IV Cas7 homolog, Csf2, has the same RNA binding capabilities (Zhou et al., 2021). Within multi-subunit Cas RNP complexes, a single Cas5 subunit caps one end of the complex (**Figure 1-4**). Csf3 is the Cas5 homolog in type IV systems. Type IV systems also contain the *csf1* gene. *csf1* is unique to type IV and is a putative component of the RNP complex. Csf1 is most similar to the Cas8 and Cas10 subunits of type I and type III Cas RNP complexes, respectively. Both Cas8 and Cas10 bind adjacent to Cas5 at one end of the RNP complex (Staals et al., 2013; Jackson et al., 2014). However, Csf1 has a much lower molecular weight than the type I and type III large subunits and a cross-linking experiment by Özcan et al. (2018) indicates that Csf1 does not interact with Csf3, the type IV Cas5 homolog. Although the type IV Csf proteins form an RNP complex, like the type I and type III Cas proteins, the stoichiometry and arrangement of the protein subunits within the RNP complex are likely unique to type IV systems.

Finally, in stark contrast to other CRISPR-Cas systems, type IV systems are almost always encoded on mobile genetic elements (MGEs) (plasmids, prophages, etc.) (Bernheim et al., 2020; Faure et al., 2019; Koonin & Makarova, 2017, 2019; Makarova et al., 2015; Pinilla-Redondo et al., 2019). This fact has spurred some tantalizing hypotheses for type IV system functions, including that type IV systems may be mobile defense systems, passed on plasmids between organisms as needed. CRISPR-Cas systems are often co-opted by MGEs. For example, CRISPR-Cas systems have been identified within

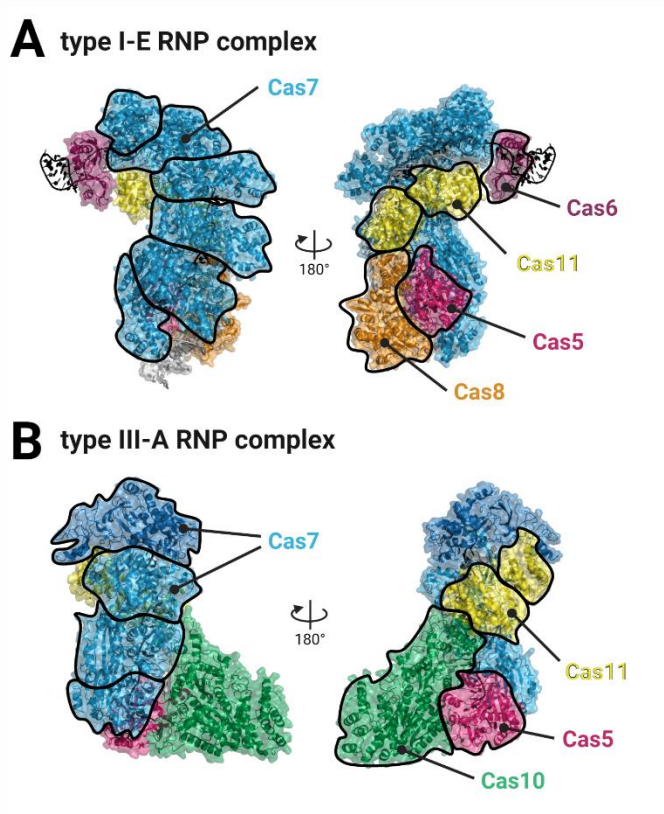

Figure 1-4. Representative class 1 RNP complexes. (**A**) Type I-E RNP complex from *E. coli*. Colored by subunits. PDBid: 5H9F. (**B**) Type III-B RNP complex from *Thermococcus onnurineus.* Colored as in (A). PDBid: 6O7I. Created with BioRender.com.

phage genomes and as critical components of transposon systems. Phages utilize CRISPR-Cas systems to by-pass host defense systems and prevent super infection by competitor bacteriophages (Al-Shayeb et al., 2020; McKitterick et al., 2019; Naser et al., 2017; Seed et al., 2013). Transposons use the RNA guided-DNA binding capabilities of the Cas RNP complex to identify genomic loci for transposition (Klompe et al., 2019; Peters et al., 2017). In the defense arms race between MGEs and prokaryotes, CRISPR systems are "guns for hire", utilized by both prokaryotes and MGEs (Peters et al., 2017). As type IV CRISPR-Cas systems are almost always encoded by MGEs, they may have a functional mechanism that benefits MGEs over prokaryotes.

TYPE IV CRISPR-CAS SYSTEMS CONTAIN DIVERSE SUBTYPES

Type IV systems are further divided into three subtypes, A-C (**Figure 1-5**). Each of these subtypes encodes a subtype-specific gene and has a distinct subtype-specific genetic architecture (Makarova et al., 2020). Unlike the other five CRISPR-Cas types, subtypes within the type IV CRISPR-Cas system are distinct enough that they each are predicted to provide a unique function and utilize different enzymes and mechanisms. When I first started my graduate studies, only the type IV-A and type IV-B CRISPR-Cas subtypes had been classified, no biochemical studies and only minimal bioinformatic studies had been done on type IV systems, and the type IV structures and functions were completely unknown. The following paragraphs give a brief overview of the type IV subtypes and will summarize the studies contained in the chapters of this dissertation.

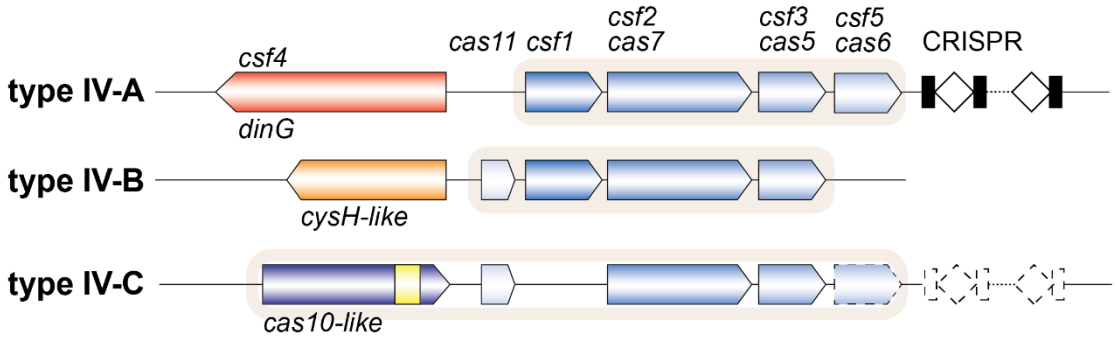

Figure 1-5. Operons of the type IV subtypes. Dashed lines indicate a component that is only sometimes present. The yellow box represents an HD nuclease domain. *dinG*, *cysH-like*, and *cas10-like* are the type IV-A, type IV-B, and type IV-C, respective, subtype-specific genes.

Type IV-A Systems are Minimal, Defense Systems

Type IV-A systems encode a *csf5/cas6* gene and a CRISPR (Makarova et al., 2015, 2020) (**Figure 1-5**). In class 1 CRISPR-Cas systems, Cas6 enzymes are responsible for providing crRNAs to the Cas RNP complex (Brouns et al., 2008). The CRISPR locus is initially transcribed into a long pre-crRNA (Carte et al., 2008) (**Figure 1-1**). The repeats of the CRISPR often contain palindromic sequences which form RNA hairpins in the transcript. Cas6 binds and cleaves the repeat sequence, thus processing the pre-crRNA into a library of crRNAs which are incorporated into Cas RNP complexes (one crRNA per RNP complex). Interestingly, type IV-A subtypes sometimes contain Cas6 enzymes and sometimes contain Csf5 enzymes (Makarova et al., 2020). To determine whether the type IV-A Cas6 protein is a crRNA processing enzyme, a structure was determined for the IV-A Cas6 from *Mahella australiensis* and biochemical assays were used to characterize its endoribonuclease activity (Taylor et al., 2019). I show that the type IV-A Cas6 is a crRNA processing endonuclease (**Chapter 3**). Additional work has shown that type IV-A Csf5 enzymes are also crRNA processing endonucleases (Özcan et al., 2018) and that Csf5 and IV-A Cas6 have unique structures, active sites, and mechanisms of actions (Taylor et al., 2019; Taylor et al., 2021) (**Chapter 3** and **Chapter 5**), indicating that type IV progenitor systems may have acquired endoribonucleases multiple times along its evolutionary history.

The subtype-specific gene of type IV-A systems is the putative helicase *dinG* (Makarova et al., 2020). Many bacteria encode *dinG* in their genomes and its gene product, DinG, has been associated with DNA damage repair (Lewis et al., 1992; Voloshin et al., 2003; Voloshin & Camerini-Otero, 2007; McRobbie et al., 2012; Thakur

et al., 2014). DinG is a helicase and some DinG variants have nuclease activity. To determine whether DinG associated with CRISPR-Cas systems (CasDinG) may have a similar enzymatic function, several amino-acid sequence alignments were performed. Results indicate that the CasDinG associated with type IV systems has likely evolved a Cas-specific function (Taylor et al., 2021) (**Chapter 5**). It is hypothesized that CasDinG may function as the ultimate destroyer of invasive nucleic acids in Type IV systems, similar to the function of Cas3 in Type I systems. In Type I systems a ribonucleoprotein complex identifies foreign dsDNA targets, then recruits Cas3 to the target dsDNA (Huo et al., 2014) (**Figure 1-3).** Cas3, a helicase and nuclease, translocates along the target dsDNA, cleaving it at certain intervals (Beloglazova et al., 2011; Mulepati & Bailey, 2011; Sinkunas et al., 2011). The role of CasDinG within type IV-A systems is a major question within the field. To determine whether type IV-A systems are defense systems, a plasmid-curing assay was performed (Crowley et al., 2019). Indeed, the type IV-A system can defend against plasmid transformation and CasDinG is required for defense (**Chapter 4**). Further studies to elucidate the mechanisms of defense indicate that the type IV system utilizes a Csf RNP complex that preferentially binds single-stranded DNA and RNAs that are complementary to the crRNA guide (**Chapter 6**). CasDinG is also further implicated as a potential nuclease.

Type IV-B Systems Likely Have a Non-Defense Function

Type IV-B CRISPR-Cas systems encode *csf1, csf2, csf3*, and a *cas11* subunit but, in addition to the absence of *cas1*, lack a *cas6/csf5* and a CRISPR locus (Makarova et al., 2015, 2020) (**Figure 1-5**). Due to the absence of a CRISPR locus and a predicted

accessory nuclease, it is unlikely that the type IV-B system is a canonical CRISPR-Cas defense system. It is hypothesized that type IV-B systems may either be a gene regulation system, as the components of CRISPR-Cas systems, which recognize and bind nucleic acids, are still present, or that type IV-B systems may be a mobile CRISPR-Cas defense system, which borrows crRNAs from other CRISPR-Cas systems to perform its function (Makarova et al., 2011; Koonin & Makarova, 2017; Shmakov et al., 2018; Faure et al., 2019). The latter hypothesis is consistent with the localization of Type IV-B systems on plasmids, therefore allowing the system to be passed between neighboring organisms as needed (Koonin & Makarova, 2019).

Type IV-B systems also encode an accessory *cysH* gene (Shmakov et al., 2018; Makarova et al., 2020). *cysH* genes encode enzymes in the PAPS reductase family that are involved in sulfonucleotide reduction during L-cysteine biosynthesis (Carroll et al., 2005). Bioinformatics studies were undertaken to determine whether CRISPR-Cas associated CysH proteins (CasCysH) could have a similar function to their non-CRISPR-Cas associated counter parts. However, the CasCysH of type IV-B CRISPR-Cas systems has a devolved sulfonucleotide reduction active site, suggesting that the CasCysH function is specific to type IV-B systems (Taylor et al., 2021) (**Chapter 5**).

The function of type IV-B systems has only been postulated. To determine whether the type IV-B system can form a Cas RNP complex and what its structure is, I purified and helped to determine the cryo-EM structure of a IV-B RNP complex, lending insights into possible functions of type IV-B CRISPR-Cas systems (**Chapter 2**).

Type IV Systems are Diverse, with New Subtypes and Variants Still Being Discovered

A hallmark of type IV systems is the diversity of its subtypes. For example, the genetic composition and predicted functions of type IV-A and type IV-B subtypes are incredibly different (Faure et al., 2019). Additionally, there are many identified variants of type IV systems which encode diverse genes, including *recD* (instead of *dinG*), a *csf1-csf3* fusion gene, and genes typically associated with type III systems (Kamruzzaman & Iredell, 2020; Pinilla-Redondo et al., 2019). Some variants entirely lack the type IV-specific gene *csf1* (Crowley et al., 2019; Pinilla-Redondo et al., 2019). The most notable variant, type IV-C, was only recently classified as a type IV subtype (Makarova et al., 2020; Pinilla-Redondo et al., 2019). Type IV-C systems encode *csf2*, *csf3*, and the type IV *cas11* genes (**Figure 1-5**). Intriguingly, type IV-C systems lack *csf1* and instead encode a *cas10-like* gene with an HD nuclease domain. The presence of the type IV Cas11 in type IV-C systems indicates that an RNP complex similar to the IV-B RNP complex may form. However, the presence of an obvious nuclease domain indicates the IV-C system could perform a defense function. Overall, it seems that the type IV-C system will have a unique function from the other type IV subtypes, illustrating the rich functional diversity within type IV CRISPR-Cas systems.

UNDERSTANDING TYPE IV SYSTEMS IS CRITICAL TO DEEPEN OUR KNOWLEDGE OF PROKARYOTIC BIOLOGY

The study of CRISPR-Cas systems is greatly expanding our understanding of interactions between prokaryotes and bacteriophages, prokaryotic evolution, and basic prokaryotic biology. CRISPR-Cas systems contain a plethora of previously unknown and

uncharacterized enzymes, many of which have been repurposed as biotechnological tools that are revolutionizing basic research, data storage/transport, nucleic acid detection/diagnostics, medicine, and more (Shipman et al., 2017; Ma et al., 2017; Gaudelli et al., 2017; Kellner et al., 2019; Yan et al., 2019; Makarova et al., 2020). Despite the paradigm shifting work described in this dissertation, the function of type IV CRISPR-Cas systems remains a glaring gap in our understanding of prokaryotic defense systems. Additionally, type IV CRISPR-Cas systems are incredibly diverse and many type IV subtypes likely have evolved non-defense functions. Therefore, the study of type IV systems will uncover novel functions and enzymatic mechanisms. An increase in knowledge is assured and the potential for world-changing discoveries is high through the study of type IV CRISPR-Cas systems.

REFERENCES

Al-Shayeb, B., Sachdeva, R., Chen, L.-X., Ward, F., Munk, P., Devoto, A., Castelle, C. J., Olm, M. R., Bouma-Gregson, K., Amano, Y., He, C., Méheust, R., Brooks, B., Thomas, A., Lavy, A., Matheus-Carnevali, P., Sun, C., Goltsman, D. S. A., Borton, M. A., … Banfield, J. F. (2020). Clades of huge phages from across Earth's ecosystems. *Nature*, 578(7795), 425–431. https://doi.org/10.1038/s41586-020-2007-4

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A., & Horvath, P. (2007). CRISPR Provides Acquired Resistance against Viruses in Prokaryotes. *Science, 315*(5819), 1709–1712.

Beloglazova, N., Petit, P., Flick, R., Brown, G., Savchenko, A., & Yakunin, A. F. (2011). Structure and activity of the Cas3 HD nuclease MJ0384, an effector enzyme of the CRISPR interference. *The EMBO Journal*, *30*(22), 4616–4627. https://doi.org/10.1038/emboj.2011.377

Bernheim, A., Bikard, D., Touchon, M., & Rocha, E. P. C. (2020). Atypical organizations and epistatic interactions of CRISPRs and cas clusters in genomes and their mobile genetic elements. *Nucleic Acids Research*, *48*(2), 748–760. https://doi.org/10.1093/nar/gkz1091

Bolotin, A., Quinquis, B., Sorokin, A., & Ehrlich, S. D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology*, *151*(8), 2551–2561. https://doi.org/10.1099/mic.0.28048-0

Brouns, S. J. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J. H., Snijders, A. P. L., Dickman, M. J., Makarova, K. S., Koonin, E. V., & van der Oost, J. (2008). Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. *Science*, *321*(5891), 960–964. https://doi.org/10.1126/science.1159689

Campbell, A. (2003). The future of bacteriophage biology. *Nature Reviews Genetics*, *4*. https://doi.org/10.1038/nrg1089

Carroll, K. S., Gao, H., Chen, H., Stout, C. D., Leary, J. A., & Bertozzi, C. R. (2005). A Conserved Mechanism for Sulfonucleotide Reduction. *PLoS Biology*, *3*(8), e250. https://doi.org/10.1371/journal.pbio.0030250

Carte, J., Wang, R., Li, H., Terns, R. M., & Terns, M. P. (2008). Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes & Development*, *22*(24), 3489–3496. https://doi.org/10.1101/gad.1742908

Crowley, V. M., Catching, A., Taylor, H. N., Borges, A. L., Metcalf, J., Bondy-Denomy, J., & Jackson, R. N. (2019). A Type IV-A CRISPR-Cas System in *Pseudomonas aeruginosa* Mediates RNA-Guided Plasmid Interference *In Vivo*. *The CRISPR Journal*, *2*(6), 434–440. https://doi.org/10.1089/crispr.2019.0048

Elmore, J., Deighan, T., Westpheling, J., Terns, R. M., & Terns, M. P. (2015). DNA targeting by the type I-G and type I-A CRISPR-Cas systems of Pyrococcus furiosus. *Nucleic Acids Research*, *43*(21), 10353–10363. https://doi.org/10.1093/nar/gkv1140

Faure, G., Shmakov, S. A., Yan, W. X., Cheng, D. R., Scott, D. A., Peters, J. E., Makarova, K. S., & Koonin, E. V. (2019). CRISPR–Cas in mobile genetic elements: Counter-defence and beyond. *Nature Reviews Microbiology*. https://doi.org/10.1038/s41579-019-0204-7

Frost, L. S., Leplae, R., Summers, A. O., & Toussaint, A. (2005). Mobile genetic elements: The agents of open source evolution. *Nature Reviews Microbiology*, *3*. https://doi.org/10.1038/nrmicro1235

Garneau, J. E., Dupuis, M.-È., Villion, M., Romero, D. A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadán, A. H., & Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*, *468*(7320), 67–71. https://doi.org/10.1038/nature09523

Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I., & Liu, D. R. (2017). Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature*, *551*(7681), 464–471. https://doi.org/10.1038/nature24644

Gómez, P., & Buckling, A. (2011). Bacteria-Phage Antagonistic Coevolution in Soil. *Science*, *332*(6025), 106–109. https://doi.org/10.1126/science.1198767

Gurney, J., Pleška, M., & Levin, B. R. (2019). Why put up with immunity when there is resistance: An excursion into the population and evolutionary dynamics of restriction–modification and CRISPR-Cas. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*(1772). https://doi.org/10.1098/rstb.2018.0096

Hale, C., Kleppe, K., Terns, R. M., & Terns, M. P. (2008). Prokaryotic silencing (psi)RNAs in Pyrococcus furiosus. *RNA*, *14*(12), 2572–2579. https://doi.org/10.1261/rna.1246808

Hale, C. R., Zhao, P., Olson, S., Duff, M. O., Graveley, B. R., Wells, L., Terns, R. M., & Terns, M. P. (2009). RNA-Guided RNA Cleavage by a CRISPR RNA-Cas Protein Complex. *Cell*, *139*(5), 945–956. https://doi.org/10.1016/j.cell.2009.07.040

Hall, A. R., Scanlan, P. D., Morgan, A. D., & Buckling, A. (2011). Host–parasite coevolutionary arms races give way to fluctuating selection. *Ecology Letters*, *14*(7), 635–642. https://doi.org/10.1111/j.1461-0248.2011.01624.x

Hampton, H. G., Watson, B. N. J., & Fineran, P. C. (2020). The arms race between bacteria and their phage foes. *Nature*, *577*(7790), 327–336. https://doi.org/10.1038/s41586-019-1894-8

Huo, Y., Nam, K. H., Ding, F., Lee, H., Wu, L., Xiao, Y., Farchione, M. D., Zhou, S., Rajashankar, K., Kurinov, I., Zhang, R., & Ke, A. (2014). Structures of CRISPR Cas3 offer mechanistic insights into Cascade-activated DNA unwinding and degradation. *Nature Structural & Molecular Biology*, *21*(9), 771–777. https://doi.org/10.1038/nsmb.2875

Jackson, R. N., Golden, S. M., Erp, P. B. G. van, Carter, J., Westra, E. R., Brouns, S. J. J., Oost, J. van der, Terwilliger, T. C., Read, R. J., & Wiedenheft, B. (2014). Crystal structure of the CRISPR RNA–guided surveillance complex from Escherichia coli. *Science*, *345*(6203), 1473–1479. https://doi.org/10.1126/science.1256328

Jore, M. M., Lundgren, M., van Duijn, E., Bultema, J. B., Westra, E. R., Waghmare, S. P., Wiedenheft, B., Pul, Ü., Wurm, R., Wagner, R., Beijer, M. R., Barendregt, A., Zhou, K., Snijders, A. P. L., Dickman, M. J., Doudna, J. A., Boekema, E. J., Heck, A. J. R., van der Oost, J., & Brouns, S. J. J. (2011). Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nature Structural & Molecular Biology*, *18*(5), 529–536. https://doi.org/10.1038/nsmb.2019

Kamruzzaman, M., & Iredell, J. R. (2020). CRISPR-Cas System in Antibiotic Resistance Plasmids in Klebsiella pneumoniae. *Frontiers in Microbiology*, *10*. https://doi.org/10.3389/fmicb.2019.02934

Kellner, M. J., Koob, J. G., Gootenberg, J. S., Abudayyeh, O. O., & Zhang, F. (2019). SHERLOCK: Nucleic acid detection with CRISPR nucleases. *Nature Protocols*, *14*(10), 2986–3012. https://doi.org/10.1038/s41596-019-0210-2

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. (2018). Cas4 Facilitates PAM-Compatible Spacer Selection during CRISPR Adaptation. *Cell Reports*, *22*(13), 3377–3384. https://doi.org/10.1016/j.celrep.2018.02.103

Klompe, S. E., Vo, P. L. H., Halpin-Healy, T. S., & Sternberg, S. H. (2019). Transposon-encoded CRISPR–Cas systems direct RNA-guided DNA integration. *Nature*. https://doi.org/10.1038/s41586-019-1323-z

Koonin, E. V., & Makarova, K. S. (2017). Mobile Genetic Elements and Evolution of CRISPR-Cas Systems: All the Way There and Back. *Genome Biology and Evolution*, *9*(10), 2812–2825. https://doi.org/10.1093/gbe/evx192

Koonin, E. V., & Makarova, K. S. (2019). Origins and evolution of CRISPR-Cas systems. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*(1772). https://doi.org/10.1098/rstb.2018.0087

Labrie, S. J., Samson, J. E., & Moineau, S. (2010). Bacteriophage resistance mechanisms. *Nature Reviews Microbiology*, *8*. https://doi.org/10.1038/nrmicro2315

Lee, H., Zhou, Y., Taylor, D. W., & Sashital, D. G. (2018). Cas4-Dependent Prespacer Processing Ensures High-Fidelity Programming of CRISPR Arrays. *Molecular Cell*, *70*(1), 48-59.e5. https://doi.org/10.1016/j.molcel.2018.03.003

Lewis, L. K., Jenkins, M. E., & Mount, D. W. (1992). Isolation of DNA damage-inducible promoters in Escherichia coli: Regulation of polB (dinA), dinG, and dinH by LexA repressor. *Journal of Bacteriology*, *174*(10), 3377–3385.

Lintner, N. G., Kerou, M., Brumfield, S. K., Graham, S., Liu, H., Naismith, J. H., Sdano, M., Peng, N., She, Q., Copié, V., Young, M. J., White, M. F., & Lawrence, C. M. (2011). Structural and Functional Characterization of an Archaeal Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-associated Complex for Antiviral Defense (CASCADE). *Journal of Biological Chemistry*, *286*(24), 21643–21656. https://doi.org/10.1074/jbc.M111.238485

Ma, H., Marti-Gutierrez, N., Park, S.-W., Wu, J., Lee, Y., Suzuki, K., Koski, A., Ji, D., Hayama, T., Ahmed, R., Darby, H., Van Dyken, C., Li, Y., Kang, E., Park, A.-R., Kim, D., Kim, S.-T., Gong, J., Gu, Y., … Mitalipov, S. (2017). Correction of a pathogenic gene mutation in human embryos. *Nature*, *548*(7668), 413–419. https://doi.org/10.1038/nature23305

Makarova, K. S., Aravind, L., Wolf, Y. I., & Koonin, E. V. (2011). Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. *Biology Direct*, *6*(1), 38. https://doi.org/10.1186/1745-6150-6-38

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., … Koonin, E. V. (2015). An updated evolutionary classification of CRISPR–Cas systems. *Nature Reviews Microbiology*, *13*(11), 722–736. https://doi.org/10.1038/nrmicro3569

Makarova, K. S., Wolf, Y. I., Iranzo, J., Shmakov, S. A., Alkhnbashi, O. S., Brouns, S. J. J., Charpentier, E., Cheng, D., Haft, D. H., Horvath, P., Moineau, S., Mojica, F. J. M., Scott, D., Shah, S. A., Siksnys, V., Terns, M. P., Venclovas, Č., White, M. F., Yakunin, A. F., … Koonin, E. V. (2020). Evolutionary classification of CRISPR–Cas systems: A burst of class 2 and derived variants. *Nature Reviews Microbiology*. https://doi.org/10.1038/s41579-019-0299-x

Makarova, K. S., Wolf, Y. I., & Koonin, E. V. (2018). Classification and Nomenclature of CRISPR-Cas Systems: Where from Here? *The CRISPR Journal*, *1*(5), 325–336. https://doi.org/10.1089/crispr.2018.0033

McKitterick, A. C., LeGault, K. N., Angermeyer, A., Alam, M., & Seed, K. D. (2019). Competition between mobile genetic elements drives optimization of a phage-encoded CRISPR-Cas system: Insights from a natural arms race. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 374(1772), 20180089. https://doi.org/10.1098/rstb.2018.0089

McRobbie, A.-M., Meyer, B., Rouillon, C., Petrovic-Stojanovska, B., Liu, H., & White, M. F. (2012). *Staphylococcus aureus* DinG, a helicase that has evolved into a nuclease. *Biochemical Journal*, *442*(1), 77–84. https://doi.org/10.1042/BJ20111903

Mojica, F. J. M., Diez-Villasenor, C., Garcia-Martinez, J., & Soria, E. (2005). Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements. *Journal of Molecular Evolution*, *60*(2), 174–182. https://doi.org/10.1007/s00239-004-0046-3

Mulepati, S., & Bailey, S. (2011). Structural and Biochemical Analysis of Nuclease Domain of Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-associated Protein 3 (Cas3). *Journal of Biological Chemistry*, *286*(36), 31896–31903. https://doi.org/10.1074/jbc.M111.270017

Mulepati, S., Héroux, A., & Bailey, S. (2014). Crystal structure of a CRISPR RNA–guided surveillance complex bound to a ssDNA target. *Science*, *345*(6203), 1479–1484. https://doi.org/10.1126/science.1256996

Naser, I. B., Hoque, M. M., Nahid, M. A., Tareq, T. M., Rocky, M. K., & Faruque, S. M. (2017). Analysis of the CRISPR-Cas system in bacteriophages active on epidemic strains of Vibrio cholerae in Bangladesh. *Scientific Reports*, 7(1), 14880. https://doi.org/10.1038/s41598-017-14839-2

Nuñez, J. K., Kranzusch, P. J., Noeske, J., Wright, A. V., Davies, C. W., & Doudna, J. A. (2014). Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. *Nature Structural & Molecular Biology*, *21*(6), 528–534. https://doi.org/10.1038/nsmb.2820

Osawa, T., Inanaga, H., Sato, C., & Numata, T. (2015). Crystal Structure of the CRISPR-Cas RNA Silencing Cmr Complex Bound to a Target Analog. *Molecular Cell*, *58*(3), 418–430. https://doi.org/10.1016/j.molcel.2015.03.018

Özcan, A., Pausch, P., Linden, A., Wulf, A., Schühle, K., Heider, J., Urlaub, H., Heimerl, T., Bange, G., & Randau, L. (2018). Type IV CRISPR RNA processing and effector complex formation in Aromatoleum aromaticum. *Nature Microbiology*. https://doi.org/10.1038/s41564-018-0274-8

Peters, J. E., Makarova, K. S., Shmakov, S., & Koonin, E. V. (2017). Recruitment of CRISPR-Cas systems by Tn7-like transposons. *Proceedings of the National Academy of Sciences of the United States of America*, 114(35), E7358–E7366. https://doi.org/10.1073/pnas.1709035114

Pinilla-Redondo, R., Mayo-Muñoz, D., Russel, J., Garrett, R. A., Randau, L., Sørensen, S. J., & Shah, S. A. (2019). Type IV CRISPR-Cas systems are highly diverse and involved in competition between plasmids. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkz1197

Pourcel, C., Salvignol, G., & Vergnaud, G. (2005). CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology*, *151*(3), 653–663. https://doi.org/10.1099/mic.0.27437-0

Rollie, C., Schneider, S., Brinkmann, A. S., Bolt, E. L., & White, M. F. (2015). Intrinsic sequence specificity of the Cas1 integrase directs new spacer acquisition. *ELife*, *4*. https://doi.org/10.7554/eLife.08716

Rostøl, J. T., & Marraffini, L. (2019). (Ph)ighting Phages: How Bacteria Resist Their Parasites. *Cell Host & Microbe*, *25*(2), 184–194. https://doi.org/10.1016/j.chom.2019.01.009

Seed, K. D., Lazinski, D. W., Calderwood, S. B., & Camilli, A. (2013). A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. *Nature*, 494(7438), 489–491. https://doi.org/10.1038/nature11927

Shipman, S. L., Nivala, J., Macklis, J. D., & Church, G. M. (2017). CRISPR–Cas encoding of a digital movie into the genomes of a population of living bacteria. *Nature*, *547*(7663), 345–349. https://doi.org/10.1038/nature23017

Shmakov, S. A., Makarova, K. S., Wolf, Y. I., Severinov, K. V., & Koonin, E. V. (2018). Systematic prediction of genes functionally linked to CRISPR-Cas systems by gene neighborhood analysis. *Proceedings of the National Academy of Sciences*, 115(23), E5307–E5316. https://doi.org/10.1073/pnas.1803440115

Sinkunas, T., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., & Siksnys, V. (2011). Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system: Cas3 nuclease/helicase. *The EMBO Journal*, *30*(7), 1335–1342. https://doi.org/10.1038/emboj.2011.41

Staals, R. H. J., Agari, Y., Maki-Yonekura, S., Zhu, Y., Taylor, D. W., van Duijn, E., Barendregt, A., Vlot, M., Koehorst, J. J., Sakamoto, K., Masuda, A., Dohmae, N., Schaap, P. J., Doudna, J. A., Heck, A. J. R., Yonekura, K., van der Oost, J., & Shinkai, A. (2013). Structure and Activity of the RNA-Targeting Type III-B CRISPR-Cas Complex of Thermus thermophilus. *Molecular Cell*, *52*(1), 135–145. https://doi.org/10.1016/j.molcel.2013.09.013

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., … van der Oost, J. (2014). RNA targeting by the type III-A CRISPR-Cas Csm complex of Thermus thermophilus. *Molecular Cell*, *56*(4), 518–530. https://doi.org/10.1016/j.molcel.2014.10.005

Taylor, H. N., Laderman, E., Armbrust, M., Hallmark, T., Keiser, D., Bondy-Denomy, J., & Jackson, R. N. (2021). Positioning Diverse Type IV Structures and Functions Within Class 1 CRISPR-Cas Systems. *Frontiers in Microbiology*, 12. https://doi.org/10.3389/fmicb.2021.671522

Taylor, H. N., Warner, E. E., Armbrust, M. J., Crowley, V. M., Olsen, K. J., & Jackson, R. N. (2019). Structural basis of Type IV CRISPR RNA biogenesis by a Cas6 endoribonuclease. *RNA Biology*. https://doi.org/10.1080/15476286.2019.1634965

Thakur, R. S., Desingu, A., Basavaraju, S., Subramanya, S., Rao, D. N., & Nagaraju, G. (2014). Mycobacterium tuberculosis DinG Is a Structure-specific Helicase That Unwinds G4 DNA Implications for Targeting G4 DNA as a Novel Therapeutic Approach. *Journal of Biological Chemistry*, *289*(36), 25112–25136. https://doi.org/10.1074/jbc.M114.563569

Voloshin, O. N., & Camerini-Otero, R. D. (2007). The DinG Protein from *Escherichia coli* Is a Structure-specific Helicase. *Journal of Biological Chemistry*, *282*(25), 18437–18447. https://doi.org/10.1074/jbc.M700376200

Voloshin, O. N., Vanevski, F., Khil, P. P., & Camerini-Otero, R. D. (2003). Characterization of the DNA Damage-inducible Helicase DinG from *Escherichia coli*. *Journal of Biological Chemistry*, *278*(30), 28284–28293. https://doi.org/10.1074/jbc.M301188200

Wang, J., Li, J., Zhao, H., Sheng, G., Wang, M., Yin, M., & Wang, Y. (2015). Structural and Mechanistic Basis of PAM-Dependent Spacer Acquisition in CRISPR-Cas Systems. *Cell*, *163*(4), 840–853. https://doi.org/10.1016/j.cell.2015.10.008

Wiedenheft, B., Zhou, K., Jinek, M., Coyle, S. M., Ma, W., & Doudna, J. A. (2009). Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense. *Structure*, *17*(6), 904–912. https://doi.org/10.1016/j.str.2009.03.019

Xiao, Y., Ng, S., Hyun Nam, K., & Ke, A. (2017). How type II CRISPR–Cas establish immunity through Cas1–Cas2-mediated spacer integration. *Nature*, *550*(7674), 137–141. https://doi.org/10.1038/nature24020

Yan, W. X., Hunnewell, P., Alfonse, L. E., Carte, J. M., Keston-Smith, E., Sothiselvam, S., Garrity, A. J., Chong, S., Makarova, K. S., Koonin, E. V., Cheng, D. R., & Scott, D. A. (2019). Functionally diverse type V CRISPR-Cas systems. *Science*, *363*(6422), 88–91. https://doi.org/10.1126/science.aav7271

Zhou, Y., Bravo, J. P. K., Taylor, H. N., Steens, J. A., Jackson, R. N., Staals, R. H. J., & Taylor, D. W. (2021). Structure of a type IV CRISPR-Cas ribonucleoprotein complex. *iScience*, *24*(3). https://doi.org/10.1016/j.isci.2021.102201

CHAPTER 2

STRUCTURE OF A TYPE IV CRISPR-CAS RIBONUCLEOPROTEIN COMPLEX[1]

ABSTRACT

We reveal the cryo-electron microscopy structure of a type IV-B CRISPR ribonucleoprotein (RNP) complex (Csf) at 3.9-Å resolution. The complex best resembles the type III-A CRISPR Csm effector complex, consisting of a Cas7-like (Csf2) filament intertwined with a small subunit (Cas11) filament, but the complex lacks subunits for RNA processing and target DNA cleavage. Surprisingly, instead of assembling around a CRISPR-derived RNA (crRNA), the complex assembles upon heterogeneous RNA of a regular length arranged in a pseudo-A-form configuration. These findings provide a high-resolution glimpse into the assembly and function of enigmatic type IV CRISPR systems, expanding our understanding of class I CRISPR-Cas system architecture, and suggesting a function for type IV-B RNPs that may be distinct from other class 1 CRISPR-associated systems.

INTRODUCTION

Bacteria and archaea employ CRISPR (Clustered Regularly Interspaced Short Palindromic Repeat)-Cas (CRISPR-associated) systems for adaptive immunity against

---

[1] Co-authored by: Yi Zhou*, Jack P. K. Bravo*, Hannah N. Taylor*, Jurre A. Steens, Ryan N. Jackson, Raymond H. J. Staals, David W. Taylor.
Published: *iScience* (2021).
Final publication can be accessed at https://doi.org/10.1016/j.isci.2021.102201
*These authors should be considered co-1st authors. HNT designed experiments, purified protein complexes, contributed to RNA sequencing and model building, created figures, and co-wrote the manuscript.

phages, plasmids and other mobile-genetic elements (MGEs) (Makarova et al., 2020). In the multi-subunit class 1 systems, the CRISPR locus is transcribed and processed into small crRNA guides (CRISPR-derived RNA), around which several Cas proteins assemble to form large ribonucleoprotein (RNP) complexes that facilitate RNA-guided surveillance and degradation of complementary targets (Hille et al., 2018). While a myriad of structures have been determined for most types of CRISPR RNA-guided complexes (types I (Jackson et al., 2014; Mulepati, Héroux and Bailey, 2014; Chowdhury et al., 2017; Xiao et al., 2018; Rollins et al., 2019), II (Jinek et al., 2014; Jiang et al., 2016; Zhu et al., 2019), III (Taylor et al., 2015; Jia et al., 2019; You et al., 2019; Sofos et al., 2020), V (Stella, Alcón and Montoya, 2017; Liu et al., 2019; Zhang et al., 2020; Li et al., 2021; Takeda et al., 2021) and IV (Yan et al., 2018; Slaymaker et al., 2019; Meeske et al., 2020)), the RNP complexes of the highly diverse type IV CRISPR systems have largely remained structurally uncharacterized (Crowley et al., 2019; Faure et al., 2019; Özcan et al., 2019; Taylor et al., 2019; Makarova et al., 2020).

Type IV CRISPR systems primarily occur within plasmid-like elements, lack genes encoding adaptation modules (*cas1*, *cas2* and *cas4*), and are classified into three distinct subtypes (IV-A, IV-B, IV-C) (Özcan et al., 2019; Pinilla-Redondo et al., 2019; Makarova et al., 2020). All type IV systems contain genes that encode for Csf2 (Cas7), Csf3 (Cas5), and Csf1 (large subunit) proteins, which assemble around an RNA to form a multi-subunit complex (Özcan et al., 2019; Pinilla-Redondo et al., 2019; Makarova et al., 2020). However, subtype-specific signature genes suggest distinct subtype functions. Type IV-A systems encode a DinG helicase shown to be essential for type IV-A mediated plasmid clearance (Crowley et al., 2019), Type IV-B systems contain the ancillary gene

cysH of the phosphoadenosine phosphosulfate reductase family, and type IV-C systems encode a large subunit that contains an HD-nuclease domain (Özcan et al., 2019; Pinilla-Redondo et al., 2019; Makarova et al., 2020) (**Figure 2-1**). Additionally, type IV-A systems encode a CRISPR array and crRNA endonuclease, while type IV-B and type IV-C systems generally do not. It has been proposed that systems lacking a CRISPR array form complexes on crRNAs generated from other CRISPR systems (e.g. type I or type III), but this hypothesis has yet to be explored experimentally. Interestingly, the two subtypes that do not contain a CRISPR array (type IV-B and type IV-C) encode a small α-helical protein (Cas11) predicted to form part of the multi-subunit complex. Thus, there are two distinct type IV multi-subunit complexes, one that contains the small Cas11 subunit (types IV-B and IV-C), and another (type IV-A) that does not contain Cas11 but contains a crRNA derived from a type IV-A CRISPR array and processed by a type IV Cas6 endonuclease. To better understand the function of type IV CRISPR systems as well as their subtype-specific similarities and differences, we isolated a type IV-B complex, analyzed the sequence of the small RNAs bound within the complex, and determined a near-atomic resolution structure.

RESULTS

The Type IV-B RNP Assembles on Non-Specific RNAs

The *Mycobacterium sp.* JS623 type IV-B CRISPR operon is encoded within a megaplasmid and lacks both a pre-crRNA maturase (Cas6/Csf5 (Taylor et al., 2019) and a CRISPR array, containing only *csf1* (Cas8-like large subunit), *cas11* (small subunit), *csf2* (Cas7) and *csf3* (Cas5) genes (**Figure 2-2A**). Interestingly, *M. sp.* JS623 also harbors

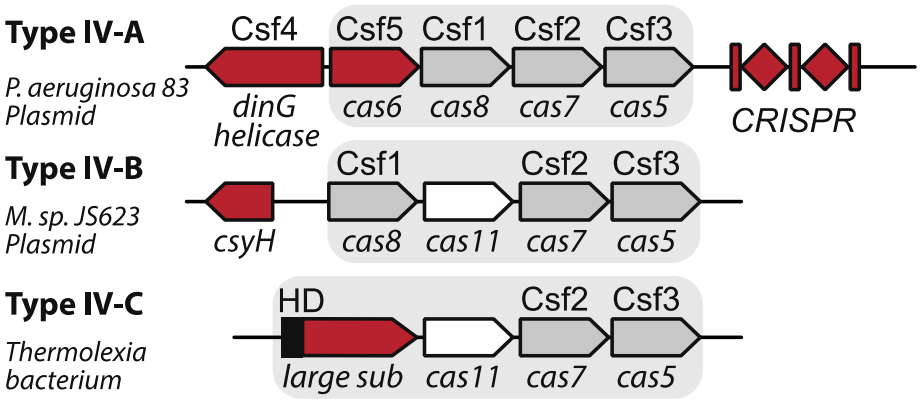

Figure 2-1. Classification of type IV subytypes. Related to Figure 2-2. Schematic of the three distinct type IV subtypes defined in (Makarova *et al.*, 2020). Genetic features (genes and CRISPRs) found primarily in a single subtype are colored red. Gray rectangles indicate genes expected to encode proteins that form RNP complexes. The *cas11* gene is colored white to highlight that it is found within the subtypes lacking a CRISPR (subtypes IV-B and IV-C). The HD nuclease domain of the cas8-like subunit of IV-C is colored black.

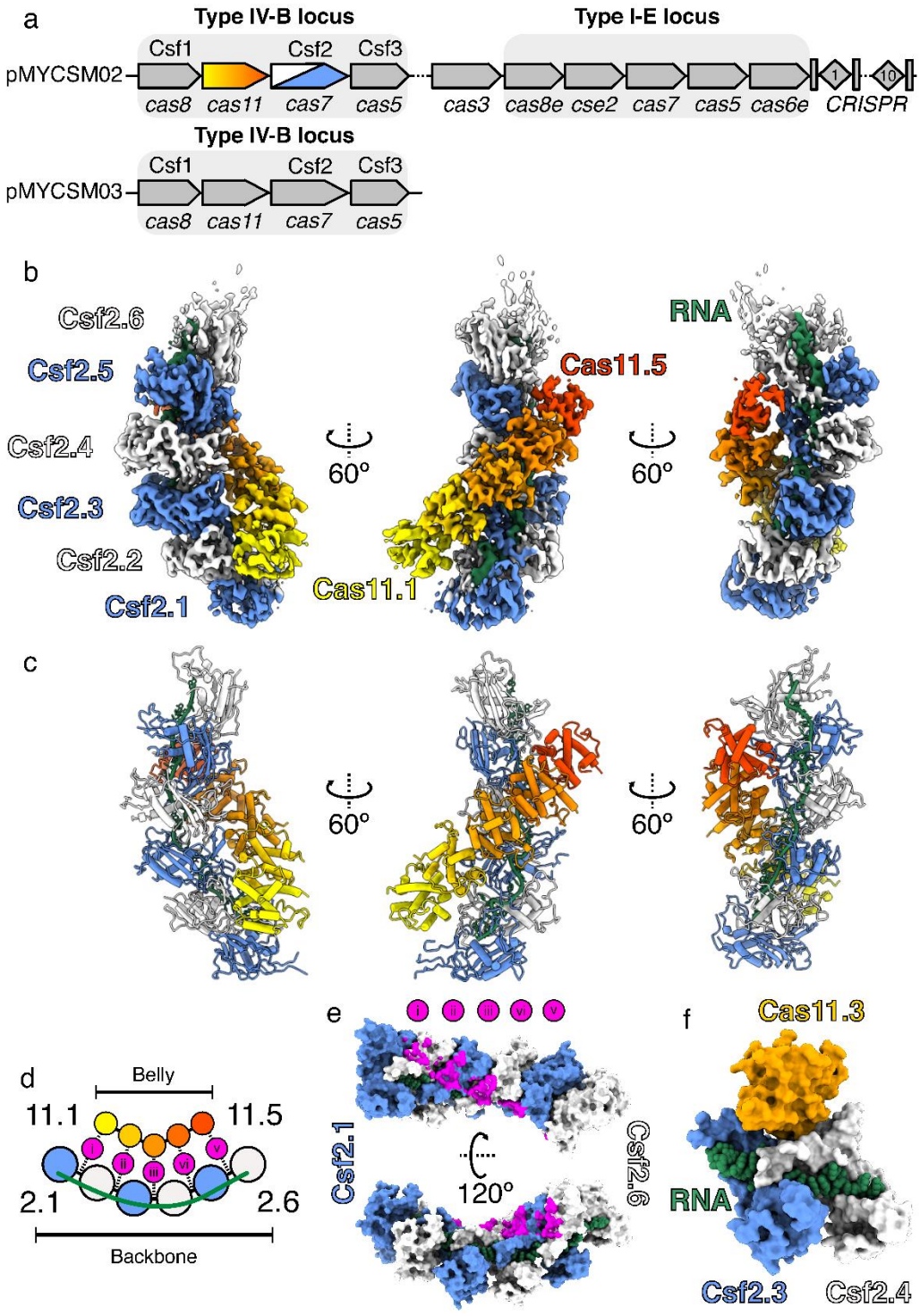

Figure 2-2. Structure of type IV-B CRISPR complex. See also Figures 2-1, 2-3, 2-4, 2-5 and Table 2-1. (A) *M. sp*. JS623 plasmid-encoded CRISPR operons. Top: Type IV-B and I-E CRISPR loci present on pMCYCM02 megaplasmid. Bottom: Additional type IV-B locus encoded by pMCYCM03 megaplasmid. Genes predicted to encode RNP complex subunits are indicated with a gray rectangle. **(B)** 3.9 Å-resolution cryo-EM reconstruction

of type IV-B CRISPR complex. Cas7 subunits are colored blue and white, and five Cas11 subunits are colored as a yellow-orange-red gradient. Csf-bound RNA is green. **(C)** Refined model for the Csf effector complex derived from the cryo-EM maps shown in B. **(D)** Schematic of Cas7-Cas11 interactions. Five Csf2-Cas11 interactions occur in this complex (labelled i – v). (**E)** Positions of Cas11 contacts on Csf2 backbone, colored magenta as shown in panel D**.** Cas11 sits upon the Csf2-Csf2 interface. **(F)** Cas11 binds at the interface with buried surface area of 505 $Å^2$ (150 $Å^2$ and 355 $Å^2$ with Csf2.3 and Csf2.4, respectively). Cas11 is completely occluded from bound RNA. Csf2 subunits are intimately connected (1021 $Å^2$) and make a network of contacts with bound RNA (~1200 $Å^2$ buried surface area per Csf2 subunit).

a type I-E system (with an associated CRISPR array) on the same megaplasmid, and another type IV-B operon encoded on a different megaplasmid (**Figure 2-2A**), suggesting that type IV-B complexes may assemble on crRNAs encoded and processed by other CRISPR systems. However, the structure and function of such hybrid complexes are unknown.

To gain mechanistic insights into the type IV-B system, we transformed E. coli BL21 cells with an expression plasmid encoding the *M. sp.* JS623 type IV-B Cas proteins, and the *M. sp.* JS623 type I-E Cas6 and associated CRISPR array (**Figure 2-3A**). Using strep-tag affinity, size exclusion chromatography, and subsequent negative stain we observed filamentous ribonucleoprotein (RNP) complexes that eluted close to the void volume and a smaller, discrete, RNA-containing species reminiscent of class 1 multi-subunit crRNA-guided complexes (**Figure 2-3**) (Makarova, Zhang and Koonin, 2017). While this latter fraction contained all four Csf subunits, Csf2 and Cas11 were the most abundant (**Figure 2-3**). Despite the appearance of a uniform band length of ~55-60

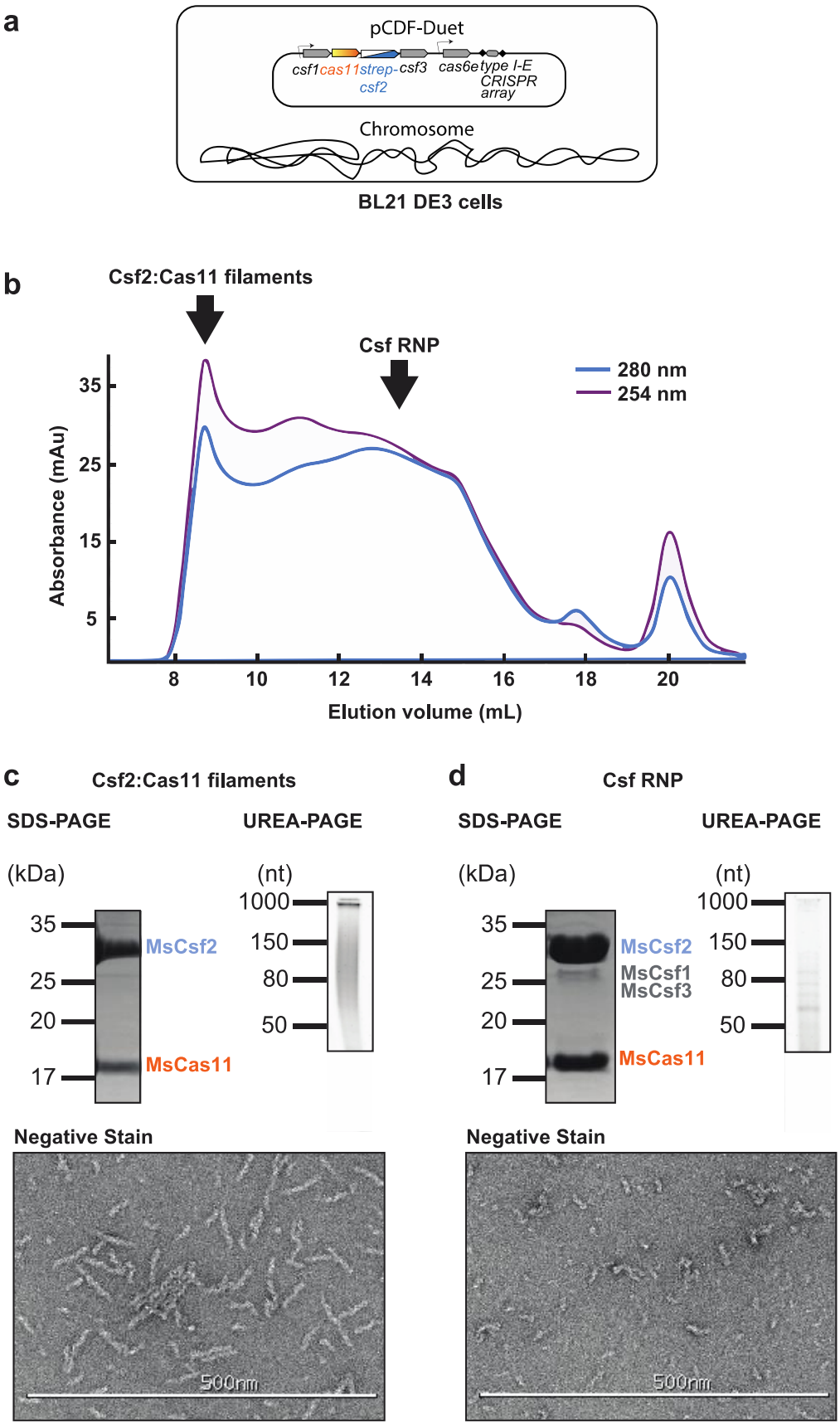
a
pCDF-Duet
csf1
cas11
strep-csf2
csf3
cas6e
type I-E CRISPR array
Chromosome
BL21 DE3 cells
b
Csf2:Cas11 filaments
Csf RNP
280 nm
254 nm
Absorbance (mAu)
35
25
15
5
8
10
12
14
16
18
20
Elution volume (mL)
c
Csf2:Cas11 filaments
SDS-PAGE
UREA-PAGE
(kDa)
(nt)
35
25
20
17
MsCsf2
MsCas11
1000
150
80
50
Negative Stain
500nm
d
Csf RNP
SDS-PAGE
UREA-PAGE
(kDa)
(nt)
35
25
20
17
MsCsf2
MsCsf1
MsCsf3
MsCas11
1000
150
80
50
Negative Stain
500nm

Figure 2-3. Purification of the *M. sp JS623* RNP complex by affinity chromatography (N-Strep-MsCsf2) and size exclusion chromatography. Related to Figure 2-2. (A) Diagram of plasmid used to express the type IV-B complex in BL21 DE3 cells. **(B)** SEC chromatogram highlighting peaks corresponding to Csf2:Cas11 filaments and the RNP complex. **(C)** SDS-PAGE, UREA-PAGE, and negative stain data indicating the presence of Csf2:Cas11 bound to long RNAs to create filamentous structures. **(D)** SDS-PAGE, UREA-PAGE, and negative stain data indicating the presence of Csf1, Csf2, Csf3, and Cas11 bound to short, distinct RNAs to create an RNP complex.

nucleotides on denaturing PAGE (**Figure 2-3D, 2-4A**), RNAseq analysis revealed bound RNAs were heterogeneous in sequence identity. Few RNAs were derived from the plasmid-encoded CRISPR array, while the majority of Csf-bound RNAs originated from the expression plasmid (63%). (**Figure 2-4B, C**). To exclude the possibility that this was due to low expression of the CRISPR array and/or lack of crRNA processing by Cas6, we repeated this analysis and compared it to an RNA-seq analysis of the total cellular population of RNAs (total RNA) extracted from the same host (**Figure 2-4D**). These results showed that the CRISPR array was indeed expressed and processed by Cas6, resulting in mature crRNAs with a typical eight nucleotide 5’ handle (a characteristic for Cas6-mediated cleavages in the repeats). However, the mature crRNAs were not enriched in the RNAs isolated from type IV RNPs and were in low abundance (~0.12% of all reads). The apparent lack of sequence specific assembly of the Csf complex on mostly non-crRNAs is different from other CRISPR-Cas systems (Makarova, Zhang and

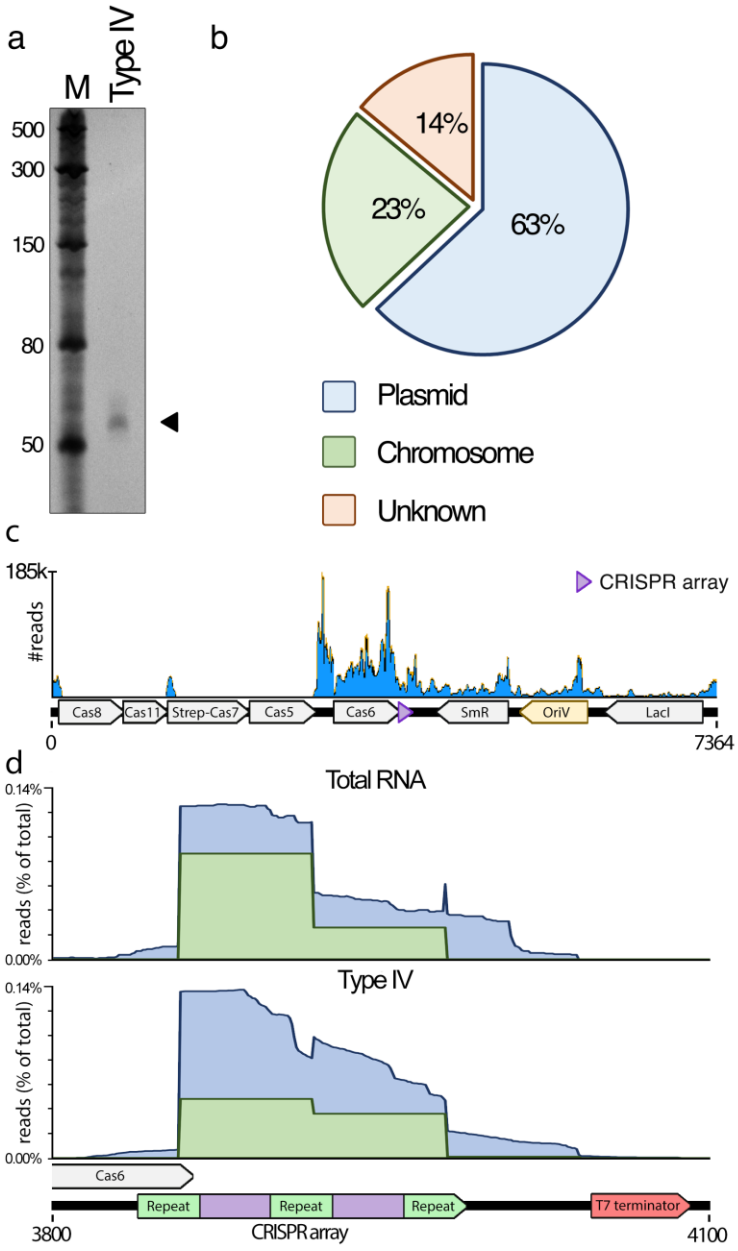

Figure 2-4. RNA sequencing on co-purifying nucleotides with type IV complex. Related to Figure 2-2. (A) UREA-PAGE gel showing nucleic acids co-purifying with the type IV-B complex. The triangle indicates the ~60nt band purified from isolated type IV-B complexes that was used for RNA sequencing analysis. **(B)** Percentage of reads mapping to either to expression plasmid, chromosome, or of unknown origin. **(C)** Distribution of reads mapping on the expression plasmid. **(D)** Comparison of repeat-containing RNAs from the total cellular RNA population ("total RNA") and type IV RNP-associated ("type IV") RNAs mapped on the CRISPR array of the expression plasmid (in blue). Reads indicated in green represent the 61-nt RNAs with a perfect repeat-derived 8-nt 5' handle, reminiscent of Cas6-mediated cleavages in the repeats.

Koonin, 2017), and might be indicative of a role of type IV CRISPR-Cas systems in functions other than antiviral defense.

The Architecture of the Type IV-B RNP Resembles Type III Effector Complexes

To compare the type IV-B RNP complex to the complexes of other class 1 systems we next determined a cryo-EM structure of the IV-B Csf complex at 3.9 Å resolution (**Figure 2-2B, Figure 2-5, Table 2-1**), allowing us to build an atomic model of the complex de novo (**Figure 2-2C**). The type IV-B complex resembles a sea cucumber, with six Csf2 (Cas7-like) subunits forming a helical 'backbone,' and five Cas11 subunits comprising a helical 'belly'. Each Cas11 subunit sits upon a Csf2-Csf2 interface (**Figure 2-2D, E & F**). The "α-helix bundle" topology of Cas11 (**Figure 2-6C**) and presence of a contiguous positively-charged patch running along the length of the minor filament (**Figure 2-7**) are typical of Cas11 small subunits in class 1 CRISPR systems (Xiao et al., 2017; Rollins et al., 2019), although the arrangement of helices within type IV Cas11 is distinct from type I and type III small subunits.

Like other class 1 Cas7 proteins, Csf2 adopts a hand-shaped structure with fingers, a palm, and a thumb. The palm makes extensive contacts with the bound RNA (buried surface area of ~1200 Å2 per Csf2 subunit) (**Figure 2-8A**), while the thumbs of neighbouring Csf2 subunits protrude into the center of the palm, inducing a kink in the RNA backbone and a 'flipped' base at six nucleotide intervals (typical of other class 1 complexes (Jackson et al., 2014; Taylor et al., 2015)). Using our atomic model of Csf2 we searched for structural homologues. Csf2 had significant similarity to the type III-A CRISPR Csm3 (i.e. Cas7) subunit (Dali Z-score of 14.1), despite a sequence identity of

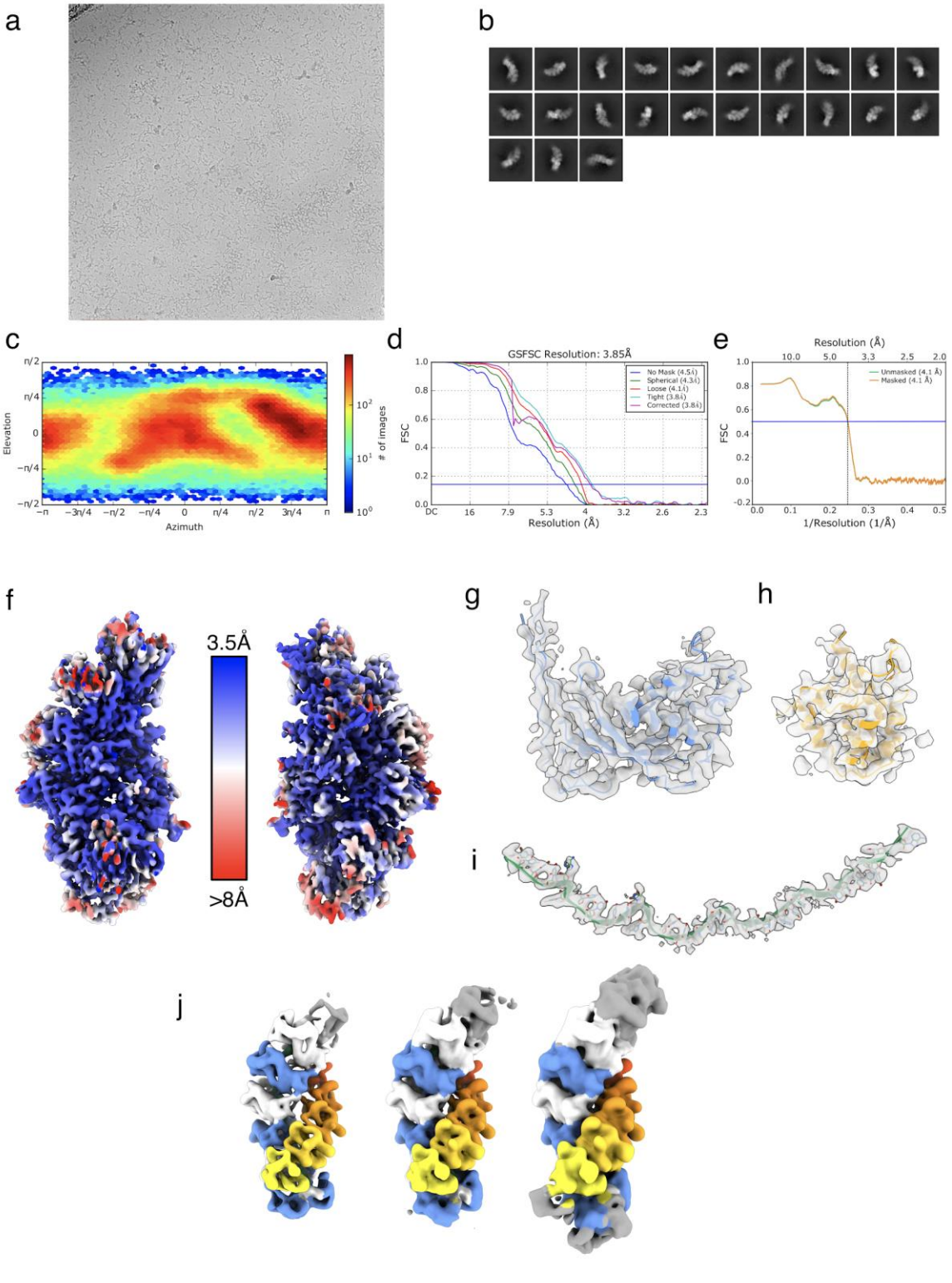
a
b
c
Elevation
Azimuth
# of images
d
GSFSC Resolution: 3.85Å
No Mask (4.5Å)
Spherical (4.3Å)
Loose (4.1Å)
Tight (3.8Å)
Corrected (3.8Å)
FSC
Resolution (Å)
e
Resolution (Å)
Unmasked (4.1 Å)
Masked (4.1 Å)
FSC
1/Resolution (1/Å)
f
3.5Å
>8Å
g
h
i
j
0.0393
0.0178
0.00396

Figure 2-5. Cryo-EM analysis of Csf. Related to Figure 2-2. (A) Representative micrograph of Csf complex particles. **(B)** 2D class averages of discrete Csf oligomers used for 3D reconstruction. **(C)** Euler angular distribution of particles contributing to final 3D reconstruction. **(D)** Fourier shell correlation (FSC) of final 3D reconstruction, with a global resolution of 3.9 Å at the 0.143 threshold. **(E)** Map-to-model FSC, with a model resolution of 4.1 Å at the 0.5 threshold. **(F)** Map of Csf complex colored by local resolution. **(G) – (I),** Representative atomic models and corresponding cryo-EM densities for Csf2 (g), Cas11 (h) and RNA (i). **(J)** Low-pass filtered (8 Å) map of Csf complex at three different isosurface thresholds. At lower thresholds (0.0178 & 0.00396), additional density appears at the top of the complex. This density may correspond to Csf1 Csf3 subunits, however the low resolution of this region makes unambiguous subunit assignment impossible.

Data collection and processing	
Magnification	22,500x
Voltage (kV)	300
Electron exposure (e^-/Å)	40
Defocus range (μM)	-1.5 to -3.0
Symmetry imposed	C1
Final particle images	296,319
Map resolution (Å)	3.9
FSC threshold	0.143
Map resolution range (Å)	3.5 to > 8
Refinement	
Model resolution (Å) (0.143 FSC)	4.1
Map sharpening *B* factor (Å^2)	130.2
Model composition	
Nonhydrogen atoms	16636
Residues (Protein/RNA)	2083/31
B factors (Å^2) (min/max/mean)	
Protein	19.75/206.87/81.59
RNA	34.71/125.88/56.91
r.m.s. deviations	
Bond lengths (Å)/bond angle (º)	0.005/1.112
Validation	
MolProbity score	1.9
Clashscore	4.72
Poor rotamers (%)	0
Ramachandran plot	
Favoured/Allowed/Disallowed (%)	84.99/15.01/0

Table 2-1. Cryo-EM data collection and processing parameters. Related to Figure 2-2.

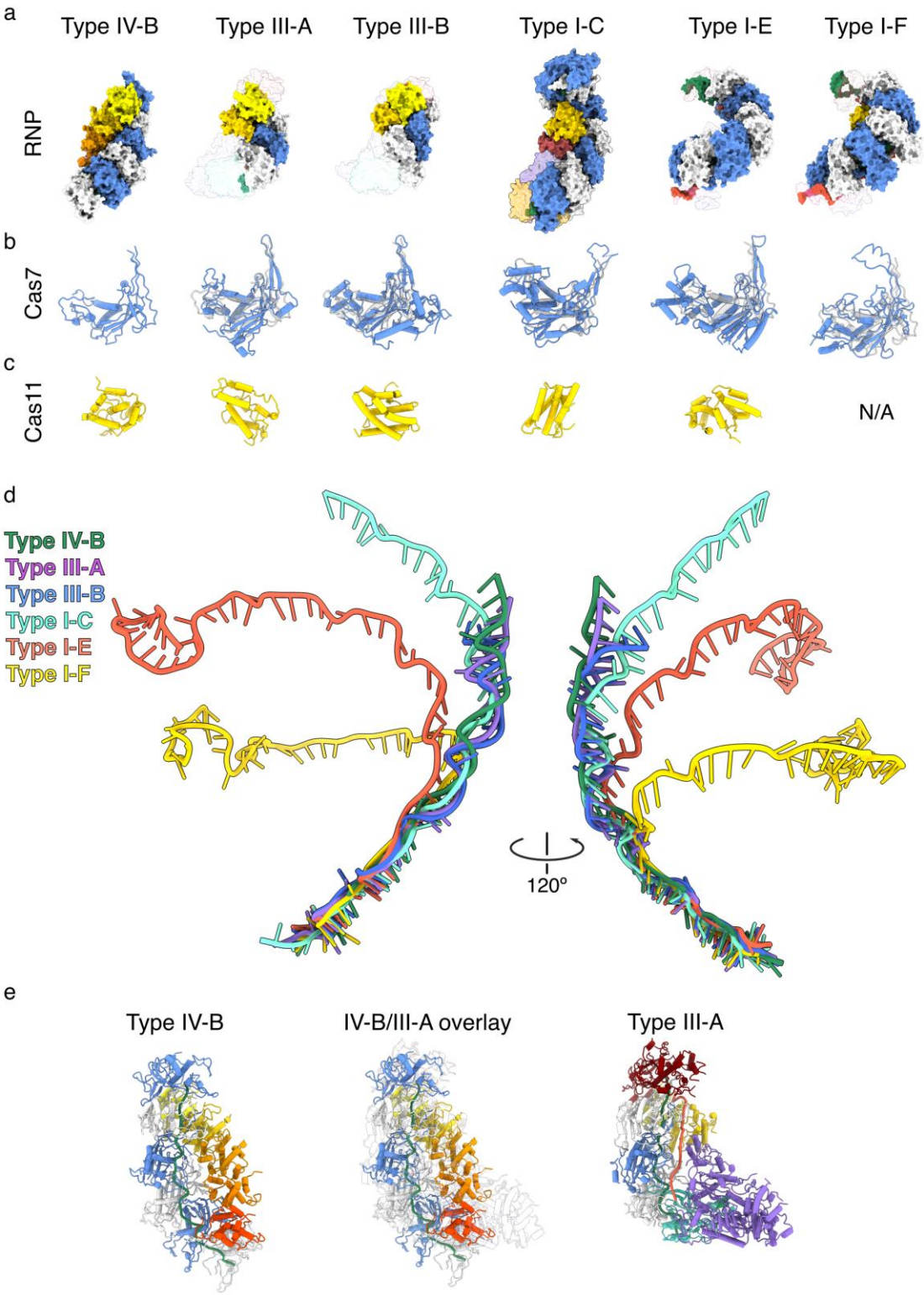

Figure 2-6. Comparison to Cascade complexes. Related to Figure 2-8. (A) Comparison of Type IV-B Csf with other effector/Cascade complexes. crRNA within Type III-A (PDB 6o7i) (Jia *et al.*, 2019), III-B (PDB 3x1l) (Osawa *et al.*, 2015) & I-C (O'Brien *et al.*, 2020)

(7kha) complexes were aligned to Csf RNA (RMSD 5.9 Å, 8.8 Å, and 12.7 Å respectively). Due to the highly curved nature of type I-E (4tvx) (Jackson *et al.*, 2014) and I-F (5uz9) (Chowdhury *et al.*, 2017) crRNA, it was not possible to perform such alignment to IV-B. Instead, individual backbone subunits were aligned to corresponding Csf2 subunits, thus aligning the top of I-E or I-F with the top of IV-B (RMSD ~24.4 Å and 25.8 Å, respectively). In all complexes, non-Cas7/Cas11 subunits are shown as transparent surfaces. All Cas7/Cas11 and RNA, crRNA and TS are colored as in **Figures 2-2 & 2-8**, with the addition of the TS in light red. **(B)** Alignment of Cas7 with Csf2. Csf2 is shown as grey, transparent cartoon. RMSD is typically ~20 Å – 25 Å, although they clearly align well by eye. The high RMSD is likely due to presence of additional residues not present in Csf2. **(C)** Cas11 subunits. All Cas11 subunits are helical bundles that resemble each other. However, due to diverse Cas11 sequences these subunits align poorly (RMSD 16 Å – 18 Å)**. (D)** Alignment of (cr)RNA from available CRISPR effector complexes. Type IV-B RNA aligns more closely to type III-A and -B crRNAs, consistent with the proposed evolutionary lineage of type IV CRISPR systems emerging from a type III-like ancestor. **(E)** Overlay of type IV-B and type III-A CRISPR complexes based on RNA alignment. Type III-A displays the strongest structural homology between Cas7 subunits.

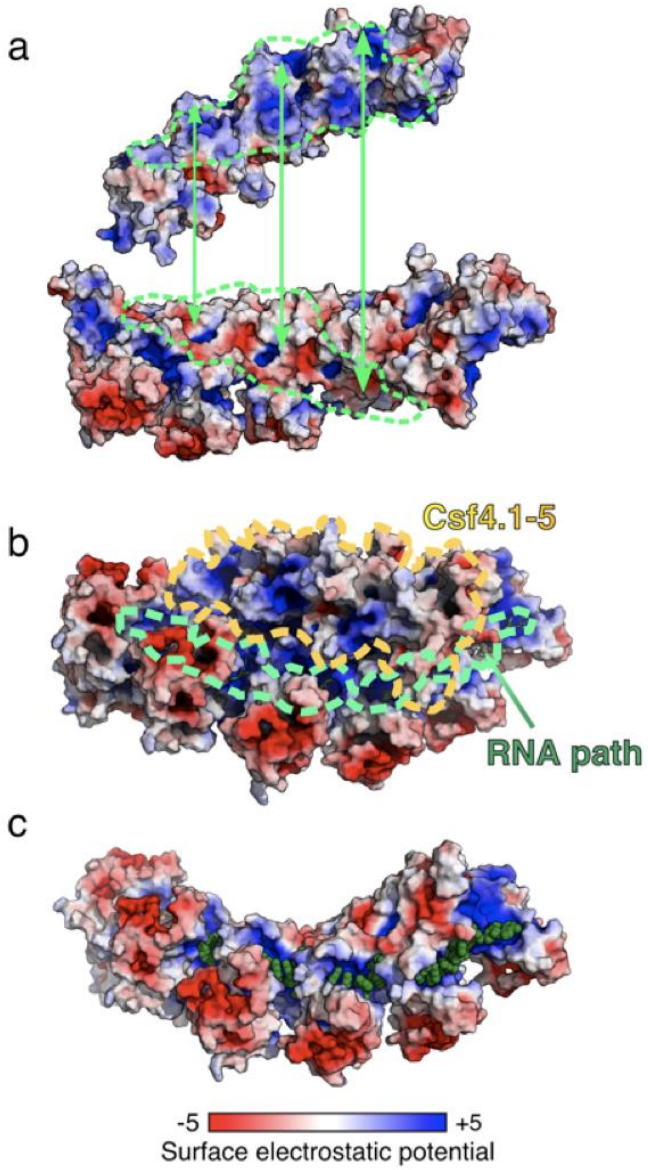

Figure 2-7. Surface electrostatics of Csf complex. Related to Figure 2-8. (A) & (B) Cas11 and Csf2 filaments. Green dashed outlines denote complementary surfaces. **(B)** Surface electrostatics of the Csf complex. Cas11 subunits and RNA path are outlined. RNA is shown as green cartoon, but it is almost completely occluded by Cas7 subunits. The overall path of the RNA bound within the Cas7 filament is outlined by green dashed lines. **(C)** Csf2 filament with RNA contacts (green spheres) shown. Cas11 minor filament is removed for clarity. RNA is bound by a contiguous positively-charged surface. The high electrostatic contribution to RNA binding by Csf2 is typical of non-specific RNA-binding proteins (Bravo *et al.*, 2018, 2020).

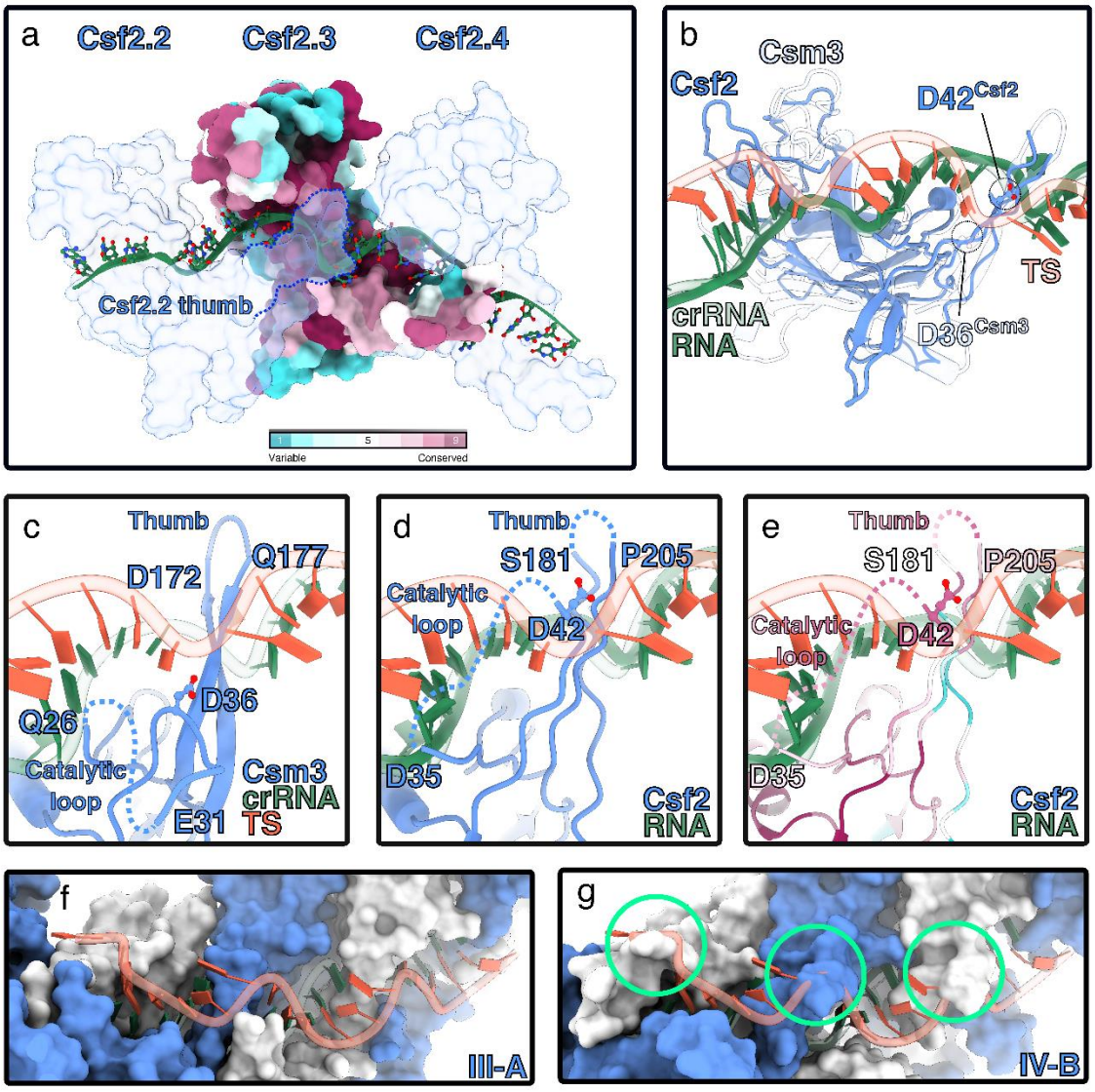

Figure 2-8. RNA-binding by type IV-B Cas7. **See also Figures 2-6, 2-7, and 2-9. (A)** RNA (green) binding site runs across the palms of Csf2 subunits. Csf2.3 is colored according to conservation. The "thumb" of the n_{-1} Csf2 (i.e. Csf2.2) protrudes into the backbone of bound RNA (solid green), inducing a kink. **(B)** Alignment of type III-A backbone subunit Csm3 (PDB 607i, transparent) with Csf2 (solid blue). Csm3 and Csf2 align with an r.m.s.d. of 2.9Å, with a Dali server Dali server Z-score 14.1. Csf2-bound RNA binds in the same conformation as crRNA (transparent green) to Csm3 (RMSD of 1.5 Å). Catalytic residue $Asp36^{Csm3}$ and putative catalytic residue $Asp42^{Csf2}$ side chains are located near the target strand (TS - transparent red), bound to the type III crRNA (transparent green). **(C)** Residues flanking unstructured catalytic loop (27-35) and apical loop of Csm3 thumb also interact with the TS. Catalytic residue D36 is shown for clarity. (**D)** Putative interactions with Csf2 and TS, based on alignments with the Csm complex.

(E) Putative interactions colored by conservation. The Csf2 thumb contains a flexible 20 residue insertion, not visible in our cryo-EM map. **(F)** Path of TS bound by type III-A Csm complex. **(G)** Putative path of TS along IV-B. Severe classes with TS and Csf2 are circled in green.

only 16%. Csf2 and Csm3 superimpose with an r.m.s.d of 2.9 Å and use equivalent interfaces to bind RNA and induce near-identical RNA backbone conformations (r.m.s.d of 1.5 Å) (**Figure 2-8A**). This supports previous bioinformatics-based hypotheses that type IV systems originated from type III-like ancestors (Özcan et al., 2019; Pinilla-Redondo et al., 2019; Makarova et al., 2020).

The type III backbone protein Csm3 cleaves the phosphodiester backbone of crRNA-bound target strand (TS) RNA at 6-nt intervals (Staals et al., 2014; Steens et al., 2021). Given that the Csm crRNA aligns almost perfectly with Csf-bound RNA, we reasoned that Csf2 might also possess RNase activity. Within our aligned structures, both the catalytic Asp36Csm3 residue and the conserved Asp42Csf2 residue are similarly positioned within an unstructured "catalytic loop" (**Figure 2-8C – E, Figure 2-9**). However, despite this similarity, structural alignment with a target-bound type III complex reveals significant steric clashes between the path of the bound nucleic acid target and the Csf2 catalytic loop (**Figure 2-8F & G**), suggesting a significant conformational rearrangement of subunits would need to occur upon target binding to place D42Csf2 in a position amenable to catalyze target RNA cleavage. Thus, additional

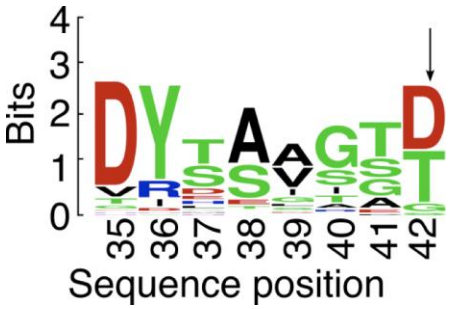

Figure 2-9. Weblogo of cleavage loop, with candidate catalytic residue (D42) denoted by arrow. Related to Figure 2-8. Multiple sequence alignment (MSA) generated using the top 100 result from a BLAST search against the Csf2 sequence from *M. sp JS623*. Output from the MSA was used to generate a sequence logo using the WebLogo server (Crooks *et al.*, 2004) and conservation score used in Figure 2-8.

substrate bound structures and in vitro functional assays are needed to more fully explore the possibility of Csf2-mediated RNase activity.

DISCUSSION

Our structure of the Csf complex provides evidence that type IV-B evolved from type III CRISPR-Cas systems, but lost its CRISPR and Cas6-based crRNA processing activity due to functional re-specialization. Although the *M. sp.* JS623 type IV-B operon contains both Csf3 (Cas5) and the putative large subunit Csf1, we did not observe corresponding densities within the high-resolution cryo-EM structure. However, bands that correspond to Csf1 and Csf3 are observed in SDS-PAGE analysis of the sample (**Figure 2-3D**) and there is unmodeled ambiguous density on the top and bottom of the complex that could represent a flexible association with Csf1 and Csf3 or additional Csf2 subunits. In type I CRISPR systems, Cas5 binds the 5' crRNA handle with high affinity and sequence specificity, nucleating complex assembly (Hochstrasser et al., 2016; Chowdhury et al., 2017; Jia et al., 2019). The lack of discernible density for the Cas5-like Csf3 subunit within our complex may explain the heterogeneous assembly of type IV-B Csf complexes around non-specific RNA (**Figure 2-4**). However, because the type IV-B system does not encode a CRISPR array, the identity of the RNA sequence that Csf3 would specifically recognize is unknown. Indeed, it remains to be determined whether Csf3 truly serves a similar role to the Cas5 subunits in other systems, binding the 5'-handle of processed crRNAs. We hypothesized that crRNAs generated from the adjacent type I-E CRISPR and Cas6 endonuclease would be bound by the type IV-B complex. However, our sequencing analysis showed no enrichment for crRNAs within the RNPs or

any other RNAs available in the total sample. Interestingly, recent bioinformatic analysis indicated a negative co-occurrence of type IV-B systems with other CRISPR systems suggesting their function is not dependent on co-occurring CRISPR arrays (Pinilla-Redondo et al., 2019). The ability of the Csf complex to assemble on non-specific RNAs of a uniform length suggests that type IV-B systems may have been functionally repurposed for a yet to be identified role.

The lack of discernible density for the Csf3 and Csf1 subunits suggests our structure may not accurately reflect the functional type IV-B Csf effector complex. However, several lines of reasoning argue that even without obvious density for Csf1 and Csf3, this complex provides important insights into understanding type IV-B system function. Superposition of the helical Cas7 backbones from type III effector complexes with our structure show that they are nearly identical in arrangement (**Figure 2-6A**). Additionally, the crRNA from the type IV RNP can be overlaid on that of the type III effector with an r.m.s.d. of 1.5 Å (**Figure 2-8A**), indicating our complex presents RNA in a conformation amenable for base pairing with complementary nucleic acid. In fact, studies have shown that there are no structural differences between filaments assembled around non-specific RNAs and correctly processed crRNAs bound to the effector (Hochstrasser et al., 2016). Importantly, the structures of all CRISPR-Cas effector complexes involve non-sequence specific interactions between the crRNA and Cas7-like backbone proteins, suggesting that there would be no structural differences between a random RNA and a crRNA bound within the Cas7 backbone of an RNP complex. Thus, our structure likely accurately represents the structure of the Cas7-like core of the effector complex even though it is bound to heterogeneous RNA, and no density is

observed for Csf1 and Csf3. Completely novel information is gleaned from our cryo-EM reconstruction of the type IV-B RNP including (1) the first structure of a type IV Cas11 protein, which adopts a novel small subunit fold, (2) the first structure of a Cas7-like Csf2 subunit, and (3) interactions between these subunits with each other and with bound RNA.

Since all type IV systems identified lack adaptation subunits and almost all (97.8%) type IV-B operons identified lack a CRISPR array, it is likely they do not participate in selective pre-spacer acquisition or adaptive immunity (Özcan et al., 2019; Pinilla-Redondo et al., 2019; Makarova et al., 2020). Instead, they may have been co-opted for an orthogonal function. While there is a precedent for the repurposing of CRISPR systems for non-defense functions (Klompe et al., 2019; Halpin-Healy et al., 2020), the role of type IV-B systems remains a mystery. A particularly tantalizing hypothesis is that type IV-B Csf complexes assemble on small RNAs, acting as non-specific RNA-sponges, and enabling IV-B-encoding megaplasmids to evade targeting by host cell RNA guided defenses (Pinilla-Redondo et al., 2019). Future experiments are essential to reveal the biological functions of type IV systems. Recent classifications have indicated that although type IV-B systems are highly diverse, they are almost always associated with an adenosine 5'-phosphosulfate reductase-family gene cysH (Özcan et al., 2019; Pinilla-Redondo et al., 2019; Makarova et al., 2020) (**Figure 2-1**). Thus, understanding the interplay between *cysH* and the type IV-B Csf RNP complex may be the key to deciphering the enigmatic role of type IV-B CRISPR systems.

LIMITATIONS OF THE STUDY

The current structure lacks discernible density for Csf1 and Csf3 proteins. The equivalent subunits in Type I systems are responsible for specific functions. Without complementary functional in vitro and in vivo data, it is impossible to unambiguously characterize the current structure as a functional effector complex.

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, David W. Taylor (dtaylor@utexas.edu).

Materials Availability

All unique/stable reagents generated in this study are available from the Lead Contact without restriction.

Data and Code Availability

The cryo-EM structure and associated atomic coordinates have been deposited in the Electron Microscopy Data Bank and the Protein Data Bank with accession codes EMD-22340 and 7JHY, respectively. The raw sequencing data has been deposited in the Sequence Read Archive (SRA) database with submission number SUB8825456.

ACKNOWLEDGEMENTS

We thank members of the Staals, Jackson, and Taylor labs for helpful discussions. This work was supported in part by Welch Foundation grant F-1938 (to D.W.T.), Army Research Office Grant W911NF-15-1-0120 (to D.W.T.), and a Robert J. Kleberg, Jr. and

Helen C. Kleberg Foundation Medical Research Award (to D.W.T.). D.W.T is a CPRIT Scholar supported by the Cancer Prevention and Research Institute of Texas (RR160088) and an Army Young Investigator supported by the Army Research Office (W911NF-19-1-0021). Research in the Jackson Lab is supported by Utah State University New Faculty Start-up funding from the Department of Chemistry and Biochemistry, the Research and Graduate Studies Office, and the College of Science as well as the National Institutes of Health (R35GM138080). Data was collected at the Sauer Structural Biology Lab at UT Austin.

AUTHOR CONTRIBUTIONS

H.N.T. and J.A.S. performed purification of complexes. Y.Z. and J.P.K.B. collected and processed cryo-EM data. Y.Z., H.N.T., R.N.J. and J.P.K.B. performed the model-building and Y.Z. and J.P.K.B. performed model refinement. J.A.S. and H.N.T. performed the RNA-seq experiments. All authors interpreted the results and wrote the manuscript. R.H.J.S., R.N.J., and D.W.T. conceived the experiments, supervised the research, and secured funding for the project.

METHODS

Expression and Purification of the *M. sp.* JS623 Csf Complex

E.coli BL21 (DE3) cells containing the pCDF-Duet1-Csf1-Cas11-Strep-Csf2-Csf3(MCS1)-Cas6 array (MCS2) expression vector were inoculated in 6 X 0.5 L lysogeny broth (LB) and grown at 37°C with 200 rpm shaking. Cells were grown to an optical density (O.D. 600 nm) between 0.6-0.7 then cold shocked on ice for 30-60 min.

Recombinant protein expression was induced with 0.8 mM IPTG (isopropyl β-D-1-thiogalactopyranoside). After induction, cells were grown at 16°C for 18-24 hours and pelleted via centrifugation. Pelleted cells were resuspended in 20-30 mL Buffer W (100 mM Tris, pH 8.0; 150 mM NaCl; 2 µM $ZnSO_4$). Protease inhibitors were added to the following final concentrations: 10 µg/mL leupeptin, 2 µg/mL aprotinin, and 170 µg/mL phenylmethylsulfonyl fluoride (PMSF). Cells were lysed by sonication and lysate was clarified by centrifugation. Polyethylenimine was added to the soluble fraction at a final concentration of 0.1% to precipitate nucleic acids and again clarified by centrifugation. The supernatant was applied to a StrepTrap HP 5 mL column (GE Healthcare) and the bound protein was eluted with Buffer E (Buffer W + 5 mL desthiobiotin). The RNP complex was further purified with a Superose6 Increase 10/300 GL column (GE Healthcare), eluting into SEC Buffer (50 mM HEPES, pH 8.0; 150 mM NaCl; 2 µM $ZnSO_4$).

RNA Sequencing and Analysis

Nucleic acids that co-purified with type IV-B complex were extracted with acid phenol:chloroform and subsequent ethanol precipitation. The resulting fraction was loaded on a 20% denaturing PAGE gel after which a band of approximately 55-60 nt (**Figure 2-4A**) was excised and purified from gel using the ZR small-RNA PAGE Recovery Kit (Zymo Research, USA). Small RNAs were prepared by GenXPro (GenXPro GmbH, Germany) using the TrueQuant smallRNA Seq kit according to the manufacturer's instructions and were sequenced on a HiSeq2000 (Illumina, USA). After quality control filtering and adapter trimming using Cutadapt (Martin, 2011), the reads

were mapped on the expression plasmid and the *E.coli* BL21 (DE3) genome (Genbank accession CP001509) with Geneious Prime 2020.10.2 (https://www.geneious.com). For comparing the abundance and processing of (mature) crRNAs of the total cellular RNA population versus those associated with the type IV complex (**Figure 2-3D**), the extracted total RNAs (NEB Monarch Total RNA Miniprep Kit) were first depleted for ribosomal RNAs (Invitrogen Ribominus Transcriptome Isolation Kit) before they were sequenced on an Illumina MiSeq (Center for Integrated Biosystems, Utah State University, USA). After quality control and adapter trimming, the resulting reads were mate-paired and merged using SeqPrep (https://github.com/jstjohn/SeqPrep), filtered for reads containing CRISPR-array repeat nucleotides, and mapped on the expression plasmid using Geneious (Langmead et al., 2009). Visualisation of the mapping results and further downstream analyses were performed using Geneious and Microsoft Excel.

Cryo-EM Data Acquisition and Processing

C-flat holy carbon grids (CF-4/2, Protochips Inc.) were glow-discharged for 30 seconds using a Gatan Solarus plasma cleaner. 2.5 µl of Type IV complex (~0.3 mg/ml) was applied onto grids, blotted for 2.5 seconds with a blotting force of 1 and rapidly plunged into liquid ethane using a FEI Vitrobot MarkIV operated at 4 ℃ and 100% humidity. Data were acquired using a FEI Titan Krios transmission electron microscope (Sauer Structural Biology Laboratory, University of Texas at Austin) operating at 300 keV at a nominal magnification of ×22,500 (1.1 Å pixel size) with defocus ranging from -1.5 to -3.0 µm. The data were collected using a total exposure of 6 s fractionated into 20 frames (300 ms per frame) with a dose rate of ~8 electrons per pixel per second and a total exposure

dose of ~40 $e^{-}Å^{-2}$. Three datasets were automatically recorded on a Gatan K2 Summit direct electron detector operated in counting mode using the MSI-Template application within the automated macromolecular microscopy software LEGINON (Suloway et al., 2005).

All image pre-processing was performed in Appion (Lander et al., 2009). Individual movie frames were aligned and averaged using 'MotionCor2' drift-correction software (Zheng et al., 2017). The contrast transfer function (CTF) of each micrograph was estimated using CTFFIND4 (Rohou & Grigorieff, 2015). Particles were picked with a template-based particle picker using a reference-free 2D class average from a small subset of manually picked particles as templates. Selected particles were extracted from micrographs using particle extraction within Relion (Scheres, 2012) and the coordinates exported from Appion. After multiple rounds of 2D classification in Relion to remove junk particles, 824,421 particles were left and imported into cryoSPARC (Punjani et al., 2017) for further processing. After multiple rounds of 2D classification and heterogeneous refinement, a final reconstruction containing 296,319 particles was determined to a global resolution of 3.9 Å (based on the gold standard 0.143 FSC criterion using two independent half-maps) using local refinement (implementing non-uniform refinement) with a mask corresponding to the entire complex.

Csf Model Building, Refinement and Structural Analysis

An atomic model for the Csf complex was built de novo in Coot (Emsley et al, 2004), and subjected to multiple iterative rounds of molecular dynamics - flexible fitting in Namdinator (Kidmose et al, 2019) and real-space refinement in Phenix (Afonine et al.,

2018). The majority of the RNA was modelled as polyU, with occasional bases modelled as A depending on the size of the cryoEM density corresponding to the nucleobase (i.e. if the density was unambiguously a purine, (Bravo et al., 2021)). The refined Csf complex model was validated using MolProbity (Chen et al., 2010) as implemented in Phenix. Protein sequence conservation analysis was performed using online ConSurf (Ashkenazy et al., 2016) server, with multiple sequence alignment (MSA) generated using the top 100 result from a BLAST search against Csf sequences. Output from the MSA was used to generate a sequence logo using the WebLogo server (G. Crooks et al., 2004)

Maps and models were visualized using ChimeraX (Goddard et al., 2018) and the electrostatic surfaces were determined using the APBS plugin (Baker et al., 2001) within PyMol. Root-mean-square deviation (r.m.s.d.) values between equivalent atoms in Csf2 and type III-A Csm3, and between type IV-B RNA and III-A crRNA were calculated using ChimeraX and PyMol.

REFERENCES

Afonine, P. V. *et al.* (2018) 'Real-space refinement in PHENIX for cryo-EM and crystallography', *Acta Crystallographica Section D: Structural Biology*. International Union of Crystallography, 74(6), pp. 531–544. doi: 10.1107/S2059798318006551.

Ashkenazy, H. *et al.* (2016) 'ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules', *Nucleic acids research*, 44(W1), pp. W344–W350. doi: 10.1093/nar/gkw408.

Baker, N. A. *et al.* (2001) 'Electrostatics of nanosystems: Application to microtubules

and the ribosome', *Proceedings of the National Academy of Sciences of the United States of America*, 98(18), pp. 10037–10041. doi: 10.1073/pnas.181342398.

Bravo, J. P. K. *et al.* (2018) 'Stability of local secondary structure determines selectivity of viral RNA chaperones', *Nucleic Acids Research*. Oxford University Press, (May), p. 293191. doi: 10.1101/293191.

Bravo, J. P. K. *et al.* (2020) 'Structural basis of rotavirus RNA chaperone displacement and RNA annealing', *bioRxiv*, p. 2020.10.26.354233. Available at: https://doi.org/10.1101/2020.10.26.354233.

Bravo, J. P. K. *et al.* (2021) 'Remdesivir is a delayed translocation inhibitor of SARS CoV-2 replication', *Molecular Cell*. Elsevier Inc. doi: 10.1016/j.molcel.2021.01.035.

Chen, V. B. *et al.* (2010) 'MolProbity: All-atom structure validation for macromolecular crystallography', *Acta Crystallographica Section D: Biological Crystallography*. International Union of Crystallography, 66(1), pp. 12–21. doi: 10.1107/S0907444909042073.

Chowdhury, S. *et al.* (2017) 'Structure Reveals Mechanisms of Viral Suppressors that Intercept a CRISPR RNA-Guided Surveillance Complex', *Cell*. Elsevier Inc., 169(1), pp. 47-57.e11. doi: 10.1016/j.cell.2017.03.012.

Crooks, G. *et al.* (2004) 'WebLogo: a sequence logo generator', *Genome Research*, 14, pp. 1188–1190. doi: 10.1101/gr.849004.1.

Crowley, V. M. *et al.* (2019) ' A Type IV-A CRISPR-Cas System in Pseudomonas aeruginosa Mediates RNA-Guided Plasmid Interference In Vivo', *The CRISPR*

Journal, 2(6), pp. 434–440. doi: 10.1089/crispr.2019.0048.

Emsley, P. & Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography.* 60, 2126–2132 (2004).

Faure, G. *et al.* (2019) 'CRISPR–Cas in mobile genetic elements: counter-defence and beyond', *Nature Reviews Microbiology*. Springer US, 17(8), pp. 513–525. doi: 10.1038/s41579-019-0204-7.

Goddard, T. D. *et al.* (2018) 'UCSF ChimeraX: Meeting modern challenges in visualization and analysis', *Protein Science*, 27(1), pp. 14–25. doi: 10.1002/pro.3235.

Halpin-Healy, T. S. *et al.* (2020) 'Structural basis of DNA targeting by a transposon-encoded CRISPR–Cas system', *Nature*. Springer US, 577(7789), pp. 271–274. doi: 10.1038/s41586-019-1849-0.

Hille, F. *et al.* (2018) 'The Biology of CRISPR-Cas: Backward and Forward', *Cell*, 172(6), pp. 1239–1259. doi: 10.1016/j.cell.2017.11.032.

Hochstrasser, M. L. *et al.* (2016) 'DNA Targeting by a Minimal CRISPR RNA-Guided Cascade', *Molecular Cell*. Elsevier Inc., 63(5), pp. 840–851. doi: 10.1016/j.molcel.2016.07.027.

Jackson, R. N. *et al.* (2014) 'Structural biology. Crystal structure of the CRISPR RNA-guided surveillance complex from Escherichia coli.', *Science (New York, N.Y.)*, 345(6203), pp. 1473–1479. doi: 10.1126/science.1256328.

Jia, N. *et al.* (2019) 'Type III-A CRISPR-Cas Csm Complexes: Assembly, Periodic RNA Cleavage, DNase Activity Regulation, and Autoimmunity', *Molecular Cell*. Elsevier Inc., 73(2), pp. 264-277.e5. doi: 10.1016/j.molcel.2018.11.007.

Jiang, F. *et al.* (2016) 'Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage', *Science*, 351(6275), pp. 867–871. doi: 10.1126/science.aad8282.

Jinek, M. *et al.* (2014) 'Structures of Cas9 Endonucleases Reveal RNA-Mediated Conformational Activation', *Science*, 343(6176), pp. 1247997–1247997. doi: 10.1126/science.1247997.

Kidmose, R. T. *et al.* Namdinator - Automatic molecular dynamics flexible fitting of structural models into cryo-EM and crystallography experimental maps. *IUCrJ* 6, 526–531 (2019).

Klompe, S. E. *et al.* (2019) 'Transposon-encoded CRISPR–Cas systems direct RNA-guided DNA integration', *Nature*. Springer US, 21. doi: 10.1038/s41586-019-1323-z.

Lander, G. C. *et al.* Appion: An integrated, database-driven pipeline to facilitate EM image processing. *Journal of Structural Biology.* 166, 95–102 (2009).

Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology.* 10, (2009).

Li, Z. *et al.* (2021) 'Cryo-EM structure of the RNA-guided ribonuclease Cas12g', *Nature Chemical Biology*. Springer US. doi: 10.1038/s41589-020-00721-2.

Liu, J. J. *et al.* (2019) 'CasX enzymes comprise a distinct family of RNA-guided genome editors', *Nature*, 566(7743), pp. 218–223. doi: 10.1038/s41586-019-0908-x.

Makarova, K. S. *et al.* (2020) 'Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants', *Nature Reviews Microbiology*, 18(2), pp. 67–83. doi: 10.1038/s41579-019-0299-x.

Makarova, K. S., Zhang, F. and Koonin, E. V. (2017) 'SnapShot: Class 1 CRISPR-Cas Systems', *Cell*. Elsevier, 168(5), pp. 946-946.e1. doi: 10.1016/j.cell.2017.02.018.

Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal [Online]* 17, 10–12 (2011).

Meeske, A. J. *et al.* (2020) 'A phage-encoded anti-CRISPR enables complete evasion of type VI-A CRISPR-Cas immunity', *Science*, pp. 1–13.

Mulepati, S., Héroux, A. and Bailey, S. (2014) 'Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target', *Science*, 345(6203).

O'Brien, R.E., Santos, I. C., Wrapp, D., Bravo, J. P. K., Schwartz, E.A., Brodbelt, J.S., Taylor, D.W. Structural basis for assembly of non-canonical small subunits into type I-C Cascade. *Nature Communications*. 1-6 (2020).

Osawa, T., Inanaga, H., Sato, C. & Numata, T. Crystal structure of the crispr-cas RNA silencing cmr complex bound to a target analog. *Molecular Cell* 58, 418–430 (2015).

Özcan, A. *et al.* (2019) 'Type IV CRISPR RNA processing and effector complex formation in Aromatoleum aromaticum', *Nature Microbiology*. Springer US, 4(1), pp. 89–96. doi: 10.1038/s41564-018-0274-8.

Pinilla-Redondo, R. *et al.* (2019) 'Type IV CRISPR–Cas systems are highly diverse and involved in competition between plasmids', *Nucleic Acids Research*, 48(4), pp. 2000–2012. doi: 10.1093/nar/gkz1197.

Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. CryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods* 14, 290–296 (2017).

Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *Journal of Structural Biology*. 192, 216–221 (2015).

Rollins, M. C. F. *et al.* (2019) 'Structure Reveals a Mechanism of CRISPR-RNA-Guided Nuclease Recruitment and Anti-CRISPR Viral Mimicry', *Molecular Cell*. Elsevier Inc., 74(1), pp. 132-142.e5. doi: 10.1016/j.molcel.2019.02.001.

Scheres, S. H. W. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *Journal of Structural Biology*. 180, 519–530 (2012).

Slaymaker, I. M. *et al.* (2019) 'High-Resolution Structure of Cas13b and Biochemical Characterization of RNA Targeting and Cleavage', *Cell Reports*. ElsevierCompany., 26(13), pp. 3741-3751.e5. doi: 10.1016/j.celrep.2019.02.094.

Sofos, N. *et al.* (2020) 'Structures of the Cmr-β Complex Reveal the Regulation of the Immunity Mechanism of Type III-B CRISPR-Cas', *Molecular Cell*. Elsevier Inc., 79(5), pp. 741-757.e7. doi: 10.1016/j.molcel.2020.07.008.

Staals, R. H. J. *et al.* (2014) 'RNA Targeting by the Type III-A CRISPR-Cas Csm Complex of Thermus thermophilus', *Molecular Cell*. Elsevier Inc., 56(4), pp. 518–530. doi: 10.1016/j.molcel.2014.10.005.

Steens, J. A. *et al.* (2021) 'SCOPE: Flexible targeting and stringent CARF activation enables type III CRISPR-Cas diagnostics', *bioRxiv*, (10), p. 2021.02.01.429135. Available at: https://doi.org/10.1101/2021.02.01.429135.

Stella, S., Alcón, P. and Montoya, G. (2017) 'Structure of the Cpf1 endonuclease R-loop complex after target DNA cleavage', *Nature Publishing Group*. Nature Publishing Group, 546(7659), pp. 559–563. doi: 10.1101/122648.

Suloway, C. *et al.* Automated molecular microscopy: The new Leginon system. *Journal*

of Structural Biology 151, 41–60 (2005).

Takeda, S. N. *et al.* (2021) 'Structure of the miniature type V-F CRISPR-Cas effector enzyme', *Molecular Cell*. Elsevier Inc., 81(3), pp. 558-570.e3. doi: 10.1016/j.molcel.2020.11.035.

Taylor, D. W. *et al.* (2015) 'Structure of the CRISPR-Cmr complex reveal mode of RNA target positioning', *Science*, 348(6234), pp. 581–586. doi: 10.1126/science.aaa4535.

Taylor, H. N. *et al.* (2019) 'Structural basis of Type IV CRISPR RNA biogenesis by a Cas6 endoribonuclease', *RNA Biology*. Taylor & Francis, 16(10), pp. 1438–1447. doi: 10.1080/15476286.2019.1634965.

Xiao, Y. *et al.* (2017) 'Structure Basis for Directional R-loop Formation and Substrate Handover Mechanisms in Type I CRISPR-Cas System', *Cell*. Elsevier, 170(1), pp. 48-60.e11. doi: 10.1016/j.cell.2017.06.012.

Xiao, Y *et al.* (2018) 'Structure basis for RNA-guided DNA degradation by Cascade and Cas3', 0839(June), pp. 1–12. doi: 10.1126/science.aat0839.

Yan, W. X. *et al.* (2018) 'Cas13d Is a Compact RNA-Targeting Type VI CRISPR Effector Positively Modulated by a WYL-Domain-Containing Accessory Protein', *Molecular Cell*. Elsevier Inc., 70(2), pp. 327-339.e5. doi: 10.1016/j.molcel.2018.02.028.

You, L. *et al.* (2019) 'Structure Studies of the CRISPR-Csm Complex Reveal Mechanism of Co-transcriptional Interference', *Cell*. Elsevier Inc., 176(1–2), pp. 239-253.e16. doi: 10.1016/j.cell.2018.10.052.

Zhang, H. *et al.* (2020) 'Mechanisms for target recognition and cleavage by the Cas12i

RNA-guided endonuclease', *Nature Structural and Molecular Biology*. Springer US, 27(11), pp. 1069–1076. doi: 10.1038/s41594-020-0499-0.

Zheng, S. Q. *et al*. MotionCor2: Anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nature Methods* 14, 331–332 (2017).

Zhu, X. *et al.* (2019) 'Cryo-EM structures reveal coordinated domain motions that govern DNA cleavage by Cas9', *Nature Structural and Molecular Biology*. Springer US, 26(8), pp. 679–685. doi: 10.1038/s41594-019-0258-2.

CHAPTER 3

STRUCTURAL BASIS OF TYPE IV CRISPR RNA BIOGENESIS BY A CAS6 ENDORIBONUCLEASE[2]

ABSTRACT

Prokaryotic CRISPR-Cas adaptive immune systems rely on small non-coding RNAs derived from CRISPR loci to recognize and destroy complementary nucleic acids. However, the mechanism of Type IV CRISPR RNA (crRNA) biogenesis is poorly understood. To dissect the mechanism of Type IV CRISPR RNA biogenesis, we determined the x-ray crystal structure of the putative Type IV CRISPR associated endoribonuclease Cas6 from Mahella australiensis (Ma Cas6-IV) and characterized its enzymatic activity with RNA cleavage assays. We show that Ma Cas6-IV specifically cleaves Type IV crRNA repeats at the 3' side of a predicted stem loop, with a metal-independent, single-turnover mechanism that relies on a histidine and a tyrosine located within the putative endonuclease active site. Structure and sequence alignments with Cas6 orthologs reveal that although Ma Cas6-IV shares little sequence homology with other Cas6 proteins, all share common structural features that bind distinct crRNA repeat sequences. This analysis of Type IV crRNA biogenesis provides a structural and

---

[2] Co-authored by: H. N. Taylor*, E. E. Warner, M. J. Armbrust, V. M. Crowley, K. J. Olsen, and R. N. Jackson.
Published: *RNA Biology* (2019).
Final publication can be accessed at https://doi.org/10.1080/15476286.2019.1634965
*HNT designed and performed all biochemical cloning and experiments, analyzed all biochemical data, created figures, and wrote the manuscript.

biochemical framework for understanding the similarities and differences of crRNA biogenesis across multi-subunit Class 1 CRISPR immune systems.

INTRODUCTION

Bacteria and archaea use small RNAs derived from CRISPR (Clustered Regularly Interspaced Short Palindromic Repeat) loci to guide CRISPR associated (Cas) proteins to complementary targets such as invasive phage and plasmid DNA [1–8]. Once bound, crRNA-guided ribonucleoprotein complexes undergo conformational rearrangements that activate trans- or cis-acting nucleases to degrade bound targets [9]. Thus, a critical step of prokaryotic crRNA-mediated immunity is crRNA biogenesis, where long, CRISPR-derived transcripts are processed into small crRNA guides.

Although crRNA biogenesis is a fundamental process of all known CRISPR adaptive immune systems, the proteins and mechanisms involved vary across the two CRISPR system classes (1 and 2), six types (I, II, III, IV, V and VI), and more than thirty-three distinct subtypes [10–13]. For example, some single-subunit Class 2 proteins rely on RNase III to cleave duplexed RNAs formed from the base pairing of crRNA repeats with a second trans-activating CRISPR RNA (e.g. Type II-A, and V-B) [14,15], while other Class 2 systems process crRNAs with endonuclease activities that reside within single subunit effector proteins (types V-A, and VI) [16,17].

Most multi-subunit Class 1 systems process crRNAs with CRISPR associated endonucleases, called Cas6, which share conserved structural motifs that bind crRNAs [11]. In general, Cas6 enzymes use a metal-ion-independent mechanism to cleave crRNAs on the 3’-side of stem-loops formed within the palindromic CRISPR repeat

[1,18]. Cleavage is generally catalyzed by stabilizing nucleophilic attack from the 2' hydroxyl located upstream from the scissile phosphate [19–22]. In spite of these similarities, Cas6 amino acid sequences are remarkably diverse [23], and several structural and mechanistic differences have been observed [11]. For example, often a histidine is used to catalyze cleavage [20,23–27], but other residues, such as lysine, have been shown to catalyze the reaction when histidine is not present (e.g. subtype I-A) [28,29]. Additionally, distinct Cas6 proteins associate differently with processed crRNAs after cleavage. In subtypes I-B, I-E, I-D and I-F, Cas6 makes structural and base specific interactions with the stable stem-loop formed by the palindromic CRISPR repeat and typically stays bound after cleavage to form a component of the multi-subunit interference complex [19,21,22,24,25,30–33]. In contrast, the repeats of subtypes I-A, III-A, and III-B are less stable, allowing Cas6 to dissociate from the processed crRNA and to perform multi-turnover crRNA cleavage [26,28,29,34].

Type IV CRISPR systems are categorized as Class 1 as they are predicted to form multi-subunit crRNA-guided complexes [6,10,13]. However, the enzymatic functions of Type IV systems remain largely unknown, including the mechanisms of crRNA biogenesis. All Type IV systems contain genes predicted to encode a multi-subunit complex consisting of a large-subunit (*csf1, cas8*-like), backbone (*csf2, cas7*-like), and tail (*csf3, cas5*-like) [10]. Type IV-A systems contain additional genes that encode for a putative helicase (*dinG*) and a *cas6* endonuclease, while Type IV-B systems lack these accessory genes, and interestingly, often lack a CRISPR locus. Distinct Type IV-A systems contain diverse *cas6* gene sequences, including genes designated as *cas6e* and

cas6f (*cas6* sequences observed in subtypes I-E and I-F), and a Type IV-specific *cas6*-like gene called *csf5* [13].

The presence of *cas6* homologs suggests that Type IV-A systems process their own crRNAs through a Cas6-mediated mechanism. Supporting this prediction, it was recently shown that in vivo assembly of a Type IV-A crRNA-protein complex required the Cas6-homolog Csf5 [35]. Independently purified Csf5 remained bound to a processed crRNA, suggesting Csf5 mediates single-turnover crRNA cleavage. However, Csf5 was unstable without a co-expressed crRNA, therefore, it could not be purified in its apo form and a mechanistic analysis of Csf5 catalytic activity was not assayed. Thus, a need remains for a mechanistic description of Type IV crRNA processing.

To better understand the mechanism of crRNA biogenesis in Type IV CRISPR systems we recombinantly expressed and purified the apo Type IV Cas6 protein from the thermophilic anaerobe *Mahella australiensis* (*Ma* Cas6-IV) [36]. The genome of *M. australiensis* harbors putative Type I-B and Type IV-A CRISPR systems at two distinct loci (**Figure 3-1A**) each with distinct CRISPR repeat sequences which we have designated CRISPR I and CRISPR IV, respectively. We analyzed the *Ma* Cas6-IV-mediated processing of Type IV crRNA repeats with cleavage assays and identified a putative nuclease active site by determining the structure of *Ma* Cas6-IV. Additionally we show that the crRNA is cleaved on the 3’ side of the predicted stem-loop structure, with nucleophilic attack on the scissile phosphate coming from the 2’ hydroxyl of base G22 of the repeat. These results provide a biochemical analysis of Type IV crRNA biogenesis and suggest that although various mechanisms exist, Cas6-mediated metal-independent processing of crRNA is a conserved process across diverse Class 1 systems.

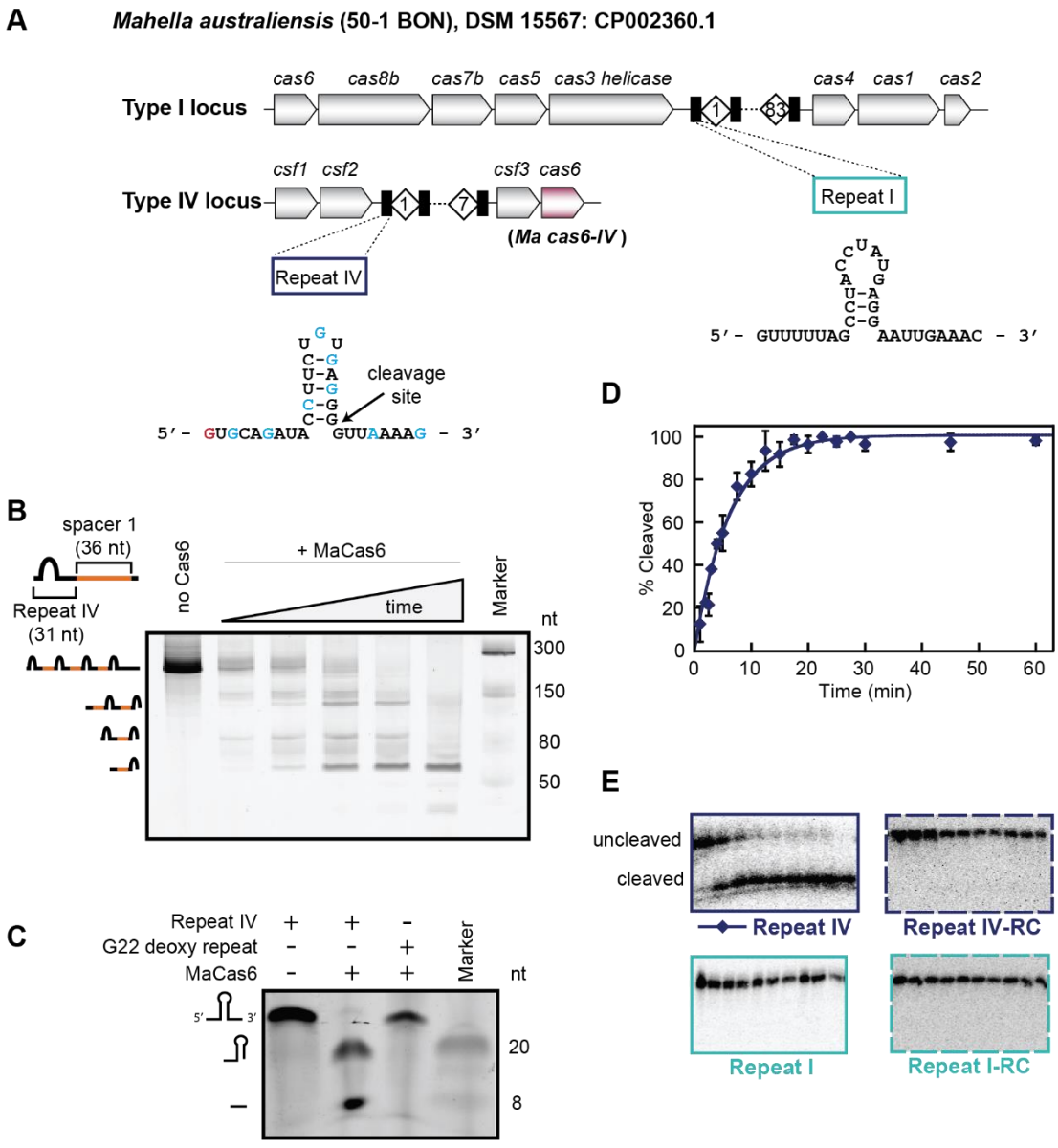

Figure 3-1. The Type IV Cas6 from *M. australiensis* (*Ma* Cas6-IV) cleaves the CRISPR repeat of the Type IV associated CRISPR. (**A**) *M. australiensis* contains a Type I CRISPR system and a Type IV CRISPR system, each with a distinct repeat sequence designated as Repeat I and Repeat IV. Nucleotides highlighted in blue indicate a purine-purine or pyrimidine-pyrimidine shift in sequence. Bases colored red indicate broader mutations. (**B**) *Ma* Cas6-IV-mediated processing of a pre-crRNA composed of four Repeat IV sequences interspersed with the first spacer sequence from the *M. australiensis* Type IV CRISPR. (**C**) SYBR Gold stained gel of the 30 nt long Repeat IV and the G22 2'-deoxy repeat incubated with *Ma* Cas6-IV. The G22 2'-deoxy repeat remains uncleaved in the presence of *Ma* Cas6-IV (**D**) Fit to cleavage data of a single

radiolabeled repeat by *Ma* Cas6-IV. Data for the Repeat IV cleavage were fit as described in (Niewoehner et al., 2014). Error bars denote standard deviation between two or more experiments. (**E**) *Ma* Cas6-IV-mediated cleavage trials of 5'-^{32}P-end labeled Repeat IV, Repeat I, and their reverse complements (Repeat IV-RC, Repeat I-RC). Data were not fit for Repeat I or the RC sequences because no cleavage was observed.

RESULTS

***Ma* Cas6-IV Processes Pre-crRNAs**

To investigate the mechanism of Type IV crRNA biogenesis, we recombinantly expressed and purified N-terminally His-tagged Cas6 protein from the *M. australiensis* Type IV CRISPR system (*Ma* Cas6-IV) (**Figure 3-1A and Figure 3-2**). To determine if *Ma* Cas6-IV processes crRNAs, we performed an RNA cleavage assay with a pre-crRNA transcribed in vitro with the same directionality as the Type IV cas genes. The pre-crRNA was designed to contain four Type IV repeat sequences, each 31 nucleotides in length, and three 36 nucleotide spacer sequences, each identical to the first spacer observed in the native CRISPR IV (**Figure 3-3**). The first three of eight repeats observed in the *Mahella* CRISPR are identical, and this sequence was used in the pre-crRNA design (**Figure 3-3**). Switches in purine or pyrimidine identity occur within the repeat at positions along the 5' and 3' arms and stem of the repeat hairpin. We note that the base switches at the base of the stem would not disrupt base pairing if both Watson Crick and G-U wobble base pairs are acceptable for binding, suggesting the stem of the repeat may be recognized primarily through shape rather than base-specific interactions (**Figure 3-**

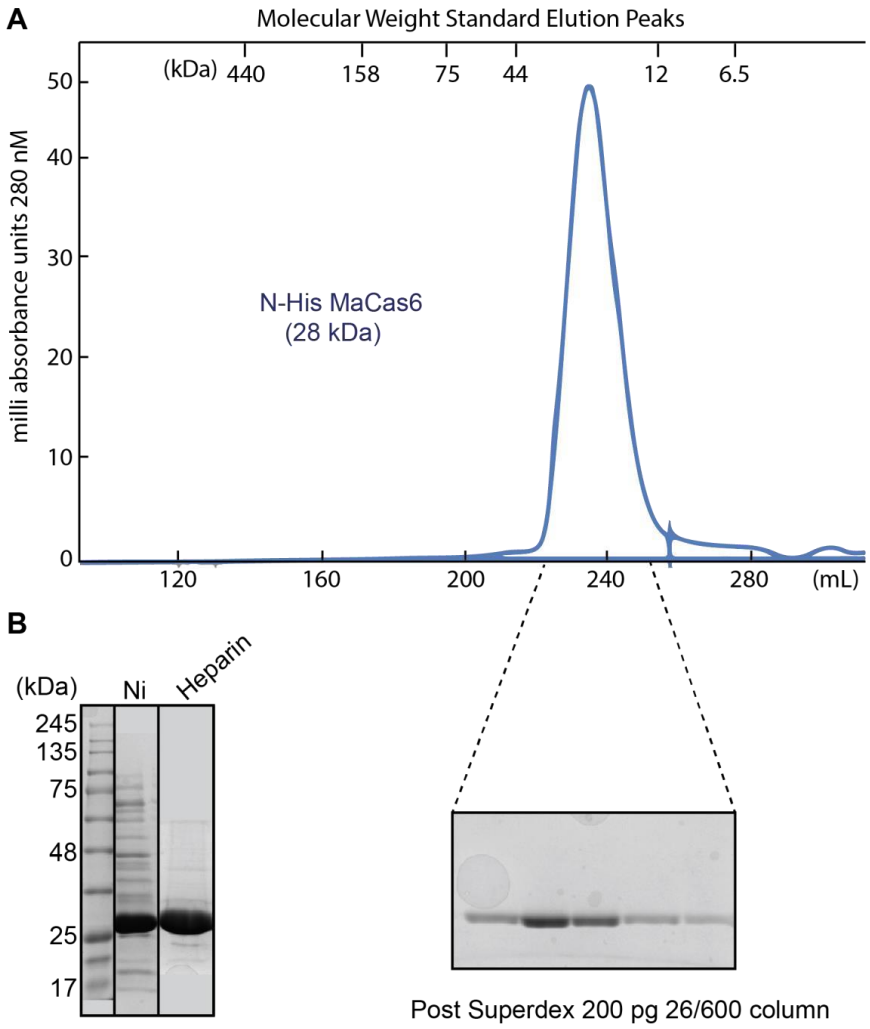

Figure 3-2. *Ma* Cas6-IV purifies as a monomer. (**A**) Size exclusion chromatogram of purified *Ma* Cas6-IV. Molecular weight standard elution peaks are indicated on top. His-tagged *Ma* Cas6–IV elution peak suggests a 28 kDa monomer is purified. (**B**) SDS-PAGE representatives after each of the steps of purification (Ni affinity, Heparin ion exchange, and Size Exclusion over a Superdex 200 pg column).

Figure 3-3. CRISPR RNA repeat sequence degeneracy and spacer sequences. (**A**) Shown is the repeat spacer repeat sequence that was repeated three times in tandem to make the in vitro transcribed pre-crRNA. Nucleotides highlighted in blue indicate a purine-purine or pyrimidine-pyrimidine shift in sequence. Bases colored red indicate broader mutations. (**B**) The repeat and spacer sequences observed in the *Mahella australiensis* genome. Repeat nucleotides that differ from the first three repeats are colored as in (A).

1A and Figure 3-3). Other switches in the arms and loop of the hairpin likewise suggest that those positions are recognized through shape, or are not necessary at all for binding.

Our initial nuclease assay showed that *Ma* Cas6-IV cleaves the pre-crRNA at regular intervals, producing several species of smaller RNAs with differences in length of ~ 67 nucleotides (the length of the repeat + spacer) (**Figure 3-1B**). This result indicated that *Ma* Cas6-IV processes the pre-crRNA into smaller crRNAs with a mechanism similar to other Cas6 nucleases that bind and cleave conserved features of the palindromic CRISPR repeat [11]. To evaluate this activity in more detail, we showed that *Ma* Cas6-IV cleaves a radio- or fluorescein-labeled 30-nucleotide RNA identical in sequence to the first crRNA repeat of CRISPR IV (**Figure 3-1C - E**). As *M. australiensis* is a thermophile [36], we were not surprised to observe that optimal *Ma* Cas6-IV cleavage occurs at temperatures around 50°C, indicating the cleavage mechanism is thermostable (**Figure 3-4**). Additionally, *Ma* Cas6-IV was able to cleave crRNAs in the presence of high concentrations of the metal chelator EDTA (see Methods section), indicating that the cleavage mechanism is metal-ion-independent, consistent with other Class 1 crRNA processing mechanisms [1,18,21,28].

To determine where on the CRISPR repeat *Ma* Cas6-IV cuts, a small RNA marker was run alongside a cleaved repeat stained with SYBR gold (**Figure 3-1C**). We observed that the smaller of two distinct cleavage products ran next to the 8 nt band, while the larger ran next to a 20 nt band, indicating *Ma* Cas6-IV cleaves the repeat asymmetrically on the 3'-side of the repeat near position 22. To confirm the location of cleavage we incubated *Ma* Cas6-IV with a repeat containing a 2'-deoxy sugar at nucleotide position G22 (G22 2'-deoxy repeat). *Ma* Cas6-IV is unable to cleave the G22

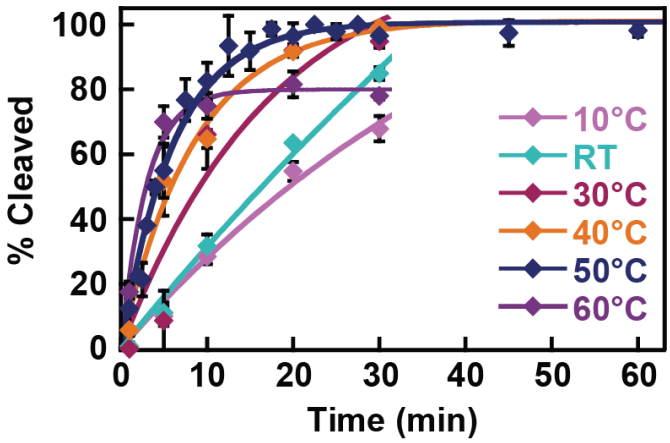

Figure 3-4. Cleavage of Repeat IV by WT *Ma* Cas6-IV at various temperatures. Cleavage of radiolabeled Type IV Repeat was monitored at different temperatures. Shown are data collected in triplicate and fit as in Figure 3-1. Cleavage occurred at all temperatures, with optimal cleavage observed at 50°C.

2'-deoxy repeat indicating the scissile phosphate resides on nucleotide 23 of the repeat, and cleavage is mediated through nucleophilic attack of the 2' hydroxyl of position G22.

Previous analysis of Cas6 endonucleases from the thermophile *Thermus thermophilus* (TtCas6A and TtCas6B) showed that some Cas6 endonucleases encoded in organisms with multiple CRISPR loci are capable of cleaving more than one CRISPR repeat sequence [24]. As *M. australiensis* contains two different CRISPR loci in its genome, each with a distinct CRISPR repeat sequence, *Ma* Cas6-IV could potentially cleave both Type IV and Type I repeats (**Figure 3-1A**). To measure the sequence specificity of *Ma* Cas6-IV, in addition to the Type IV repeat, we tested for nuclease activity against the Type I repeat, the Type I repeat-reverse complement, and the Type IV repeat-reverse complement (**Figure 3-1C**). Unlike the Cas6 enzymes from *Thermus*

thermophilus that cleave multiple repeat sequences, *Ma* Cas6-IV cleaves only the Type IV crRNA, indicating *Ma* Cas6-IV is specific for the Type IV repeat sequence.

Structure of *Ma* Cas6-IV

Our initial analysis showed that *Ma* Cas6-IV specifically cleaves Type IV crRNA repeats, but mechanistic details, such as the turnover kinetics for this enzyme and the features important for binding and cleaving the crRNA, remained unknown. Sequence alignments of *Ma* Cas6-IV with Type I and Type III Cas6 orthologs, as well as the recently investigated Type IV-specific Csf5 protein revealed low sequence identity (~ 16%) (**Table 3-1**), including in the region of the putative active site. Thus, the mechanistic details of *Ma* Cas6-IV could not be deduced through simple amino acid sequence alignments with previously characterized proteins.

To better understand *Ma* Cas6-IV function, we determined the x-ray crystal structure of apo *Ma* Cas6-IV at 1.76 Å resolution (**Figure 3-5 and Table 3-2**). Crystals of N-terminally His-tagged *Ma* Cas6-IV were grown using sitting-drop vapor diffusion and formed in a hexagonal space group (P61) (**Figure 3-6**). Our initial attempts to solve the crystal structure using existing Cas6 structures as molecular replacement models failed. We hypothesized that this was due to structural differences between *Ma* Cas6-IV and the available Cas6 models, so we instead solved the structure using Single-wavelength Anomalous Dispersion methods with crystals soaked overnight in Potassium Tetrachloroplatinate (See Methods and **Table 3-2**).

The structure of *Ma* Cas6-IV reveals that two protein chains reside in the asymmetric unit. The electron density is well resolved and continuous, allowing all 229

Table 3-1. Structural alignment data for Cas6 homologs from multiple subtypes of CRISPR-Cas systems.

Subtype	Organism	Other Names	PDB ID[a]	RMSD (Å)	Aligned/Moving Residues	% Identity	References
I-A	*Sulfolobus solfataricus*	Sso2004	4ILR	2.8495	187/279	15.508	Shao, Y. and Li, H. (2013)
			4ILL	2.8417	185/278	15.1351	
			4ILM	3.2212	188/279	12.766	
		Sso1437	3ZFV	2.7694	165/259	16.9697	Reeks, J. *et al.* (2013)
	Methanocaldococcus jannaschii	Cas6	5YI6	3.4796	175/241	10.8571	Lee, M.C. *et al.* (2018)
I-B	*Thermus thermophilus*	TTHB231	4C98	2.7742	172/242	17.4419	Niewoehner, O. *et al.* (2014)
			4C9D	2.5348	189/262	16.9312	
	Methanococcus maripaludis	Mm Cas6b	*4Z7K*	3.2812	167/218	16.1677	Shao, Y. *et al.* (2016)
I-C	*Bacillus halodurans*	Cas5d	4F3M	3.0815	67/213	8.9552	Nam, K.H. *et al.* (2012); Koo, Y. *et al.* (2013)
	Mannheimia succiniciproducens		3KG4	4.8526	43/184	13.9535	
	Streptococcus pyogenes		3VZH	5.005	52/194	11.5385	
	Xanthomonas oryzae		3VZI	2.4844	74/214	9.4595	
I-E	*Thermus thermophilus*	TTHB192, Cse3	1WJ9	3.4209	134/188	10.4478	Sashital, D.G. *et al.* (2011); Gesner, E.M. *et al.* (2011); Ebihara, A. *et al.* (2006)
			2Y8W	3.6472	150/215	11.3333	
			3QRP	3.609	151/214	10.596	
	Escherichia coli	CasE	4DZD	3.5718	147/193	9.5238	
I-F	*Pseudomonas aeruginosa*	Csy4	2XLI	3.5108	85/167	3.5294	Haurwitz, R.E. *et al.* (2010); Haurwitz, R.E. *et al.* (2012)
			4AL5	3.4897	89/189	2.2472	
III-B	*Pyrococcus furiosus*	PfCas6	3I4H	3.3531	180/232	13.3333	Carte, J. *et al.* (2008); Wang, R. *et al.* (2011)
			3PKM	3.4297	175/229	12	
			3UFC	3.5745	181/243	9.3923	Park, H.M. *et al.* (2012)
	Marinomonas mediterranea	Cas6	6DD5	3.3147	188/652	14.3617	Mohr, G. *et al.* (2018)
IV-A	*Aromatoleum aromaticum*	Csf5	6H9H	3.9062	108/250	8.3333	Özcan, Ahsen, *et al.* (2019)
			6H9I	3.7411	156/250	8.3333	
Orphan	*Thermus thermophilus*	TTHB78	4C97	2.338	173/233	16.763	Niewoehner, O. et al. (2014)
			4C8Y	2.3557	180/238	16.6667	
			4C8Z	2.4531	184/238	16.8478	

[a]PDB ID key: apo (plain text), *substrate bound* (italics), **product bound** (bold).

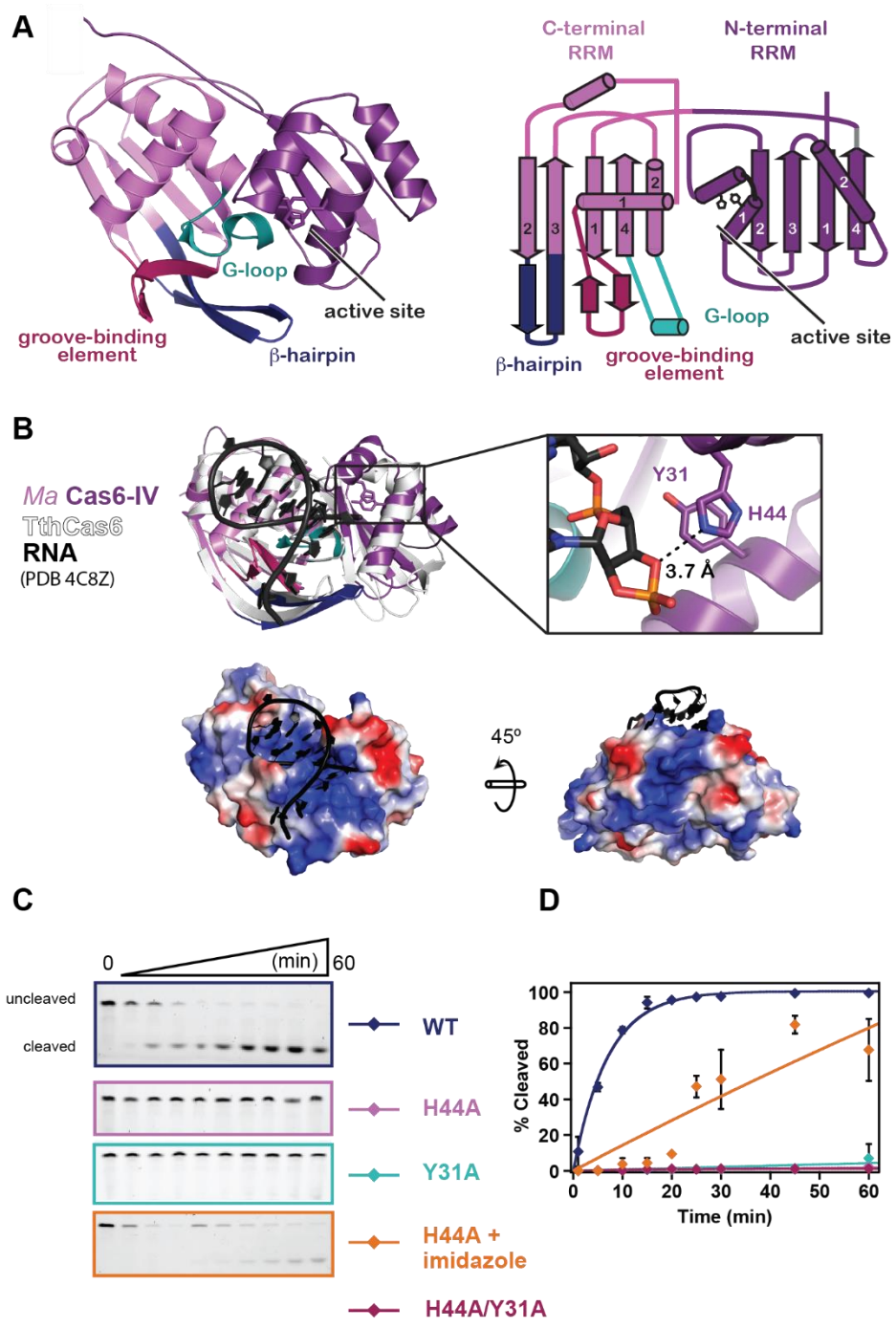

Figure 3-5. Structure of apo *Ma* Cas6-IV and identification of the active site residues, His44 and Tyr31. (**A**) Ribbon model (left) and topology diagram (right) of the *Ma* Cas6-IV structure. The two RRM domains, C-terminal structural motifs involved in binding the crRNA, and putative active site are indicated. (**B**) Alignment of apo *Ma* Cas6-IV to RNA-bound TthCas6 (PDB:4C8Z) with a zoomed in view of the *Ma* Cas6-IV active site containing His44 and Tyr31 residues. RMSD of the alignment is 1.26 Å over 95 out of 257 Cα carbons. The bottom panels show surface renditions of *Ma* Cas6-IV

aligned with the RNA of PDB 4C8Z. Predicted positively charged surface is colored blue and negatively charged surface is colored red. (**C**) Cleavage assays of Repeat IV by predicted active site mutants of *Ma* Cas6-IV (left) with cleavage percentages fit to a pseudo-first order rate equation (right). Cleavage activity of H44A *Ma* Cas6-IV in the presence of 500 mM imidazole is also shown.

Table 3-2. Data collection and refinement statistics.

Dataset	Native	SAD (K_2• Pt Cl_4 – soak)
Data collection		
Beamline	SSRL 12-2	SSRL 12-2
Space group	$P6_1$	$P6_1$
Cell dimensions		
a, *b*, *c* (Å)	85.5, 85.5, 142.6	85.9, 85.9, 144.5
α, β, γ (deg)	90.0, 90.0, 120.0	90.0, 90.0, 120.0
Wavelength (Å)	1.55	1.0718
Resolution (Å)	40-1.76 (1.82-1.76)*	40-2.95 (3.06-2.95)
R_{merge}(%)	5.3 (72.7)	6.5 (22.3)
$CC_{1/2}$	0.995 (0.736)	0.994 (0.941)
I / σI	25.3 (2.2)	15.4 (3.5)
Observations	350790 (23595)	95034 (5484)
Unique reflections	58149 (5618)	24773 (2359)
Multiplicity	6.0 (4.2)	3.8 (2.3)
Completeness (%)	99.6 (96.3)	99.3 (94.9)
Refinement		
Resolution† (Å)	40-1.76 (1.82-1.76)	
No. reflections	148794	
R_{work}/R_{free} (%)	17.1/19.6	
No. atoms	4202	
Protein	3861	
Water	246	
Ligands	95	
B-factors		
mean	43.84	
Protein	42.84	
Water	50.51	
Ligands	78.20	
R.m.s. deviations		
Bond lengths (Å)	0.008	
Bond angles (deg)	1.20	
Ramachandran		
Favored (%)	98.72	
Allowed (%)	1.28	
Outliers (%)	0	
Clashscore	9.30	

*Values in parentheses are for highest-resolution shell.
†Resolution limits use the criterion of I/σI > 2.0

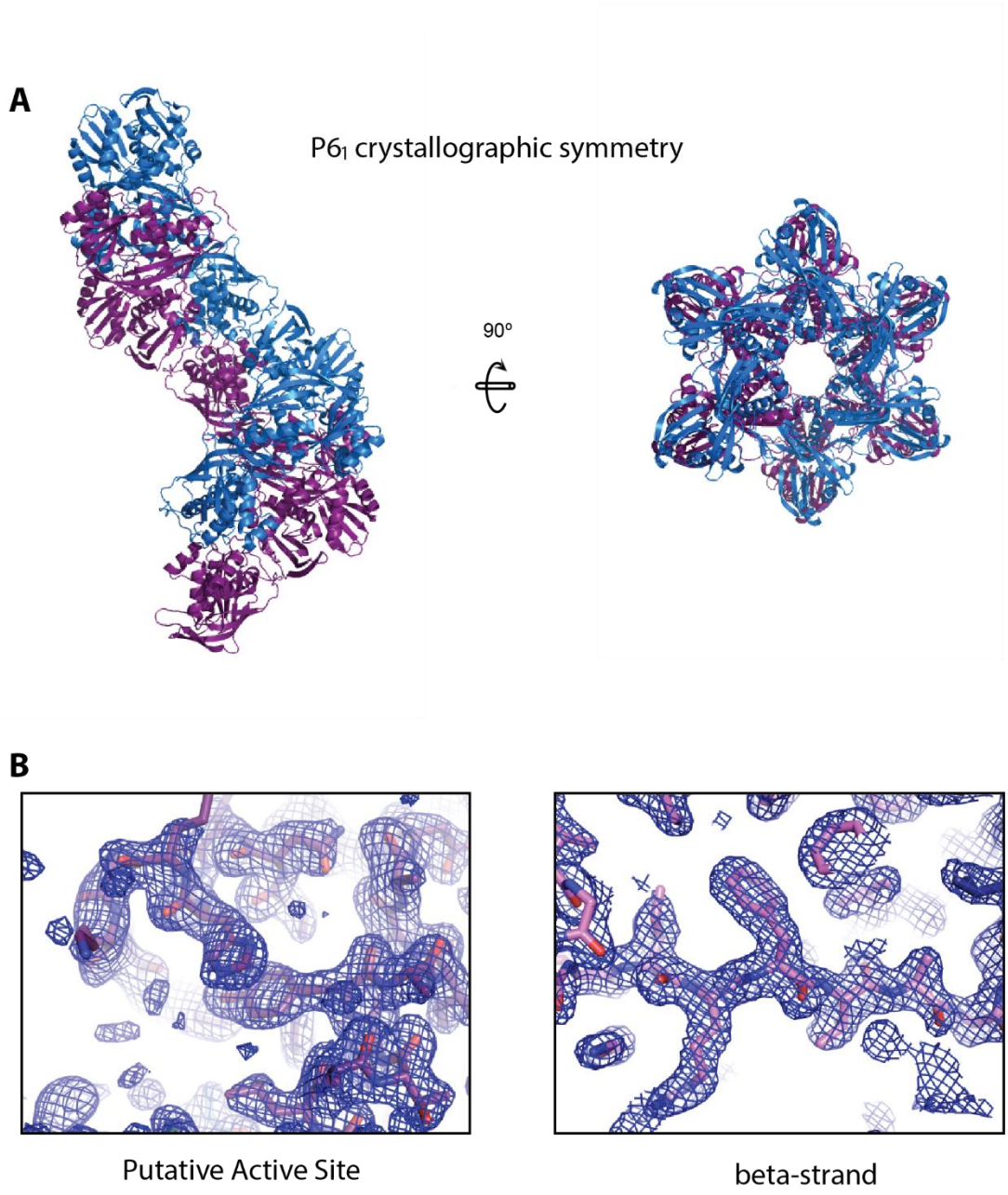

Figure 3-6. Crystallographic symmetry and electron density of *Ma* Cas6. (**A**) Side and top views of the P61 symmetry of Cas6 dimers in the crystal. (**B**) Examples of electron density observed at the putative active site and within the C-terminal RRM (RNA Recognition Motif).

amino acids of both proteins to be modeled along with 7 additional nucleotides of the histidine affinity tag (**Figure 3-6**). The monomers are arranged with two-fold symmetry and are superimposable with an R.M.S.D. < 0.29 Å over 227 of 237 Cα carbons. Although the structure indicates *Ma* Cas6-IV is capable of forming a dimer, size exclusion profiles indicate it purifies as a monomer (**Figure 3-2**). Additionally, analysis with the PDBePISA server reveals the dimer interface is 1067 $Å_2$ with a Complex Formation Significance Score (CSS) of 0.066 [37] (**Figure 3-7**). The low CSS score (scores range from 0 to 1) implies the interface is not significant for complex formation. These results, along with the observation that the dimer conformation is fundamentally different from other Cas6 dimers shown to be functionally relevant [24,29], suggest the observed dimer may solely be the result of crystal packing.

Each monomer is composed of two modified RRM (RNA recognition motif) domains, also known as ferredoxin-like folds (**Figure 3-5A**). Canonical RRM domains adopt β_1, α_1, β_2, β_3, α_2, β_4 fold, where the two helices pack against the concave face of an anti-parallel beta-sheet formed by the four beta-strands [38]. The two RRM domains of *Ma* Cas6-IV express this secondary structure, but in each domain the first alpha-helix has been modified into a helix-turn-helix. Additionally, the C-terminal RRM contains extended features between β_1 and α_1, β_2 and β_3, and α_2 and β_4. To identify the function of these motifs we aligned our structure with the available Cas6 structures in the PDB using the secondary structure matching (SSM) tool in the molecular modeling program Coot (**Figure 3-5**, **Table 3-1**, and **Figure 3-8**) [39]. Our structural alignments reveal that the C-terminal RRM extensions align with conserved features important for binding the crRNA and positioning the scissile phosphate into the nuclease active site. These motifs

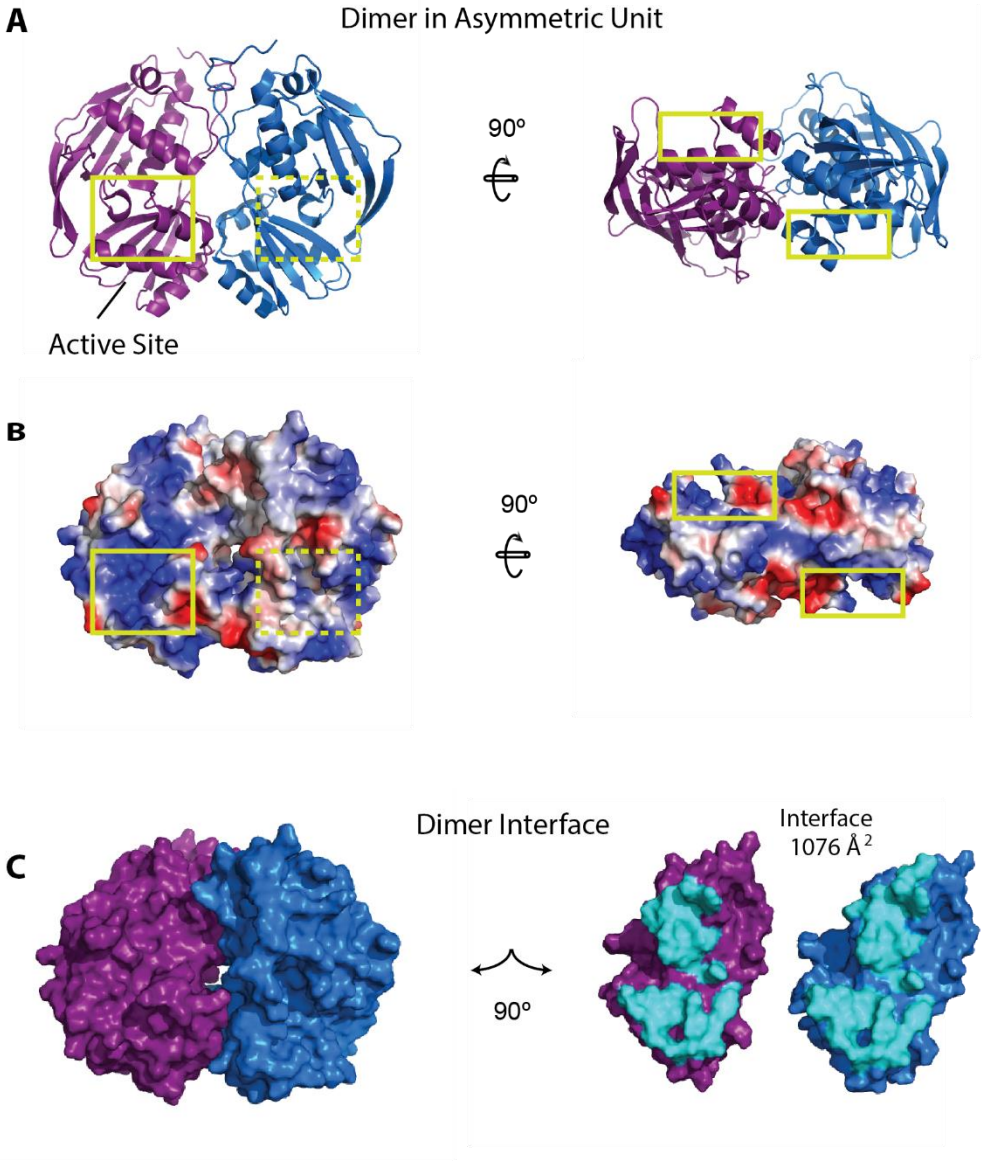

Figure 3-7. Structural analysis of the crystallographic dimer. (**A**) Ribbon views of the side and bottom of the crystallographic dimer. A yellow square line indicates the location of the active site on proximal side, and a dashed square line indicates the active site on the opposite side (**B**) Surface charge is shown on side and bottom views of dimer. Active sites are indicated as in A. (**C**) The dimers are shown as a surface. The interface is shown as cyan where the monomers of the dimer have been rotated 90 degrees. The calculated surface area of the interface is 1076 $Å^2$.

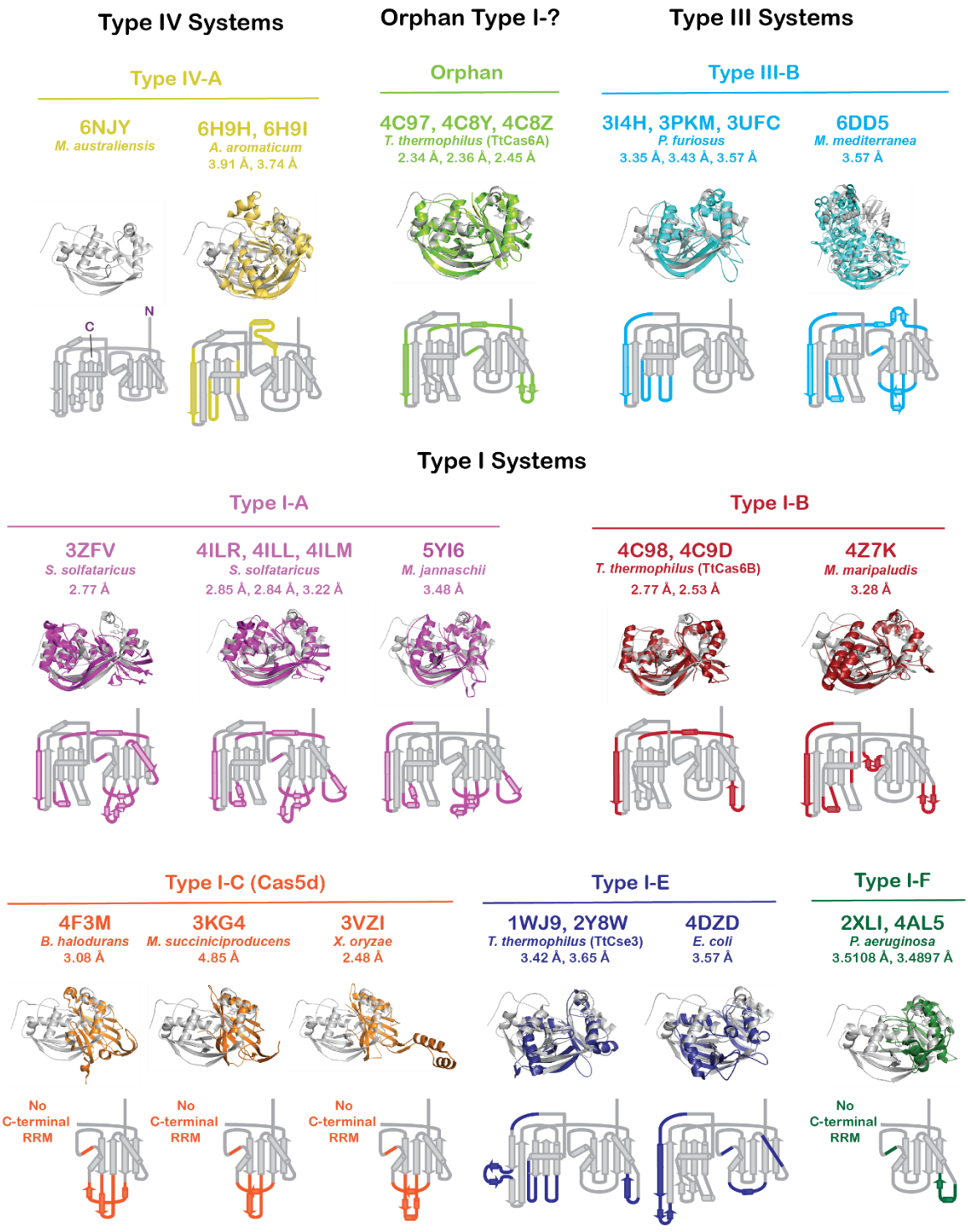

Figure 3-8. Alignment of *Ma* Cas6-IV with Cas6 orthologs. *Ma* Cas6-IV was aligned with Cas6 and Cas5d models available in the PDB using the SSM (Secondary Structure Matching) tool in Coot. The structure of *Ma* Cas6-IV is shown in the top left corner. PDB codes, RMSDs, and organism names of aligned structures are indicated. A topology map

of each structure is shown. It is indicated in color where the topology of the aligned structure is different from *Ma* Cas6-IV.

have been previously described as the groove-binding element (GBE), the β-hairpin, and the glycine-rich loop (G-loop) (**Figure 3-5A**) [11]. In *Ma* Cas6-IV the GBE forms a β-hairpin. In RNA-bound structures of other Cas6 proteins, the GBE typically makes sequence and shape specific contacts within the major groove of the crRNA stem-loop [21,24,25,40]. An alignment of 184 Cα carbons with the RNA-bound TthCas6A protein from Thermus thermophilus (PDB 4C8Z, RMSD 2.45 Å) positions the GBE from *Ma* Cas6-IV into the major groove of the bound RNA, suggesting that the GBE of *Ma* Cas6-IV binds the major groove of the crRNA stem-loop (**Figure 3-5B** and **Figure 3-9**). The β-hairpin motif occurs in the majority of Cas6 enzymes, and typically contacts the base of the crRNA stem-loop and positions the scissile phosphate into the active site [21,25,28]. Our alignment with the RNA-bound TthCas6A shows the tip of the *Ma* Cas6-IV β-hairpin pointing away from the scissile phosphate, suggesting this feature may undergo a conformational change upon binding the crRNA [25]. The G-loop, in the *Ma* Cas6-IV structure resides between α_2 and β_4 and forms a small loop-helix-loop structure. In other Cas6 enzymes, this motif is involved in binding the crRNA through ionic interactions along the phosphate backbone [23–25,41]. The G-loop in *Ma* Cas6-IV contains two lysine residues that help form a large positively charged patch on the surface of the protein (**Figure 3-5B**). In our alignment, the crRNA of the RNA-bound structure is positioned on top of this positively charged surface, suggesting the lysines within the G-

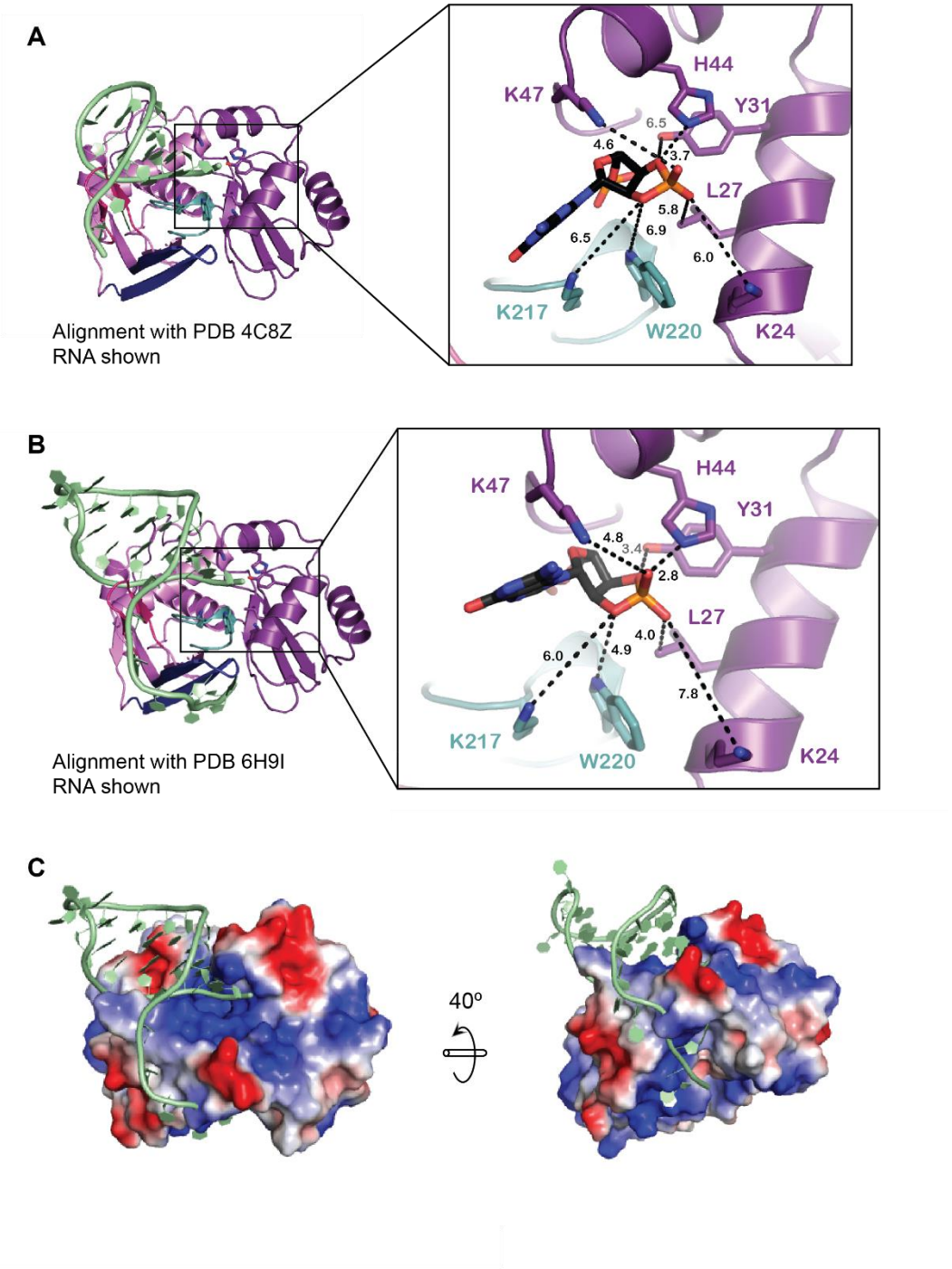

Figure 3-9. Overlays of *Ma* Cas6 with Cas6 homologs bound to RNA. (**A**) Overlay of the RNA of PDB 4C8Z. Inset shows residues in putative active sites and distances to 2'-3' cyclic phosphate. (**B**) Second overlay with RNA from Type IV Cas-homolog Csf5. Putative active site is shown in the inset. (**C**) Overlay of Csf5 on *Ma* Cas6 rendered as an electrostatic surface.

loop of *Ma* Cas6-IV may form ionic interactions with the negatively charged backbone of the CRISPR IV crRNA. Interestingly, this positive patch extends towards the back of the C-terminal RRM, suggesting a possible trajectory of the crRNA that wraps around the protein as seen in Type III-A Cas6 enzymes (**Figure 3-5B** bottom and **Figure 3-9**) [26,28,34].

His44 and Tyr31 are Catalytic Residues of the *Ma* Cas6-IV Active Site

Cas6 nuclease active sites are typically located in the cleft between the two RRM folds [11] but, without a substrate bound, our structure of apo *Ma* Cas6-IV was not sufficient to determine the location of the *Ma* Cas6-IV active site. In our alignment with TthCas6A bound to RNA, *Ma* Cas6-IV residues His44 and Tyr31 are located within 4 Å of the aligned scissile phosphate, suggesting they could be involved in catalysis. Indeed, in the TthCas6A protein, a histidine in this region was shown to catalyze cleavage [24], and histidine residues in this region have been shown to have a role in catalysis of several other Cas6 enzymes [20,23,25–27]. To determine if the *Ma* Cas6-IV residues His44 and Tyr31 are responsible for catalysis of RNA cleavage, we created alanine mutants of each residue, alone and in tandem, and assayed their ability to cleave Repeat IV. Mutation of either residue to alanine severely reduced cleavage (**Figure 3-5C** and **3-5D**). Previous analyses of Cas6 catalytic activity showed that, in some cases, nuclease activity lost upon histidine mutation could be restored with the addition of imidazole [24]. To better understand how the histidine catalyzes cleavage, we added 500 mM imidazole to our H44A *Ma* Cas6-IV mutant cleavage reaction. Under these conditions, H44A *Ma* Cas6-IV partially regained its cleavage activity (**Figure 3-5C** and **3-5D**), suggesting the imidazole

ring compensates for the histidine mutant, consistent with a model where the active site histidine plays a role in catalysis [20,42].

***Ma* Cas6-IV is a Single-Turnover Enzyme to the CRISPR Repeat RNA Substrate**

Many Cas6 enzymes exhibit single-turnover characteristics, remaining bound to their cleaved crRNA products, while others dissociate from cleaved crRNAs allowing for multi-turnover activity [11]. To determine the turnover number of the *Ma* Cas6-IV enzyme, we performed nuclease assays with constant concentrations of Repeat IV and varying concentrations of *Ma* Cas6-IV corresponding to the following ratios of Cas6 to RNA substrate; 2:1, 1:1, 1:2, 1:4, and 1:8. In both the 2:1 and 1:1 conditions nearly all of the RNA substrate was cleaved. In each successive condition, a cleavage amplitude was reached where less total RNA was cleaved than in the previous condition (**Figure 3-10**). When excess substrate is present, cleavage does not go to completion, but rather corresponds to the cleavage of one RNA molecule per Cas6 active site. Thus, we conclude that *Ma* Cas6-IV is a single-turnover enzyme.

Structural Comparison of *Ma* Cas6-IV to Csf5

Although Type IV systems are similar in gene arrangement and identity, Type IV Cas6 proteins are unexpectedly diverse in sequence (**Figure 3-11** and **Table 3-1**). Type IV systems contain several Cas6 variants, including sequences that resemble Cas6e and Cas6f, and others that encode a Type IV-specific Cas6 homolog called Csf5 [13]. Structural comparisons of *Ma* Cas6-IV with Csf5 from *Aromatoleum aromaticum* reveal they both contain the dual RRM domain scaffold generally observed in Cas6 proteins

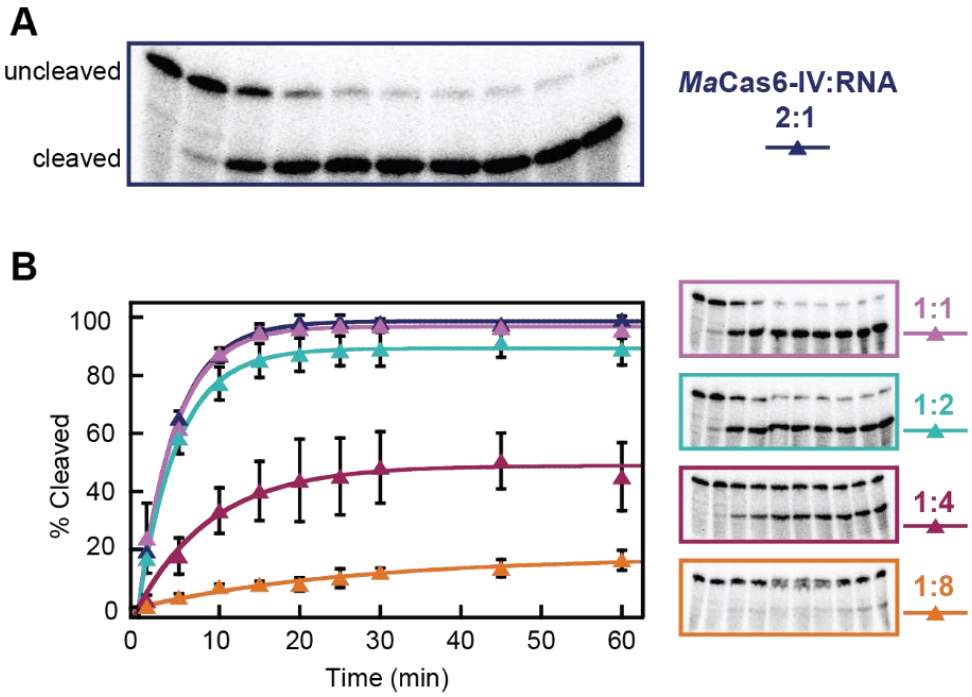

Figure 3-10. *Ma* Cas6-IV exhibits single-turnover characteristics for Repeat IV. (**A**) Cleavage of Repeat 4 by *Ma* Cas6-IV with a Cas6:Repeat IV ratio of 2:1. (**B**) Isotherms and electrophoresis gels of additional ratios of Cas6:Repeat are shown. The data were fit as described in (Niewoehner et al., 2014). Error bars denote standard deviation between three experiments.

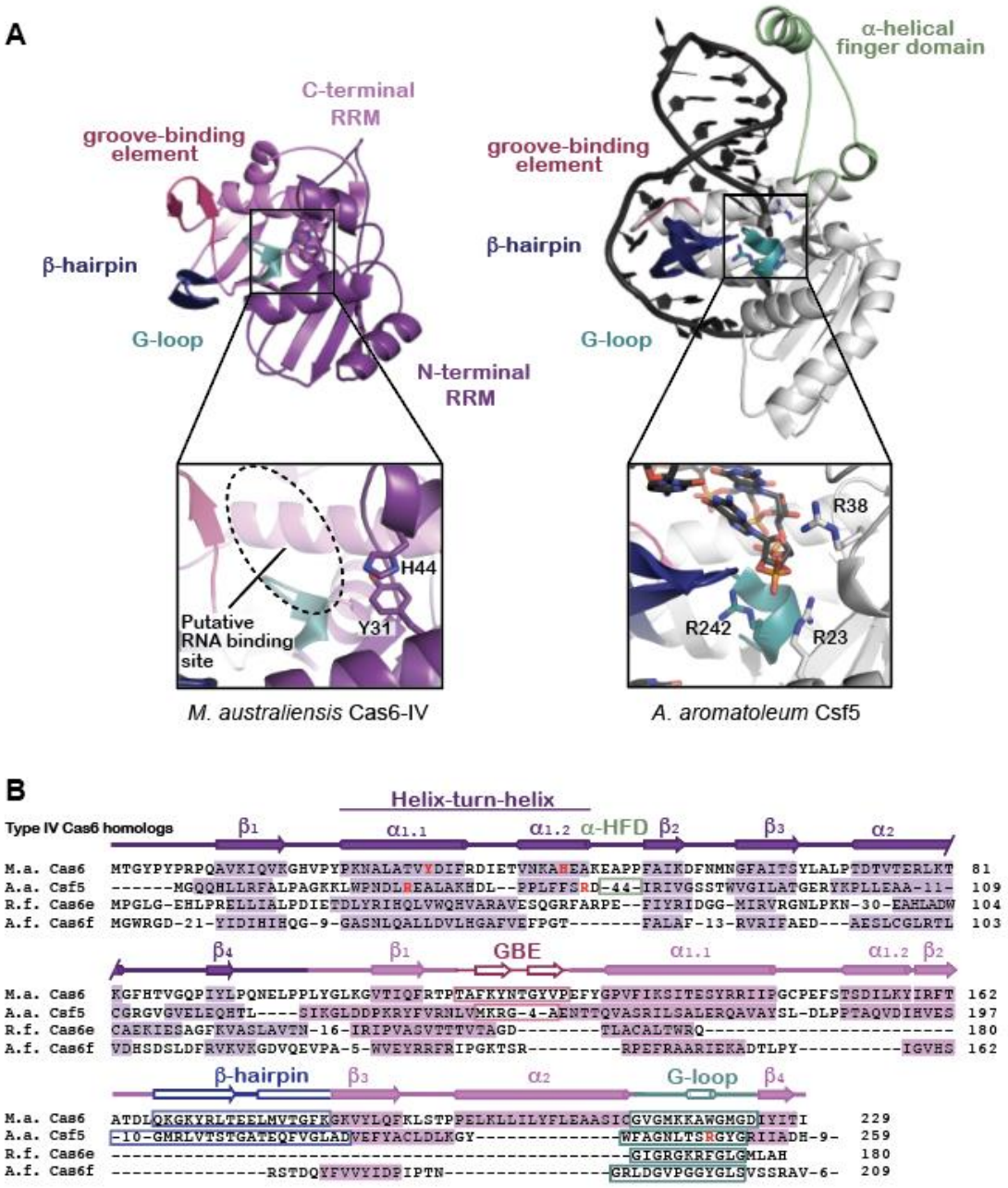

Figure 3-11. Structure and sequence alignments of *Ma* Cas6-IV with other Type IV RNA endonucleases. (**A**) A structural comparison of *Ma* Cas6-IV with the Cas6-homolog Csf5 from *Aromatoleum aromaticum* (PDB 6H9I). Features involved in binding crRNAs are indicated. The Csf5 protein contains a large insert called the alpha-helical finger domain (light green) that is not observed in *Ma* Cas6-IV. Residues predicted to activate cleavage of the crRNA are indicated in the inset below. (**B**) Sequence alignment of *Ma* Cas6-IV with other RNA endonucleases observed in Type IV systems. The N- and C-terminal RRM secondary structure elements are indicated, as well as features which

bind crRNA, including the groove-binding element (GBE), beta-hairpin, and glycine-rich loop (G-loop). The alpha-helical finger domain (α-HFD) insert of Csf5 is also indicated. Active site residues of *Ma* Cas6-IV and Csf5 are bolded in red. Cas6e and Cas6f sequences are noticeably shorter than *Ma* Cas6-IV and Csf5, lacking large portions of the C-terminal RRM.

(**Figure 3-11A**). The C-terminal RRM domains of both enzymes contain the motifs that bind crRNA (GBE, β-hairpin, and G-loop), but the C-terminal domain of Csf5 differs from *Ma* Cas6-IV in that the second alpha helix (α_2) of the canonical RRM fold is absent (**Figure 3-8**). In both Csf5 and *Ma* Cas6-IV the α1 helices of the N-terminal RRM domains have been replaced with helix-turn-helix motifs that house putative active-site residues (**Figure 3-11B**). However, instead of the small loop sequence observed in *Ma* Cas6-IV that connects the helix-loop-helix to β_2, Csf5 has an insertion of ~40 amino acids called the α-helical finger domain that contains two additional helices. One of these helices interacts with the minor groove of the crRNA stem-loop, providing additional contacts for binding the crRNA that may provide additional specificity toward Type IV crRNA repeats.

Interestingly, despite structural similarity the mechanisms of cleavage utilized by these protein homologs are diverse. Both Csf5 and *Ma* Cas6-IV contain a histidine in the N-terminal RRM at the same sequence position (H44). In *Ma* Cas6-IV the histidine resides in the helix-turn-helix and is within H-bonding distance of the scissile phosphate (**Figure 3-9**). However, in Csf5 this histidine resides within the 40 amino acid insert several Ångstroms away from the scissile phosphate, suggesting it does not participate in nuclease activity. In support of this hypothesis, an H44A mutant of Csf5 did not impair cleavage [35]. Rather, mutation of arginine residues located on the helix-turn-helix and the G-loop (R23A, R38A, R242A) impaired cleavage (**Figure 3-11A**). Notably several of these arginines are located in similar positions to the active site residues of *Ma* Cas6-IV (His44 and Tyr31) (**Figure 3-11B** and **Figure 3-9**), revealing that although these Type IV

crRNA processing enzymes are diverse in sequence, they rely on similar structural themes to bind and cleave crRNA substrates.

DISCUSSION

Because Type IV CRISPR systems have just recently been discovered, the biological function of these systems remains largely unknown. We hypothesized that pre-crRNA transcripts from adjacent CRISPR loci would be processed into small crRNA-guides to direct a biological function. To understand the feasibility and mechanism of Type IV crRNA biogenesis, we characterized the structure and activity of a Type IV-A Cas6 crRNA endonuclease from the microbe *Mahella australiensis* (*Ma* Cas6-IV). We observed that *Ma* Cas6-IV processes pre-crRNAs in vitro by cleaving within the CRISPR repeat sequence, through a mechanism that is metal-independent and facilitated by residues His44 and Tyr31. Analysis of repeat cleavage products suggests that the repeat is cleaved asymmetrically on the 3' side of the predicted stem loop with the 5'-phosphate of the 23rd nucleotide as the scissile phosphate. We show that *Ma* Cas6-IV cleaves crRNAs with single-turnover kinetics, suggesting the enzyme stays bound after cleavage to one side of the repeat. Binding assays and crystal structures show that in several Type I systems Cas6 remains bound to its crRNA product, forming part of the interference complex that uses the crRNA as a guide to bind complementary targets [30–32,43]. We hypothesize that the single-turnover kinetics we observe are due to *Ma* Cas6-IV remaining bound to the 5' side of its cleaved crRNA product, suggesting *Ma* Cas6-IV may be a component of a multi-subunit complex that assembles on the processed crRNA. Alternatively *Ma* Cas6-IV could undergo a conformational change upon release of both

sides of the crRNA that does not allow for additional repeats to be cleaved. Further work is needed to determine how *Ma* Cas6-IV interacts with the repeat cleavage products and whether it forms a part of the predicted multi-subunit complex.

Phylogenetic work previously predicted that multi-subunit complexes would form in Type IV-A systems and would be composed of Csf2 (Cas7-like backbone protein), Csf3 (Cas5-like tail protein) and the Type IV signature protein Csf1 [6,10,13]. Supporting this prediction, Özcan and colleagues recombinantly expressed and isolated a Type IV-A complex containing these exact subunits and the Cas6-homolog Csf5 [35]. This recent work, coupled with our observations, suggest that in Type IV-A systems Cas6 mediates cleavage of the pre-crRNA, creating small crRNA guides that are then bound by Csf1, Csf2, and Csf3 to form a multi-subunit crRNA-guided complex. It appears that in some cases Cas6 forms a part of the multi-subunit complex, but it remains unknown if this is a conserved feature of all Type IV-A systems.

The function of Type IV systems also remains unknown. A recent phylogenetic analysis reported that many Type IV system spacers are complementary to known phage sequences [6], suggesting the system acts as an immune system. However, analysis of the spacers of *Mahella australiensis* using the CRISPRTarget tool [44] reported no plausible matches, with only three spacer sequences retaining some complementarity to phage sequences, but with multiple mismatches (**Figure 3-12**). Even if it were clear that Type IV systems target phage or plasmid sequences the function of such targeting would remain unknown, as no nuclease domains have yet been clearly identified in any of the Type IV systems. Thus, a more detailed analysis of the structure and function of these systems is needed.

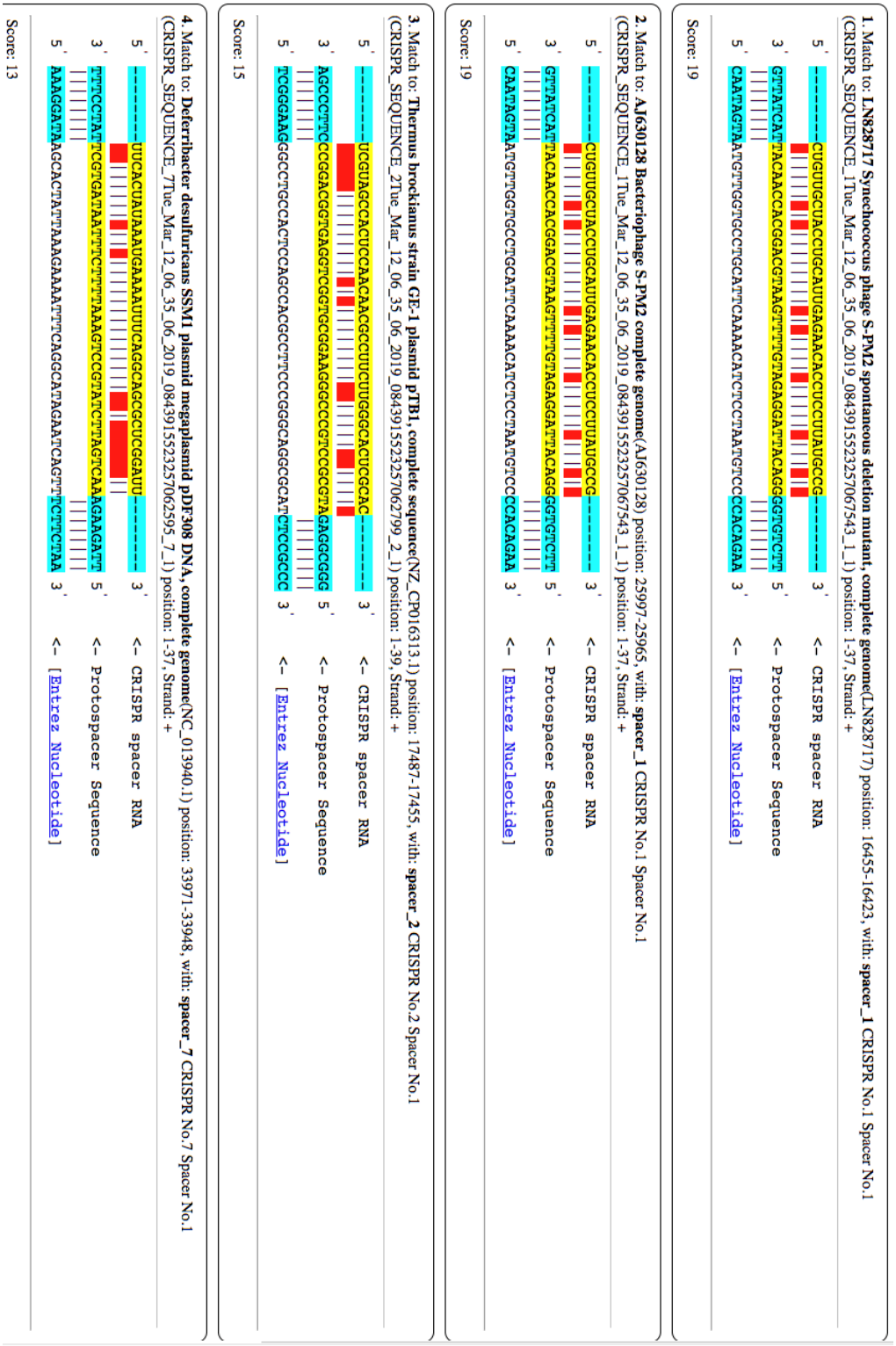

Figure 3-12. CRISPRTarget results with *Mahella australiensis* spacer sequences as search criteria. Matches 1 and 2 are of the same phage. Mismatches are highlighted in red.

In contrast to Type IV-A systems, Type IV-B systems do not encode Cas6 enzymes and are not associated with CRISPR loci. However, Type IV-B systems do encode the Csf1, Csf2, and Csf3 proteins, suggesting they may be capable of forming a complex on processed crRNA guides. It has been proposed that Type IV-B systems may utilize crRNAs generated from other CRISPR systems to form complexes and may have a non-defense function [45]. However, it has yet to be shown that Type IV-B proteins can form a complex on a crRNA.

Cas6-like RNA endonucleases associated with Type IV-A systems are quite diverse in amino acid sequence. A small group of systems encode for the Type IV-specific Csf5 protein, but others encode variants of Cas6 that are dissimilar in sequence but retain the dual RRM architecture observed in Cas6 proteins (**Figure 3-11B**). The wide range of Cas6 sequences across Type IV-A systems suggests that perhaps Cas6 enzymes were acquired by Type IV-B systems at different times in evolutionary history. We postulate that such fusions also included a CRISPR locus, providing a crRNA-guide upon which the Csf1, Csf2, and Csf3 proteins could assemble. Interestingly, the protein for which *Ma* Cas6-IV is most similar in structure and sequence is an orphaned Cas6-homolog from *Thermus thermophilus* (TthCas6A) (**Figure 3-5**, **Figure 3-3**, and **Table 3-1**). TthCas6A is not found within a CRISPR system but is adjacent to a CRISPR locus, suggesting it could combine with other systems in tandem with its CRISPR locus. We speculate that perhaps a similar orphaned Cas6 sequence combined with a Type IV-B system to produce the Type IV-A system of *Mahella australiensis.*

This work and a recent study on the Type IV system from *A. aromaticum* [35] have displayed that Type IV-A systems process crRNA guides through a Cas6-mediated

mechanism. Additionally it has been shown that multi-protein complexes assemble onto these crRNAs [35]. However, the biological function of both Type IV-A and Type IV-B systems remains unknown and awaits additional biochemical, *in vivo*, and structural studies.

METHODS

Expression and Purification of *Ma* Cas6-IV

The full-length *M. australiensis* Cas6 gene sequence (Uniprot AEE97687.1) was obtained from IDT (Integrated DNA Technologies) as a gBlock. The sequence was PCR amplified with primers containing LIC (Ligation Independent Cloning) overhangs, and was subcloned, using ligation independent cloning, into the 2B-T transfer vector. *Ma* Cas6-IV mutants were created using the Q5 Site-Directed Mutagenesis kit (New England BioLabs (NEB)).

Recombinant protein expression was induced in BL21 DE3 cells with 0.1 mM of IPTG (isopropyl β-D-1-thiogalactopyranoside) at an optical cell density (O. D. 600 nM) of ~0.3-0.4. After induction, cells were grown at 16°C for 24 hours and then pelleted via centrifugation. Pelleted cells were added at a ratio of 1:8 (g of cells : mL of buffer) to 25mM $NaPO_4$ (pH 7.5), 500mM NaCl, 25mM imidazole, 5% glycerol, 0.01% Triton X-100, 1 mM Tris [2-carboxyethyl] phosphine hydrochloride (TCEP), 0.5 mM phenylmethylsulphonyl fluoride (PMSF), and 0.1 mM lysozyme. Cells were lysed by sonication. Sonicated lysate was clarified via centrifugation. Clarified lysate was placed over a HisTrap FF column (GE Healthcare) and the bound protein was eluted with a high imidazole buffer (25mM $NaPO_4$ (pH 7.5), 500 mM NaCl, 700 mM imidazole, 5%

glycerol, 1 mM TCEP) then desalted into 25 mM $NaPO_4$ (pH 7.5), 150 mM NaCl, 1mM TCEP, 5% glycerol with a HiPrep 26/10 desalting column (GE healthcare). Desalted protein was placed over a HiTrap SP FF column (GE healthcare) and eluted with a NaCl gradient. The protein was further purified with a Superdex 200 pg 26/600 column (GE Healthcare) and eluted in 100 mM HEPES (pH 7.5), 150 KCl, 5% glycerol, 1 mM TCEP. The protein was then concentrated to 6-8 mg/ml using Vivaspin centrifugation concentrators. Wild type and mutant protein constructs were expressed and purified with the same protocol.

Generation of RNA Substrates

The *Ma* pre-crRNA was generated using *in vitro* transcription with the HiScribe T7 Quick kit (NEB) and a plasmid linearized immediately after the encoded *Ma* CRISPR under control of a T7 promoter. All small RNA oligonucleotides were synthesized by IDT with or without a 3' fluorescein label. Radiolabeled substrates were 5'-end labeled with (γ-^{32}P)-ATP (Perkin Elmer) and T4 polynucleotide kinase (NEB). Labeled RNAs were separated from excess ATP with a MicroSpin G-25 column (GE Healthcare), then gel purified on a 12% denaturing (7M urea) polyacrylamide gel, ethanol precipitated, and recovered in water.

Pre-crRNA Cleavage Assay

15 µM Cas6 and 0.25 µM *Ma* pre-crRNA were incubated at 50°C in 20 mM HEPES (pH 7.5), 100 mM KCl, 1 mM β-Mercaptoethanol, and 50 mM EDTA. 10 µl aliquots were removed at indicated timepoints and quenched with 50 µl acid phenol

chloroform and briefly centrifuged. Eight µl of the aqueous layer was mixed with formamide RNA loading buffer and resolved on a 12% denaturing (7M urea) polyacrylamide gel with a Low Range ssRNA ladder (NEB). The gel was stained with SYBR Gold Nucleic Acid Gel stain (Invitrogen) and imaged with a ChemiDoc MP Imaging system (Bio-Rad).

Repeat Cleavage and Turnover Assays

2.5 µM Cas6 and 5 nM 5'-end ^{32}P or 1 µM 3'-end fluorescein labeled RNA substrate were incubated at 50°C in 20 mM HEPES (pH 7.5), 100 mM KCl, 1 mM β-Mercaptoethanol, and 50 mM EDTA. Time points were collected and resolved on a gel as described for the pre-crRNA cleavage assay. Fluorescein labeled substrates were imaged and quantified with a ChemiDoc MP Imaging system (Bio-Rad). Gels containing ^{32}P labeled substrates were dried, exposed to a phosphor storage screen, and scanned with a Typhoon (GE Healthcare) phosphorimager. Cleaved and uncleaved fractions were quantified using ImageQuant (GE Healthcare) software. All data were fit as described by Niewoehner and colleagues [24]. Briefly, using Kaleidagraph (Synergy Software), the cleavage assay data were fit to the equation:

$$fraction\ cleaved = curve\ amplitude \times (1 - \exp(-rate\ constant \times time))$$

Reported data is the average of three experiments and error bars represent standard deviations. To recover the cleavage activity of the *Ma* Cas6-IV H63A mutant, 500 mM imidazole was added to the reaction buffer.

The turnover kinetics of *Ma* Cas6-IV were tested by performing cleavage assays as described above, except with 1000 nM substrate (5 nM of which was radiolabeled) and varying concentrations of *Ma* Cas6-IV (2000, 1000, 500, 250, and 125 nM).

Crystal Growth and Structure Determination

Ma Cas6-IV was crystallized at 22°C using the sitting drop vapor diffusion method. Drops were made by mixing equal volumes of protein (6-8 mg/ml in 100 mM HEPES (pH 7.5), 150 KCl, 5% glycerol, 1 mM TCEP) and reservoir solution (0.1 M sodium acetate trihydrate (pH 4.8-5.1) and 1.60 - 1.75 M ammonium sulphate). Most crystals grew to a size large enough for data collection in about a week.

As initial attempts at solving the crystal structure using molecular replacement with existing models failed, and because we were unable to grow crystals of appreciable size with selenomethionine derivatized protein, we solved the structure by soaking in heavy metals. Crystals were soaked in tetrachloroplatinate overnight and then back soaked into cryo-protectant containing the mother liquor solution and 30% dextrose followed by flash freezing. Native and anomalous diffraction data were collected remotely at SSRL beamline 12-2 with high redundancy and completeness (Table 3-1). The Phenix tool autosol was used to identify an initial substructure, phase the data, density modify initial maps, and autobuild an initial model [46–50]. Although incomplete, the initial autobuild model was sufficient for molecular replacement for phasing the native data [51]. Iterative rounds of model building and refinement using Coot and Phenix.refine tools produced the final model [39,52].

Structure and Sequence Alignments

Ma Cas6-IV was aligned with Cas6, and Cas5d models available in the PDB using the SSM (Secondary Structure Matching) tool in Coot [39]. After alignment Coot spits out statistics that list the number of Cα carbons aligned, the R.M.S.D. of the alignment, and the percentage of residues aligned that are identical (see Table B1). Alignment of diverse Type IV Cas6 sequences that had structures available (*Ma* Cas6-IV and *Aa* Csf5) was done in Coot using SSM. However the Cas6e and Cas6f sequences were aligned manually using secondary structure predicted by the PSIPRED tool as a guide [53].

ACKNOWLEDGMENTS

Research in the Jackson Lab is supported by Utah State University New Faculty Start-up funding from the Department of Chemistry and Biochemistry, the Research and Graduate Studies Office, and the College of Science. The Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory is supported by the U.S. Department of Energy, Contract No. DE-AC02-76SF00515, and the National Institutes of Health NIGMS (P41GM103393).

DATA DEPOSTITION

The model coordinates and structure factors for the apo *Ma* Cas6-IV structure have been deposited in the Protein Data Bank under PDB code: 6NJY.

REFERENCES

[1] Brouns SJJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJH, Snijders APL, et al. Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. Science 2008;321:960–4. doi:10.1126/science.1159689.

[2] Barrangou R. Diversity of CRISPR-Cas immune systems and molecular machines. Genome Biology 2015;16:247. doi:10.1186/s13059-015-0816-9.

[3] Wright AV, Nuñez JK, Doudna JA. Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. Cell 2016;164:29–44. doi:10.1016/j.cell.2015.12.035.

[4] Marraffini LA. CRISPR-Cas immunity in prokaryotes. Nature 2015;526:55–61. doi:10.1038/nature15386.

[5] Mohanraju P, Makarova KS, Zetsche B, Zhang F, Koonin EV, van der Oost J. Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. Science 2016;353:aad5147. doi:10.1126/science.aad5147.

[6] Koonin EV, Makarova KS, Zhang F. Diversity, classification and evolution of CRISPR-Cas systems. Current Opinion in Microbiology 2017;37:67–78. doi:10.1016/j.mib.2017.05.008.

[7] Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, et al. CRISPR Provides Acquired Resistance against Viruses in Prokaryotes. Science, New Series 2007;315:1709–12.

[8] Gasiunas G, Sinkunas T, Siksnys V. Molecular mechanisms of CRISPR-mediated microbial immunity. Cellular and Molecular Life Sciences 2014;71:449–65. doi:10.1007/s00018-013-1438-6.

[9] Jackson RN, van Erp PB, Sternberg SH, Wiedenheft B. Conformational regulation of CRISPR-associated nucleases. Current Opinion in Microbiology 2017;37:110–9. doi:10.1016/j.mib.2017.05.010.

[10] Makarova KS, Wolf YI, Koonin EV. Classification and Nomenclature of CRISPR-Cas Systems: Where from Here? The CRISPR Journal 2018;1:325–36. doi:10.1089/crispr.2018.0033.

[11] Hochstrasser ML, Doudna JA. Cutting it close: CRISPR-associated endoribonuclease structure and function. Trends in Biochemical Sciences 2015;40:58–66. doi:10.1016/j.tibs.2014.10.007.

[12] Charpentier E, Richter H, van der Oost J, White MF. Biogenesis pathways of RNA guides in archaeal and bacterial CRISPR-Cas adaptive immunity. FEMS Microbiology Reviews 2015;39:428–41. doi:10.1093/femsre/fuv023.

[13] Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ, et al. An updated evolutionary classification of CRISPR–Cas systems. Nature Reviews Microbiology 2015;13:722–36. doi:10.1038/nrmicro3569.

[14] Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature 2011;471:602–7. doi:10.1038/nature09886.

[15] Shmakov S, Abudayyeh OO, Makarova KS, Wolf YI, Gootenberg JS, Semenova E, et al. Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. Molecular Cell 2015;60:385–97. doi:10.1016/j.molcel.2015.10.008.

[16] Fonfara I, Richter H, Bratovič M, Le Rhun A, Charpentier E. The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. Nature 2016;532:517–21. doi:10.1038/nature17945.

[17] East-Seletsky A, O'Connell MR, Knight SC, Burstein D, Cate JHD, Tjian R, et al. Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. Nature 2016;538:270–3. doi:10.1038/nature19802.

[18] Carte J, Wang R, Li H, Terns RM, Terns MP. Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. Genes & Development 2008;22:3489–96. doi:10.1101/gad.1742908.

[19] Carte J, Pfister NT, Compton MM, Terns RM, Terns MP. Binding and cleavage of CRISPR RNA by Cas6. RNA 2010;16:2181–8. doi:10.1261/rna.2230110.

[20] Haurwitz RE, Sternberg SH, Doudna JA. Csy4 relies on an unusual catalytic dyad to position and cleave CRISPR RNA: Mechanism of CRISPR RNA cleavage. The EMBO Journal 2012;31:2824–32. doi:10.1038/emboj.2012.107.

[21] Haurwitz RE, Jinek M, Wiedenheft B, Zhou K, Doudna JA. Sequence- and Structure-Specific RNA Processing by a CRISPR Endonuclease. Science 2010;329:1355–8. doi:10.1126/science.1192272.

[22] Sternberg SH, Haurwitz RE, Doudna JA. Mechanism of substrate selection by a highly specific CRISPR endoribonuclease. RNA 2012;18:661–72. doi:10.1261/rna.030882.111.

[23] Jesser R, Behler J, Benda C, Reimann V, Hess WR. Biochemical analysis of the Cas6-1 RNA endonuclease associated with the subtype I-D CRISPR-Cas system in

Synechocystis sp. PCC 6803. RNA Biology 2018:1–11. doi:10.1080/15476286.2018.1447742.

[24] Niewoehner O, Jinek M, Doudna JA. Evolution of CRISPR RNA recognition and processing by Cas6 endonucleases. Nucleic Acids Research 2014;42:1341–53. doi:10.1093/nar/gkt922.

[25] Sashital DG, Jinek M, Doudna JA. An RNA-induced conformational change required for CRISPR RNA cleavage by the endoribonuclease Cse3. Nature Structural & Molecular Biology 2011;18:680–7. doi:10.1038/nsmb.2043.

[26] Wang R, Preamplume G, Terns MP, Terns RM, Li H. Interaction of the Cas6 Riboendonuclease with CRISPR RNAs: Recognition and Cleavage. Structure 2011;19:257–64. doi:10.1016/j.str.2010.11.014.

[27] Lee M, Tseng S, Yang J, Hsieh T, Wu S, Kuan S, et al. Expression, Purification, Crystallization, and X-ray Structural Analysis of CRISPR-Associated Protein Cas6 from Methanocaldococcus jannaschii. Crystals 2017;7:344. doi:10.3390/cryst7110344.

[28] Shao Y, Li H. Recognition and Cleavage of a Nonstructured CRISPR RNA by Its Processing Endoribonuclease Cas6. Structure 2013;21:385–93. doi:10.1016/j.str.2013.01.010.

[29] Reeks J, Sokolowski RD, Graham S, Liu H, Naismith JH, White MF. Structure of a dimeric crenarchaeal Cas6 enzyme with an atypical active site for CRISPR RNA processing. Biochemical Journal 2013;452:223–30. doi:10.1042/BJ20130269.

[30] Jackson RN, Golden SM, Erp PBG van, Carter J, Westra ER, Brouns SJJ, et al. Crystal structure of the CRISPR RNA–guided surveillance complex from Escherichia coli. Science 2014;345:1473–9. doi:10.1126/science.1256328.

[31] Zhao H, Sheng G, Wang J, Wang M, Bunkoczi G, Gong W, et al. Crystal structure of the RNA-guided immune surveillance Cascade complex in *Escherichia coli*. Nature 2014;515:147–50. doi:10.1038/nature13733.

[32] Mulepati S, Héroux A, Bailey S. Crystal structure of a CRISPR RNA–guided surveillance complex bound to a ssDNA target. Science 2014;345:1479–84. doi:10.1126/science.1256996.

[33] Pausch P, Müller-Esparza H, Gleditzsch D, Altegoer F, Randau L, Bange G. Structural Variation of Type I-F CRISPR RNA Guided DNA Surveillance. Mol Cell 2017;67:622-632.e4. doi:10.1016/j.molcel.2017.06.036.

[34] Shao Y, Richter H, Sun S, Sharma K, Urlaub H, Randau L, et al. A Non-Stem-Loop CRISPR RNA Is Processed by Dual Binding Cas6. Structure 2016;24:547–54. doi:10.1016/j.str.2016.02.009.

[35] Özcan A, Pausch P, Linden A, Wulf A, Schühle K, Heider J, et al. Type IV CRISPR RNA processing and effector complex formation in Aromatoleum aromaticum. Nature Microbiology 2018. doi:10.1038/s41564-018-0274-8.

[36] Salinas MB. Mahella australiensis gen. nov., sp. nov., a moderately thermophilic anaerobic bacterium isolated from an Australian oil well. International Journal of Systematic and Evolutionary Microbiology 2004;54:2169–73. doi:10.1099/ijs.0.02926-0.

[37] Krissinel E, Henrick K. Inference of Macromolecular Assemblies from Crystalline State. Journal of Molecular Biology 2007;372:774–97. doi:10.1016/j.jmb.2007.05.022.

[38] Maris C, Dominguez C, Allain FH-T. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression: The RRM domain, a plastic RNA-binding platform. FEBS Journal 2005;272:2118–31. doi:10.1111/j.1742-4658.2005.04653.x.

[39] Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of *Coot*. Acta Crystallographica Section D Biological Crystallography 2010;66:486–501. doi:10.1107/S0907444910007493.

[40] Gesner EM, Schellenberg MJ, Garside EL, George MM, MacMillan AM. Recognition and maturation of effector RNAs in a CRISPR interference pathway. Nature Structural & Molecular Biology 2011;18:688–92. doi:10.1038/nsmb.2042.

[41] Wei W, Zhang S, Fleming J, Chen Y, Li Z, Fan S, et al. *Mycobacterium tuberculosis* type III-A CRISPR/Cas system crRNA and its maturation have atypical features. The FASEB Journal 2018:fj.201800557RR. doi:10.1096/fj.201800557RR.

[42] Lee HY, Haurwitz RE, Apffel A, Zhou K, Smart B, Wenger CD, et al. RNA-protein analysis using a conditional CRISPR nuclease. Proceedings of the National Academy of Sciences 2013;110:5416–21. doi:10.1073/pnas.1302807110.

[43] Pausch P, Müller-Esparza H, Gleditzsch D, Altegoer F, Randau L, Bange G. Structural Variation of Type I-F CRISPR RNA Guided DNA Surveillance. Molecular Cell 2017;67:622-632.e4. doi:10.1016/j.molcel.2017.06.036.

[44] Biswas A, Gagnon JN, Brouns SJJ, Fineran PC, Brown CM. CRISPRTarget. RNA Biology 2013;10:817–27. doi:10.4161/rna.24046.

[45] Faure G, Makarova KS, Koonin EV. CRISPR–Cas: Complex Functional Networks and Multiple Roles beyond Adaptive Immunity. Journal of Molecular Biology 2018. doi:10.1016/j.jmb.2018.08.030.

[46] Terwilliger TC, Adams PD, Read RJ, McCoy AJ, Moriarty NW, Grosse-Kunstleve RW, et al. Decision-making in structure solution using Bayesian estimates of map quality: the *PHENIX AutoSol* wizard. Acta Crystallographica Section D Biological Crystallography 2009;65:582–601. doi:10.1107/S0907444909012098.

[47] Grosse-Kunstleve RW, Adams PD. Substructure search procedures for macromolecular structures. Acta Crystallographica Section D Biological Crystallography 2003;59:1966–73. doi:10.1107/S0907444903018043.

[48] Terwilliger TC. Maximum-likelihood density modification. Acta Crystallographica Section D Biological Crystallography 2000;56:965–72. doi:10.1107/S0907444900005072.

[49] McCoy AJ, Storoni LC, Read RJ. Simple algorithm for a maximum-likelihood SAD function. Acta Crystallographica Section D Biological Crystallography 2004;60:1220–8. doi:10.1107/S0907444904009990.

[50] Terwilliger TC. Automated main-chain model building by template matching and iterative fragment extension. Acta Crystallographica Section D Biological Crystallography 2003;59:38–44. doi:10.1107/S0907444902018036.

[51] Bunkóczi G, Echols N, McCoy AJ, Oeffner RD, Adams PD, Read RJ. *Phaser.MRage* : automated molecular replacement. Acta Crystallographica Section D Biological Crystallography 2013;69:2276–86. doi:10.1107/S0907444913022750.

[52] Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, et al. Towards automated crystallographic structure refinement with *phenix.refine*. Acta Crystallographica Section D Biological Crystallography 2012;68:352–67. doi:10.1107/S0907444912001308.

[53] Buchan DWA, Minneci F, Nugent TCO, Bryson K, Jones DT. Scalable web services for the PSIPRED Protein Analysis Workbench. Nucleic Acids Research 2013;41:W349–57. doi:10.1093/nar/gkt381

CHAPTER 4

A TYPE IV-A CRISPR-CAS SYSTEM IN *PSEUDOMONAS AERUGINOSA* STRAIN PA83 MEDIATES RNA-GUIDED INTERFERENCE *IN VIVO*[3]

ABSTRACT

Bacteria and archaea use CRISPR-Cas adaptive immune systems to destroy complementary nucleic acids using RNAs derived from CRISPR loci. Here we provide the first functional evidence for Type IV CRISPR-Cas, demonstrating that the system from *Pseudomonas aeruginosa* strain PA83 mediates RNA-guided interference against a plasmid *in vivo*, both clearing the plasmid and inhibiting its uptake. This interference depends on the putative NTP-dependent helicase activity of Csf4/DinG.

INTRODUCTION

Prokaryotes use small non-coding RNAs derived from CRISPR (Clustered Regularly Interspaced Short Palindromic Repeat) arrays to guide CRISPR associated (Cas) proteins to invasive complementary nucleic acids such as phage and plasmid DNA for destruction[1–5].

Computational analyses have grouped CRISPR-Cas systems into two classes, 6 types and at least 33 subtypes[6]. The function of Class 1, Type IV systems, remains largely unknown, in part because this system is comparatively rare and considered a

---

[3] Co-authors: Valerie M. Crowley, Adam Catching, Hannah N. Taylor*, Adair L. Borges, Josie Metcalf, Joseph Bondy-Denomy, and Ryan N. Jackson.
Published: *The CRISPR Journal* (2019).
Final publication is available from Mary Ann Liebert, Inc.: http://dx.doi.org/10.1089/crispr.2019.0048
*HNT designed experiments, cloned plasmids, and provided critical manuscript feedback.

minimal CRISPR-Cas variant[6]. All Type IV systems contain genes predicted to encode a multi-subunit complex composed of large (*csf1, cas8*-like), backbone (*csf2, cas7*-like) and tail subunits (*csf3, cas5*-like). To date, Type IV systems have been subdivided into IV-A and IV-B subtypes. Type IV-A systems contain genes that encode for a putative helicase (*dinG*) and a *cas6*-like endonuclease, whereas Type IV-B system lack these genes and are often missing a CRISPR locus[6]. It has been shown that the Type IV-A Cas6 from *Mahella australiensis* and a Cas6-like homolog, Csf5 from *Aromatoleum aromaticum*, are involved in biogenesis of CRISPR derived RNAs (crRNA)[7,8]. Additionally, the Type IV-A proteins from *A. aromaticum* (AaCsf1, AaCsf2, AaCsf3 and AaCsf5) were shown to form a complex that assembles on a crRNA[8]. It has been hypothesized that the Type IV-A ribonucleoprotein complex enables RNA-guided target detection, the function of which remains unknown[8]. Here we investigate the Type IV CRISPR-Cas systems in *Pseudomonas aeruginosa* and demonstrate that a Type IV-A system mediates interference against plasmids in vivo.

MATERIALS AND METHODS

Bioinformatic Analysis

A Psi-Blast was performed on January 8, 2018 on the NCBI protein database with the seed WP_018940624.1 (Csf1), WP_013006553.1 (Csf2), and WP_018940624.1 (Csf3) of Thioalkalivibrio sp. K90Mix. Csf1, Csf2, and Csf3 returned 389, 1134, and 737 hits, of which 38, 48, and 41 hits were from *P. aeruginosa*, respectively. We proceeded by downloading amino acid sequences solely from *P. aeruginosa* for later bioinformatic analysis. Each hit from the Csf1 and Csf3 searches were found in the Csf2 search. In

addition, genes upstream and downstream of these hits were downloaded and compared by amino acid identity. Position relative to Csf2 hits was used to compare genes, revealing 4 conserved architectures. Each system, including intergenic regions and 7 genes in each direction of Csf2, in a given architecture was saved for further analysis. In some of these systems the downstream region of Csf3 contained putative CRISPR arrays already annotated, while the rest did not have annotation. An in-house script was written to search for non-identical repeats, with a search seed of the palindromic region CCCCGC or GCCGCC. By accounting for a hamming distance up to 3 between nearby putative repeats, unannotated regions were found to have CRISPR arrays. This was assisted with the inference of direction, using the rate of divergence from the consensus sequence over the array to find the conserved orientation between systems. Spacers between repeats were declared to have originated from a *Pseudomonas* plasmid if they appeared at least once on an annotated plasmid. Spacers were used to search for protospacers that originated from phages/prophages or plasmids. This was done using BLASTn to find matches in prokaryotes, which were then determined to be either located on a plasmid or a prophage (using the online tool PHASTER[9]). To rule out that BLASTn hits were spacer sequences in CRISPR arrays, the protospacer was confirmed to be within a gene, or where intergenic, confirmed to not neighbor a *csf* gene.

Accession Numbers

NZ_CP017294 – P. aeruginosa PA83 plasmid unnamed1

Annotation	Feature	GenPept	Genomic position
Putative ATP-dependent DNA helicase DinG	Csf4/DinG	PSA83_06667	299162..301341
Hypothetical protein	Csf5/Cas6	PSA83_06666	298385..299113 (complen
CRISPR type AFERR-associated protein Csf1	Csf1	PSA83_06665	297670..298401 (compler
CRISPR type AFERR-associated protein Csf2	Csf2	PSA83_06664	296619..297665 (complen
CRISPR type AFERR-associated protein Csf3	Csf3	PSA83_06663	295957..296619 (complen

CRISPR Array Used in Plasmid Transformation Efficiency and Maintenance Assays

We used the direct repeat found on the PA83 plasmid upstream of spacers that had hits to phage (5'-GTGTTCCCCGCATACGCGGGGGTGAACGG-3'). The CRISPR arrays used in our experiments contain a single spacer flanked by two of the same direct repeats. We evaluated two targeting spacers (TS1, TS2) and one non-targeting spacer (NTS). The spacer sequences we used in our experiments were TS1: 5'-TGGAGCAACACCTGAAGGAAGGCTTGATGAGC -3', TS2: 5-CTCAACCGAGGGTGGTTTTGTCTA-3' and NTS 5'-CTGAGTGTGATCGATGCCATCAGCGAAGGGCC-3'. The targeting spacers TS1 and TS2 target a sequence specific to the CAO1 gene from *Oryza sativa* (rice). These target sequences were picked for future experiments unrelated to this manuscript.

Plasmids

All *P. aeruginosa* genes and CRISPRs were synthesized by Twist Bioscience (San Francisco). PaCsf1was codon optimized for synthesis (see DNA sequence in Supplementary Table 1). Primers are listed in Supplementary Table 1. *csf1*, *csf2* and

csf5/*cas6* were placed on the same plasmid, with ribosome binding sites in between each gene, and was initially obtained in the pTWIST CMV expression vector. The CMV enhancer was removed and a T7 promoter was inserted upstream of the three gene polycistronic block, via two separate mutagenesis reactions. The cassette was PCR amplified and subcloned into pCDF using NcoI and PacI sites for origin of replication compatibility with other plasmids expressing the system. The CRISPR array with TS1, *csf3* and *csf4*/*dinG* were synthesized and inserted into the pTWIST Amp High Copy expression vector. The CRISPR was PCR amplified and subcloned into multiple cloning site 1 of pACYC via EcoRI and SacI. *csf3* and *csf4*/*dinG* were PCR amplified and subcloned into multiple cloning site 2 of pACYC containing the CRISPR via NdeI and EcoRV sites. CRISPR arrays with TS2 and NTS were subcloned into pACYC plasmids containing *csf3* and *csf4*/*dinG* via EcoRI and SacI. To test each targeting spacer independently, TS1 and TS2 were tested with distinct vectors.

Plasmid Transformation Efficiency Assay

Plasmid transformation efficiency assays were adapted from work previously described[10]. In brief, *E. coli* BL21-AI cells were co-transformed with pCDF-*csf1*-*csf2*-*csf5*/*cas6* and pACYC-Pa83CRTS1-*csf3*-*csf4*/*dinG* to reconstitute the components of the PA83 Type IV-A1 operon. The Type IV system genes and CRISPR were placed on two different plasmids as has been previously done to express multi-subunit class 1 systems[1,11,12]. This strategy makes it easier to clone, and later mutate systems containing multiple genes and a CRISPR locus and reduces the overall size of expression plasmids. This cell line was made chemically competent using standard methods. This cell line was

transformed with 25ng of target or non-target plasmid. The target and non-target plasmid are in pET27b(+), and the non-target differs from the target in that it is truncated and does not contain the protospacer sequence. After transformation with the target or non-target plasmid, the cells were grown overnight at 37°C in 5mL of LB-media supplemented with 1 mM IPTG, 0.2% L-arabinose, 50, 25 and 50 μg/ml of kanamycin, chloramphenicol and streptomycin respectively, shaking at 200 rpm. Serially diluted cells were plated on LB-agar plates supplemented with 1 mM IPTG, 0.2% L-arabinose, 50, 25 and 50 μg/ml of kanamycin, chloramphenicol and streptomycin respectively. The next day, colony forming units were counted for analysis. The error bars represent standard error of the mean calculated from three independent transformations. The experiment was performed with three biological replicates that produced the same results.

Plasmid Maintenance Assay

Plasmid maintenance assays were adapted from work previously described[10,13]. We transformed BL21-AI cells to generate individual cell lines expressing either the full or partial Type IV-A operons. We transformed these cells lines with an equimolar mixture of target and non-target plasmid. After transformation the cells were grown for 1 hour at 37°C in 500μL of LB-media supplemented with 1 mM IPTG and 0.2% L-arabinose for induction, shaking at 200 rpm. Each experiment contained an uninduced control for comparison, where cells were grown solely in LB. Induced cells were plated on LB-agar supplemented with 1 mM IPTG, 0.2% L-arabinose, 50, 25 and 50 μg/ml of kanamycin, chloramphenicol and streptomycin respectively and uninduced cells were plated on the same LB-agar conditions without IPTG or L-arabinose. The next day, 48

colonies were randomly screened for the presence of target or non-target plasmid by colony PCR using forward (5'-GAGTTCTGGCTGGCTAGCC-3') and reverse (5'-GGAATTGTGAGCGGATAACAA-3') primers that amplify a 679 bp region of the target or 384 bp of the non-target plasmid. PCR reactions were separated by agarose gel electrophoresis and stained with ethidium bromide. The number of PCR products corresponding to the target and non-target were counted and the data is expressed as a ratio of target to non-target bands. The error bars represent standard error of the mean calculated from three independent experiments. Values significantly below one indicate CRISPR interference. *csf4*/*dinG* mutants were made by site-directed mutagenesis using the pACYC-Pa83CRTS1-*csf3-csf4*/*dinG* plasmid as a template. All individual components of the Type IV operon were removed from either pCDF-*csf1-csf2-csf5*/*cas6* or pACYC-Pa83CRTS1-*csf3-csf4*/*dinG* plasmids via site-directed mutagenesis. All site-directed mutagenesis reactions were performed with the Q5® Site-Directed Mutagenesis Kit (New England BioLabs Inc.). Primers are listed in Supplementary Table 1.

RESULTS

Using the Type IV CRISPR-Cas system from *Thioalkalivibrio sp*. K90Mix[11] as a seed, we searched *P. aeruginosa* genomes for putative Type IV-A CRISPR-Cas systems using PSI-BLAST[12]. We observed 4 variants of Type IV-A systems based on amino acid sequence identity, gene arrangement, and CRISPR repeat sequence (**Figure 4-1A**, Supplementary Figure 1, Supplementary Table 2). Variant systems were grouped based on the shared sequence identity between conserved genes (*csf2*, *csf3*, *csf4*, *csf5*/*cas6*, Supplementary Figure 1). The architecture of the gene position and direction was

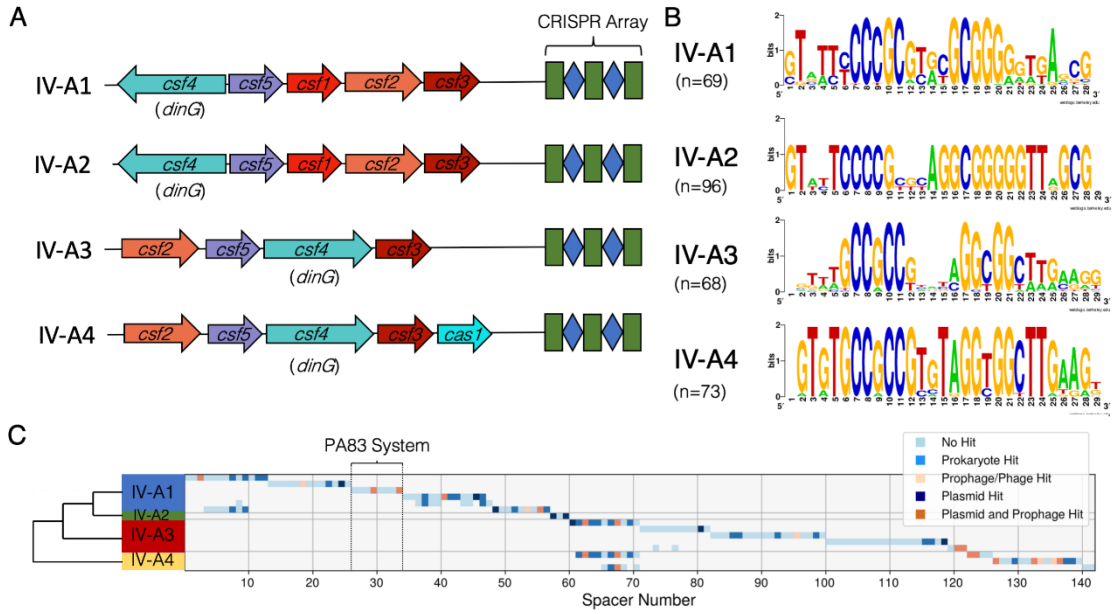

Figure 4-1. Type IV-A CRISPR-Cas systems found in *P. aeruginosa*. (**A**) Classification of Type IV-A CRISPR-Cas systems found in *P. aeruginosa*. Conserved architectures of these four variants is observed. All four variants include *csf2*, *csf3*, *csf4/dinG*, and *csf5*/*cas6*. Despite the similar architecture between Type IV-A1 and Type-IV-A2, they represent two separate groups based on differences in direct repeat sequences and amino acid divergence (Supplementary Figure 1). (**B**) Consensus sequences for direct repeats found in Type IV-A CRISPR arrays in *P. aeruginosa*. Palindromic regions are in the center of the repeat. (**C**) Table of spacers found in each system. Systems are grouped by variant, with group 2 consolidated, as all their CRISPR arrays are identical. Spacers are coloured based on mapping results according to the legend. A dendrogram was generated based on representative *csf2* sequences to illustrate the relationship between variants.

conserved within each variant (**Figure 4-1A**). Some variants had annotated CRISPR arrays, however, visual inspection of the region downstream of *csf3* found unannotated palindromic repeats. An in-house script was written to take mutant repeats into account so that CRISPRs containing repeats that differ slightly in sequence from one another were found. Using hamming distance of 3 mismatches between a system's repeats a total of 335 repeats were found, with their respective consensus sequences shown in **Figure 4-1B**. From identifying these repeats a total of 195 spacers were found, of which 142 were unique (**Figure 4-1C**). While redundant spacers were removed from consideration, all repeats were used to create consensus motifs (**Figure 4-1B**). Out of the 142 spacers, 52 are complementary to putative protospacers. 25 of these protospacers are found in prokaryotes without prophage or plasmid match, while 27 of the protospacers match either plasmids (8), prophages (5), or elements with signatures of both prophages and plasmids (14) (**Figure 4-1C**, Supplementary Tables 3 and 4).

Given that Type IV systems are often found on plasmids, we selected a plasmid-encoded system from *P. aeruginosa* strain PA83 as a model. This strain has a CRISPR array with a spacer matching a phage, suggesting it may be active. Owing to the large number of anti-CRISPR proteins found in *P. aeruginosa*[13], we investigated the interference function heterologously in *E. coli* using distinct plasmid transformation efficiency and maintenance assays.

In the plasmid transformation efficiency assay, we compared the number of colonies generated when the cells harboring the full *P. aeruginosa* PA83 Type IV-A1 operon were transformed with a target or non-target plasmid (**Figure 4-2A**). Successful inhibition of transformation was observed with the target plasmid, but not with the non-

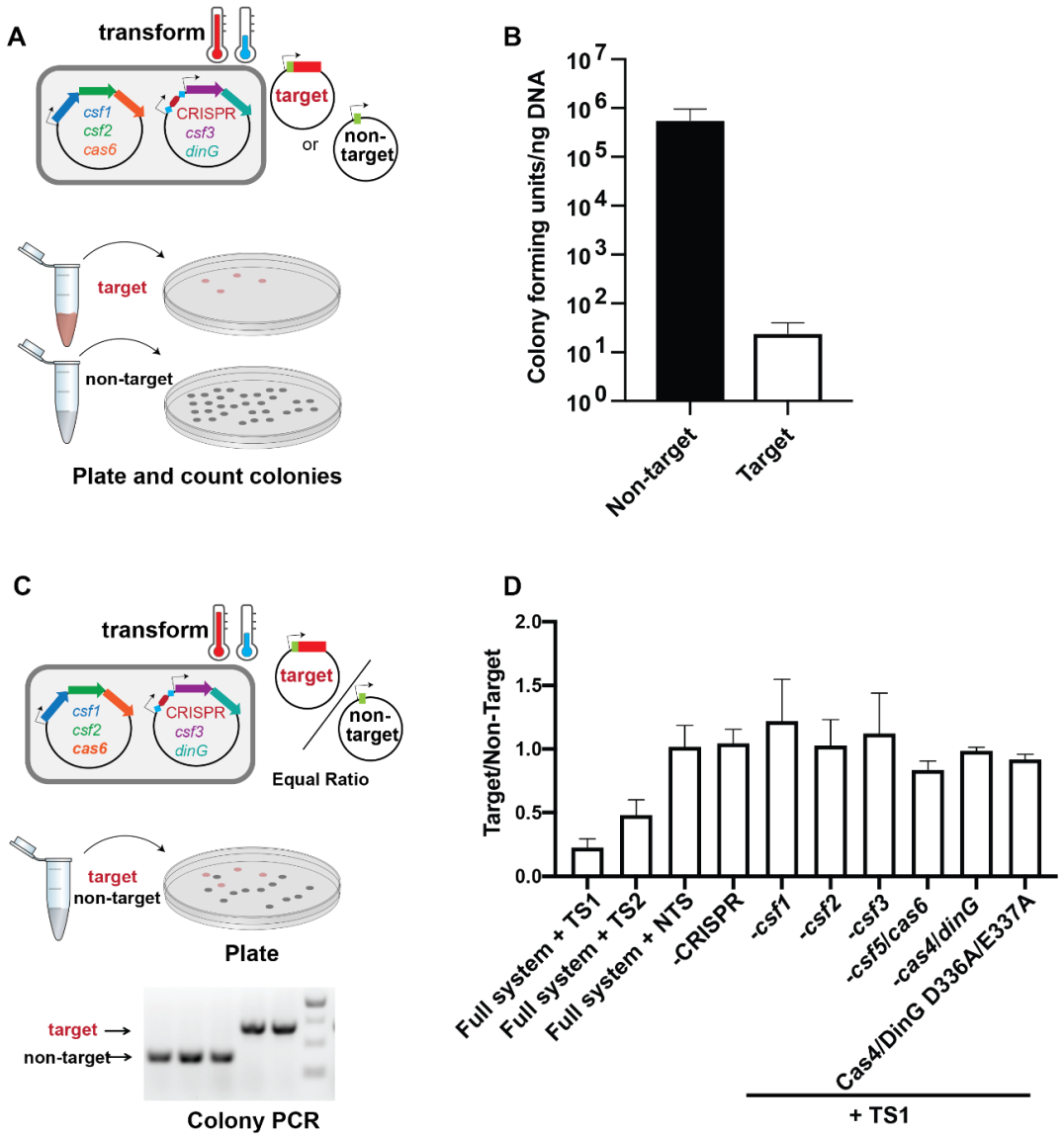

Figure 4-2. The Type IV-A1 CRISPR-Cas variant from *Pseudomonas aeruginosa* PA83 mediates RNA-guided interference *in vivo.* (**A**) Plasmid transformation efficiency assay described in the methods section. Small arrows on plasmids indicate a T7 promoter. (**B**) A reduction in plasmid transformation efficiency was observed when cells harboring a CRISPR array with TS1, *csf1*, *csf2*, *csf5*/*cas6*, *csf3*, *csf4*/*dinG* were transformed with the target compared to those transformed with the non-target plasmid. (**C**) Plasmid maintenance assay described in the methods section. Small arrows on plasmids indicate a T7 promoter. (**D**) An interference defect is observed when any component is removed and when the Csf4/DinG DEAH-box is mutated (D336A/E337A). TS1 was used in each deletion line and the Csf4/DinG mutant. As each condition represents a different competent cell line, an uninduced control was always included (Supplementary Figure 2). TS, targeting spacer. NTS, non-targeting spacer.

target plasmid, leading to an ~4 order of magnitude decrease in transformation efficiency (**Figure 4-2B**).

We next evaluated the interference requirements of the Type IV-A1 operon from *P. aeruginosa* PA83 by systematically deleting each component and evaluating interference with plasmid maintenance assays (**Figure 4-2C**). As each deletion required comparing different competent cell lines, we switched to plasmid maintenance assays to evaluate interference differences by PCR to reduce variation that could be attributed to differences between the competencies of prepared cell lines. In this experiment equimolar amounts of plasmids with target or non-target sequences were mixed and transformed into cells expressing the Type IV-A1 system. The ratio of colonies containing target vs non-target plasmid was assessed with colony PCR. A ratio less than 1 of target / non-target indicates interference (**Figure 4-2**). An interference deficiency was observed when any component of the Type IV-A1 operon was removed, including a point mutation in the genes encoding the putative DEAH-box helicase, Csf4/DinG (**Figure 4-2D**). DEAH-box helicases are proteins that unwind double-stranded nucleic acids using the energy released by NTP hydrolysis14. The putative activity of the DEAH-box of the Csf4/DinG helicase was inactivated by mutating the first two residues of the DEAH-box (D336A/E337A), and indeed, we found that these mutations resulted in an interference defect (**Figure 4-2D**). The target and non-target plasmids both contain a T7 promoter upstream of a 5'-CTTTC-3' sequence that lies adjacent to the protospacer. In the case of the target plasmid, this sequence is directly upstream of the protospacer. It is unclear at this time whether this is a protospacer adjacent motif (PAM) utilized by the Type IV-A1 system or if it is using another sequence in the vicinity of the protospacer. It

is also uncertain whether DNA or RNA is targeted, however the crRNA is not expected to base pair with the RNA transcribed from the target.

DISCUSSION

Here we provide the first functional evidence that a Type IV CRISPR-Cas system mediates RNA-guided interference against plasmids *in vivo* and that this activity requires the putative NTP-dependent helicase activity of Csf4/DinG.

Although we show Type IV CRISPR system interference requires the putative DinG helicase, the function of CRISPR associated DinG proteins remains unclear. However, studies on DinG proteins that are not associated with CRISPR systems show that related DinG helicases are involved in recombinational DNA repair and the re-initiation of replication after DNA damage[15]. Also it has been shown that non-CRISPR associated DinG unwinds R-and D-loops, forked substrates and 5' single-stranded overhangs with 5' to 3' polarity[15,17] allowing other proteins to access nucleic acid or modify nucleic acid structure. Additionally there is evidence that non-CRISPR associated DinG is recruited to disturbances in duplex DNA via changes in redox potential, and that in turn DinG recruits a nuclease to the disturbance site[20]. The Type IV associated and non-CRISPR associated DinG proteins are somewhat similar in sequence suggesting the CRISPR associated DinG protein may express similar functions. However such activities remain to be confirmed in CRISPR associated DinG proteins.

Unlike Type I systems, which encode for the helicase-nuclease Cas3, the protein sequences of the Type IV system appear to not contain any obvious nuclease domains. Thus, it is not clear how Type IV-A systems protect against plasmid targets and whether

the plasmid targets are cleaved. We speculate that similar to Type I and Type III systems, multi-subunit complexes composed of Csf proteins and a crRNA use the crRNA as a guide to bind complementary nucleic acid forming R-loops. We expect DinG, is recruited to the R-loops and then that DinG either acts directly to destroy the plasmid through an unknown mechanism or recruits an endogenous nuclease to mediate RNA-guided interference. However, these hypotheses remain to be tested. As Type IV-A systems lack a predicted effector nuclease and often lack a Cas1-Cas2 adaptation module, it has been reasoned that this CRISPR-Cas system cannot function as an independent adaptive immune system[18]. Here we show that a Type IV-A system from *P. aeruginosa* is capable of interference. However, it remains unknown if Type IV-A systems are capable of acquiring their own spacer sequences and whether DinG or the putative Csf crRNA complex are involved.

Here we provide the first evidence that a Type IV-A CRISPR-Cas system from *P. aeruginosa* mediates RNA-guided interference of plasmid DNA, and that such activity requires a putative multi-subunit crRNA-guided complex and a putative NTP-dependent helicase Csf4/DinG. However, many outstanding questions remain for this area of CRISPR-Cas biology: Do Type IV-A systems act alone as distinct RNA-guided adaptive immune systems[21,22] or does interference require additional host factors? Do Type IV-A systems associate with the acquisition machinery of other CRISPR-Cas systems to acquire immunity[21,22,8]? Do Type IV-A systems have roles apart from immunity altogether such as regulating gene expression or genome stability?

CONCLUSION

We conclude that the Type IV-A1 system from *P. aeruginosa* strain PA83 mediates RNA-guided interference against plasmids *in vivo*. This activity requires all components of the Type IV-A1 operon, including the accessory protein Csf4/DinG and its putative NTP-dependent helicase activity.

Supplementary material can be found at https://www.liebertpub.com/doi/10.1089/crispr.2019.0048

ACKNOWLEDGMENTS

The authors wish to thank two anonymous reviewers for their helpful comments during manuscript editing. Research in the Jackson Lab is supported by Utah State University RC grant, and New Faculty Start-up funding from the Department of Chemistry and Biochemistry, the Research and Graduate Studies Office, and the College of Science. The Bondy-Denomy Lab was supported by the UCSF Program for Breakthrough Biomedical Research funded in part by the Sandler Foundation, an NIH Director's Early Independence Award DP5-OD021344, and R01GM127489.

AUTHOR DISCLOSURE STATEMENT

J.B.-D. is a scientific advisory board member of SNIPR Biome and Excision Biotherapeutics and a scientific advisory board member and co-founder of Acrigen Biosciences. No additional competing financial interests exist.

REFERENCES

1. Brouns SJJ, Jore MM, Lundgren M, et al. Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. Science. 2008;321(5891):960-964. doi:10.1126/science.1159689
2. Barrangou R, Fremaux C, Deveau H, et al. CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. Science. 2007;315(5819):1709-1712. doi:10.1126/science.1138140
3. Barrangou R. Diversity of CRISPR-Cas immune systems and molecular machines. Genome Biol. 2015;16(1):247. doi:10.1186/s13059-015-0816-9
4. Marraffini LA. CRISPR-Cas immunity in prokaryotes. Nature. 2015;526(7571):55-61. doi:10.1038/nature15386
5. Jackson RN, van Erp PB, Sternberg SH, Wiedenheft B. Conformational regulation of CRISPR-associated nucleases. Curr Opin Microbiol. 2017;37:110-119. doi:10.1016/j.mib.2017.05.010
6. Makarova KS, Wolf YI, Koonin EV. Classification and Nomenclature of CRISPR-Cas Systems: Where from Here? CRISPR J. 2018;1(5):325-336. doi:10.1089/crispr.2018.0033
7. Taylor HN, Warner EE, Armbrust MJ, Crowley VM, Olsen KJ, Jackson RN. Structural basis of Type IV CRISPR RNA biogenesis by a Cas6 endoribonuclease. RNA Biol. doi:10.1080/15476286.2019.1634965
8. Özcan A, Pausch P, Linden A, et al. Type IV CRISPR RNA processing and effector complex formation in Aromatoleum aromaticum. Nat Microbiol. 2019;4(1):89-96. doi:10.1038/s41564-018-0274-8

9. Arndt D, Grant JR, Marcu A, et al. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res. 2016;44(W1):W16-W21. doi:10.1093/nar/gkw387

10. Almendros C, Guzmán NM, Díez-Villaseñor C, García-Martínez J, Mojica FJM. Target Motifs Affecting Natural Immunity by a Constitutive CRISPR-Cas System in Escherichia coli. Mokrousov I, ed. PLoS ONE. 2012;7(11):e50797. doi:10.1371/journal.pone.0050797

11. Jackson RN, Golden SM, van Erp PBG, et al. Crystal structure of the CRISPR RNA-guided surveillance complex from Escherichia coli. Science. 2014;345(6203):1473-1479. doi:10.1126/science.1256328

12. Gasiunas G, Sinkunas T, Siksnys V. Molecular mechanisms of CRISPR-mediated microbial immunity. Cell Mol Life Sci. 2014;71(3):449-465. doi:10.1007/s00018-013-1438-6

13. van Erp PBG, Jackson RN, Carter J, Golden SM, Bailey S, Wiedenheft B. Mechanism of CRISPR-RNA guided recognition of DNA targets in Escherichia coli. Nucleic Acids Res. 2015;43(17):8381-8391. doi:10.1093/nar/gkv793

14. Koonin EV, Makarova KS, Zhang F. Diversity, classification and evolution of CRISPR-Cas systems. Curr Opin Microbiol. 2017;37:67-78. doi:10.1016/j.mib.2017.05.008

15. Altschul S. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25(17):3389-3402. doi:10.1093/nar/25.17.3389

16. Borges AL, Davidson AR, Bondy-Denomy J. The Discovery, Mechanisms, and Evolutionary Impact of Anti-CRISPRs. Annu Rev Virol. 2017;4(1):37-59. doi:10.1146/annurev-virology-101416-041616

17. Byrd AK, Raney KD. Superfamily 2 helicases. Front Biosci Landmark Ed. 2012;17:2070-2088.

18. Voloshin ON, Camerini-Otero RD. The DinG Protein from Escherichia coli Is a Structure-specific Helicase. J Biol Chem. 2007;282(25):18437-18447. doi:10.1074/jbc.M700376200

19. Voloshin ON, Vanevski F, Khil PP, Camerini-Otero RD. Characterization of the DNA Damage-inducible Helicase DinG from Escherichia coli. J Biol Chem. 2003;278(30):28284-28293. doi:10.1074/jbc.M301188200

20. Grodick MA, Segal HM, Zwang TJ, Barton JK. DNA-Mediated Signaling by Proteins with 4Fe–4S Clusters Is Necessary for Genomic Integrity. J Am Chem Soc. 2014;136(17):6470-6478. doi:10.1021/ja501973c

21. Koonin EV, Krupovic M. Evolution of adaptive immunity from transposable elements combined with innate immune systems. Nat Rev Genet. 2015;16(3):184-192. doi:10.1038/nrg3859

22. Makarova KS, Koonin EV. Annotation and Classification of CRISPR-Cas Systems. In: Lundgren M, Charpentier E, Fineran PC, eds. CRISPR. Vol 1311. New York, NY: Springer New York; 2015:47-75. doi:10.1007/978-1-4939-2687-9_4

CHAPTER 5

POSITIONING DIVERSE TYPE IV STRUCTURES AND FUNCTIONS WITHIN CLASS 1 CRISPR-CAS SYSTEMS[4]

ABSTRACT

Type IV CRISPR systems encode CRISPR associated (Cas) -like proteins that combine with small RNAs to form multi-subunit ribonucleoprotein complexes. However, the lack of Cas nucleases, integrases, and other genetic features commonly observed in most CRISPR systems has made it difficult to predict type IV mechanisms of action and biological function. Here we summarize recent bioinformatic and experimental advancements that collectively provide the first glimpses into the function of specific type IV subtypes. We also provide a bioinformatic and structural analysis of type IV-specific proteins within the context of multi-subunit (class 1) CRISPR systems, informing future studies aimed at elucidating the function of these cryptic systems.

INTRODUCTION

CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repeats-CRISPR associated) prokaryotic defense systems utilize Cas1 and Cas2 proteins, along with system-specific proteins such as Cas4, IHF, Csn2, and Cas9, to integrate foreign genetic material into the CRISPR locus, immunizing the cell against viruses and plasmids

---

[4] Co-authors: Hannah N. Taylor*, Eric Laderman, Matt Armbrust, Thomson Hallmark, Dylan Keiser, Joseph Bondy-Denomy, and Ryan N. Jackson.
Published: *Frontiers in Microbiology* (2021).
Final publication available at https://doi.org/10.3389/fmicb.2021.671522
*HNT performed sequence alignments, made the figures, and wrote the manuscript.

(Datsenko et al., 2012; Heler et al., 2015; S. A. Jackson et al., 2017; Kieper et al., 2018; Lee et al., 2018; Nuñez et al., 2014, 2016; Rollie et al., 2015; Sternberg et al., 2016; J. Wang et al., 2015; Wei et al., 2015; Yosef et al., 2012). To provide immunity, the CRISPR locus is transcribed and processed by RNA nucleases into CRISPR derived RNAs (crRNAs) (Brouns et al., 2008b; Deltcheva et al., 2011; Haurwitz et al., 2010; Marraffini & Sontheimer, 2008). The crRNAs combine with Cas proteins to form ribonucleoprotein (RNP) complexes, which recognize and bind complementary nucleic acids. Binding induces cleavage of the foreign nucleic acid, protecting the cell (Brouns et al., 2008b; Carte et al., 2008; Garneau et al., 2010; C. Hale et al., 2008; C. R. Hale et al., 2009; Hille et al., 2018; R. N. Jackson et al., 2017; Marraffini & Sontheimer, 2008).

Although all CRISPR systems use these general mechanisms to achieve immunity, the systems themselves are remarkably diverse, comprising two classes (1-2), six types (I-VI), and at least 33 subtypes (Makarova et al., 2020; Yan et al., 2019). In class 2 systems (types II, V, VI) a single Cas protein binds the crRNA to form the RNP complex, while class 1 RNP complexes (types I, III, IV) bind the crRNA with several proteins. Of the six CRISPR-Cas types, the least understood is type IV. Recent bioinformatic, biochemical, and structural studies of type IV CRISPR-Cas systems have provided valuable insights into type IV system function. Here we compile known data on type IV systems, highlight recent advances in type IV system biology and biochemistry, and indicate questions concerning type IV systems that need to be addressed. Additionally, we provide phylogenetic analyses that suggest ancillary proteins associated with type IV systems have evolved Cas-specific functions.

TYPE IV SYSTEMS ARE MINIMAL, MOBILE CRISPR-Cas SYSTEMS

Distinguishable from other CRISPR-Cas systems, Type IV systems encode a distinct *cas7*-like gene (*csf2)*, lack adaptation genes, rarely encode an obvious nuclease, and are primarily found on plasmids (Koonin & Makarova, 2017, 2019; Pinilla-Redondo et al., 2019). These unique features of type IV systems have made it difficult to predict the function of type IV systems.

All type IV systems encode homologs of proteins known to form multi-subunit RNP complexes, explaining their class 1 designation. However, the presence of specific genes, gene arrangements, and differences in gene sequences have been used to further classify type IV systems into three distinct subtypes (IV-A, IV-B, and IV-C) (Makarova et al., 2011, 2015; Makarova et al., 2020). Types IV-A, IV-B, and IV-C each contain a subtype-specific gene (*dinG, cysH-like,* and *cas10-like,* respectively) and subtype-specific features (**Figure 5-1A**). Type IV-A operons encode the three core type IV genes (*csf1*, *csf2*, and *csf3*), an endoribonuclease (*cas6*/*csf5*), a CRISPR array, and a putative helicase (*dinG).* Type IV-B operons encode the three core type IV genes and a *cas11-like* gene but lack a CRISPR locus. Additionally, type IV-B operons encode an ancillary *cysH-like* gene, labeled such because its predicted secondary structure resembles the *cysH* gene (Faure et al., 2019; Shmakov et al., 2018). Type IV-C systems encode *csf2* and *csf3*, but in place of the *csf1* gene they encode a *cas10-like* gene with a putative HD-nuclease domain. They also encode the *cas11-like* gene observed in IV-B systems, and sometimes a *cas6* RNA endonuclease and CRISPR array (Makarova et al., 2020; Pinilla-Redondo et al., 2019).

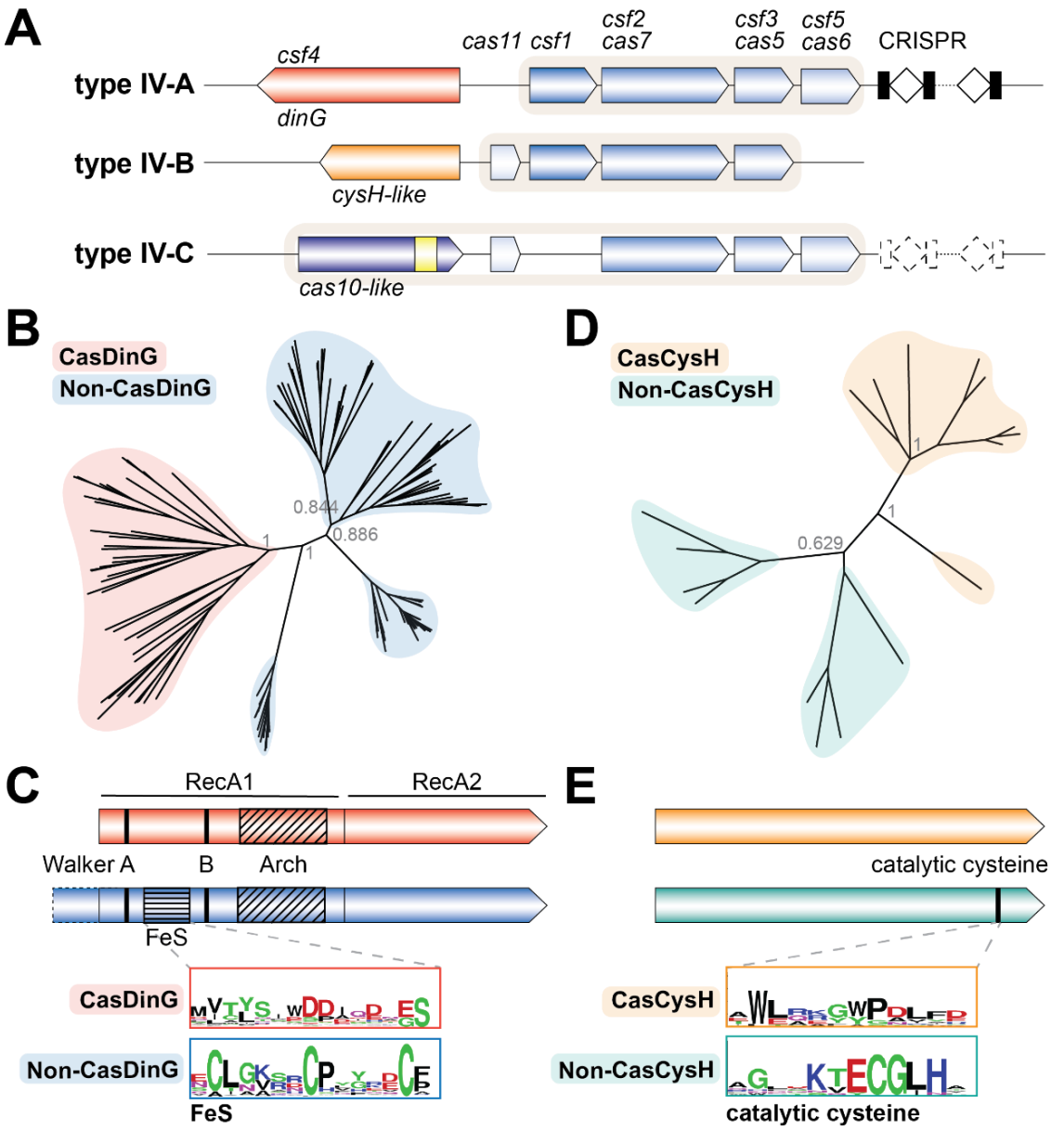

Figure 5-1. The type IV Cas accessory proteins have evolved a Cas specific function. (**A**) Classification schematic of type IV CRISPR-Cas systems. A typical locus is represented for each type IV subtype. Dashed lines indicate components that are sometimes not encoded by the subtype. Shaded backgrounds highlight which gene products form the ribonucleoprotein (RNP) complex. The yellow square in the IV-C *cas10-like* large subunit represents an HD nuclease domain. (**B**) Phylogenetic tree of Cas- and non-CasDinG sequences. Posterior probabilities are shown. (**C**) Cartoons of Cas- and non-CasDinG sequences indicating positions of certain helicase motifs and domain architecture. Weblogos (Crooks, 2004) of the FeS cluster region in non-CasDinG (below, blue outline) and CasDinG (top, red outline) are shown. (**D**) Phylogenetic tree of Cas- and non-CasCysH sequences. Posterior probabilities are shown. (**E**) Cartoons of Cas- and

non-CasCysH sequences. CasCysH is predicted to adopt the Rossman α–β–α fold observed in non-CasCysH structures. Positions and sequences of P- and PP-loops are indicated. Weblogos of the catalytic cysteine in non-CasCysH (bottom, teal outline) and CasCysH (top, orange outline) are shown.

It is curious that the type IV systems that encode CRISPR loci do so in the absence of adaptation genes. It has been hypothesized that these type IV systems commandeer adaptation machinery from other CRISPR-Cas types to maintain their CRISPRs, similar to some type III systems (Staals et al., 2013, 2014; J. Elmore et al., 2015; Bernheim et al., 2020). Supporting this hypothesis, recent bioinformatic work showed that some type IV-A subtypes co-localize with certain type I systems (e.g. I-F, I-E), and that spacers found within co-localized type IV CRISPR loci appeared to be selected with the same criteria utilized by the type I system adaptation machinery (e.g. both I-E and IV-A protospacers are flanked with an 5'-AAG-3' protospacer adjacent motif (PAM)) suggesting there may be functional cross-talk between these systems (Pinilla-Redondo et al., 2019). *In vivo* and *in vitro* experimental work that examines adaptation in type IV systems with adaptation proteins from co-localized systems is needed to confirm this proposed cooperation.

THE *cas7*-LIKE GENE, *csf2*, DISTINGUISHES TYPE IV FROM OTHER CLASS 1 CRISPR-Cas SYSTEMS

Initial bioinformatic analyses proposed *csf1* as the type IV *cas* signature gene (Makarova et al., 2015). However, some type IV systems lacking *csf1* have been

identified, necessitating that the type IV *cas7* homolog, *csf2,* be used to classify type IV systems (Pinilla-Redondo et al., 2019). In type I and type III systems, Cas7-like proteins bind the crRNA guide within a helical backbone of a multi-subunit RNP complex and make direct interactions with other protein subunits (Jore et al., 2011; Lintner et al., 2011; Staals et al., 2013; R. N. Jackson, Golden, Erp, et al., 2014; Mulepati et al., 2014; Osawa et al., 2015b). Similarly, a recent cryo-EM structure of the type IV-B RNP complex revealed that Csf2 proteins bind RNA within a helical backbone, indicating a conserved function for Cas7-like proteins in all class 1 systems (Y. Zhou et al., 2021). Despite this conservation, the sequence and structure of Csf2 is distinguishable from other Cas7 proteins (Makarova et al., 2011) (**Figure 5-2**). For example, when representative Cas7 sequences from all class 1 subtypes were aligned and a phylogenetic tree created, Csf2 sequences clustered on a separate branch from type I and type III Cas7 sequences (**Figure 5-3A** and Methods). Csf2 is distinct from other Cas7 homologs but appears to be most closely related to type III, supporting evolutionary hypotheses that type IV systems diverged from type III systems (Koonin & Makarova, 2019; Makarova et al., 2020; Pinilla-Redondo et al., 2019). Interestingly, an alignment of only Csf2 sequences shows clustering of Csf2 from each type IV subtype on its own branch, illustrating the intrinsic diversity of type IV subtypes and suggesting subtype-specific functional distinctions (**Figure 5-3B**). It is worth noting that the type IV-B Csf2 subunit structure is most similar to the structure of the Cas7 homolog in type III-A systems, Csm3 (Y. Zhou et al., 2021). Csm3 contains a catalytic aspartate that cleaves RNA targets (Tamulaitis et al., 2014). Alignment of target-bound Csm3 with Csf2 indicates that, although Csf2 also contains a conserved aspartate residue in a similar location, it is not in a position amenable for target

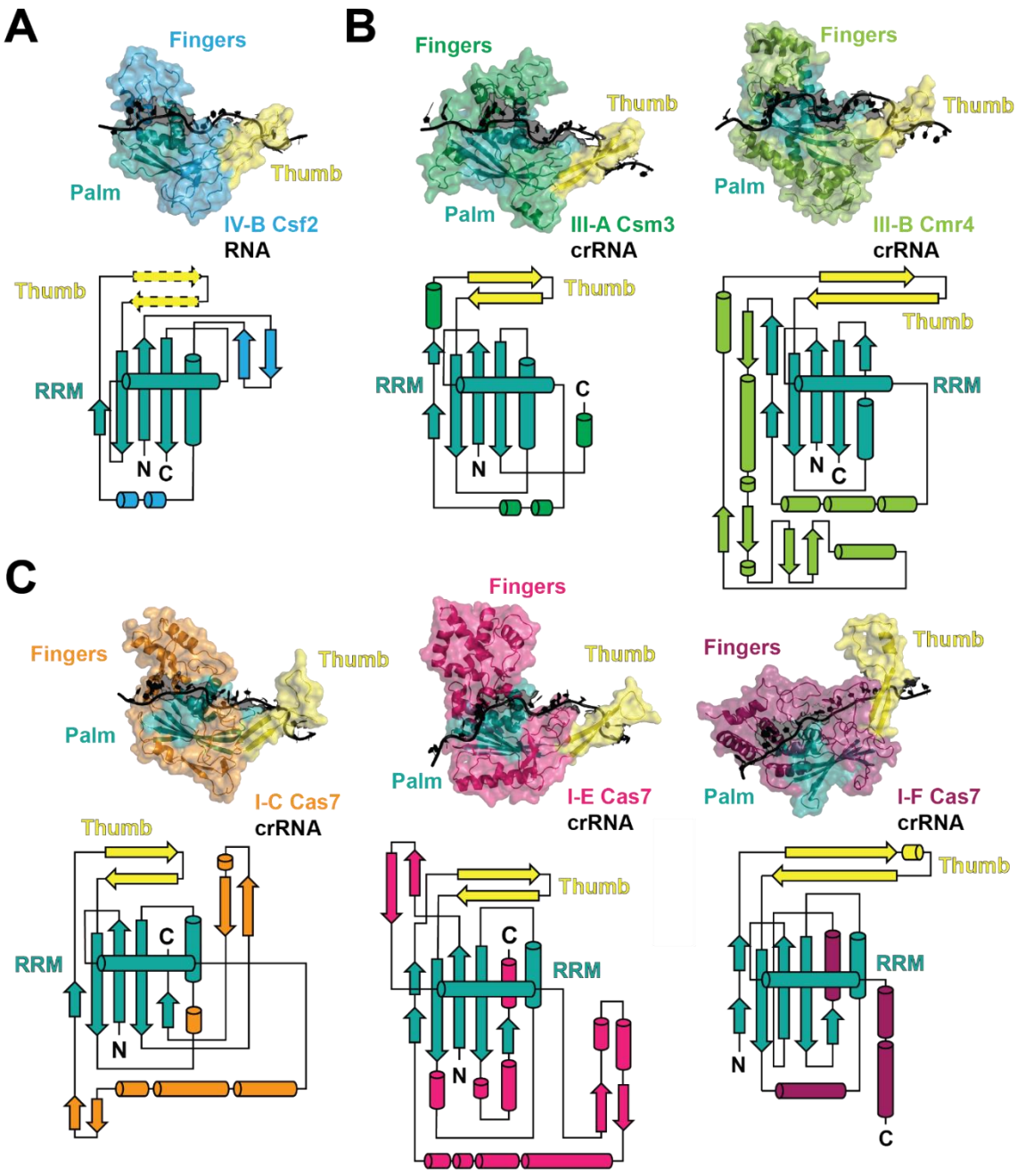

Figure 5-2. Csf2 is a unique Cas7-like backbone subunit. (**A**) Structure of IV-B Csf2 (PDBid 7JHY) and protein topology map highlighting the hand-like domains typical of Cas7-like proteins. Dashed lines for the thumb indicate that secondary structure is not obvious in the model. (**B**) Structures of III-A Csm3 (PDBid 6O7I) and III-B Cmr4 (PDBid 3X1L) depicted as in (A). (**C**) Structures of Cas7 from I-C (PDBid 7KHA), I-E (PDBid 5H9F), and I-F (PDBid 6B44) depicted as in (A).

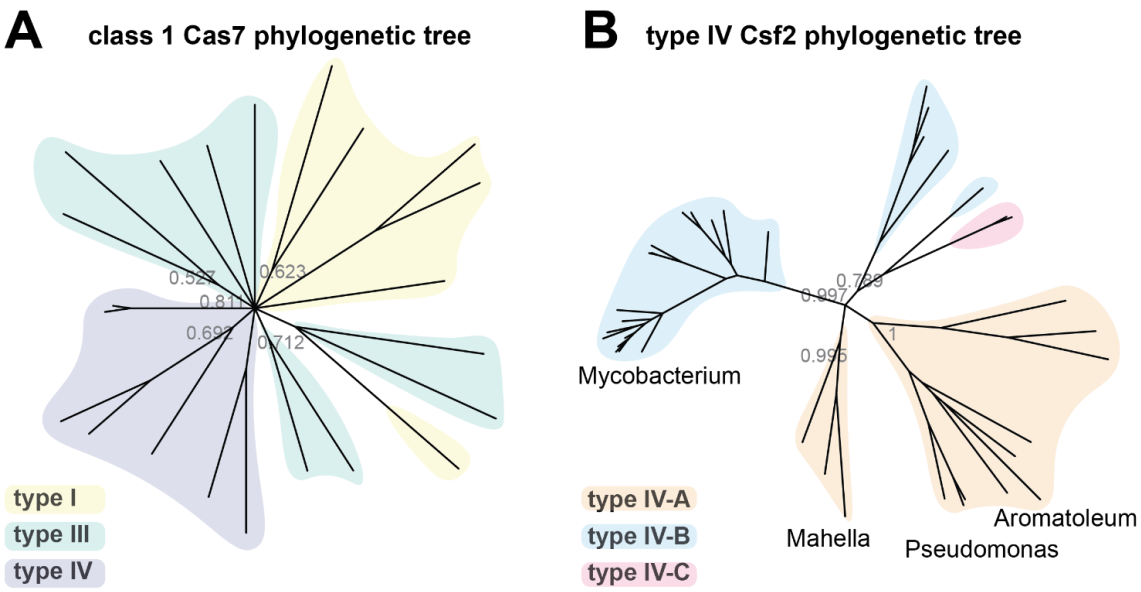

Figure 5-3. Csf2 is unique from other Cas7-like subunits and distinct within the type IV subtypes. (**A**) A phylogenetic tree of Csf2 sequences from all three type IV subtypes and a selection of Cas7 sequences from each of the several type I and type III subtypes. (**B**) A phylogenetic tree of Csf2 sequences from all three type IV subtypes. Csf2 sequences from *Mycobacterium JS623* (Zhou et al., 2021), *Mahella australiensis* (Taylor et al., 2019), *Pseudomonas aeruginosa* (Crowley et al., 2019), and *Aromatoleum aromaticum* (Özcan et al., 2018) are indicated. Sequences for both trees were selected from (Makarova et al., 2020).

cleavage (Y. Zhou et al., 2021). Additional structural studies of type IV complexes bound to nucleic acid targets and complementary biochemical assays are needed to determine whether Csf2 is capable of RNA nuclease activity.

TYPE IV-A SYSTEMS ARE DEFENSE SYSTEMS WITH AN UNKNOWN MECHANIM OF ACTION INVOLVING A DinG HELICASE

Recently, a type IV-A system from *Pseudomonas aeruginosa* was shown to exhibit crRNA-guided defense against plasmids (Crowley et al., 2019), consistent with an analysis of type IV CRISPR spacers that suggested type IV-A systems disproportionately target plasmids (Pinilla-Redondo et al., 2019). Notably, earlier bioinformatic work indicated that many type IV-A spacers target viruses and prophage sequences encoding putative anti-CRISPRs, suggesting type IV-A systems also actively target viruses (Nobrega et al., 2020; Shmakov et al., 2017; Yin et al., 2019). However, direct data, such as viral plaque assays, are needed to confirm that type IV-A systems protect against viral attack.

Structural and biochemical work on a type IV-A complex from *Aromatoleum aromaticum* and IV-A Cas6 from *Mahella australiensis* demonstrated that the RNA endonuclease Csf5/Cas6 processes a crRNA upon which Csf1, Csf2, Csf3, and Csf5 form an RNP complex (Özcan et al., 2018; Taylor et al., 2019). At least three distinct crRNA processing endoribonucleases are encoded by Type IV-A systems (Cas6, Csf5, and Cas6e) (Makarova et al., 2020) (**Figure 5-4A**). Sequence alignments between biochemically characterized and putative type IV Csf5/Cas6 enzymes revealed Csf5 enzymes cleave RNA with arginine active site residues, while type IV Cas6 and Cas6e enzymes utilize histidine/tyrosine active site residues (**Figures 5-4B and 5-5**). Despite

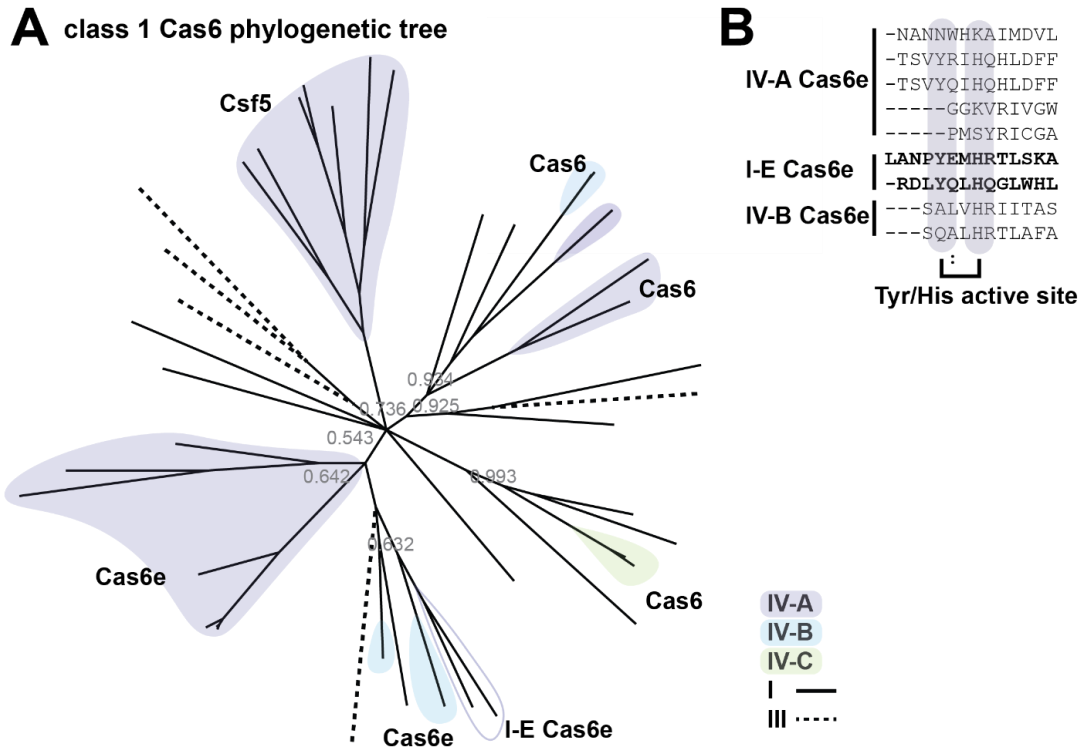

Figure 5-4. Type IV Cas6/Csf5 subunits are distinct from Cas6 homologs of other CRISPR systems. (**A**) A phylogenetic tree of type IV Cas6/Csf5 sequences and select type I and type III Cas6 sequences. Sequences were selected from (Makarova et al., 2020). (**B**) A Clustal Omega (Madeira et al., 2019) amino acid sequence alignment of type IV-A and IV-B Cas6e and type I-E Cas6e, highlighting the active site residues. The Cas6e active site is not highly conserved among type IV Cas6e sequences. Bolded sequences have experimentally determined active sites (Jackson et al., 2014; Sashital et al., 2011).

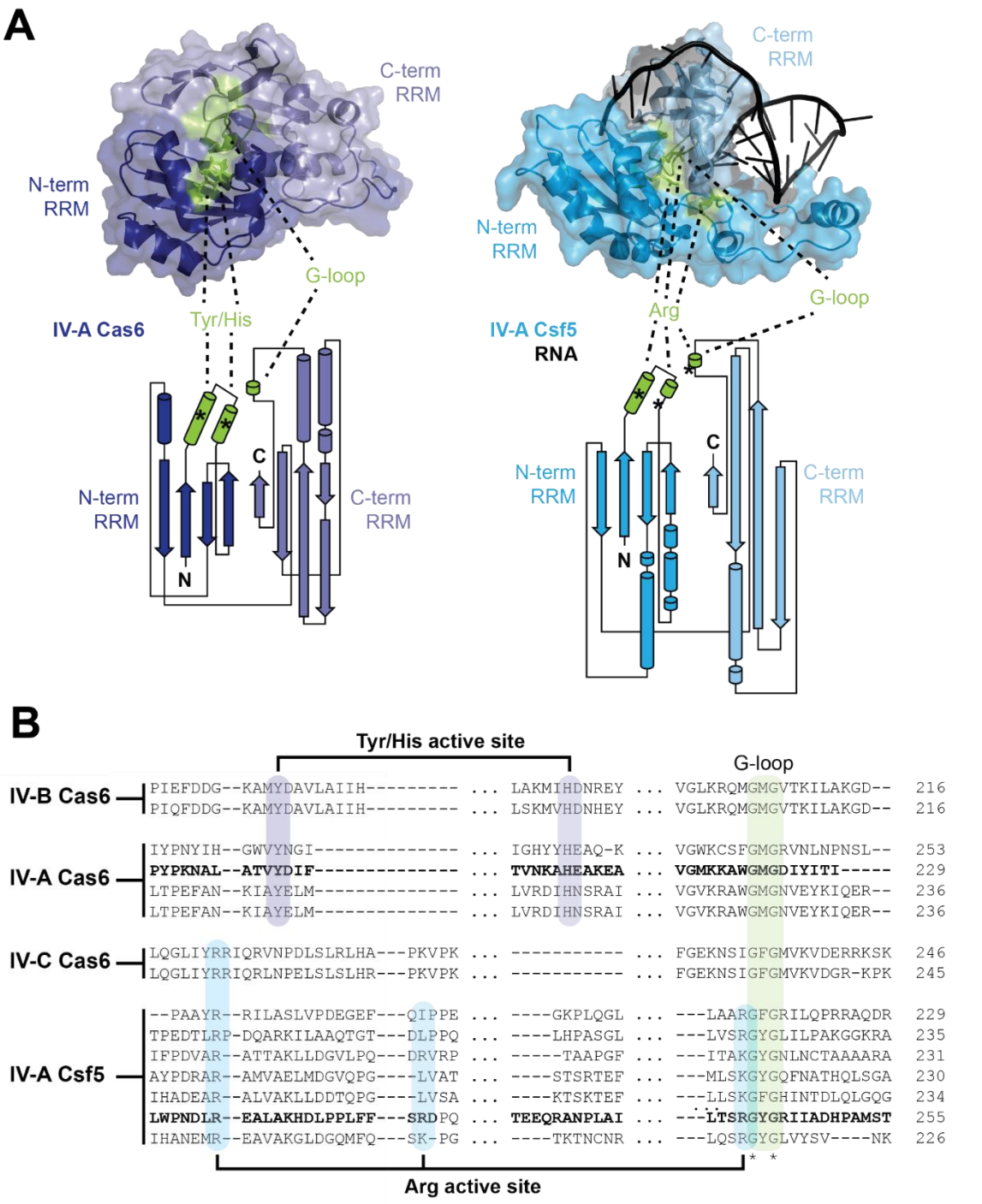

Figure 5-5. Type IV Cas6/Csf5 structures and active sites. (**A**) Structural comparisons of apo IV-A Cas6 from Mahella australiensis (PDBid 6NJY) and substrate-bound IV-B Csf5 from Aromatoleum aromaticum (PDBid 6H9I). Both structures are composed of two RRM folds with the active site positioned in the cleft between the folds. (**B**) A Clustal Omega (Madeira et al., 2019) alignment of Csf5 sequences from subtype IV-A and Cas6 sequences from subtypes IV-A, IV-B, and IV-C. The two different active sites

architectures (Tyr/His or Arg) are indicated. All aligned sequences share the conserved G-loop motif. The bolded sequences have experimentally confirmed active sites: IV-A Cas6 from *M. australiensis* (Taylor et al., 2019) and IV-A Csf5 from *A. aromaticum* (Özcan et al., 2018).

these obvious differences in endoribonucleases, we hypothesize that in all type IV-A systems the Csf1, Csf2, Csf3, and Csf5/Cas6 proteins bind to the processed crRNA to form a multi-subunit complex that binds complementary nucleic acid.

It remains unclear if type IV RNP complexes bind single stranded RNA (like the type III Csm and Cmr complexes (C. R. Hale et al., 2009; Samai et al., 2015)) or double stranded DNA (like the type I Cascade complexes (Brouns et al., 2008b)) and how type IV complexes distinguish self from non-self in its binding targets. RNPs that target dsDNA usually rely on a protein-mediated binding event with a specific non-self sequence adjacent to the complementary target, called a PAM (protospacer adjacent motif) (Mojica et al., 2009; Westra et al., 2012, 2013). PAM binding provides the energy for target duplex unwinding and interrogation of the DNA by the crRNA-guide. Work by Pinilla-Redondo et al. (2019) identified a consensus PAM (5'-AAG-3') flanking protospacers targeted by a subset of type IV-A systems, suggesting type IV-A systems rely on PAM recognition to license binding. However, the consensus PAM may only reflect a preference of the acquisition machinery, which may explain why consensus PAM sequences have not been identified in all IV-A systems. Reliance on a specific PAM sequence for type IV-A RNP interference remains to be confirmed experimentally,

but it should be noted that a promiscuous PAM recognition mechanism may indicate that the type IV complexes have evolved to accommodate the preferences of diverse Cas1 and Cas2 proteins that use different PAM sequences in spacer acquisition.

Additionally, the structural similarities of the type IV-B complex to the type III Csm complex suggest that type IV complexes may target RNA (Y. Zhou et al., 2021). Instead of recognizing a "non-self" PAM to license base pairing with a double-stranded DNA target, RNPs that bind RNA generally use a "self recognition" mechanism to distinguish self from non-self sequences (Marraffini & Sontheimer, 2010). Self-sequence located in the flanking regions of a bound RNA can base pair with the direct repeat of the crRNA disrupting downstream activation of effector nucleases (L. You et al., 2019). Self-sequences are inhibitory to overall immune function (J. R. Elmore et al., 2016; Estrella et al., 2016; Han et al., 2017; Kazlauskiene et al., 2016; Liu et al., 2017; Marraffini & Sontheimer, 2010; Zhang et al., 2016), but in some systems only a subset of non-self protospacer flanking sequences (called RNA-PAMs (rPAM) in type III systems or protospacer flanking sites (PFS) in type VI systems) are activating (Abudayyeh et al., 2016; J. R. Elmore et al., 2016; Marraffini & Sontheimer, 2010). We suspect that one or more Csf subunits may be responsible for PAM recognition to license DNA binding or rPAM recognition to activate immunity. We anticipate that *in vivo* PAM screens and biochemical binding assays with purified type IV-A RNPs will reveal the type IV-A self vs. non-self recognition mechanism.

Type IV-mediated plasmid clearance required all type IV-A system genes (*csf1, csf2, csf3, csf5, and dinG/csf4*) and a CRISPR containing a spacer complementary to a target plasmid sequence adjacent to a 5'-TTC-3' PAM (Crowley et al., 2019). Because

deleting the *dinG* gene or mutating the ATPase active site residues (DEAH-box) fully disrupted plasmid clearance, we hypothesize that RNP complex binding recruits the type IV-associated DinG (CasDinG) helicase to the bound target and CasDinG-mediated ATP binding and hydrolysis performs work, such as duplex unwinding, that is essential for plasmid clearance. Such a mechanism is similar to the more extensively studied type I Cas3 helicase-nuclease that unwinds and degrades dsDNA targets bound by the type I Cascade RNP complex (Beloglazova et al., 2011; Mulepati & Bailey, 2011; Sinkunas et al., 2011).

Both DinG and Cas3 classify as superfamily 2 helicases but, unlike Cas3, CasDinG proteins have no identifiable nuclease domains and have yet to be biochemically or structurally characterized (Fairman-Williams et al., 2010; Makarova et al., 2020). DinG helicases are generally involved in DNA recombination and repair, and are classified by amino acid sequence motifs involved in ATP binding and hydrolysis and nucleic acid binding and translocation (Cheng & Wigley, 2018; Lewis et al., 1992; McRobbie et al., 2012; Thakur et al., 2014; Voloshin et al., 2003b; Voloshin & Camerini-Otero, 2007b; Wu & Brosh, 2012). The motifs are located across two RecA helicase domains (**Figure 5-6**). The first helicase domain also harbors two insertions, an iron sulfur cluster domain, and an arch domain, which are both important for duplex strand splitting (Peissert et al., 2020; Ren et al., 2009).

Since non-CasDinG helicases and their homologs have been extensively studied biochemically and structurally, we hypothesized that an in-depth comparison of CasDinG with non-CasDinG sequences would provide insight to CasDinG function. To investigate the relationship of CasDinG to other DinG helicases, we compiled CasDinG and non-

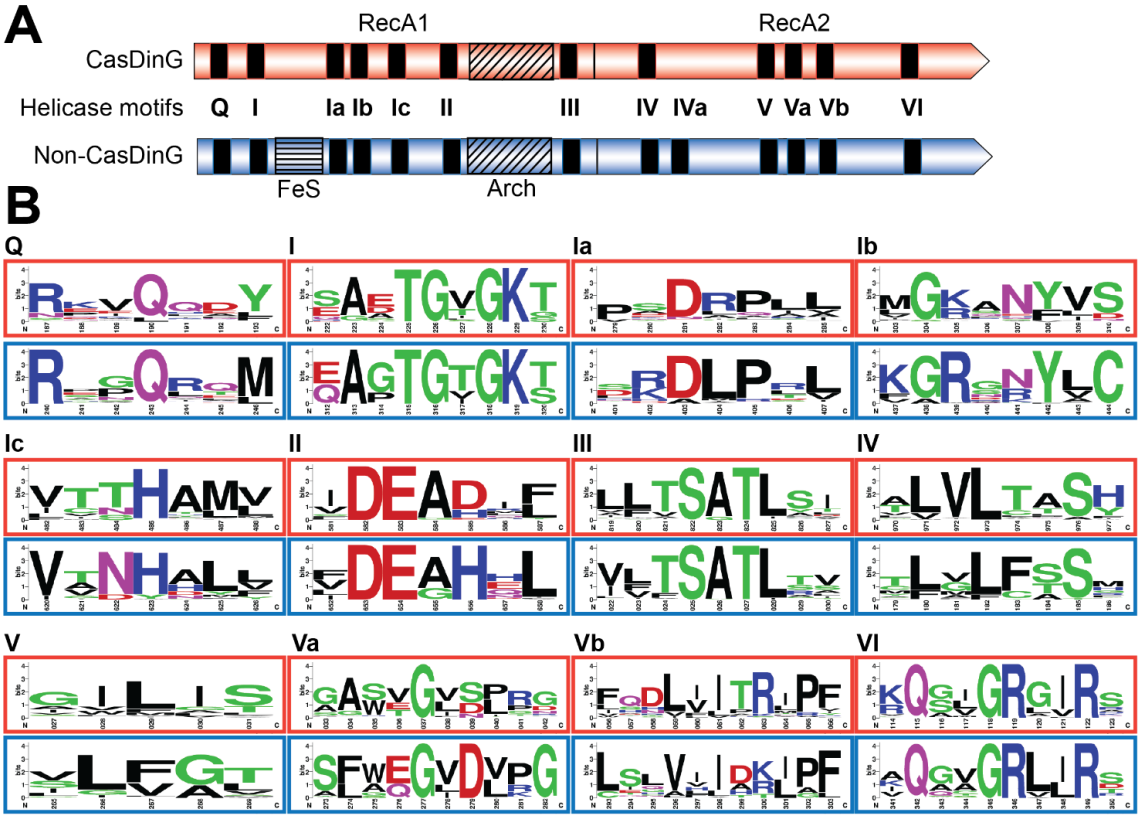

Figure 5-6. The conserved helicase motifs of CasDinG and non-CasDinG. (**A**) Cartoon depiction of a Cas- and non-CasDinG primary amino acid sequence indicating the positions of the helicase motifs. (**B**) Weblogos (Crooks, 2004) of the helicase motifs of Cas- and non-CasDinG helicases. CasDinG motifs are outlined in red boxes and non-CasDinG motifs are outlined in blue boxes.

CasDinG sequences from organisms containing a type IV-A system and generated a phylogenetic tree (**Figure 5-1B** and Methods). Interestingly, CasDinG and non-CasDinG sequences clustered separately even when the sequences were retrieved from the same organism, suggesting CasDinG is functionally distinct from non-CasDinG. Notably, CasDinG helicases contain insertions within the first RecA domain of the same lengths as the iron-sulfur and arch insertions, but they lack homology with non-CasDinG sequences, including the residues important for coordinating the iron-sulfur cluster (**Figure 5-1C**). Sequence differences in these regions suggest these inserts may be a source of functional distinctions important for defense activities. Many functions for CasDinG have been hypothesized, including a role in displacing bound RNP complexes, cleaving bound targets with an unidentified nuclease activity (perhaps housed within an insert), or recruitment of endogenous nucleases to bound targets (Grodick et al., 2014b). Notably, DinG helicases have been observed in a few type I and III systems (Dwarakanath et al., 2015; Makarova et al., 2020), indicating an evolutionary link and suggesting that some CasDinG activities essential for type IV immunity may have been co-opted by other class 1 systems.

In summary, recent bioinformatic and *in vivo* studies have indicated type IV-A systems protect prokaryotes from plasmids and viruses, but the mechanisms that underpin how the Csf RNP complex and CasDinG work together to provide immunity remain to be determined.

TYPE IV-B SYSTEMS FORM AN RNP COMPLEX OF UNKNOWN FUNCTION AND A SPECIALIZED CysH-LIKE PROTEIN WITH PUTATIVE ATP α-HYDROLASE ACTIVITY

Unlike type IV-A and IV-C subtypes, type IV-B systems lack a CRISPR locus and a crRNA processing enzyme, and are associated with an ancillary gene identified as *cysH-like* by the HHpred secondary structure prediction and alignment tool (Makarova et al., 2020; Zimmermann et al., 2018) (**Figure 5-1A**). A recent structural study recombinantly expressed and purified a *Mycobacterium sp. JS623* IV-B Csf RNP complex containing four type IV-B proteins (Csf1, Csf2, Csf3, and Cas11) (Y. Zhou et al., 2021). Interestingly, RNA sequencing revealed the type IV-B Csf complex bound small heterogeneous RNAs, instead of co-expressed type I-E crRNAs from the *Mycobacterium sp. JS623* plasmid, suggesting a possible function other than CRISPR-mediated defense. A high resolution cryo-EM structure of the complex revealed several Csf2 subunits bind an RNA within a helical filament, while Cas11 subunits form a minor filament that contacts the larger filament at Csf2 dimer interfaces (Y. Zhou et al., 2021). This structure of intertwined large and small protein filaments is similar to other class 1 RNP complexes, suggesting similar function as an RNA-guided complex that binds complementary targets (Jore et al., 2011; Lintner et al., 2011; Staals et al., 2013; R. N. Jackson, Golden, Erp, et al., 2014; Mulepati et al., 2014; Osawa et al., 2015b) (**Figure 5-7**).

Several observations are currently confounding an understanding of the type IV-B complex function. First, electron density for Csf1 and Csf3 subunits was not clearly observed within the structure, although SDS-PAGE indicated their presence in the purified complex. Thus, the structure and function of these important proteins remains

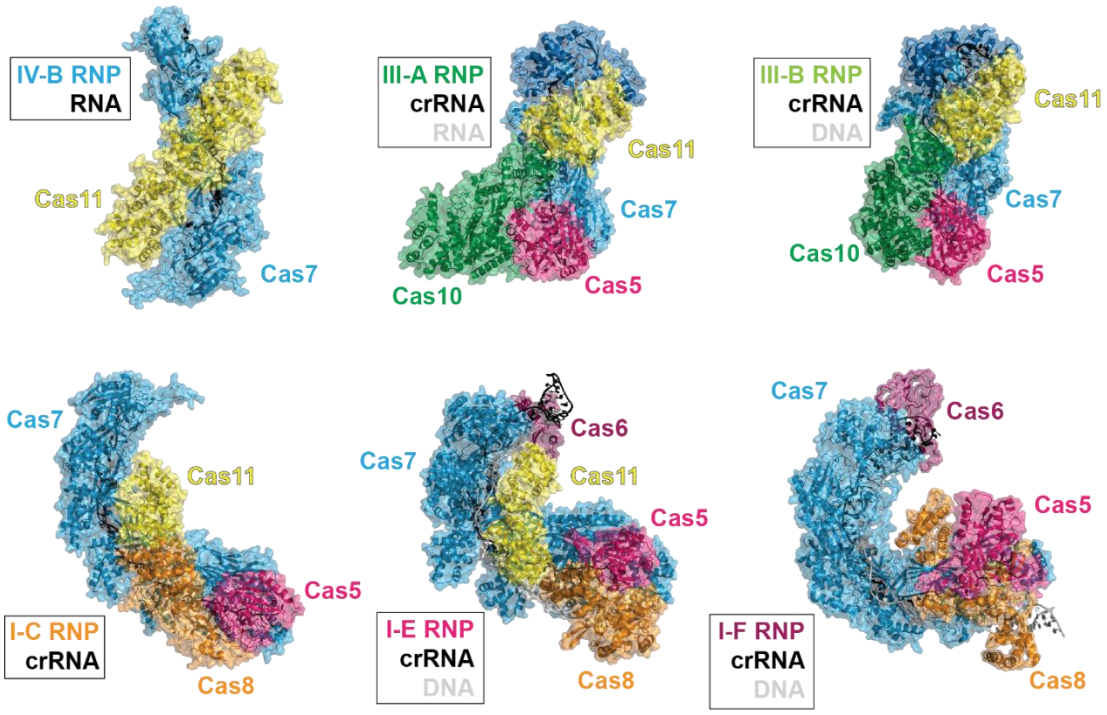

Figure 5-7. Structural comparison of class 1 RNP complexes from type IV-B (PDBid 7JHY) (Zhou et al., 2021), type III-A (PDBid 6O7I) (Jia et al., 2019), type III-B (PDBid 3X1L) (Osawa et al., 2015), type I-C (PDBid 7KHA) (O'Brien et al., 2020), type I-E (PDBid 5H9F) (Hayes et al., 2016), and type I-F (PDBid 6B44) (Guo et al., 2017). Complexes are colored by homologous subunits.

unknown. Second, because the IV-B Csf complex bound heterogenous RNA, it remains unknown whether the Csf complex lacks sequence-specific preference for small RNAs or if the RNA(s) that the complex would normally bind were not available in the recombinant expression conditions. Finally, the role of the strictly conserved ancillary CysH-like protein and how it may interact with the complex is unknown.

The key to understanding the function of type IV-B systems likely lies with the uncharacterized, but ubiquitous, type IV-B accessory *cysH*-like gene (Faure et al., 2019; Shmakov et al., 2018). Typical CysH proteins are phosphoadenosine phosphosulfate (PAPS) reductases involved in sulfate assimilation. Structures reveal CysH proteins fit within a family of enzymes that adopt a Rossman-like α–β–α sandwich fold that binds the nucleotides (InterPro IPR014729) (Blum et al., 2020). CysH proteins also contain a P-loop motif (GXXGXGKT/S consensus sequence) that binds nucleotide phosphates, and a conserved C-terminal cysteine that performs nucleophilic attack on the PAPS β–sulfate, hydrolyzing PAPS at the α-phosphate and forming a covalent thiosulfanate intermediate during sulfur reduction (Carroll et al., 2005; Chartron et al., 2007; Savage et al., 1997). Interestingly, the DndC protein from the recently discovered DND bacterial immune system also belongs to the PAPS reductase family, and uses a similar mechanism to incorporate sulfur into the backbone of chromosomal DNA through a disulfide cysteine (Faure et al., 2019; L. Wang et al., 2018; D. You et al., 2007). These phosphorothioate modifications serve as an epigenetic signature that allows the DND system to distinguish self from non-self DNA (Wang et al., 2019). The predicted structural homology between the type IV-B CysH (CasCysH) and DndC proteins justifies speculation that CasCysH proteins perform a similar function. However, a closer analysis of type IV-B CasCysH

sequences suggests that if CasCysH does epigenetically modify DNA, it will not be through the formation of phosphorothioates. Although HHPred predicts CasCysH adopts a Rossman-like α–β–α sandwich fold, the catalytic cysteine important for sulfonate reduction in non-CasCysH and phosphothiolation of DNA by DndC is absent. Interestingly, the P-loop sequence of CasCysH is more similar to the PP-motif (pyrophospohatase motif) (SGGXDS/T consensus sequence) observed in ATP PPases (Bork & Koonin, 1994) (**Figure 5-1E**).

To better understand CasCysH activity and to explore the relationship between non-CasCysH and CasCysH proteins, sequences from organisms encoding both Cas- and non-CasCysH were aligned and phylogenetic trees determined. As was seen with Cas- and non-CasDinG, CasCysH sequences cluster separately from non-CasCysH sequences even when the sequences were retrieved from the same organisms (**Figure 5-1D**). Together these differences suggest CasCysH evolved to preserve nucleotide binding without sulfonucleotide reduction.

Non-CasCysH enzymes fall within the larger classification of ATP α-hydrolases, which include N-type ATP PPases (Savage et al., 1997). However, unlike non-CasCysH and DndC, N-type ATP PPases catalyze sequential reactions involving substrate AMPylation, instead of the formation of covalent enzyme substrate intermediates requiring nucleophilic attack from a catalytic cysteine (Wang et al., 2019; Chartron et al., 2007). The absence of a catalytic cysteine suggests that the role of CasCysH is to stabilize the AMPylation of specific substrates through catalysis of ATP α–hydrolase activity. We hypothesize that such an activity could be used to modify nucleic acids bound by the type IV-B RNP for immune system purposes, gene regulation, or the

formation of secondary messengers. Future biochemical studies aimed at defining the function of CasCysH and its interactions with the IV-B Csf RNP complex will be critical for understanding type IV-B systems.

Several hypotheses exist concerning the function of type IV-B CRISPR-Cas systems. As they lack both a CRISPR array and an obvious nuclease it seems unlikely that type IV-B systems function as independent CRISPR-Cas defense systems (Faure et al., 2019; Makarova et al., 2011; Y. Zhou et al., 2021). It has been suggested that type IV-B systems could bind crRNAs derived from other CRISPR systems, forming IV-B RNP complexes that perform RNA-guided defense (Makarova et al., 2011, 2015; Koonin & Makarova, 2019). As type IV systems are generally encoded on plasmids, such a crRNA scavenging system could be passed between organisms, acting as a mobile defense system. Interestingly, it was recently shown that sometimes type IV-B systems colocalize with specific class 1 systems, suggesting a cooperative function (Pinilla-Redondo et al., 2019). However, the same study showed that type IV-B systems are most often observed without other CRISPR systems, supporting an alternative hypothesis that proposes IV-B systems may protect plasmids from RNA-guided defense mechanisms by sponging up and inactivating small guide RNAs, including crRNAs (Faure et al., 2019; Koonin & Makarova, 2017; Pinilla-Redondo et al., 2019). Such an anti-guide-RNA activity could give plasmids containing a type IV-B system a selective advantage (Shmakov et al., 2018; Koonin & Makarova, 2019). Although intriguing, neither of these hypotheses explain the role of the highly conserved ancillary protein CasCysH, suggesting the true function of IV-B systems may be more intricate than has so far been proposed.

THE NEWLY CLASSIFIED TYPE IV-C SYSTEM HIGHLIGHTS THE DIVERSE NATURE OF TYPE IV CRISPR-Cas SYSTEMS

Only recently did bioinformatics studies classify the subtype IV-C CRISPR-Cas system (Makarova et al., 2020; Pinilla-Redondo et al., 2019). Type IV-C systems lack a Csf1 subunit, instead encoding a Cas10-like subunit with an HD nuclease domain (**Figure 5-1A**). Type III CRISPR-Cas systems also encode Cas10, which is the large subunit of the RNP complex. In type III systems Cas10 has nuclease activity and synthesizes cyclic oligoadenylate second messengers (Jung et al., 2015; Kazlauskiene et al., 2017; Niewoehner et al., 2017). The type IV Cas10 contains an HD nuclease domain but not a nucleotide cyclase motif 'GGDD', suggesting it has nuclease but not cyclic adenylate synthetase activity (Pinilla-Redondo et al., 2019). Interestingly, the HD domain motifs of type IV Cas10 are more similar to the HD domain of Cas3 than the type III Cas10 (Aravind & Koonin, 1998; Makarova et al., 2020). The presence of a *cas10-like* gene in type IV-C systems and the similarity between the type III-A and type IV-B RNP complexes support proposals that type IV and type III CRISPR-Cas systems share a common ancestor (Pinilla-Redondo et al., 2019; Makarova et al., 2020; Y. Zhou et al., 2021). Experimental work is needed to better understand the function of these fascinating systems.

Several variants of type IV systems have been identified in bioinformatics studies and clinical samples which include type IV systems; without a *csf1*, with a *csf1-csf3* fusion, with a *recD* helicase instead of *dinG*, and in association with IncH1b plasmids (Crowley et al., 2019; Pinilla-Redondo et al., 2019; Kamruzzaman & Iredell, 2020; Newire et al., 2020). We expect further study and analysis of these diverse systems will

reveal unique mechanisms and functions that may expand the CRISPR-based genome editing toolbox.

UNANSWERED QUESTIONS CONCERNING TYPE IV BIOCHEMISTRY AND BIOLOGICAL FUNCTION

Throughout this perspective we have highlighted pressing questions concerning type IV CRISPR-Cas system structures and functions. Here we suggest models for the function of each type IV subtype and indicate areas which require further understanding. Type IV-A systems have been shown to form RNP complexes and prevent targeted plasmid transformation, but they have not been shown to target viruses nor is the mechanism of crRNA-guided defense clear (**Figure 5-8A**). Understanding the targets of the type IV-A system is critical to understanding the full scope of its defense activity. The presence of a helicase within the system suggests the need to unwind a duplex substrate. We suspect that the type IV-A system targets dsDNA, as it does defend against invasive plasmids (Crowley et al., 2019). However, CasDinG could also be important for unwinding duplex secondary structure within ssRNA targets or for targeting dsRNA phages (Poranen & Tuma, 2004). Remaining questions include the targeting parameters of the complex (DNA vs RNA, seed sequence, mismatch tolerance), how the complex distinguishes self from non-self, the role of CasDinG in immunity, and how targets are cleared without an identifiable nuclease domain within the system. We speculate that the IV-A RNP complex will bind to a dsDNA target and recruit CasDinG to the resulting R-loop, allowing CasDinG to unwind the target. To clear the target from the cell, we hypothesize that either an endogenous nuclease will degrade the unwound nucleic acid, or CasDinG harbors an intrinsic nuclease activity not predicted by the sequence.

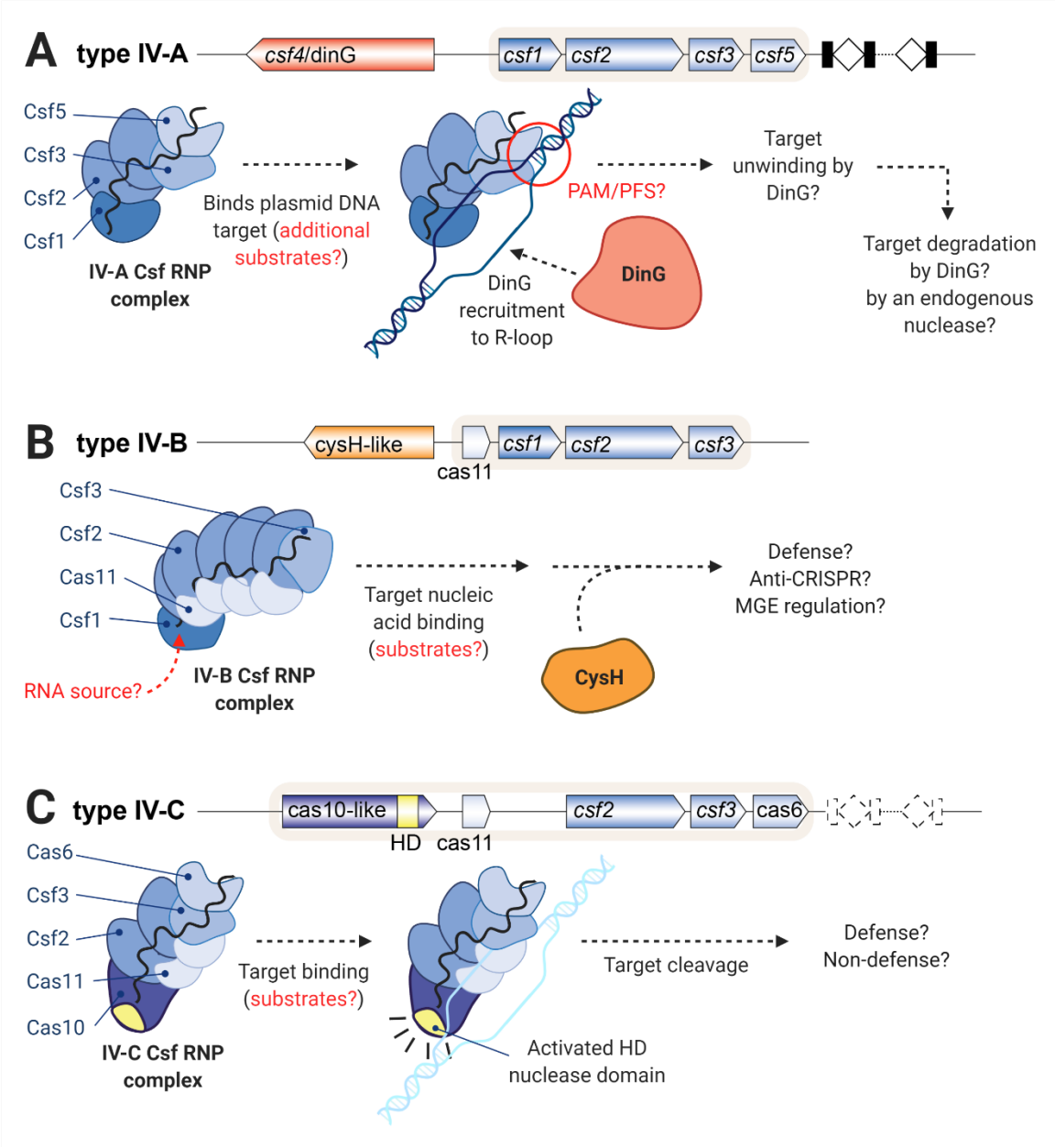

Figure 5-8. Models of type IV system functions highlighting questions that remain to be answered. (**A**) IV-A RNP complexes likely bind DNA targets and recruit CasDinG for target unwinding and degradation. (**B**) IV-B RNP complexes likely interact with CasCysH to perform an unknown function. (**C**) The putative IV-C RNP complex likely binds a nucleic acid target and cleaves that target with the HD nuclease domain. Created with BioRender.com.

The ultimate function of type IV-B Csf RNP complexes is still unknown (**Figure 5-8B**). Many questions of type IV-B system function will be answered as the source of the RNA component of the IV-B RNP complex is discovered and the function of the accessory protein CasCysH is understood. We propose that the Csf RNP complex will bind a nucleic acid target and recruit CasCysH to modify the nucleic acid.

No biochemical studies have been done on type IV-C systems, to date. We hypothesize that the IV-C Csf proteins will form an RNP complex with a crRNA and the Cas10-like subunit (**Figure 5-8C**). The IV-C Csf RNP complex will bind a nucleic acid target complementary to the crRNA and the HD nuclease domain of the Cas10-like subunit will cleave the target. Some IV-C systems have a CRISPR and a crRNA processing endonuclease and others do not, suggesting some IV-C systems may serve a crRNA-guided defense function while others may employ Cas proteins to perform an entirely different, non-defense function. Future studies should seek to understand the role of Cas10 within the type IV Csf RNP complex and the overall function of type IV-C CRISPR-Cas systems.

To understand the function of type IV CRISPR-Cas systems, it is critical that we determine the structures and biochemical functions of the type IV subtype specific proteins: CasDinG, CasCysH, and Cas10-like. Phylogenetic trees suggest that the IV-A DinG and IV-B CysH have evolved to support a putative Cas specific function. The IV-C Cas10 also has a unique domain composition that likely supports a distinct function. We propose that, due to the different accessory proteins and subtype specific proteins encoded by the three subtypes, each type IV subtype will have a distinct mechanism of

action and possibly distinct function. We highly anticipate future work detailing the mechanisms and functions of type IV RNP complexes and their accessory proteins.

METHODS

Sequence Alignment and Phylogenetic Tree Generation for DinG and CysH

DinG sequences were identified by searching all bacterial genomes from NCBI (downloaded on 12/03/2020) for homologs of a list of manually curated non-Cas associated dinGs using BLAST (S. F. Altschul et al., 1990), as well as matches to HMM profiles for cas-associated dinGs (csf4) using hmmsearch3 (hmmer.org). DinG homologs were classified as cas-associated if they were within 10 genes of a Type IV-A CRISPR-Cas locus. Type IV-A CRISPR-Cas loci were identified by searching for loci which contained matches to HMM profiles for csf2, csf3, csf4, and csf5 within 10 genes of each other. All HMM profiles were obtained from TIGRFAM and Makarova et al., 2020 and searching was performed using hmmsearch3 with an evalue cutoff of 0.001. All BLAST searches were performed with an evalue cutoff of 0.001.

CysH sequences were identified in the same way as dinG sequences. Type IV-B CRISPR-Cas loci were identified by searching for loci which contained homologs of csf2, csf3, and cysH. CysH sequences were considered cas-associated if they were within 15 genes of a Type IV-B CRISPR-Cas locus.

All dinG and cysH sequences were then clustered using MMseqs2 (Steinegger & Söding, 2017) with a minimum sequence identity of 0.80 within a cluster. It was confirmed that no clusters contained both cas and non-cas associated sequences and one representative sequence from each cluster was included in the final tree. Sequences were

then aligned with MAFFT (L-INS-i option) and phylogenies were reconstructed using MrBayes (Katoh & Standley, 2013; Ronquist et al., 2012). MrBayes was run until the standard deviation of split frequencies remained below 0.02 for 100000 generations.

Sequence Alignment and Phylogenetic Tree Generation for Cas7 and Cas6 Homologs

Alignments were performed and phylogenetic trees generated as described for DinG and CysH. Only sequences identified in Makarova et al., 2020 were included.

CONFLICT OF INTEREST

J.B.-D. is a scientific advisory board member of SNIPR Biome and Excision Biotherapeutics, and a scientific advisory board member and co-founder of Acrigen Biosciences.

AUTHOR CONTRIBUTIONS

H.N.T. and R.N.J conceived of and wrote the manuscript. H.N.T., E.L., M.A., T. H., and D. K. performed alignments and generated phylogenetic trees. All authors provided critical feedback on the manuscript.

FUNDING STATEMENT

J.B.-D. and the Bondy-Denomy lab were supported by the UCSF Program for Breakthrough Biomedical Research funded in part by the Sandler Foundation, the Searle Fellowship, the Vallee Foundation, the Innovative Genomics Institute, and the NIH

[DP5-OD021344, R01GM127489]. Research in the Jackson Lab is supported by the National Institutes of Health National Institute of General Medical Sciences Maximizing Investigators' Research Award (R35GM138080).

ACKNOWLEDGMENTS

We thank members of the Jackson and Bondy-Denomy labs for helpful discussions.

REFERENCES

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. (2016). C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science*, *353*(6299). https://doi.org/10.1126/science.aaf5573

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Aravind, L., & Koonin, E. V. (1998). The HD domain defines a new superfamily of metal-dependent phosphohydrolases. *Trends in Biochemical Sciences*, *23*(12), 469–472. https://doi.org/10.1016/S0968-0004(98)01293-6

Beloglazova, N., Petit, P., Flick, R., Brown, G., Savchenko, A., & Yakunin, A. F. (2011). Structure and activity of the Cas3 HD nuclease MJ0384, an effector enzyme of

the CRISPR interference. *The EMBO Journal*, *30*(22), 4616–4627. https://doi.org/10.1038/emboj.2011.377

Bernheim, A., Bikard, D., Touchon, M., & Rocha, E. P. C. (2020). Atypical organizations and epistatic interactions of CRISPRs and cas clusters in genomes and their mobile genetic elements. *Nucleic Acids Research*, *48*(2), 748–760. https://doi.org/10.1093/nar/gkz1091

Blum, M., Chang, H.-Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., Nuka, G., Paysan-Lafosse, T., Qureshi, M., Raj, S., Richardson, L., Salazar, G. A., Williams, L., Bork, P., Bridge, A., Gough, J., Haft, D. H., Letunic, I., Marchler-Bauer, A., … Finn, R. D. (2020). The InterPro protein families and domains database: 20 years on. *Nucleic Acids Research*, *49*(D1), D344–D354. https://doi.org/10.1093/nar/gkaa977

Bork, P., & Koonin, E. V. (1994). A P-loop-like motif in a widespread ATP pyrophosphatase domain: Implications for the evolution of sequence motifs and enzyme activity. *Proteins*, *20*(4), 347–355. https://doi.org/10.1002/prot.340200407

Brouns, S. J. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J. H., Snijders, A. P. L., Dickman, M. J., Makarova, K. S., Koonin, E. V., & van der Oost, J. (2008). Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. *Science*, *321*(5891), 960–964. https://doi.org/10.1126/science.1159689

Carroll, K. S., Gao, H., Chen, H., Stout, C. D., Leary, J. A., & Bertozzi, C. R. (2005). A Conserved Mechanism for Sulfonucleotide Reduction. *PLoS Biology*, *3*(8). https://doi.org/10.1371/journal.pbio.0030250

Carte, J., Wang, R., Li, H., Terns, R. M., & Terns, M. P. (2008). Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes & Development*, *22*(24), 3489–3496. https://doi.org/10.1101/gad.1742908

Chartron, J., Shiau, C., Stout, C. D., & Carroll, K. S. (2007). 3'-Phosphoadenosine-5'-phosphosulfate reductase in complex with thioredoxin: A structural snapshot in the catalytic cycle. *Biochemistry*, *46*(13), 3942–3951. https://doi.org/10.1021/bi700130e

Cheng, K., & Wigley, D. B. (2018). DNA translocation mechanism of an XPD family helicase. *ELife*. https://doi.org/10.7554/elife.42400.001

Crooks, G. E. (2004). WebLogo: A Sequence Logo Generator. *Genome Research*, *14*(6), 1188–1190. https://doi.org/10.1101/gr.849004

Crowley, V. M., Catching, A., Taylor, H. N., Borges, A. L., Metcalf, J., Bondy-Denomy, J., & Jackson, R. N. (2019). A Type IV-A CRISPR-Cas System in *Pseudomonas aeruginosa* Mediates RNA-Guided Plasmid Interference *In Vivo*. *The CRISPR Journal*, *2*(6), 434–440. https://doi.org/10.1089/crispr.2019.0048

Datsenko, K. A., Pougach, K., Tikhonov, A., Wanner, B. L., Severinov, K., & Semenova, E. (2012). Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nature Communications*, *3*(1), 945. https://doi.org/10.1038/ncomms1937

Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A., Eckert, M. R., Vogel, J., & Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, *471*(7340), 602–607. https://doi.org/10.1038/nature09886

Dwarakanath, S., Brenzinger, S., Gleditzsch, D., Plagens, A., Klingl, A., Thormann, K., & Randau, L. (2015). Interference activity of a minimal Type I CRISPR–Cas system from Shewanella putrefaciens. *Nucleic Acids Research, 43*(18), 8913–8923. https://doi.org/10.1093/nar/gkv882

Elmore, J., Deighan, T., Westpheling, J., Terns, R. M., & Terns, M. P. (2015). DNA targeting by the type I-G and type I-A CRISPR-Cas systems of Pyrococcus furiosus. *Nucleic Acids Research*, *43*(21), 10353–10363. https://doi.org/10.1093/nar/gkv1140

Elmore, J. R., Sheppard, N. F., Ramia, N., Deighan, T., Li, H., Terns, R. M., & Terns, M. P. (2016). Bipartite recognition of target RNAs activates DNA cleavage by the Type III-B CRISPR–Cas system. *Genes & Development*, *30*(4), 447–459. https://doi.org/10.1101/gad.272153.115

Estrella, M. A., Kuo, F.-T., & Bailey, S. (2016). RNA-activated DNA cleavage by the Type III-B CRISPR-Cas effector complex. *Genes & Development*, *30*(4), 460–470. https://doi.org/10.1101/gad.273722.115

Fairman-Williams, M. E., Guenther, U.-P., & Jankowsky, E. (2010). SF1 and SF2 helicases: Family matters. *Current Opinion in Structural Biology*, *20*(3), 313–324. https://doi.org/10.1016/j.sbi.2010.03.011

Faure, G., Shmakov, S. A., Yan, W. X., Cheng, D. R., Scott, D. A., Peters, J. E., Makarova, K. S., & Koonin, E. V. (2019). CRISPR–Cas in mobile genetic elements: Counter-defence and beyond. *Nature Reviews Microbiology*. https://doi.org/10.1038/s41579-019-0204-7

Garneau, J. E., Dupuis, M.-È., Villion, M., Romero, D. A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadán, A. H., & Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*, *468*(7320), 67–71. https://doi.org/10.1038/nature09523

Grodick, M. A., Segal, H. M., Zwang, T. J., & Barton, J. K. (2014). DNA-Mediated Signaling by Proteins with 4Fe–4S Clusters Is Necessary for Genomic Integrity. *Journal of the American Chemical Society*, *136*(17), 6470–6478. https://doi.org/10.1021/ja501973c

Guo, T. W., Bartesaghi, A., Yang, H., Falconieri, V., Rao, P., Merk, A., Eng, E. T., Raczkowski, A. M., Fox, T., Earl, L. A., Patel, D., & Subramaniam, S. (2017). Cryo-EM Structures Reveal Mechanism and Inhibition of DNA Targeting by a CRISPR-Cas Surveillance Complex. *Cell*, *171*(2), 414-426.e12. https://doi.org/10.1016/j.cell.2017.09.006

Hale, C., Kleppe, K., Terns, R. M., & Terns, M. P. (2008). Prokaryotic silencing (psi)RNAs in Pyrococcus furiosus. *RNA*, *14*(12), 2572–2579. https://doi.org/10.1261/rna.1246808

Hale, C. R., Zhao, P., Olson, S., Duff, M. O., Graveley, B. R., Wells, L., Terns, R. M., & Terns, M. P. (2009). RNA-Guided RNA Cleavage by a CRISPR RNA-Cas Protein Complex. *Cell*, *139*(5), 945–956. https://doi.org/10.1016/j.cell.2009.07.040

Han, W., Li, Y., Deng, L., Feng, M., Peng, W., Hallstrøm, S., Zhang, J., Peng, N., Liang, Y. X., White, M. F., & She, Q. (2017). A type III-B CRISPR-Cas effector

complex mediating massive target DNA destruction. *Nucleic Acids Research*, *45*(4), 1983–1993. https://doi.org/10.1093/nar/gkw1274

Haurwitz, R. E., Jinek, M., Wiedenheft, B., Zhou, K., & Doudna, J. A. (2010). Sequence- and Structure-Specific RNA Processing by a CRISPR Endonuclease. *Science*, *329*(5997), 1355–1358. https://doi.org/10.1126/science.1192272

Hayes, R. P., Xiao, Y., Ding, F., van Erp, P. B. G., Rajashankar, K., Bailey, S., Wiedenheft, B., & Ke, A. (2016). Structural basis for promiscuous PAM recognition in Type I-E Cascade from E. coli. *Nature*, *530*(7591), 499–503. https://doi.org/10.1038/nature16995Heler, R., Samai, P., Modell, J. W., Weiner, C., Goldberg, G. W., Bikard, D., & Marraffini, L. A. (2015). Cas9 specifies functional viral targets during CRISPR–Cas adaptation. *Nature*, *519*(7542), 199–202. https://doi.org/10.1038/nature14245

Hille, F., Richter, H., Wong, S. P., Bratovič, M., Ressel, S., & Charpentier, E. (2018). The Biology of CRISPR-Cas: Backward and Forward. *Cell*, *172*(6), 1239–1259. https://doi.org/10.1016/j.cell.2017.11.032

Jackson, R. N., Golden, S. M., Erp, P. B. G. van, Carter, J., Westra, E. R., Brouns, S. J. J., Oost, J. van der, Terwilliger, T. C., Read, R. J., & Wiedenheft, B. (2014). Crystal structure of the CRISPR RNA–guided surveillance complex from Escherichia coli. *Science*, *345*(6203), 1473–1479. https://doi.org/10.1126/science.1256328

Jackson, R. N., van Erp, P. B., Sternberg, S. H., & Wiedenheft, B. (2017). Conformational regulation of CRISPR-associated nucleases. *Current Opinion in Microbiology*, *37*, 110–119. https://doi.org/10.1016/j.mib.2017.05.010

Jackson, S. A., McKenzie, R. E., Fagerlund, R. D., Kieper, S. N., Fineran, P. C., & Brouns, S. J. J. (2017). CRISPR-Cas: Adapting to change. *Science*, *356*(6333). https://doi.org/10.1126/science.aal5056

Jia, N., Jones, R., Sukenick, G., & Patel, D. J. (2019). Second Messenger cA4 Formation within the Composite Csm1 Palm Pocket of Type III-A CRISPR-Cas Csm Complex and Its Release Path. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2019.06.013

Jore, M. M., Lundgren, M., van Duijn, E., Bultema, J. B., Westra, E. R., Waghmare, S. P., Wiedenheft, B., Pul, Ü., Wurm, R., Wagner, R., Beijer, M. R., Barendregt, A., Zhou, K., Snijders, A. P. L., Dickman, M. J., Doudna, J. A., Boekema, E. J., Heck, A. J. R., van der Oost, J., & Brouns, S. J. J. (2011). Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nature Structural & Molecular Biology*, *18*(5), 529–536. https://doi.org/10.1038/nsmb.2019

Jung, T.-Y., An, Y., Park, K.-H., Lee, M.-H., Oh, B.-H., & Woo, E. (2015). Crystal Structure of the Csm1 Subunit of the Csm Complex and Its Single-Stranded DNA-Specific Nuclease Activity. *Structure*, *23*(4), 782–790. https://doi.org/10.1016/j.str.2015.01.021

Kamruzzaman, M., & Iredell, J. R. (2020). CRISPR-Cas System in Antibiotic Resistance Plasmids in Klebsiella pneumoniae. *Frontiers in Microbiology*, *10*. https://doi.org/10.3389/fmicb.2019.02934

Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Kazlauskiene, M., Kostiuk, G., Venclovas, Č., Tamulaitis, G., & Siksnys, V. (2017). A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems. *Science, 357*(6351), 605–609. https://doi.org/10.1126/science.aao0100

Kazlauskiene, M., Tamulaitis, G., Kostiuk, G., Venclovas, Č., & Siksnys, V. (2016). Spatiotemporal Control of Type III-A CRISPR-Cas Immunity: Coupling DNA Degradation with the Target RNA Recognition. *Molecular Cell, 62*(2), 295–306. https://doi.org/10.1016/j.molcel.2016.03.024

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. (2018). Cas4 Facilitates PAM-Compatible Spacer Selection during CRISPR Adaptation. *Cell Reports, 22*(13), 3377–3384. https://doi.org/10.1016/j.celrep.2018.02.103

Koonin, E. V., & Makarova, K. S. (2017). Mobile Genetic Elements and Evolution of CRISPR-Cas Systems: All the Way There and Back. *Genome Biology and Evolution, 9*(10), 2812–2825. https://doi.org/10.1093/gbe/evx192

Koonin, E. V., & Makarova, K. S. (2019). Origins and evolution of CRISPR-Cas systems. *Philosophical Transactions of the Royal Society B: Biological Sciences, 374*(1772). https://doi.org/10.1098/rstb.2018.0087

Lee, H., Zhou, Y., Taylor, D. W., & Sashital, D. G. (2018). Cas4-Dependent Prespacer Processing Ensures High-Fidelity Programming of CRISPR Arrays. *Molecular Cell*, *70*(1), 48-59.e5. https://doi.org/10.1016/j.molcel.2018.03.003

Lewis, L. K., Jenkins, M. E., & Mount, D. W. (1992). Isolation of DNA damage-inducible promoters in Escherichia coli: Regulation of polB (dinA), dinG, and dinH by LexA repressor. *Journal of Bacteriology, 174*(10), 3377–3385.

Lintner, N. G., Kerou, M., Brumfield, S. K., Graham, S., Liu, H., Naismith, J. H., Sdano, M., Peng, N., She, Q., Copié, V., Young, M. J., White, M. F., & Lawrence, C. M. (2011). Structural and Functional Characterization of an Archaeal Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-associated Complex for Antiviral Defense (CASCADE). *Journal of Biological Chemistry*, *286*(24), 21643–21656. https://doi.org/10.1074/jbc.M111.238485

Liu, T. Y., Iavarone, A. T., & Doudna, J. A. (2017). RNA and DNA Targeting by a Reconstituted Thermus thermophilus Type III-A CRISPR-Cas System. *PLOS ONE*, *12*(1), e0170552. https://doi.org/10.1371/journal.pone.0170552

Madeira, F., Park, Y. mi, Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A. R. N., Potter, S. C., Finn, R. D., & Lopez, R. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research*, *47*(W1), W636–W641. https://doi.org/10.1093/nar/gkz268

Makarova, K. S., Aravind, L., Wolf, Y. I., & Koonin, E. V. (2011). Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. *Biology Direct*, *6*(1), 38. https://doi.org/10.1186/1745-6150-6-38

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., … Koonin, E. V. (2015). An updated evolutionary classification of CRISPR–Cas systems. *Nature Reviews Microbiology*, *13*(11), 722–736. https://doi.org/10.1038/nrmicro3569

Makarova, K. S., Wolf, Y. I., Iranzo, J., Shmakov, S. A., Alkhnbashi, O. S., Brouns, S. J. J., Charpentier, E., Cheng, D., Haft, D. H., Horvath, P., Moineau, S., Mojica, F. J. M., Scott, D., Shah, S. A., Siksnys, V., Terns, M. P., Venclovas, Č., White, M. F., Yakunin, A. F., … Koonin, E. V. (2020). Evolutionary classification of CRISPR–Cas systems: A burst of class 2 and derived variants. *Nature Reviews Microbiology*. https://doi.org/10.1038/s41579-019-0299-x

Marraffini, L. A., & Sontheimer, E. J. (2008). CRISPR Interference Limits Horizontal Gene Transfer in Staphylococci by Targeting DNA. *Science*, *322*(5909), 1843–1845. https://doi.org/10.1126/science.1165771

Marraffini, L. A., & Sontheimer, E. J. (2010). Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature*, *463*(7280), 568–571. https://doi.org/10.1038/nature08703

McRobbie, A.-M., Meyer, B., Rouillon, C., Petrovic-Stojanovska, B., Liu, H., & White, M. F. (2012). *Staphylococcus aureus* DinG, a helicase that has evolved into a nuclease. *Biochemical Journal*, *442*(1), 77–84. https://doi.org/10.1042/BJ20111903

Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J., & Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology*, *155*(3), 733–740. https://doi.org/10.1099/mic.0.023960-0

Mulepati, S., & Bailey, S. (2011). Structural and Biochemical Analysis of Nuclease Domain of Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-associated Protein 3 (Cas3). *Journal of Biological Chemistry*, *286*(36), 31896–31903. https://doi.org/10.1074/jbc.M111.270017

Mulepati, S., Héroux, A., & Bailey, S. (2014). Crystal structure of a CRISPR RNA–guided surveillance complex bound to a ssDNA target. *Science*, *345*(6203), 1479–1484. https://doi.org/10.1126/science.1256996

Newire, E., Aydin, A., Juma, S., Enne, V. I., & Roberts, A. P. (2020). Identification of a Type IV-A CRISPR-Cas System Located Exclusively on IncHI1B/IncFIB Plasmids in Enterobacteriaceae. *Frontiers in Microbiology*, *11*. https://doi.org/10.3389/fmicb.2020.01937

Niewoehner, O., Garcia-Doval, C., Rostøl, J. T., Berk, C., Schwede, F., Bigler, L., Hall, J., Marraffini, L. A., & Jinek, M. (2017). Type III CRISPR–Cas systems produce cyclic oligoadenylate second messengers. *Nature*, *548*(7669), 543–548. https://doi.org/10.1038/nature23467

Nobrega, F. L., Walinga, H., Dutilh, B. E., & Brouns, S. J. J. (2020). Prophages are associated with extensive CRISPR–Cas auto-immunity. *Nucleic Acids Research*, gkaa1071. https://doi.org/10.1093/nar/gkaa1071

Nuñez, J. K., Bai, L., Harrington, L. B., Hinder, T. L., & Doudna, J. A. (2016). CRISPR Immunological Memory Requires a Host Factor for Specificity. *Molecular Cell*, *62*(6), 824–833. https://doi.org/10.1016/j.molcel.2016.04.027

Nuñez, J. K., Kranzusch, P. J., Noeske, J., Wright, A. V., Davies, C. W., & Doudna, J. A. (2014). Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. *Nature Structural & Molecular Biology*, *21*(6), 528–534. https://doi.org/10.1038/nsmb.2820

O'Brien, R. E., Santos, I. C., Wrapp, D., Bravo, J. P. K., Schwartz, E. A., Brodbelt, J. S., & Taylor, D. W. (2020). Structural basis for assembly of non-canonical small

subunits into type I-C Cascade. *Nature Communications*, *11*. https://doi.org/10.1038/s41467-020-19785-8

Osawa, T., Inanaga, H., Sato, C., & Numata, T. (2015). Crystal Structure of the CRISPR-Cas RNA Silencing Cmr Complex Bound to a Target Analog. *Molecular Cell*, *58*(3), 418–430. https://doi.org/10.1016/j.molcel.2015.03.018

Özcan, A., Pausch, P., Linden, A., Wulf, A., Schühle, K., Heider, J., Urlaub, H., Heimerl, T., Bange, G., & Randau, L. (2018). Type IV CRISPR RNA processing and effector complex formation in Aromatoleum aromaticum. *Nature Microbiology*. https://doi.org/10.1038/s41564-018-0274-8

Peissert, S., Sauer, F., Grabarczyk, D. B., Braun, C., Sander, G., Poterszman, A., Egly, J.-M., Kuper, J., & Kisker, C. (2020). In TFIIH the Arch domain of XPD is mechanistically essential for transcription and DNA repair. *Nature Communications*, *11*(1), 1667. https://doi.org/10.1038/s41467-020-15241-9

Pinilla-Redondo, R., Mayo-Muñoz, D., Russel, J., Garrett, R. A., Randau, L., Sørensen, S. J., & Shah, S. A. (2019). Type IV CRISPR–Cas systems are highly diverse and involved in competition between plasmids. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkz1197

Poranen, M. M., & Tuma, R. (2004). Self-assembly of double-stranded RNA bacteriophages. *Virus Research*, *101*(1), 93–100. https://doi.org/10.1016/j.virusres.2003.12.009

Ren, B., Duan, X., & Ding, H. (2009). Redox control of the DNA damage-inducible protein DinG helicase activity via its iron-sulfur cluster. *The Journal of Biological Chemistry*, *284*(8), 4829–4835. https://doi.org/10.1074/jbc.M807943200

Rollie, C., Schneider, S., Brinkmann, A. S., Bolt, E. L., & White, M. F. (2015). Intrinsic sequence specificity of the Cas1 integrase directs new spacer acquisition. *ELife*, *4*. https://doi.org/10.7554/eLife.08716

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A., & Huelsenbeck, J. P. (2012). MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Systematic Biology*, *61*(3), 539–542. https://doi.org/10.1093/sysbio/sys029

Samai, P., Pyenson, N., Jiang, W., Goldberg, G. W., Hatoum-Aslan, A., & Marraffini, L. A. (2015). Co-transcriptional DNA and RNA Cleavage during Type III CRISPR-Cas Immunity. *Cell*, *161*(5), 1164–1174. https://doi.org/10.1016/j.cell.2015.04.027

Sashital, D. G., Jinek, M., & Doudna, J. A. (2011). An RNA-induced conformational change required for CRISPR RNA cleavage by the endoribonuclease Cse3. *Nature Structural & Molecular Biology*, *18*(6), 680–687. https://doi.org/10.1038/nsmb.2043

Savage, H., Montoya, G., Svensson, C., Schwenn, J. D., & Sinning, I. (1997). Crystal structure of phosphoadenylyl sulphate (PAPS) reductase: A new family of adenine nucleotide α hydrolases. *Structure*, *5*(7), 895–906. https://doi.org/10.1016/S0969-2126(97)00244-X

Shmakov, S. A., Makarova, K. S., Wolf, Y. I., Severinov, K. V., & Koonin, E. V. (2018). Systematic prediction of genes functionally linked to CRISPR-Cas systems by

gene neighborhood analysis. *Proceedings of the National Academy of Sciences*, *115*(23), E5307–E5316. https://doi.org/10.1073/pnas.1803440115

Shmakov, S. A., Sitnik, V., Makarova, K. S., Wolf, Y. I., Severinov, K. V., & Koonin, E. V. (2017). The CRISPR Spacer Space Is Dominated by Sequences from Species-Specific Mobilomes. *MBio*, *8*(5). https://doi.org/10.1128/mBio.01397-17

Sinkunas, T., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., & Siksnys, V. (2011). Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system: Cas3 nuclease/helicase. *The EMBO Journal*, *30*(7), 1335–1342. https://doi.org/10.1038/emboj.2011.41

Staals, R. H. J., Agari, Y., Maki-Yonekura, S., Zhu, Y., Taylor, D. W., van Duijn, E., Barendregt, A., Vlot, M., Koehorst, J. J., Sakamoto, K., Masuda, A., Dohmae, N., Schaap, P. J., Doudna, J. A., Heck, A. J. R., Yonekura, K., van der Oost, J., & Shinkai, A. (2013). Structure and Activity of the RNA-Targeting Type III-B CRISPR-Cas Complex of Thermus thermophilus. *Molecular Cell*, *52*(1), 135–145. https://doi.org/10.1016/j.molcel.2013.09.013

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., … van der Oost, J. (2014). RNA targeting by the type III-A CRISPR-Cas Csm complex of Thermus thermophilus. *Molecular Cell*, *56*(4), 518–530. https://doi.org/10.1016/j.molcel.2014.10.005

Steinegger, M., & Söding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotechnology*, *35*(11), 1026–1028. https://doi.org/10.1038/nbt.3988

Sternberg, S. H., Richter, H., Charpentier, E., & Qimron, U. (2016). Adaptation in CRISPR-Cas Systems. *Molecular Cell*, *61*(6), 797–808. https://doi.org/10.1016/j.molcel.2016.01.030

Tamulaitis, G., Kazlauskiene, M., Manakova, E., Venclovas, Č., Nwokeoji, A. O., Dickman, M. J., Horvath, P., & Siksnys, V. (2014). Programmable RNA Shredding by the Type III-A CRISPR-Cas System of Streptococcus thermophilus. *Molecular Cell*, *56*(4), 506–517. https://doi.org/10.1016/j.molcel.2014.09.027

Taylor, H. N., Warner, E. E., Armbrust, M. J., Crowley, V. M., Olsen, K. J., & Jackson, R. N. (2019). Structural basis of Type IV CRISPR RNA biogenesis by a Cas6 endoribonuclease. *RNA Biology*. https://doi.org/10.1080/15476286.2019.1634965

Thakur, R. S., Desingu, A., Basavaraju, S., Subramanya, S., Rao, D. N., & Nagaraju, G. (2014). Mycobacterium tuberculosis DinG Is a Structure-specific Helicase That Unwinds G4 DNA Implications for Targeting G4 DNA as a Novel Therapeutic Approach. *Journal of Biological Chemistry*, *289*(36), 25112–25136. https://doi.org/10.1074/jbc.M114.563569

Voloshin, O. N., & Camerini-Otero, R. D. (2007). The DinG Protein from *Escherichia coli* Is a Structure-specific Helicase. *Journal of Biological Chemistry*, *282*(25), 18437–18447. https://doi.org/10.1074/jbc.M700376200

Voloshin, O. N., Vanevski, F., Khil, P. P., & Camerini-Otero, R. D. (2003). Characterization of the DNA Damage-inducible Helicase DinG from *Escherichia*

coli. *Journal of Biological Chemistry*, *278*(30), 28284–28293. https://doi.org/10.1074/jbc.M301188200

Wang, J., Li, J., Zhao, H., Sheng, G., Wang, M., Yin, M., & Wang, Y. (2015). Structural and Mechanistic Basis of PAM-Dependent Spacer Acquisition in CRISPR-Cas Systems. *Cell*, *163*(4), 840–853. https://doi.org/10.1016/j.cell.2015.10.008

Wang, L., Jiang, S., Deng, Z., Dedon, P. C., & Chen, S. (2018). DNA phosphorothioate modification—A new multi-functional epigenetic system in bacteria. *FEMS Microbiology Reviews*, *43*(2), 109–122. https://doi.org/10.1093/femsre/fuy036

Wei, Y., Terns, R. M., & Terns, M. P. (2015). Cas9 function and host genome sampling in Type II-A CRISPR–Cas adaptation. *Genes & Development*, *29*(4), 356–361. https://doi.org/10.1101/gad.257550.114

Westra, E. R., Semenova, E., Datsenko, K. A., Jackson, R. N., Wiedenheft, B., Severinov, K., & Brouns, S. J. J. (2013). Type I-E CRISPR-Cas Systems Discriminate Target from Non-Target DNA through Base Pairing-Independent PAM Recognition. *PLOS Genetics*, *9*(9), e1003742. https://doi.org/10.1371/journal.pgen.1003742

Westra, E. R., van Erp, P. B. G., Künne, T., Wong, S. P., Staals, R. H. J., Seegers, C. L. C., Bollen, S., Jore, M. M., Semenova, E., Severinov, K., de Vos, W. M., Dame, R. T., de Vries, R., Brouns, S. J. J., & van der Oost, J. (2012). CRISPR Immunity Relies on the Consecutive Binding and Degradation of Negatively Supercoiled Invader DNA by Cascade and Cas3. *Molecular Cell*, *46*(5), 595–605. https://doi.org/10.1016/j.molcel.2012.03.018

Wu, Y., & Brosh, R. M., Jr. (2012). DNA helicase and helicase–nuclease enzymes with a conserved iron–sulfur cluster. *Nucleic Acids Research*, *40*(10), 4247–4260. https://doi.org/10.1093/nar/gks039

Yan, W. X., Hunnewell, P., Alfonse, L. E., Carte, J. M., Keston-Smith, E., Sothiselvam, S., Garrity, A. J., Chong, S., Makarova, K. S., Koonin, E. V., Cheng, D. R., & Scott, D. A. (2019). Functionally diverse type V CRISPR-Cas systems. *Science*, *363*(6422), 88–91. https://doi.org/10.1126/science.aav7271

Yin, Y., Yang, B., & Entwistle, S. (2019). Bioinformatics Identification of Anti-CRISPR Loci by Using Homology, Guilt-by-Association, and CRISPR Self-Targeting Spacer Approaches. *MSystems*, *4*(5), e00455-19, /msystems/4/5/msys.00455-19.atom. https://doi.org/10.1128/mSystems.00455-19

Yosef, I., Goren, M. G., & Qimron, U. (2012). Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. *Nucleic Acids Research*, *40*(12), 5569–5576. https://doi.org/10.1093/nar/gks216

You, D., Wang, L., Yao, F., Zhou, X., & Deng, Z. (2007). A Novel DNA Modification by Sulfur: DndA Is a NifS-like Cysteine Desulfurase Capable of Assembling DndC as an Iron−Sulfur Cluster Protein in *Streptomyces lividans* †. *Biochemistry*, *46*(20), 6126–6133. https://doi.org/10.1021/bi602615k

You, L., Ma, J., Wang, J., Artamonova, D., Wang, M., Liu, L., Xiang, H., Severinov, K., Zhang, X., & Wang, Y. (2019). Structure Studies of the CRISPR-Csm Complex Reveal Mechanism of Co-transcriptional Interference. *Cell*, *176*(1–2), 239-253.e16. https://doi.org/10.1016/j.cell.2018.10.052

Zhang, J., Graham, S., Tello, A., Liu, H., & White, M. F. (2016). Multiple nucleic acid cleavage modes in divergent type III CRISPR systems. *Nucleic Acids Research*, *44*(4), 1789–1799. https://doi.org/10.1093/nar/gkw020

Zhou, Y., Bravo, J. P. K., Taylor, H. N., Steens, J. A., Jackson, R. N., Staals, R. H. J., & Taylor, D. W. (2021). Structure of a type IV CRISPR-Cas ribonucleoprotein complex. *iScience*, 102201. https://doi.org/10.1016/j.isci.2021.102201

Zimmermann, L., Stephens, A., Nam, S.-Z., Rau, D., Kübler, J., Lozajic, M., Gabler, F., Söding, J., Lupas, A. N., & Alva, V. (2018). A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *Journal of Molecular Biology*, *430*(15), 2237–2243. https://doi.org/10.1016/j.jmb.2017.12.007

CHAPTER 6

crRNA GUIDED NUCLEIC ACID BINDING OF A TYPE IV-A RNP EFFECTOR COMPLEX[5]

ABSTRACT

The type IV-A CRISPR-Cas system from *Pseudomonas aeruginosa* has been shown to provide defense against invasive plasmids. However, the mechanisms of defense are unknown. To determine possible defense mechanisms, the *P. aeruginosa* type IV-A system was expressed and purified from *E. coli* and shown to form an RNP complex with a crRNA. EMSAs were used to show that the IV-A RNP complex preferentially binds nucleic acids complementary to the crRNA. Surprisingly, the IV-A RNP complex did not bind dsDNA. Preliminary data suggests that CasDinG may be a nuclease. We propose that the type IV-A CRISPR-Cas system provides plasmid defense through RNP binding to transcribed plasmid RNA which recruits CasDinG to degrade nearby nucleic acids, including the plasmid.

INTRODUCTION

Bacteria and archaea use CRISPR-Cas systems to defend themselves from invasion by mobile genetic elements (MGEs) such as bacteriophage and plasmids (Barrangou et al., 2007). The CRISPR locus within prokaryotic genomes contains sequences (called spacers) comprised of MGE DNA co-opted during a previous infection

---

[5] Co-authors: Hannah N. Taylor*, Jack P. K. Bravo, Jurre Steens, Hannah Domgaard, Raymond H. J. Staals, David Taylor, and Ryan N. Jackson.
Unpublished.
*HNT designed experiments, purified protein complexes, prepared and analyzed RNA sequencing, performed EMSAs, made the figures, and wrote the manuscript.

by the MGE (Bolotin et al., 2005; Mojica et al., 2005; Pourcel et al., 2005). Transcription and processing of the CRISPR locus generates CRISPR-derived RNAs (crRNAs) which are complementary to the MGE (Brouns et al., 2008; Carte et al., 2008; C. Hale et al., 2008). Cas proteins form a ribonucleoprotein (RNP) complex with the crRNA to surveil the cell for invasive nucleic acids and induce cleavage of the MGE upon binding (Brouns et al., 2008; C. R. Hale et al., 2009; Garneau et al., 2010). Two classes, six types, and over 33 subtypes of CRISPR-Cas systems have been classified, with each CRISPR-Cas type performing its defense function via unique Cas proteins and mechanisms (Makarova et al., 2015, 2020). The type IV-A CRISPR-Cas subtype defends against invasive plasmids (Crowley et al., 2019). However, the mechanisms used by the type IV-A system are unclear.

Type IV-A systems encode *csf1* (*cas8*-like), *csf2* (*cas7*-like), *csf3* (*cas5*-like), *csf5* (*cas6*-like), *dinG/csf4,* and a CRISPR locus (Makarova et al., 2015, 2020) (**Figure 6-1A**). In *Aromatoleum aromaticum* the IV-A Csf proteins (Csf1, Csf2, Csf3, and Csf5) form a Csf RNP complex with a crRNA that is processed by Csf5 (Özcan et al., 2018). The stoichiometry of this complex has not been determined, although it appears that Csf2 is present in excess to the other subunits (Özcan et al., 2018). Additionally, the type IV-A encoded DinG (CasDinG) has not been biochemically characterized nor studied in tandem with the IV-A Csf RNP complex. Non-CasDinG enzymes are helicases which have been implicated in damage inducible DNA repair (Lewis et al., 1992; Voloshin & Camerini-Otero, 2007; Thakur et al., 2014). Some non-CasDinG helicases also have nuclease activity (McRobbie et al., 2012). CasDinG contains helicase motifs but does not contain an identifiable nuclease domain, indicating that CasDinG may not have nuclease

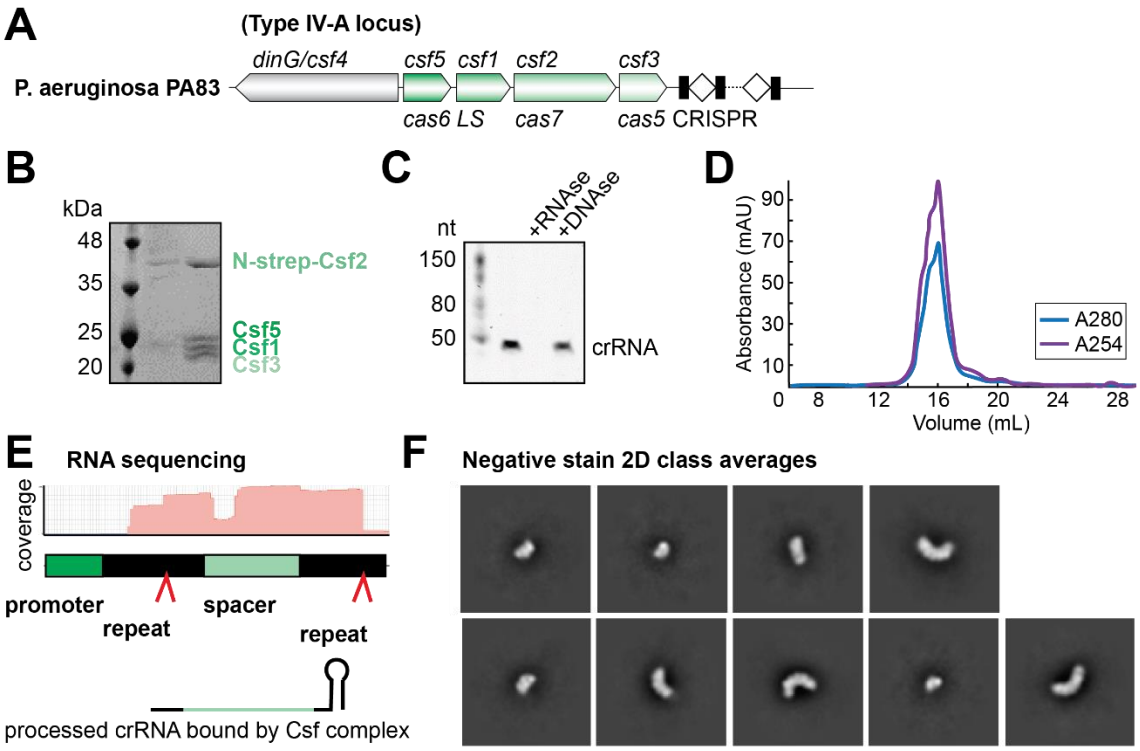

Figure 6-1. The IV-A Csf proteins form an RNP complex with a crRNA guide. (**A**) Type IV-A operon in *P. aeruginosa*. (**B**) SDS-PAGE analysis of the purified Csf RNP complex. (**C**) UREA-PAGE analysis of nucleic acids extracted from the purified Csf RNP complex. (**D**) Gel filtration chromatogram of the Csf RNP complex. (**E**) Sequenced miRNAs extracted from the purified Csf RNP complex. Red arrows indicate cleavage sites. (**F**) 2D class averages from negative stain cryo-EM of the Csf RNP complex.

activity or may have evolved a novel nuclease domain (Taylor et al., 2021). A recent study showed that the type IV-A CRISPR-Cas system from *Pseudomonas aeruginosa* was able to protect BL21 cells expressing the IV-A system from transformation by plasmids containing a sequence complementary to crRNAs derived from the CRISPR spacer (Crowley et al., 2019). The IV-A system genes *csf1, csf2, csf3, csf5,* and the CRISPR locus are all required for plasmid defense, indicating that a Csf RNP complex could be essential for defense activity (Crowley et al., 2019). *dinG/csf4* is also required for defense, and mutation of the CasDinG DEAD-box motif, which is required for helicase activity, abolishes defense activity (Crowley et al., 2019). This data indicates that CasDinG is likely a functional helicase and its helicase activity is crucial for the ultimate degradation of invasive nucleic acids.

To better understand the type IV-A CRISPR-Cas mechanisms of action, we expressed and purified the type IV-A *csf* genes from *P. aeruginosa*. We determined that a *P. aeruginosa* IV-A Csf RNP complex forms around a crRNA and binds complementary nucleic acid substrates. Mass spectrometry and negative stain experiments confirm the stoichiometry and shape of the Csf RNP complex. Preliminary data from the expression and purification of CasDinG from *P. aeruginosa* suggests that CasDinG may possess nuclease activity, though further replications and controls are required to confirm this activity. Altogether, the data presented here provide a more in-depth view of the mechanisms of defense against MGEs by the type IV-A CRISPR-Cas system and highlight important remaining questions.

RESULTS

The Type IV-A Csf Proteins from *Pseudomonas Aeruginosa* Form an RNP Complex with a crRNA

To determine whether a Csf RNP complex forms in the *P. aeruginosa* type IV-A CRISPR-Cas system, the *csf* genes were co-expressed and purified from pCDF-Duet and pACYC-Duet vectors (**Figure 6-2**). The encoded CRISPR consists of a repeat-spacer-repeat sequence. Based on the composition of a type IV-A Csf RNP complex in *Aromatoleum aromaticum* (Özcan et al., 2018), *casdinG* was not included in the co-expression. An N-terminal Strep-tag present on Csf2 was sufficient to pull-down a Csf RNP complex through affinity chromatography (**Figure 6-1**). All Csf proteins and an approximagely 50 bp RNA are present in the Csf RNP complex, as seen via SDS-PAGE and nucleic acid extractions. Negative stain cryo-electron microscopy (cryoEM) was performed to determine the shape of the complex (**Figure 6-1F**). The *P. aeruginosa* Csf RNP complex forms a curved filament of a distinct length, typical of muti-subunit Cas RNP complexes and reminiscent of the *A. aromaticum* Csf RNP complex (Jore et al., 2011; Staals et al., 2013; Jackson et al., 2014; Özcan et al., 2018). Native mass spectrometry (native-MS) confirms the stoichiometry of the complex 1:2:1:1 for Csf1:Csf2:Csf3:Csf5 (**Figure 6-3**). However, multiple complexes with molecular masses larger and smaller than the 1:2:1:1 complex were present in the native mass spectra. Negative stain and native-MS data combined indicate that the major RNP complex species has the 1:2:1:1 stoichiometry, though other stoichiometries may be possible. Some subunits of the Csf RNP complex may have flexible interactions with the complex, quickly dissociating and associating with the RNP complex.

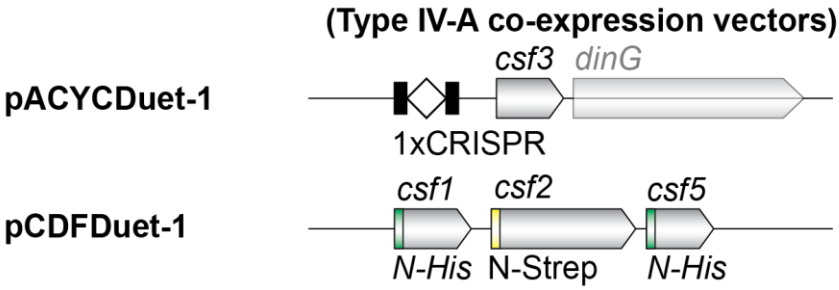

Figure 6-2. Expression vectors used to co-express the components of the IV-A Csf RNP complex. Tags included were the N-Strep on Csf2 alone or in combination with either an N-His on Csf1 or N-His on Csf5.

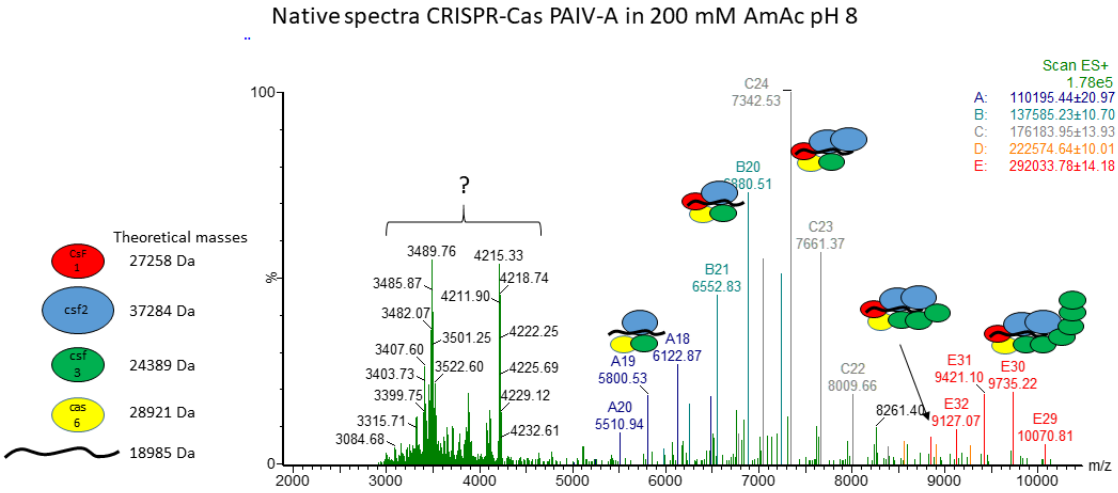

Figure 6-3. Native mass spectra of the Csf RNP complex. The spectra indicates that the most represented Csf RNP complex has a stoichiometry of 1:2:1:1 for Csf1:Csf2:Csf3:Csf5(Cas6).

To confirm that the RNA component of the Csf RNP complex is a processed crRNA, miRNAs were extracted from the Csf RNP complex and sequenced. The RNA sequencing data indicates the presence of a processed crRNA containing a spacer sequence flanked by repeat sequences which have been cleaved at the base of the repeat hairpin (**Figure 6-1E** and **Figure 6-4**).

The Type IV-A Csf RNP Complex Exhibits crRNA Guided Binding of Nucleic Acid Substrates

To determine whether the IV-A Csf RNP complex binds nucleic acid substrates, we performed electromobility shift assays (EMSAs) with the Csf RNP complex and several nucleic acid substrates both complementary to and non-complementary to the crRNA spacer (**Figure 6-5** and **Table 6-1**). To prevent contamination of the Csf RNP complex with excess Csf2 subunits, a dual affinity tag purification was performed, in which a secondary N-His affinity tag on Csf5 was used to purify the RNP complex after purification with the N-Strep affinity tag on Csf2. ssDNA, dsDNA, and RNA oligonucleotide substrates were designed to contain the same sequence, including the putative protospacer binding motif (PAM), used to demonstrate plasmid defense with the *P. aeruginosa* IV-A system (Crowley et al., 2019). A bubble DNA oligo, which is double stranded except at the target sequence, was also designed (**Figure 6-5A**). The Csf RNP complex bound the ssDNA target, bubble DNA target, and RNA target oligos with an apparent K_D of 20.1 nM ± 3.2, 48.4 nM ± 17.4, and 14.9 nM ± 3.8, respectively (**Figure 6-5B**).

At least two bands corresponding to bound oligos are present for all three of these oligos, suggesting that multiple binding events are occurring. This could be due, in part,

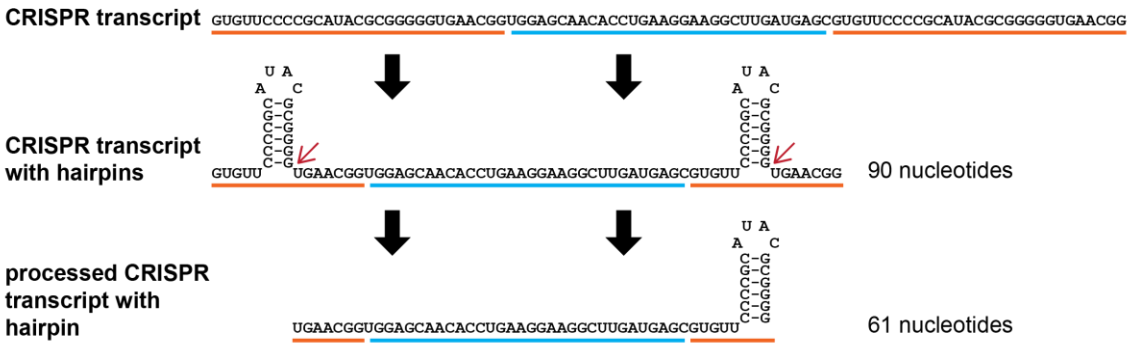

Figure 6-4. crRNA processing of the CRISPR transcript used during the expression of the Csf RNP complex. Hairpins form within the repeat sequence (orange). Red arrows indicate cleavage sites at the base of the hairpin. The 61 nucleotide fully processed crRNA contains one repeat hairpin and the spacer sequence (blue). The processed crRNA was confirmed by RNA sequencing (**Figure 6-1**).

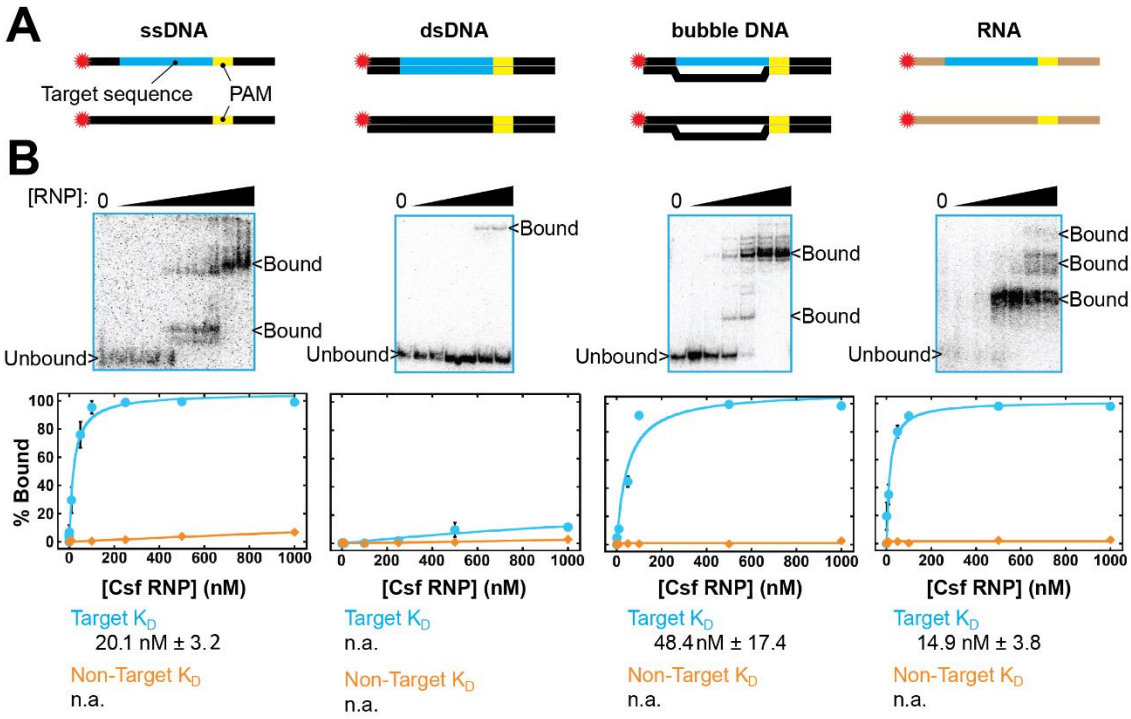

Figure 6-5. The Csf RNP complex binds single ssDNA, bubble DNA, and RNA, but not dsDNA. (**A**) Cartoons of oligos used to test Csf RNP complex binding. Red stars indicate 5'-^{32}P – radiolabels. (**B**) EMSAs and binding curves with the Csf RNP complex up to 1000 nM and oligos in (A). Data was collected in triplicate for each data point. Non-visible error bars are too tight to be seen behind the data point marker.

Table 6-1. Sequences of oligonucleotides used to perform EMSAs.

Oligo	Sequence	Length
ssDNA Target	TCTTTGCTGCCGCACTTGCTCATCAAGCCTTCCTTCAGGTGTTGCTCCAGAAAGGTGAGTTCTTCTTGTTGTT	73
ssDNA Non-Target	CGACGGGGTCGAGAGATAAGGGGCGCGCGGCGAGTGGCGCGCCACGTCGGAAATCTAGAGGGGAATTGTTATC	73
dsDNA Target	TCTTTGCTGCCGCACTTGCTCATCAAGCCTTCCTTCAGGTGTTGCTCCAGAAAGGTGAGTTCTTCTTGTTGTT	73
	AACAACAAGAAGAACTCACCTTTCTGGAGCAACACCTGAAGGAAGGCTTGATGAGCAAGTGCGGCAGCAAAGA	73
dsDNA Non-Target	CGACGGGGTCGAGAGATAAGGGGCGCGCGGCGAGTGGCGCGCCACGTCGGAAATCTAGAGGGGAATTGTTATC	73
	GATAACAATTCCCCTCTAGATTTCCGACGTGGCGCGCCACTCGCCGCGCGCCCCTTATCTCTCGACCCCGTCG	73
bubble DNA Target	TCTTTGCTGCCGCACTTGCTCATCAAGCCTTCCTTCAGGTGTTGCTCCAGAAAGGTGAGTTCTTCTTGTTGTT	73
	AACAACAAGAAGAACTCACCTTTCGCTTATCGATTAGCTTTACGATTAGCTGAAGCAAGTGCGGCAGCAAAGA	73
bubble DNA Non-Target	CGACGGGGTCGAGAGATAAGGGGCGCGCGGCGAGTGGCGCGCCACGTCGGAAATCTAGAGGGGAATTGTTATC	73
	GATAACAATTCCCCTCTAGATTTCAACCTTAGTGTTAAACGTTCAACAATTTACTTATCTCTCGACCCCGTCG	73
RNA Target	CGCACUUGCUCAUCAAGCCUUCCUUCAGGUGUUGCUCCAGAAAGGUGAGUUCU	53
RNA Non-Target	GAGAGAUAAGGGGCGCGCGGCGAGUGGCGCGCCACGUCGGAAAUCUAGAGGGG	53

to the length of the oligonucleotide (73 nucleotides for DNA substrates, 53 nucleotides for RNA substrates) (**Table 6-1**). As the concentration of RNP complex increases, the RNP complex may be binding the oligo in a non-sequence specific manner, thus shifting the bound substrate higher on the gel. However, even at the highest concentrations of RNP complex used, very little non-sequence specific binding occurred with substrates lacking complementarity to the crRNA, indicating that non-sequence specific binding is rare. The super-shift seen may be due to non-sequence specific binding by additional RNP complexes only if this secondary binding event is stabilized by the RNP complex bound to the substrate through crRNA sequence specific binding. A more tantalizing hypothesis suggests that the super-shift may be due to a dynamic RNP complex. After the RNP complex binds the substrate, additional Csf subunits may attach to the RNP complex, thus increasing its size. This hypothesis is supported by native mass spectrometry data, which shows that complexes of multiple sizes are present in the sample (**Figure 6-3**). Native mass spectrometry data collected on a RNP complex bound to a substrate would lend insight into the composition of the RNP complex after binding occurs.

Interestingly, the Csf RNP complex did not bind the dsDNA target oligo to an appreciable degree within the concentrations of Csf RNP tested (**Figure 6-5B**). For each nucleic acid oligo tested the Csf RNP complex did not bind the oligo containing the non-target sequence within the range of concentrations that were tested, indicating that nucleic acid binding by the Csf RNP complex is crRNA guided (**Figure 6-5B**).

CasDinG May Possess Nuclease Activity

As it is known that CasDinG is required for the *P. aeruginosa* IV-A defense activity, we next performed EMSAs with *P. aeruginosa* CasDinG (purified with a N-Strep tag) and the nucleic acid oligos described in the previous section (**Figure 6-6**). CasDinG EMSAs were performed with both the target and non-target versions of each oligo. However, since CasDinG was expressed without a CRISPR present and it has not been shown that CasDinG binds crRNAs, we expect that the presence of the target or non-target sequence within the nucleic acid oligo will not affect CasDinG binding. Bands corresponding to oligos bound by CasDinG are present only on the ssDNA target and non-target oligos, bubble DNA target oligo, and RNA target and non-target oligos (**Figure 6-6B**). Unexpectedly, CasDinG displayed cleavage activity with all oligos tested. This nuclease activity is visualized either as the appearance of a cleavage band or disappearance of the oligo altogether (**Figure 6-6B**). This ubiquitous cleavage activity was not expected, and gels were trimmed below the dye front before imaging the gel. I suspect that cleavage bands are present in all the gels but were often cut off the gel before visualization. Interestingly, multiple cleavage events may be occurring, as multiple cleavage bands are sometimes apparent and some cleavage bands disappear as additional cleavage events shift the bands lower in the gel where they were trimmed off and not visualized. Apparent K_D values were calculated if enough data points were obtained for a full binding curve. However, as is apparent by the high deviation values, these K_D values are skewed by the cleavage activity of CasDinG and are not accurate (**Figure 6-6B**). Two bound bands are observed for the ssDNA non-target oligo and the RNA target oligo, indicating multiple binding states are present.

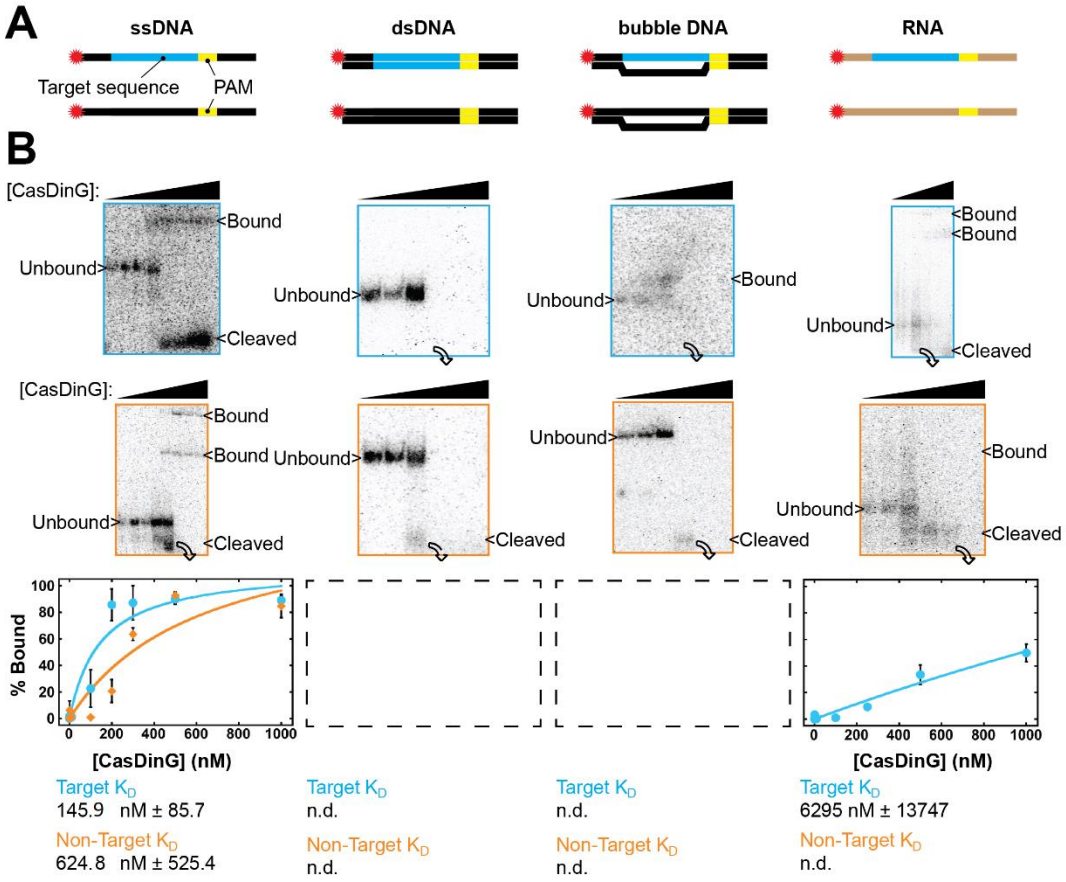

Figure 6-6. CasDinG may be a nuclease. (**A**) Cartoons of oligos used to test CasDinG binding. Same as in **Figure 6-5**. (**B**) EMSAs and binding curves for CasDinG up to 1000 nM and oligos in (A). Data was collected in triplicate for each data point. Non-visible error bars are too tight to be seen behind the data point marker. Binding curves could not be calculated for all oligos tested. Curved arrows on gels indicate that cleavage bands may be found below the area of gel that could be imaged.

DISCUSSION AND FUTURE DIRECTIONS

We have shown here that the *P. aeruginosa* type IV-A Csf proteins combine with a crRNA guide to form an RNP complex which is able to bind single stranded DNA and RNA targets containing a sequence complementary to the spacer of the crRNA. We also show that the *P. aeruginosa* CasDinG is a potential nuclease. Altogether, the data presented here imply the type IV-A system provides defense through a mechanism mediated by Csf RNP complex binding to complementary single stranded DNA or RNA molecules of invasive MGEs. We hypothesize that CasDinG is recruited to invasive nucleic acids by the Csf RNP complex and will degrade the invasive nucleic acids to provide defense. To better understand the mechanisms governing the type IV-A CRISPR-Cas system, a high-resolution structure of the Csf RNP complex and a biochemical characterization of DinG and its interactions with the Csf RNP complex are needed.

The First Description of a Csf RNP Complex Stoichiometry

As a type IV-A Csf RNP complex has been previously described in *Aromatoleum aromaticum* (Özcan et al., 2018), it was not surprising that the *P. aeruginosa* type IV-A Csf proteins also formed an RNP complex with a crRNA guide. We have reported here the first description of the stoichiometry of a IV-A Csf RNP complex as 1:2:1:1 for Csf1:Csf2:Csf3:Csf5. It is interesting that the Csf5 subunit is a component of the Csf RNP complex. The structure of a type IV-B Csf complex indicates that type IV systems may be evolutionarily closely related to type III CRISPR-Cas systems (Zhou et al., 2021). However, in type III systems the Cas6 subunit processes the crRNA then dissociates and is not a component of the RNP complex (Sokolowski et al., 2014; Staals et al., 2013). In

contrast, the Cas6 enzyme of type I systems is a component of the RNP complex (Brouns et al., 2008; Hochstrasser & Doudna, 2015). The presence of the Cas6-like protein Csf5 in the type IV-A RNP complex further supports the hypothesis that the type IV-B system may have acquired a type I *cas6*-like gene and eventually evolved to become the type IV-A CRISPR-Cas system (Taylor et al., 2019). This hypothesis explains the dual nature of type IV systems, as they contain some features reminiscent of type I systems and other features reminiscent of type III systems. To better understand the components and interactions within the Csf RNP complex, a structure should be obtained. Studies are underway to determine a cryo-electron microscopy structure of the Csf RNP complex in *apo* form and bound to nucleic acid substrates.

The Csf RNP Complex Binds Nucleic Acids Complementary to the crRNA, though its Preferred Substrate is Unclear

We also report here the first evidence of crRNA guided binding of nucleic acid substrates. A fundamental tenant of CRISPR-Cas defense is crRNA guided binding of nucleic acid substrates (Barrangou et al., 2007; Garneau et al., 2010; Gasiunas et al., 2012; Jinek et al., 2012). As the type IV-A system was recently shown to be a defense system (Crowley et al., 2019), we expected that the type IV-A Csf RNP complex would exhibit crRNA guided binding of a nucleic acid substrate. In the Crowley et al. study (2019), the type IV-A system provided defense against transformation by a plasmid. Thus, we also expected that the Csf RNP complex would specifically bind dsDNA containing a target (complementary to the crRNA spacer) sequence. Surprisingly, we found that the Csf RNP complex bound ssDNA and RNA substrates with a target

sequence, but not dsDNA (**Figure 6-5**). We have two hypotheses that could explain this discrepancy.

The first hypothesis is that the oligos used in this study do not contain an appropriate protospacer adjacent motif (PAM). The PAM is a 2-5 nucleotide motif in the target nucleic acid adjacent to the sequence complementary to the crRNA spacer (**Figure 6-5A**). PAM recognition by the RNP complex is a common method for identifying self vs. non-self by CRISPR-Cas systems during both adaptation (the acquiring of new spacers in the CRISPR locus) and interference (binding of target nucleic acids for defense), as the PAM is present in invasive MGEs but not in the CRISPR locus (Deveau et al., 2008; Horvath et al., 2008; Marraffini & Sontheimer, 2010). As the crRNA spacer is complementary to the CRISPR sequence from which it is derived, binding the PAM before inducing a cleavage response is critical for avoiding autoimmunity. PAMs are involved in target searching and R-loop formation in CRISPR-Cas systems that target dsDNA (Redding et al., 2015; Sternberg et al., 2015). The RNP complex scans dsDNA until a PAM is identified through Cas protein interactions which disrupt the hydrogen bonds between the dsDNA of the PAM. The dsDNA is further unzipped as the crRNA binds to the complementary strand, forming the R-loop (Anders et al., 2014; Mulepati et al., 2014; Sternberg et al., 2014; Leenay & Beisel, 2017). Thus, if the PAM is not present, the RNP complex may not bind to a complementary target. This may explain why the Csf RNP complex is unable to bind a dsDNA oligo containing a target sequence. All the other oligos tested, which the Csf RNP complex did bind when a target sequence was present, are single stranded at the target sequence, thus negating the need for a PAM to initiate melting of the dsDNA. However, there is little evidence to support the view that type IV-

A systems do utilize PAM sequences (Pinilla-Redondo et al., 2019). Further studies are needed to investigate whether type IV-A systems require PAMs for interference and what the specific requirements of the PAM are. Despite unknowns regarding the need for a PAM, the PAM, target sequence, and other flanking sequences of the oligos used to perform EMSAs in this study are identical to those used by Crowley et al. (2019) to demonstrate that the type IV-A CRISPR-Cas system can defend against plasmids. Thus, we know that when a cell expressing the *P. aeruginosa* type IV-A CRISPR-Cas system is invaded by dsDNA (plasmid) complementary to the spacer of the crRNA, a defense response is induced, leading to the destruction of the dsDNA or death of the cell. If the Csf RNP complex is not directly binding the target sequence on the dsDNA, then some other mechanism must be at play.

The second hypothesis that may explain why the Csf RNP complex does not bind dsDNA, despite being able to defend against invasive plasmids, is that the preferred substrate of the type IV-A Csf RNP complex is RNA, not dsDNA. Several CRISPR-Cas systems preferentially target RNA over DNA, mainly type III and type VI CRISPR-Cas systems (Elmore et al., 2016; Kazlauskiene et al., 2016; Abudayyeh et al., 2016; East-Seletsky et al., 2016). Type III systems, specifically, have been shown to defend against plasmid invasion by targeting RNA transcripts (Marraffini & Sontheimer, 2008; Tamulaitis et al., 2014). In these systems, binding of the RNP complex to the target RNA, induces tightly controlled, non-specific DNase and RNase activity in various enzymes that leads to either destruction of the invasive MGE or death of the cell, thus preventing propagation of the MGE (Athukoralage et al., 2020). A similar mechanism could be used by the type IV-A CRISPR system to provide defense. In this scenario, dsDNA would not

be bound by the Csf RNP complex, consistent with our EMSA data. When binding nucleic acids, the Csf RNP complex is unable to distinguish ssDNA from RNA, binding each with similar affinities (**Figure 6-5B**). It is possible that either, or both, ssDNA and/or RNA could be the preferred substrate for the Csf RNP complex. It should be noted that type III RNP complexes contain DNase activity within the Cas10 subunit that cleaves non-specific ssDNA and some also contain RNase activity within the Cas7-like subunits that cleaves the bound RNA (C. R. Hale et al., 2012; Jung et al., 2015; Staals et al., 2013). Although no cleavage activity was observed in the Csf RNP EMSAs we report, cleavage assays should be run under similar conditions to determine whether the Csf RNP complex exhibits any cleavage activity towards its target nucleic acid or spectator nucleic acids. Additionally, the type IV-A system may provide defense through cell death. When the type IV-A system is activated by Csf RNP complex binding, downstream effects result in cell death, thus protecting the bacterial community from the invasive nucleic acid present in the affected cell. This hypothesis should be explored through bacterial growth studies post infection/transformation.

The Potential Cleavage Activity of CasDinG Needs to be Confirmed

CasDinG does not contain a predicted nuclease domain (Taylor et al., 2021), so we were surprised to see such robust cleavage activity when attempting to perform EMSAs with CasDinG. Cleavage activity was observed for all oligos tested, regardless of sequence (**Figure 6-6**). The CasDinG EMSAs presented here are preliminary results and need to be validated. Firstly, the cleavage activity seen could be the result of a contaminating nuclease and not CasDinG. To confirm that CasDinG is the nuclease, the

CasDinG nuclease active site needs to be identified, then mutated. If the mutated CasDinG no longer exhibits nuclease activity, then the nuclease activity can be attributed to CasDinG. We suspect that the CasDinG nuclease active site will exist within its arch domain (Taylor et al., 2021) and studies are underway to remove chunks of the arch domain and test for nuclease activity. CasDinG is a predicted ATP-dependent helicase and its ATPase activity is required for defense (Voloshin et al., 2003; Voloshin & Camerini-Otero, 2007; McRobbie et al., 2012; Thakur et al., 2014; Crowley et al., 2019). Interestingly, ATP was not added to the CasDinG EMSAs, so any CasDinG nuclease activity occurs independently of ATPase/helicase activity. Equilibrium dissociation constants (K_D) could not be determined for CasDinG and the oligos tested, as binding was not completely visualized due to the cleavage activity. EMSAs should be repeated with CasDinG either in a mutated form where nuclease activity will not occur or in a buffer condition that allows substrate binding but will prevent nuclease activity. For example, if the nuclease activity is metal dependent, adding high concentrations of EDTA will inhibit the cleavage activity. In any of these cases where cleavage activity is suppressed or inhibited, binding may also be affected, and this should be taken into consideration when analyzing the assay. Oligonucleotide binding by CasDinG should also be determined in the presence of the Csf RNP complex. If CasDinG is recruited to invasive nucleic acids by the Csf RNP complex, CasDinG should have a higher affinity to oligonucleotides bound by the Csf RNP complex. If this recruitment event is confirmed, a structure of the Csf RNP complex bound to a substrate and CasDinG should be obtained to better understand the mechanisms governing the recruitment.

We have shown conclusively that the *P. aeruginosa* type IV-A Csf RNP complex binds single stranded nucleic acid substrates in a crRNA dependent manner. Binding of the target nucleic acid will likely result in downstream effects that eventually lead to destruction of the bound target. To better understand these downstream effects, it is critical that we understand the activities of CasDinG and its relationship to the Csf RNP complex. We know that CasDinG, in addition to the Csf RNP complex, is required for defense to occur (Crowley et al., 2019). Additionally, all CRISPR-Cas systems encode a nuclease that is responsible for the ultimate destruction of targeted MGEs (Makarova et al., 2020). Thus, it is not without precedent that CasDinG may have nuclease activity and this activity needs to be confirmed before we can understand the full mechanism of type IV-A CRISPR-Cas defense. As the type IV-A CRISPR-Cas system can defend against invasive plasmids (Crowley et al., 2019), we propose that CasDinG has nuclease activity and CasDinG nuclease activity is required for defense.

A Putative Model for Type IV-A Mediated Defense

We hypothesize that the type IV-A CRISPR-Cas system provides defense via the following steps: the Csf RNP complex binds a target nucleic acid, CasDinG is recruited to the nucleic acid by the Csf RNP complex, and CasDinG unwinds and cleaves the targeted nucleic acid to prevent its propagation (**Figure 6-7**). As discussed above, the preferred substrate for the Csf RNP complex is still unclear. It may utilize a strict PAM to bind dsDNA and recruit CasDinG to unwind and cleave the dsDNA. Or the Csf RNP complex may bind an RNA target and recruit CasDinG to unwind and cleave nucleic acids in the vicinity of the RNA target. The Csf RNP complex may bind RNA transcripts,

thus recruiting CasDinG to a transcription bubble, where it may non-specifically degrade DNA and RNA substrates. Altogether, the results discussed here further our understanding of type IV-A defense mechanisms and make clear many questions that need answers in order to gain a more complete understanding of the type IV-A CRISPR-Cas defense activity.

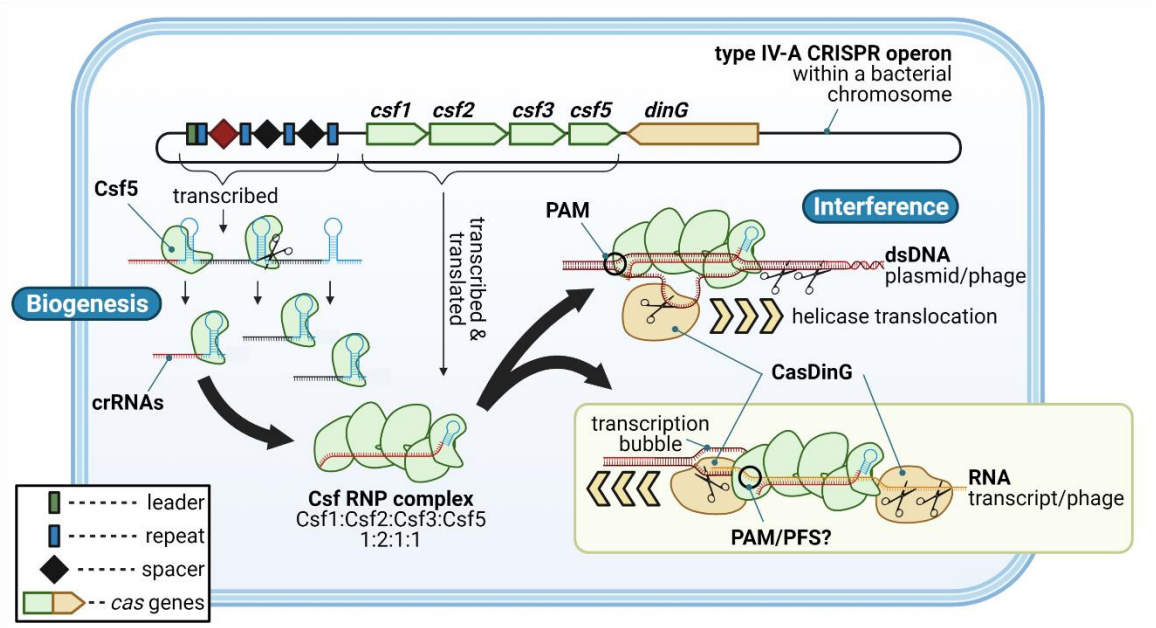

Figure 6-7. Putative model for the defense activity of type IV-A CRISPR-Cas systems. Type IV-A Csf subunits form an RNP complex with a CRISPR-derived RNA (crRNA). The RNP complex may bind dsDNA of RNA complementary to the crRNA. CasDinG may be recruited by the RNP complex to the target nucleic acid, where it will unwind and/or degrade nucleic acids. Created with BioRender.com.

COLLABORATORS AND CONTRIBUTIONS

Hannah N. Taylor, Jack P. K. Bravo, Jurre Steens, Hannah Domgaard, Raymond H. J. Staals, David Taylor, and Ryan N. Jackson

H.N.T. and R.N.J. conceived of experiments. J.P.K.B. performed negative stain analysis. H.D. expressed and purified CasDinG. J.S. performed and analyzed mass spectrometry experiments in collaboration with Arjan Barendregt at the Albert Heck lab. J.S. and R.H.J.S. analyzed RNA sequencing data. H.N.T. performed all other protein purifications, experiments, and data analysis. R.H.J.S., D.T., and R.N.J. supervised the work. H.N.T. wrote the chapter and made the figures.

MATERIALS AND METHODS

Csf RNP Complex Expression and Purification

Several growth conditions were tested to determine the optimal expression parameters for the *Pseudomonas aeruginosa* PA83 type IV-A Csf proteins individually and tangentially (**Table 6-2**, end of methods section). To express the *P. aeruginosa* PA83 type IV-A Csf RNP complex, chemically competent BL21-AI cells were transformed simultaneously with two plasmids encoding the csf genes and a 1X CRISPR: pCDFDuet_Pa_(N-His-)csf1_N-Strep-csf2_(N-His-)cas6 (secondary affinity tags were placed either on Csf1 or Cas6) and pACYCDuet_Pa83CR_csf3. An LB (lysogeny broth) starter with appropriate antibiotics was inoculated with one colony from the transformation and grown with shaking at 37°C overnight. 1 L of LB with appropriate antibiotics in a 2800 mL Fernbach culture flask was inoculated with 25 mL overnight starter and grown with shaking at 37°C until an optical density at 600 nm (OD600) of

0.5-0.7 was reached. The growth was then cold shocked without shaking at ~4°C for 45-60 minutes. The growth was then induced with 1mM IPTG and 0.2% L-arabinose and grown with shaking at 16°C overnight. Cells were harvested by centrifugation and the pellet stored at -80°C. Growths were scaled up by adding additional 1L growths.

Cell pellet was resuspended in Buffer W (100 mM Tris, pH 8.0; 150 mM NaCl; 10 mM EDTA) with 10 µg/mL leupeptin, 2 µg/mL aprotinin, and 170 µg/mL phenylmethylsulfonyl fluoride (PMSF). Cells were lysed by sonication on ice at power 3/30 for 5-10 minutes with intermittent 15 second rests. Lysate was clarified by centrifugation at 15000 RPM for 15 minutes at 4°C. Polyethylenimine (PEI) diluted in Buffer W was added to the soluble lysate at a final concentration of 0.1% PEI. Lysate was again clarified by centrifugation at 15000 RPM for 15 minutes at 4°C. Soluble lysate was combined with 4-5 mL Strep resin (IBA LifeSciences Strep-Tactin XT Superflow) previously equilibrated with Buffer W. Resin and lysate were incubated at 4°C overnight with continuous turning. Resin was settled in a column and the flow through collected. The resin was washed with 5 X 1 CV Buffer W. Pa IV-A RNP complex was eluted from the resin with 4 X 1 CV Buffer BXT (Buffer W + 50 mM biotin). Strep resin elutions were pooled and applied to a Ni column (constructs with a secondary affinity tag) or Q column (constructs with only the Strep-tag) equilibrated with Buffer W (HisTrap, crude 5 mL, GE Healthcare). Column was washed with 5 CV Buffer W and Pa IV-A RNP complex was eluted with a 0-100% 5 CV gradient up to 1 M imidazole or 1 M NaCl. The elution peak was pooled and concentrated to ≤0.5 mL with a Spin-X UF Corning 5k MWCO concentrator. Concentrated Pa Csf RNP complex was further purified by size exclusion chromatography on a HiLoad 16/600 Superose6 pg column (GE Healthcare,

Cytiva). The Superose6 column was equilibrated with 1.5 CVs SEC Buffer (50 mM HEPES, 150 mM NaCl), protein sample applied, and purified Pa Csf RNP complex eluted in SEC Buffer. Pa Csf RNP complex was further concentrated for downstream applications.

RNA Sequencing

miRNAs were extracted from the Csf RNP complex using the PureLink RNA Mini Kit (Thermo Fisher Scientific). Extracted RNAs were treated with T4 PNK (NEB) and ethanol precipitated before being sequenced on an Illumina MiSeq (Center for Integrated Biosystems, Utah State University, USA).

Native Mass Spectrometry

The *P. aeruginosa* type IV-A Csf RNP complex was buffer exchanged into 200 mM ammonium acetate pH 8.0 by seven consecutive dilution and concentration rounds on a centrifugal filter with a cut-off of 10 kDa (Sartorius) at 4°C. An aliquot of 1–2 µl with a concentration of approximately 1 uM was loaded into an in-house prepared borosilicate capillary (Kwik-Fil, World Precision Instruments, Sarasota, FL) on a P97 puller (SutterInstruments, Novato, CA). Using an Edwards Scancoat six pirani 501 sputter coater (Edwards Laboratories, Milpitas, USA), the needles were gold coated for proper conductivity. Samples were analyzed in positive ion mode on a standard commercial Q Exactive-UHMR instrument (Thermo Fisher Scientific). Ion transfer optics and voltage gradients throughout the instrument were optimized to transfer ions from different complexes through the instrument. Main focus was taken care of settings that

could either act as a mass filter or cause fragmentation. By applying 1.4 kV on the gold coated emitter ions are formed in the source. Ions are passing through a transfer tube to an S-lens, towards the exit lens, by applying RF on the electrodes. To further optimize the transfer of ions the detector optimization was set to low and the radio frequency amplitudes for the injection flatapole, bent flatapole, transfer multipole and higher-energy collisional dissociation (HCD) cell were set to 700, 940 and 900 V, respectively. In-source trapping (inject flatapole) was left at default setting. The setting for the ion transfer optics (injection flatapole, inter-flatapole lens, bent flatapole, transfer multipole and C-trap entrance lens), were put to 8,7,6, 4 and 5.8 V, respectively. Instead of trapping ions in the C-trap, ions were allowed to enter a Nitrogen filled HCD cell to allow sufficient collisional focusing. Increasing the collision cell pressure to approximately 2,3 × 10−10 mbar (readout in the Orbitrap analyzer) and the use of low HCD voltage (1.0V) improved detection sensitivity drastically. Before measurement, the instrument was manually mass-calibrated by direct infusion using solutions of 25 mg/ml CsI. All spectra were acquired by a resolving power of 8750 (transient length of 32 ms) at m/z 200 and processed using the standard Xcalibur 4.4 package (ThermoFisher Scientific).

Negative Stain Analysis

For negative-stain grid preparation, 4 µl of sample (at various concentrations ranging from 100 – 500 nM) was incubated on glow-discharged carbon-coated Formvar 300-mesh Cu grid for 90 seconds prior to blotting and stained twice with 20 µl 2% uranyl acetate (first stain immediately blotted, second stain incubated for 20 seconds prior to blotting) and allowed to dry. Grids were imaged using a FEI Talos Arctica microscope

operating at 200 kV, at a nominal magnification of 75,000x. Single-particle data processing was performed using CryoSparc.

CasDinG Expression and Purification

To express N-Strep tagged CasDinG, 1 L of LB with appropriate antibiotics is inoculated with 15 mL of overnight growth and the addition of 1000X metals mix (0.1 M $FeCl_3$-$6H_2O$, 1 M $CaCl_2$, 1 M $MnCl_2$-$4H_2O$, 1 M $ZnSO_4$-$7H_2O$, 0.2 M $CoCl_2$-$6H_2O$, 0.1 M $CuCl_2$-$2H_2O$, 0.2 M $NiCl_2$-$6H_2O$, 0.1 M Na_2MoO_4-$2H_2O$, 0.1 M Na_2SeO_3-$5H_2O$, 0.1 M H_3BO_3) and 1000X 1 M $MgSO_4$. Cells are grown to an optical density of 0.6 at 37°C. When OD is reached, the cells are induced with 0.5 mM IPTG and the temperature is dropped to 20°C for 5 hours. The cells are then harvested by centrifugation and the cell pellet stored at -80°C.

Cells are homogenized on ice with lysis buffer (100 mM Tris Base, 150 mM NaCl, pH 8.0). 0.5 µg/ml aprotinin, 0.5 µg/ml leupeptin, 0.7 µg/ml pepstatin A, and 170 µg/mL PMSF are added prior to sonication. Sonication is performed on ice at settings of 3/50 for 30 second intervals with 30 second rests for a total of 5 minutes or until the viscosity is close to that of water. The lysate is then centrifuged at 16000 RPM for 45 minutes. All purification steps are performed at 4°C. The supernatant is then poured gently over 1 mL Strep Tactin XT SuperFlow High Capacity resin (IBA Life Sciences). The flow through is collected and the resin is washed with 10 CV lysis buffer. Protein is eluted in elution buffer (100 mM Tris Base, 150 mM NaCl, 50 mM biotin, pH 8.0. Fractions containing DinG are pooled and run over a desalting column with low salt buffer (100 mM Tris Base, 10 mM NaCl, pH 8.0). Protein elution is collected and run

over a HP 5mL Heparin column (GE Healthcare) using a linear gradient between low salt buffer and high salt buffer (100 mM Tris Base, 1 M NaCl, pH 8.0). The protein elutions are then run over the desalting column with 100 mM Tris Base, 500 mM NaCl, pH 8.0. Elutions are concentrated to greater than 1 mg/ml, mixed with an equal volume of 100 mM Tris Base, 500 mM NaCl, 40% glycerol, pH 8.0, and flash frozen using liquid nitrogen for storage at -80°C.

Oligonucleotide Substrate Preparation

DNA and RNA oligonucleotides were ordered from Eurofins Genomics. Oligonucleotides were 5' radiolabeled with (γ-32P)-ATP (Perkin Elmer) and T4 polynucleotide kinase (NEB). Labeled RNAs were separated from excess ATP with a MicroSpin G-25 column (GE Healthcare), then gel purified on a denaturing (7M urea) 12% polyacrylamide gel, ethanol precipitated, and recovered in water or hybridization buffer (20 mM HEPES pH 7.5, 75 mM NaCl, 2mM EDTA, 10% glycerol, and 0.01% bromophenol blue).

dsDNA was prepared as previously described (van Erp et al., 2015) by mixing labeled oligonucleotides with five-fold molar excess of the complementary oligonucleotide in hybridization buffer, incubating at 95°C for 5 min, and gradually cooling to 25°C in a thermocycler. Oligonucleotide duplexes were gel purified on a non-denaturing 12% polyacrylamide gel, ethanol precipitated, and recovered in water or hybridization buffer.

Electromobility Shift Assays

Radiolabeled oligonucleotide substrates were mixed with 0-1000 nM Csf RNP or 0-1000 nM CasDinG in hybridization buffer with 1 mM TCEP and 5 mM $MgCl_2$. Reactions were incubated for 15-30 minutes at 37°C then separated on a non-denaturing 6% polyacrylamide gel. Gels were dried, exposed to a phosphor storage screen, and scanned with a Typhoon (GE Healthcare) phosphorimager. Bound and unbound fractions were quantified using ImageQuant (GE Healthcare) software. All data were fit using Kaleidagraph software to the equation:

$$Fraction\ bound\ oligo = \frac{M1\ *\ [protein]}{K_D +\ [protein]}$$

Where M1 is the amplitude of the binding curve and protein refers to either Csf RNP complex or CasDinG. Reported data is the average of three or more replicates and error bars represent standard deviations.

Table 6-2. *P. aeruginosa* type IV-A Csf protein expression and purification (pages 197-200).

Pseudomonas aeruginosa PA83 **type IV-A Cas gene expression and purification**

	Genes	Tags	Plasmid numbers	Expression strategy	Expression cell line	Proposed purification strategy	Date	Results	Additional details	Data location	Next steps
1	csf1	Strep-Csf2	**1287**	Co-expression	BL21 AI	Strep>SEC	4/10/2019	No detectable expression		Lab book #3 pg. 102	
	csf2		**1290**		BL21 DE3		4/25/2019	No detectable expression by Western	Expression tests	Lab book #3 pg. 112	
	csf3						5/15/2019	No detectable expression by Western	Expression tests	Lab book #3 pg. 122	
	cas6						5/21/2019	No detectable expression by Western	Expression tests	Lab book #3 pg. 123	None. Pursue other options
	dinG										
	CRISPR										
2	csf1	N-Strep-Csf2	**1287**	Co-expression	**BL21 AI**	N/A	5/15/2019	No detectable expression by Western	Expression tests	Lab book #3 pg. 122	None. Pursue other options
	csf2		**1293**		BL21 DE3	Strep>SEC	2/17/2020	Good expression. Complex formation detected	From fresh colony	Lab book #4 pg. 41	Optimize expression/purification to get more protein
	csf3						2/29/2020	Poor expression. Strep only	Bioreactor growth. From glycerol stock (GS)	Lab book #4 pg. 44	Optimize expression/purification to get more protein
	cas6						3/19/2020	Poor expression. Strep only	6 L growth. From GS	Lab book #4 pg. 45	Optimize expression/purification to get more protein
	CRISPR				(BL21 DE3)		4/3/2020	Poor expression	TB and autoinduction growths. From GS	Lab book #4 pg. 51	Optimize expression/purification to get more protein
							4/24/2020	Poor expression	Used Ozcan growth and purification protocols, From GS	Lab book #4 pg. 56	Optimize expression/purification to get more protein
							5/5/2020	Inconclusive	Expression tests. From GS	Lab book #4 pg. 58	Optimize expression/purification to get more protein
							5/26/2020	Apparent complex formation. Strep column acted weird	6L growth. From fresh colony	Lab book #4 pg. 62	Great option for purification of the complex
							6/25/2020		use Strep resin		
3	csf1	N-Strep-Csf1	**1343**	Single protein	BL21 AI	Strep	6/18/2019	Minimal expression		Lab book #3 pg. 131	None. Pursue other options
4	csf2	N-Strep-Csf2	**1347**	Single protein	BL21 AI	Strep>SEC	6/19/2019	Excellent expression, good purification over Strep		Lab book #3 pg. 131	
							6/26/2019	Purified over Strep & SEC. Stored sample		Lab book #3 pg. 134	
						Strep>Heparin>SEC	4/15/2020	Purified and stored		Lab book #4 pg. 53	Purify bucket loads and set up crystal trays

5 csf3	N-Strep-Csf3	**1217**	Single protein	BL21 AI	Strep	6/20/2019	Good expression, precipitation off of Strep		Lab book #3 pg. 131	Optimize purification to prevent precipitation
6 cas6	N-Strep-Cas6	**1352**	Single protein	BL21 AI	Strep	6/21/2019	Minimal expression		Lab book #3 pg. 131	None. Pursue other options
7 csf1	N-Strep-Csf1	**1343**	Individual expression, co-purification	BL21 AI	Strep>SEC	6/28/2019	Good Strep elution peak, no identifiable complex over SEC. Possibly only Csf2 and Csf3 were present		Lab book #3 pg. 136	None. Pursue other options
csf2	N-Strep-Csf2	**1347**								
csf3	N-Strep-Csf3	**1217**								
cas6	N-Strep-Cas6	**1352**								
8 csf1	N-His-Csf1	**1341**	Single protein	BL21 AI	Ni resin	7/9/2019	No apparent expression	Expression tests	Lab book #3 pg. 140	None. Pursue other options
9 csf1	C-His-Csf1	**1345**		BL21 AI	Ni resin	7/9/2019	Possible expression	Expression tests	Lab book #3 pg. 140	
					Ni>SEC	9/9/2019	Precipitated after Ni, but rescued protein purified well over sizing. Stored sample		Lab book #4 pg. 10	
					Ni	1/16/2020	Precipitated severely after Ni		Lab book #4 pg. 31	Optimize purification to prevent precipitation
10 cas6	N-His-Cas6	**1359**		BL21 DE3	Ni resin	7/9/2019	Possible expression	Expression tests	Lab book #3 pg. 140	Try a larger purification
11 cas6	C-His-Cas6	**1355**		BL21 AI	Ni resin	7/9/2019	Possible expression	Expression tests	Lab book #3 pg. 140	
					Ni>Q	8/26/2019			Lab Book #4 pg. 6	
					Ni>Heparin	9/9/2019	Cas6 does not stick to heparin		Lab book #4 pg. 10	Try a larger purification and alternative columns (Q, S, etc.)
12 csf1	N-MBP-Csf1	**1342**	Single protein	BL21 DE3	MBP>TEV	7/30/2019	Good expression, weird TEV cleavage	Expression tests	Lab #3 pg. 150	
					MBP>Q>TEV	8/15/2019	Struggled to purify Csf1 away from MBP & MBP-Csf1	Order of columns were varied with different samples	Lab #3 pg. 157	
					MBP>TEV>MBP	9/9/2019	Csf1 still contaminated after post-cleavage MBP		Lab book #4 pg. 10	Could try to optimize purification. His-Csf1 is more promising

13	csf1	C-MBP-Csf1	**1346**	Single protein	BL21 DE3	MBP	7/30/2019	Good expression	Expression tests	Lab #3 pg. 150	Could try to optimize purification. His-Csf1 is more promising
14	cas6	N-MBP-Cas6	**1351**	Single protein	BL21 DE3	MBP>TEV	7/30/2019	Good expression, degraded/precipitated after TEV cleavage	Expression tests	Lab #3 pg. 150	
						MBP>TEV>MBP	8/19/2019	Cas6 precipitated/degraded after TEV cleavage		Lab Book #4 pg. 1	
						MBP>TEV	8/26/2019			Lab Book #4 pg. 6	
						MBP>TEV>MBP	9/9/2019	Cas6 precipitated/degraded after TEV cleavage		Lab book #4 pg. 10	Purification needs to be optimized to stop precipitation after cleavage. Try different buffers, pH's, etc.
15	cas6	C-MBP-Cas6	**1356**	Single protein	BL21 DE3	MBP	7/30/2019	Less good expression	Expression tests	Lab #3 pg. 150	None. Pursue other options
16	csf1	C-His-Csf1	**1345**	Individual expression, co-purification (Cas6 & CRISPR coexpressed)	BL21 AI	Strep>MBP>SEC	9/25/2019	Apparent complex formation. Very little protein obtained - nothing visible on a gel after SEC		Lab book #4 pg. 23	Results need to be replicated, then purification optimized. Currently trying to get complex with fewer tags involved
	csf2	N-Strep-Csf2	**1347**		BL21 DE3						
	csf3	N-Strep-Csf3	**1217**								
	cas6	N-MBP-Cas6	**1351**								
	CRISPR		**1289**								
17	csf1	N-Strep-Csf2	**1287**	Co-expression	BL21 AI	Strep>SEC	2/6/2020	Good expression. Apparent complex with Csf2 and Cas6		Lab book #4 pg. 36	Purify Csf3 and DinG alone, combine with this prep to form a complex
	csf2		**1289**								
	cas6										
	CRISPR										
18	csf1	N-Strep-Csf2	**1351**	Co-expression	BL21 AI	Strep>SEC	2/13/2020	Good expression. crRNAs present. Clear peaks off SEC		Lab book #4 pg. 39	None. Pursue other options (no tag on cas6)
	csf2	N-MBP-Cas6	**1293**								
	csf3		**1496**								
	cas6										
	CRISPR										
19	csf1	N-Strep-Csf2	**1651**	Co-expression	BL21 AI	Strep>Ni>SEC	11.17.2020	Good expression. Full complex formation.	Csf2 from contaminating the complex	Lab book #4	Great option for purification of the complex
	csf2	N-His-Csf1	**1293**								
	csf3										
	cas6										
	CRISPR										

20	csf1	N-Strep-Csf2	**1651**	Co-expression	BL21 AI	Strep>Ni>SEC	12.23.2020	Great expression. Full complex formation.	Prevents excess Csf2 from contaminating the complex	Lab book #4	Great option for purification of the complex
	csf2	N-His-Cas6	**1293**								
	csf3										
	cas6										
	CRISPR										

REFERENCES

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. (2016). C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science*, *353*(6299). https://doi.org/10.1126/science.aaf5573

Anders, C., Niewoehner, O., Duerst, A., & Jinek, M. (2014). Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature*, *513*(7519), 569–573. https://doi.org/10.1038/nature13579

Athukoralage, J. S., Grüschow, S., Graham, S., Czekster, C. M., Rouillon, C., & White, M. F. (2020). The dynamic interplay of host and viral enzymes in type III CRISPR-mediated cyclic nucleotide signalling. *ELife*, *9*. https://doi.org/10.7554/eLife.55852

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A., & Horvath, P. (2007). CRISPR Provides Acquired Resistance against Viruses in Prokaryotes. *Science, 315*(5819), 1709–1712.

Bernheim, A., Bikard, D., Touchon, M., & Rocha, E. P. C. (2020). Atypical organizations and epistatic interactions of CRISPRs and cas clusters in genomes and their mobile genetic elements. *Nucleic Acids Research*, *48*(2), 748–760. https://doi.org/10.1093/nar/gkz1091

Bolotin, A., Quinquis, B., Sorokin, A., & Ehrlich, S. D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of

extrachromosomal origin. *Microbiology*, *151*(8), 2551–2561. https://doi.org/10.1099/mic.0.28048-0

Brouns, S. J. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J. H., Snijders, A. P. L., Dickman, M. J., Makarova, K. S., Koonin, E. V., & van der Oost, J. (2008). Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. *Science*, *321*(5891), 960–964. https://doi.org/10.1126/science.1159689

Carte, J., Wang, R., Li, H., Terns, R. M., & Terns, M. P. (2008). Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes & Development*, *22*(24), 3489–3496. https://doi.org/10.1101/gad.1742908

Crowley, V. M., Catching, A., Taylor, H. N., Borges, A. L., Metcalf, J., Bondy-Denomy, J., & Jackson, R. N. (2019). A Type IV-A CRISPR-Cas System in *Pseudomonas aeruginosa* Mediates RNA-Guided Plasmid Interference *In Vivo*. *The CRISPR Journal*, *2*(6), 434–440. https://doi.org/10.1089/crispr.2019.0048

Deveau, H., Barrangou, R., Garneau, J. E., Labonté, J., Fremaux, C., Boyaval, P., Romero, D. A., Horvath, P., & Moineau, S. (2008). Phage Response to CRISPR-Encoded Resistance in Streptococcus thermophilus. *Journal of Bacteriology*, *190*(4), 1390–1400. https://doi.org/10.1128/JB.01412-07

East-Seletsky, A., O'Connell, M. R., Knight, S. C., Burstein, D., Cate, J. H. D., Tjian, R., & Doudna, J. A. (2016). Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. *Nature*, *538*(7624), 270–273. https://doi.org/10.1038/nature19802

Elmore, J. R., Sheppard, N. F., Ramia, N., Deighan, T., Li, H., Terns, R. M., & Terns, M. P. (2016). Bipartite recognition of target RNAs activates DNA cleavage by the

Type III-B CRISPR–Cas system. *Genes & Development*, *30*(4), 447–459. https://doi.org/10.1101/gad.272153.115

Garneau, J. E., Dupuis, M.-È., Villion, M., Romero, D. A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadán, A. H., & Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*, *468*(7320), 67–71. https://doi.org/10.1038/nature09523

Gasiunas, G., Barrangou, R., Horvath, P., & Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences*, *109*(39), E2579–E2586. https://doi.org/10.1073/pnas.1208507109

Hale, C., Kleppe, K., Terns, R. M., & Terns, M. P. (2008). Prokaryotic silencing (psi)RNAs in Pyrococcus furiosus. *RNA*, *14*(12), 2572–2579. https://doi.org/10.1261/rna.1246808

Hale, C. R., Majumdar, S., Elmore, J., Pfister, N., Compton, M., Olson, S., Resch, A. M., Glover, C. V. C., Graveley, B. R., Terns, R. M., & Terns, M. P. (2012). Essential Features and Rational Design of CRISPR RNAs that Function with the Cas RAMP Module Complex to Cleave RNAs. *Molecular Cell*, *45*(3), 292–302. https://doi.org/10.1016/j.molcel.2011.10.023

Hale, C. R., Zhao, P., Olson, S., Duff, M. O., Graveley, B. R., Wells, L., Terns, R. M., & Terns, M. P. (2009). RNA-Guided RNA Cleavage by a CRISPR RNA-Cas Protein Complex. *Cell*, *139*(5), 945–956. https://doi.org/10.1016/j.cell.2009.07.040

Hochstrasser, M. L., & Doudna, J. A. (2015). Cutting it close: CRISPR-associated endoribonuclease structure and function. *Trends in Biochemical Sciences*, *40*(1), 58–66. https://doi.org/10.1016/j.tibs.2014.10.007

Horvath, P., Romero, D. A., Coûté-Monvoisin, A.-C., Richards, M., Deveau, H., Moineau, S., Boyaval, P., Fremaux, C., & Barrangou, R. (2008). Diversity, Activity, and Evolution of CRISPR Loci in Streptococcus thermophilus. *Journal of Bacteriology*, *190*(4), 1401–1412. https://doi.org/10.1128/JB.01415-07

Jackson, R. N., Golden, S. M., Erp, P. B. G. van, Carter, J., Westra, E. R., Brouns, S. J. J., Oost, J. van der, Terwilliger, T. C., Read, R. J., & Wiedenheft, B. (2014). Crystal structure of the CRISPR RNA–guided surveillance complex from Escherichia coli. *Science*, *345*(6203), 1473–1479. https://doi.org/10.1126/science.1256328

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*, *337*(6096), 816–821. https://doi.org/10.1126/science.1225829

Jore, M. M., Lundgren, M., van Duijn, E., Bultema, J. B., Westra, E. R., Waghmare, S. P., Wiedenheft, B., Pul, Ü., Wurm, R., Wagner, R., Beijer, M. R., Barendregt, A., Zhou, K., Snijders, A. P. L., Dickman, M. J., Doudna, J. A., Boekema, E. J., Heck, A. J. R., van der Oost, J., & Brouns, S. J. J. (2011). Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nature Structural & Molecular Biology*, *18*(5), 529–536. https://doi.org/10.1038/nsmb.2019

Jung, T.-Y., An, Y., Park, K.-H., Lee, M.-H., Oh, B.-H., & Woo, E. (2015). Crystal Structure of the Csm1 Subunit of the Csm Complex and Its Single-Stranded

DNA-Specific Nuclease Activity. *Structure*, *23*(4), 782–790. https://doi.org/10.1016/j.str.2015.01.021

Kazlauskiene, M., Tamulaitis, G., Kostiuk, G., Venclovas, Č., & Siksnys, V. (2016). Spatiotemporal Control of Type III-A CRISPR-Cas Immunity: Coupling DNA Degradation with the Target RNA Recognition. *Molecular Cell*, *62*(2), 295–306. https://doi.org/10.1016/j.molcel.2016.03.024

Leenay, R. T., & Beisel, C. L. (2017). Deciphering, Communicating, and Engineering the CRISPR PAM. *Journal of Molecular Biology*, *429*(2), 177–191. https://doi.org/10.1016/j.jmb.2016.11.024

Lewis, L. K., Jenkins, M. E., & Mount, D. W. (1992). Isolation of DNA damage-inducible promoters in Escherichia coli: Regulation of polB (dinA), dinG, and dinH by LexA repressor. *Journal of Bacteriology*, *174*(10), 3377–3385.

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., … Koonin, E. V. (2015). An updated evolutionary classification of CRISPR–Cas systems. *Nature Reviews Microbiology*, *13*(11), 722–736. https://doi.org/10.1038/nrmicro3569

Makarova, K. S., Wolf, Y. I., Iranzo, J., Shmakov, S. A., Alkhnbashi, O. S., Brouns, S. J. J., Charpentier, E., Cheng, D., Haft, D. H., Horvath, P., Moineau, S., Mojica, F. J. M., Scott, D., Shah, S. A., Siksnys, V., Terns, M. P., Venclovas, Č., White, M. F., Yakunin, A. F., … Koonin, E. V. (2020). Evolutionary classification of CRISPR–

Cas systems: A burst of class 2 and derived variants. *Nature Reviews Microbiology*. https://doi.org/10.1038/s41579-019-0299-x

Marraffini, L. A., & Sontheimer, E. J. (2008). CRISPR Interference Limits Horizontal Gene Transfer in Staphylococci by Targeting DNA. *Science*, *322*(5909), 1843–1845. https://doi.org/10.1126/science.1165771

Marraffini, L. A., & Sontheimer, E. J. (2010). Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature*, *463*(7280), 568–571. https://doi.org/10.1038/nature08703

McRobbie, A.-M., Meyer, B., Rouillon, C., Petrovic-Stojanovska, B., Liu, H., & White, M. F. (2012). *Staphylococcus aureus* DinG, a helicase that has evolved into a nuclease. *Biochemical Journal*, *442*(1), 77–84. https://doi.org/10.1042/BJ20111903

Mojica, F. J. M., Diez-Villasenor, C., Garcia-Martinez, J., & Soria, E. (2005). Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements. *Journal of Molecular Evolution*, *60*(2), 174–182. https://doi.org/10.1007/s00239-004-0046-3

Mulepati, S., Héroux, A., & Bailey, S. (2014). Crystal structure of a CRISPR RNA–guided surveillance complex bound to a ssDNA target. *Science*, *345*(6203), 1479–1484. https://doi.org/10.1126/science.1256996

Özcan, A., Pausch, P., Linden, A., Wulf, A., Schühle, K., Heider, J., Urlaub, H., Heimerl, T., Bange, G., & Randau, L. (2018). Type IV CRISPR RNA processing and effector complex formation in Aromatoleum aromaticum. *Nature Microbiology*. https://doi.org/10.1038/s41564-018-0274-8

Pinilla-Redondo, R., Mayo-Muñoz, D., Russel, J., Garrett, R. A., Randau, L., Sørensen, S. J., & Shah, S. A. (2019). Type IV CRISPR–Cas systems are highly diverse and involved in competition between plasmids. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkz1197

Pourcel, C., Salvignol, G., & Vergnaud, G. (2005). CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology, 151*(3), 653–663. https://doi.org/10.1099/mic.0.27437-0

Redding, S., Sternberg, S. H., Marshall, M., Gibb, B., Bhat, P., Guegler, C. K., Wiedenheft, B., Doudna, J. A., & Greene, E. C. (2015). Surveillance and Processing of Foreign DNA by the Escherichia coli CRISPR-Cas System. *Cell*, 163(4), 854–865. https://doi.org/10.1016/j.cell.2015.10.003

Sokolowski, R. D., Graham, S., & White, M. F. (2014). Cas6 specificity and CRISPR RNA loading in a complex CRISPR-Cas system. *Nucleic Acids Research, 42*(10), 6532–6541. https://doi.org/10.1093/nar/gku308

Staals, R. H. J., Agari, Y., Maki-Yonekura, S., Zhu, Y., Taylor, D. W., van Duijn, E., Barendregt, A., Vlot, M., Koehorst, J. J., Sakamoto, K., Masuda, A., Dohmae, N., Schaap, P. J., Doudna, J. A., Heck, A. J. R., Yonekura, K., van der Oost, J., & Shinkai, A. (2013). Structure and Activity of the RNA-Targeting Type III-B CRISPR-Cas Complex of Thermus thermophilus. *Molecular Cell*, *52*(1), 135–145. https://doi.org/10.1016/j.molcel.2013.09.013

Sternberg, S. H., LaFrance, B., Kaplan, M., & Doudna, J. A. (2015). Conformational control of DNA target cleavage by CRISPR–Cas9. *Nature*, 527(7576), 110–113. https://doi.org/10.1038/nature15544

Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C., & Doudna, J. A. (2014). DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*, *507*(7490), 62–67. https://doi.org/10.1038/nature13011

Tamulaitis, G., Kazlauskiene, M., Manakova, E., Venclovas, Č., Nwokeoji, A. O., Dickman, M. J., Horvath, P., & Siksnys, V. (2014). Programmable RNA Shredding by the Type III-A CRISPR-Cas System of Streptococcus thermophilus. *Molecular Cell*, *56*(4), 506–517. https://doi.org/10.1016/j.molcel.2014.09.027

Taylor, H. N., Laderman, E., Armbrust, M. J., Hallmark, T., Keiser, D., Bondy-Denomy, J., & Jackson, R. N. (2021). Positioning diverse type IV structures and functions within class 1 CRISPR-Cas systems. *Frontiers in Microbiology*, [Under review].

Taylor, H. N., Warner, E. E., Armbrust, M. J., Crowley, V. M., Olsen, K. J., & Jackson, R. N. (2019). Structural basis of Type IV CRISPR RNA biogenesis by a Cas6 endoribonuclease. *RNA Biology*. https://doi.org/10.1080/15476286.2019.1634965

Thakur, R. S., Desingu, A., Basavaraju, S., Subramanya, S., Rao, D. N., & Nagaraju, G. (2014). Mycobacterium tuberculosis DinG Is a Structure-specific Helicase That Unwinds G4 DNA Implications for Targeting G4 DNA as a Novel Therapeutic Approach. *Journal of Biological Chemistry*, *289*(36), 25112–25136. https://doi.org/10.1074/jbc.M114.563569

van Erp, P. B. G., Jackson, R. N., Carter, J., Golden, S. M., Bailey, S., & Wiedenheft, B. (2015). Mechanism of CRISPR-RNA guided recognition of DNA targets in

Escherichia coli. *Nucleic Acids Research*, *43*(17), 8381–8391. https://doi.org/10.1093/nar/gkv793

Voloshin, O. N., & Camerini-Otero, R. D. (2007). The DinG Protein from *Escherichia coli* Is a Structure-specific Helicase. *Journal of Biological Chemistry*, *282*(25), 18437–18447. https://doi.org/10.1074/jbc.M700376200

Voloshin, O. N., Vanevski, F., Khil, P. P., & Camerini-Otero, R. D. (2003). Characterization of the DNA Damage-inducible Helicase DinG from *Escherichia coli*. *Journal of Biological Chemistry*, *278*(30), 28284–28293. https://doi.org/10.1074/jbc.M301188200

Zhou, Y., Bravo, J. P. K., Taylor, H. N., Steens, J. A., Jackson, R. N., Staals, R. H. J., & Taylor, D. W. (2021). Structure of a type IV CRISPR-Cas ribonucleoprotein complex. *iScience*, *24*(3), 102201. https://doi.org/10.1016/j.isci.2021.102201

CHAPTER 7

SUMMARY AND FUTURE DIRECTIONS

INTRODUCTION

The defense and anti-defense systems employed by prokaryotes and the mobile genetic elements that infect them have long been exploited to gain insights into prokaryotic biology and develop biotechnology to advance basic research, medicine, and more (Campbell, 2003; Westra et al., 2012; Doron et al., 2018; Rostøl & Marraffini, 2019). The prokaryotic CRISPR-Cas defense systems are no different, and have spurred the development of several gene editing, therapeutic, and diagnostic tools (Adli, 2018; Pickar-Oliver & Gersbach, 2019). To provide a timely example, CRISPR-Cas systems have been repurposed as SARS-CoV-2 detection and diagnostic tools, as well as COVID-19 therapeutics (Atzrodt et al., 2020; Hou et al., 2020; Nalawansha & Samarasinghe, 2020; Nemudryi et al., 2020). Beyond their use as biotechnological tools, the study of CRISPR-Cas systems has broadened our knowledge of the scope of bacterial immune systems and allowed for the discovery of novel and intricate biological mechanisms (Barrangou et al., 2007; Barrangou, 2015; Doron et al., 2018). Two classes, six types, and over 33 subtypes of CRISPR-Cas systems have been discovered to date (Makarova et al., 2020). Of the many diverse CRISPR-Cas types, type IV CRISR-Cas systems have been the least studied.

Although originally thought to be rare and only encoded in extremophiles, type IV CRISPR-Cas systems are widespread and present in intensively studied disease causing organisms, such as *Pseudomonas aeruginosa, Mycobacterium tuberculosis,* and

Klebsiella phenumoniae (Makarova et al., 2015, 2020; Crowley et al., 2019; Kamruzzaman & Iredell, 2020; Newire et al., 2020). Type IV CRISPR-Cas systems are typified by their lack of common CRISPR-Cas elements, making them both difficult to study and incredibly interesting (Makarova et al., 2015). In fact, some type IV CRISPR-Cas systems lack a CRISPR locus, thus making them CRISPR-less CRISPR-Cas systems. Additionally, type IV systems are almost exclusively encoded on plasmids and prophages (Makarova et al., 2015; Koonin & Makarova, 2017, 2019; Shmakov et al., 2018; Pinilla-Redondo et al., 2019; Faure et al., 2019). Type IV CRISPR-Cas systems have been subdivided into three subtypes (IV-A, IV-B, and IV-C), each of which encodes a subtype specific accessory protein (CasDinG, CasCysH, and Cas10-like, respectively) (Pinilla-Redondo et al., 2019; Makarova et al., 2020) (**Figure 7-1**). The minimal size, mobile nature, and eclectic collection of genes within type IV operons, present challenges to their study. However, the same reasons that make type IV systems difficult to study make it likely that type IV systems will reveal novel and impactful insights into prokaryotic biology.

This dissertation presents some of the first structural and biochemical studies of type IV CRISPR-Cas systems. The studies described here have revealed the structures of a type IV-B Csf RNP complex and a type IV Cas6 endoribonuclease, uncovered the defense function of the type IV-A CRISPR-Cas system, and elucidated many mechanisms surrounding type IV crRNA processing and type IV-A Csf RNP complex binding to nucleic acid substrates. This research has answered some of the fundamental questions surrounding type IV systems,

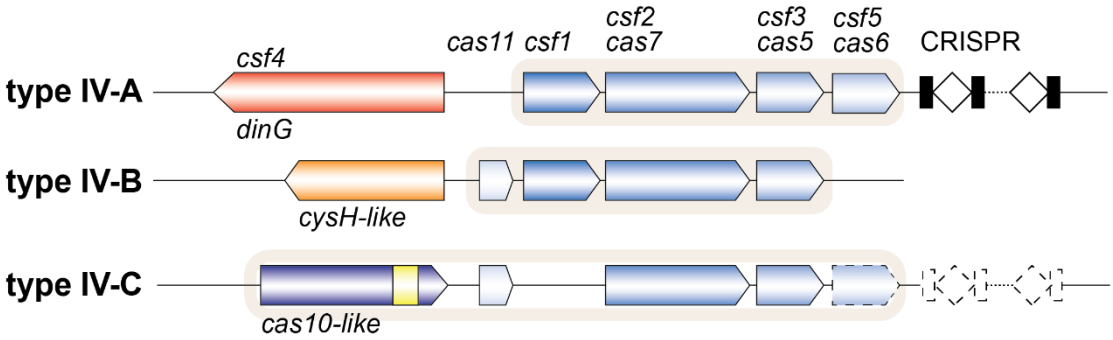

Figure 7-1. Operons of the type IV subtypes. Dashed lines indicate a component that is only sometimes present. The yellow box represents an HD nuclease domain. *dinG*, *cysH-like*, and *cas10-like* are the respective subtype-specific genes.

presented new avenues of study, and offered questions that need answered to deepen our understanding of type IV CRISPR-Cas systems and their relevance in prokaryotic biology.

This chapter will highlight the importance of each study reported in previous chapters of this dissertation and suggest future studies to further elucidate the structures, functions, and mechanisms of type IV CRISPR-Cas systems. As each of the type IV subtypes encode unique genes and have unique hypothesized functions, I will discuss each subtype in turn.

ADVANCES IN TYPE IV-A SYSTEM STRUCTURE AND FUNCTION

The Type IV Cas6/Csf5 Enzymes Process crRNAs

Within class 1 CRISPR-Cas systems, Cas6 endoribonucleases are primarily responsible for processing crRNAs (CRISPR-derived RNAs) which are bound by the Cas RNP (ribonucleoprotein) complex and used to identify complementary, and thus invasive, nucleic acids (Brouns et al., 2008; Carte et al., 2008; C. Hale et al., 2008). The Cas6 endoribonuclease encoded by the type IV-A system of *Mahella australiensis* was biochemically characterized (Taylor et al., 2019) (**Chapter 3**) (**Figure 7-2**). MaCas6 was determined to have endonuclease activity specific to the repeat sequence of the pre-crRNA transcribed from a *M. australiensis* type IV associated CRISPR locus. The crystal structure of MaCas6 was determined and used to identify the active site residues (His44 and Tyr31) required for its endoribonuclease activity. MaCas6 mutants harboring alanine mutations at either of the two active site residues had nearly abolished cleavage activity. Interestingly, these biochemical Cas6 characteristics are commonly seen in type I

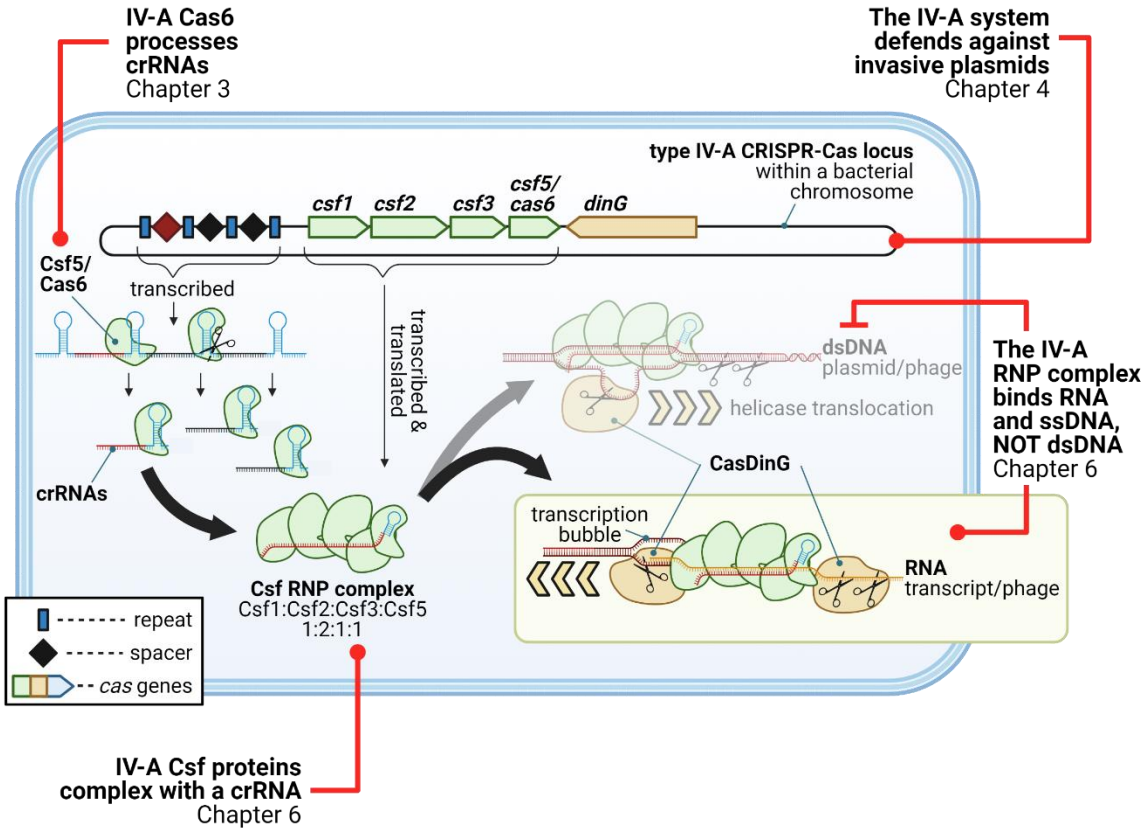

Figure 7-2. Model of type IV-A CRISPR-Cas defense. The type IV Csf5/Cas6 endoribonuclease processes the CRISPR transcript into crRNAs which are bound by the type IV-A Csf proteins to form a Csf RNP complex. The Csf RNP complex binds RNA complementary to the crRNA and may recruit DinG to degrade nearby nucleic acids. Critical advances are indicated with their respective chapter from this dissertation.

associated Cas6 enzymes (Gesner et al., 2011; Przybilski et al., 2011; Sashital et al., 2011). This led to the hypothesis that ancestral type IV systems may have acquired a type I Cas6, leading to the formation of the type IV-A operon (Taylor et al., 2019). This is further supported by the fact that MaCas6 is most structurally similar to an orphan type I Cas6 enzyme.

However, some type IV-A systems encode a Cas6-like enzyme, Csf5, that is different from the type IV Cas6 in several aspects (Makarova et al., 2020). Csf5 has been studied in the context of the type IV-A system from *Aromatoleum aromaticum* and the type IV-A system from *Pseudomonas aeruginosa* (Özcan et al., 2018; Crowley et al., 2019). Importantly, although the *P. aeruginosa* Csf RNP complex has been shown to bind crRNAs (**Chapter 6**), crRNA processing has not been directly attributed to Csf5. The PaCsf5 and AaCsf5 amino acid sequences, especially at the active site residues, align well, indicating that PaCsf5 likely has a similar activity (Taylor et al., 2021) (**Chapter 5**). Although AaCsf5 and MaCas6 have similar structural arrangements, the AaCsf5 active site is composed of three arginine residues, not a histidine and tyrosine (Özcan et al., 2018; Taylor et al., 2019, 2021).

Type IV systems encode three different types of Cas6 enzymes: Cas6, Csf5, and Cas6e (Makarova et al., 2020; Taylor et al., 2021). Cas6e is the Cas6 enzyme associated with type I-E CRISPR-Cas systems, though the type IV Cas6e amino acid sequence is distinct from the Cas6e amino acid sequence (Taylor et al., 2021). Interestingly, although the type IV Cas6e has a histidine/tyrosine active site, the sequence is different enough from type IV Cas6 enzymes (like MaCas6) that the active sites do not match in a sequence alignment (Taylor et al., 2021). The occurrence of three different Cas6 enzymes

in type IV CRISPR-Cas systems, leads to the hypothesis that type IV systems may have acquired Cas6 enzymes through three separate events. Supporting this hypothesis is the fact that type IV-B systems sometimes encode Cas6 proteins (Pinilla-Redondo et al., 2019; Makarova et al., 2020). Type IV-B systems encode the three Csf core proteins (Csf1, Csf2, Csf3), but no CRISPR. Without a CRISPR, there should not be a need to encode a Cas6 enzyme in the system. The presence of Cas6 in a type IV-B system may indicate that a common ancestor of the type IV-A and IV-B systems encoded the three Csf core proteins and acquired a Cas6 enzyme. Two lineages may have stemmed from this ancestral lineage, one becoming type IV-A systems and the other type IV-B systems. In this model, Cas6 may be a "vestigial" component of the type IV-B system.

Further work on type IV systems should continue to characterize the Cas6 enzymes from type IV-A systems, especially PaCsf5 and the uncharacterized type IV Cas6e. Work should also be done to determine the evolutionary history of type IV systems, including how, why, and when the Cas6 enzymes were acquired by type IV systems. Finally, the Cas6 proteins associated with type IV-B systems should be characterized to determine if they are functional and what their substrate is. As type IV-B systems have a hypothesized non-defense function and an unknown RNA component, IV-B associated Cas6 enzymes may have evolved to process a non-CRISPR RNA for the type IV-B Csf RNP complex to bind.

The Type IV-A Csf Proteins Form an RNP Complex and Perform crRNA Guided Nucleic Acid Binding to Induce a Downstream Defense Response

RNP effector complex formation around a crRNA is a central tenet of all CRISPR-Cas systems (Brouns et al., 2008; Jore et al., 2011; Gasiunas et al., 2012; C. R.

Hale et al., 2012; Jinek et al., 2012; Zetsche et al., 2015; Yan et al., 2018). The formation of a *Pseudomonas aeruginosa* type IV-A Csf RNP complex was confirmed by SEC, negative stain, and native mass spectrometry (**Chapter 6**) (**Figure 7-2**). This is only the second description of a type IV-A Csf RNP complex, the first coming from *Aromatoleum aromaticum* (Özcan et al., 2018). However, we are the first to determine the stoichiometry of the Csf RNP complex and to show that it is a true Cas effector complex that binds nucleic acids complementary to the crRNA. Additionally, by showing that the *P. aeruginosa* type IV-A CRISPR-Cas system can prevent the transformation of invasive plasmids (**Chapter 4**) (**Figure 7-2**), we provided the first evidence that the type IV-A CRISPR-Cas system is a defense system (Crowley et al., 2019). Interestingly both the putative helicase, CasDinG, and the ATPase active site of CasDinG are required to achieve defense. Thus, we have shown that the type IV-A CRISPR-Cas system utilizes a Csf RNP effector complex to bind target nucleic acids, leading to downstream effects involving CasDinG, which eventually provide defense against the identified nucleic acid. Unexpectedly, the type IV-A Csf RNP complex bound only single stranded oligonucleotides and did not show any preference between DNA and RNA. This could indicate that, to defend against an invasive plasmid, the IV-A Csf RNP complex binds RNA transcripts and induces downstream effects, such as plasmid cleavage or cell death, which eventually destroy the plasmid DNA.

Several questions remain to better understand the structure, function, and mechanisms of type IV-A defense. Firstly, structures of the Csf RNP complex are needed to better understand how the complex assembles, how subunits interact with each other and the crRNA, and how target nucleic acids are bound. Studies are currently underway

to obtain cryo-electron microscopy (cryo-EM) structures of the apo Csf RNP complex, the ssDNA bound Csf RNP complex, RNA bound Csf RNP complex, and dsDNA bound Csf RNP complex. These structures may also be useful in determining whether the Csf RNP complex recognizes a PAM on the target nucleic acid by examining the contacts the Csf subunits have with the target nucleic acid. Structures will also be invaluable in comparing the RNP complexes of other class 1 CRISPR-Cas systems to the type IV Csf RNP complex. Based on RNP complex stoichiometry and negative stain analysis, I hypothesize that the type IV Csf RNP complex has a shape unique for Cas RNP complexes, but has homologous subunit:subunit and subunit:crRNA interactions.

If the Csf RNP complex can recruit CasDinG, as hypothesized, then binding interactions between the Csf RNP complex and CasDinG should be explored biochemically and structurally.

Preliminary studies also describe the potential nuclease activity of CasDinG (**Chapter 6**). Studies are underway to confirm and characterize the cleavage activity of CasDinG. Determining the functional mechanisms of CasDinG and its exact role within the type IV-A CRISPR-Cas system will be critical to understanding how defense occurs after target binding by the Csf RNP complex.

Plasmids are the only experimentally determined target of type IV-A CRISPR-Cas defense systems. Bioinformatics studies have concluded that plasmids are the major target of type IV systems (Koonin & Makarova, 2019; Pinilla-Redondo et al., 2019). An important question of type IV systems is whether they can defend against bacteriophages. Some evidence suggests that phages may be targeted by type IV CRISPR-Cas systems: certain spacers within type IV associated CRISPRs match phage sequences and some

phage-encoded anti-CRISPRs proteins may block type IV systems (Shmakov et al., 2017; Crowley et al., 2019; Pinilla-Redondo et al., 2019; Yin et al., 2019; Nobrega et al., 2020). Further studies should be performed using plaque assays and other phage challenge assays with bacteria expressing the type IV-A CRISPR-Cas system to determine if the type IV-A system can defend against bacteriophage.

THE COMPLETE STRUCTURE AND FUNCTION OF TYPE IV-B SYSTEMS ARE STILL UNKNOWN

To date, the only published structural data on type IV-B systems is described by Zhou et al. (2021) (**Chapter 2**). No biochemical assays have been published on type IV-B systems. The IV-B structure is a helical ribonucleoprotein complex composed of several Csf2 backbone subunits and several Cas11 belly subunits wound around an RNA. This structure provides many important insights, including showing the similarities and differences between Csf2 and its Cas7 homologs in other class 1 CRISPR-Cas systems (Taylor et al., 2021; Zhou et al., 2021) (**Chapter 5**). The RNA component within the Csf RNP complex was sequenced and determined to be a random RNA. Interestingly, the Csf RNP complex did not bind any available I-E crRNAs, despite the fact that many CRISPR-Cas systems share and commandeer crRNAs from each other (Staals et al., 2013, 2014; Elmore et al., 2015). A major remaining question is whether the type IV-B Csf RNP complex binds RNAs non-specifically or if the preferred RNA substrate of the IV-B RNP complex was unavailable in the conditions of the study. One possible experiment that could determine the RNA component of the type IV-B system is to add an affinity tag to the *csf2* gene of an organism, such as *Mycobacterium sp.* JS623, then pull down the tagged IV-B Csf RNP complex and sequence the RNA component. If the

RNA component is indeed not a crRNA, it seems unlikely that the type IV-B systems function could be defense.

Many potential functions of the type IV-B CRISPR-Cas system are discussed in Chapter 5 (Taylor et al., 2021). However, none address how the accessory protein CasCysH might be involved in type IV-B function. Understanding the function of CasCysH will inform our ideas concerning the overall function of type IV-B systems. Thus, determining the structure, function, and mechanisms of CasCysH is critical.

Finally, a future direction is to determine a more complete cryoEM structure of the type IV-B Csf RNP complex. The available IV-B Csf RNP complex structure is missing the Csf1 and Csf3 subunits, because the electron density at the ends of the complex was not resolved enough to allow modeling of any subunits (Zhou et al., 2021). I attempted to co-express the IV-B Csf proteins with an MBP-tag on either Csf1 or Csf3 to determine a higher resolution cryoEM structure. The MBP-tag is ~40 kDa, increasing the size of the linked subunit. This could make the combined subunit and tag more "visible" in the data set. However, the bulky MBP-tag could also interfere with the formation of the RNP complex and alter its stoichiometry. I have made a promising start and future work in the Jackson lab should continue this experiment. Another potential route to obtaining a more complete cryoEM structure of the IV-B Csf RNP complex is to purify a less heterogenous sample of the complex. The Csf RNP complex we used to obtain a cryoEM structure was purified by affinity chromatography followed by size exclusion chromatography (Zhou et al., 2021). Future studies should add an additional step, such as anion exchange chromatography. The type IV-A Csf RNP complex was purified away from many contaminants using an anion exchange column. Because the

IV-A and IV-B Csf RNP complexes have some homologous subunits, I hypothesize that the IV-B Csf RNP complex would be better purified if an anion exchange column were added to the purification protocol after affinity chromatography.

STUDY OF THE NEWLY CLASSIFIED TYPE IV-C SYSTEM WILL LIKELY REVEAL NEW DEFENSE OR NON-DEFENSE FUNCTIONS

The Type IV-C subtype was very recently classified and has only been characterized using bioinformatics. Type IV-C systems encode the type IV signature gene, *csf2*, as well as a *csf3* gene and a type IV-specific *cas11* gene (Makarova et al., 2020; Pinilla-Redondo et al., 2019) (**Figure 7-1**). Strikingly, type IV-C systems do not encode a *csf1* gene. In fact, *csf1* was initially characterized as the type IV signature gene until type IV-C and other type IV variants were discovered without *csf1* (Makarova et al., 2015). In place of a *csf1* gene, Type IV-C systems encode a *cas10*-like gene. Cas10 is an enzymatic subunit of type III systems. It contains an HD nuclease domain, is a subunit of the type III RNP complex, and cleaves ssDNAs and, in some cases, produces cyclic oligoadenylates after a complementary nucleic acid is bound by the RNP complex (Jung et al., 2015; Osawa et al., 2015; Niewoehner et al., 2017; Kazlauskiene et al., 2017).

Although the IV-C Cas10-like protein is related to the type III Cas10, there are some notable differences between the two proteins. First, the IV-C Cas10-like protein does not contain an active site for cyclic oligoadenylate synthesis (Pinilla-Redondo et al., 2019). Second, the HD nuclease domain motifs of the IV-C Cas10-like protein are arranged most like the HD motifs of the type I Cas3 enzyme rather than those of the type III Cas10 enzyme (Aravind & Koonin, 1998; Makarova et al., 2020). Interestingly, I found that when IV-C Cas10-like sequences were aligned to select type III-A Cas10

sequences and type I Cas3 sequences using Clustal Omega (Madeira et al., 2019), the HD nuclease motifs did not align (**Figure 7-3**). An HHPred search on the type IV-C Cas10-like sequence from *Thermococcus onnurineus* returned a match for the type I-D Cas10d structure from *Sulfolobus islandicus* (PDBid: 6THH) (Manav et al., 2020; Zimmermann et al., 2018). An alignment of type IV-C Cas10-like sequences with type I-D Cas10d sequences shows high sequence similarity around an HD domain motif (**Figure 7-4**). Type I-D systems encode a Cas10 large subunit (Cas10d) instead of the Cas8 large subunit typically encoded by type I CRISPR-Cas systems (Makarova et al., 2015, 2020). Cas10d subunits cleave double stranded and single stranded DNA with an HD domain similar to type I Cas3 HD domains (Lin et al., 2020; Makarova et al., 2015; Manav et al., 2020). Type I-D and type IV-C systems illustrate the crossover and evolutionary relationships between CRISPR-Cas systems. Although scientists impose hierarchical classifications upon biological systems, the systems themselves are more fluid.

The genetic architecture of type IV-C systems indicates that they may have a unique function among type IV subtypes and almost certainly utilize unique mechanisms of action. I anticipate future work will reveal the structure and function of type IV-C RNP complexes.

EXPANDING STUDY OF TYPE IV SYSTEMS WILL ADVANCE UNDERSTANDING OF TYPE IV FUNCTIONS AND MECHANISMS

Type IV systems have only been studied in a few organisms. To fully understand the functions, structures, and mechanisms of type IV systems, we must study type IV systems in many different organisms. A broad expansion of studied type IV systems will confirm generalizable aspects of type IV biology and uncover exceptions and their

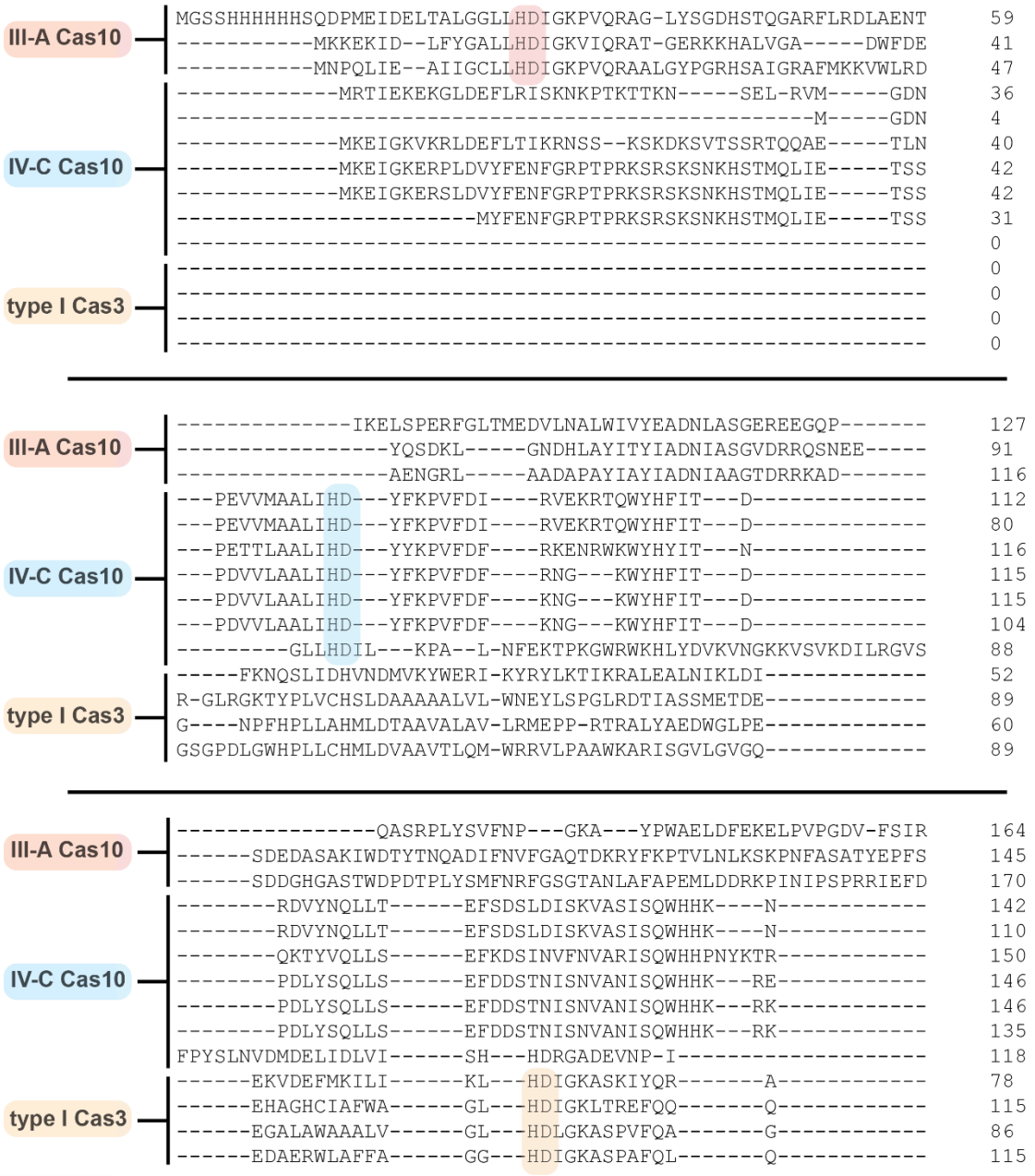

Figure 7-3. Sequence alignment of IV-C Cas10 with III-A Cas10 and type I Cas3. The HD nuclease motif is not shared between these three groups of proteins.

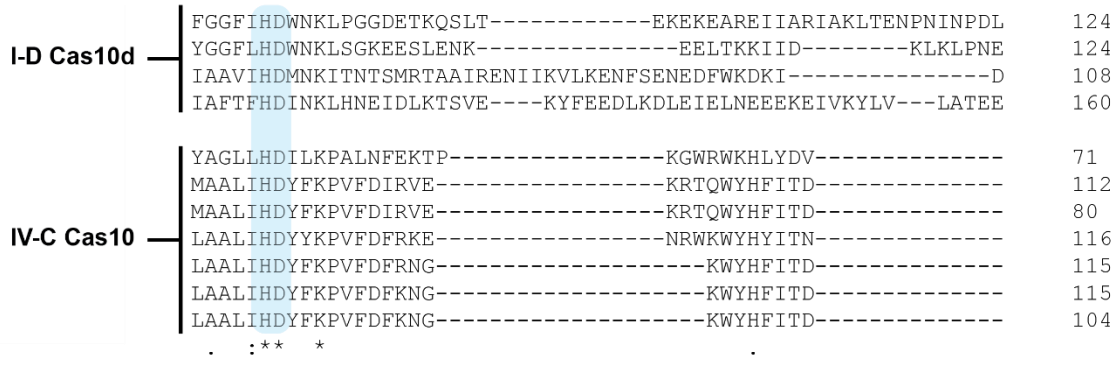

Figure 7-4. Sequence alignment of IV-C Cas10 with I-D Cas10d. The HD nuclease motifs of the IV-C Cas10 and I-D Cas10d proteins do align.

importance. I have cloned several IV-A and IV-B *csf* genes from six organisms into individual and co-expression vectors. I have also performed expression tests and determined purification protocols for many IV-A and IV-B Csf proteins and RNP complexes (**Table 7-1**). Future work in the Jackson lab should utilize the expression vectors and preliminary expression and purification data I have generated to continue study of type IV systems. To better understand the structure and function of type IV RNP complexex, we should continue to elucidate the structures and functions of the type IV-B system from *M. sp.* JS623 and the type IV-A system from *P. aeruginosa* at a deeper level and focus on determining new structures and functions of Csf proteins and RNP complexes from organisms that have not yet been studied, including *Acidithiobacillus ferrooxidans* and *Mycobacterium sp. MCS*. I am especially interested in determining the structure and function of the RNP complex from *M. australiensis*. We characterized the pre-crRNA processing activity of Cas6 from *M. australiensis* but have not determined the structure and function of the other Csf proteins nor the RNP complex. Interestingly, the *M. australiensis* type IV-A system does not contain CasDinG and instead encodes a putative helicase/ATPase and a DUF4359 domain containing protein (Makarova et al., 2020) (**Figure 7-5**). This exchange of CasDinG for a different helicase could change the function of the system, or simply allow the system to provide defense through an alternative mechanism. This ancillary protein exchange is commonly seen in prokaryotic defense systems (Gao et al., 2020; Lowey et al., 2020; Millman et al., 2020).

Future work in the Jackson lab and others in the CRISPR field should also seek new organisms with type IV systems to study. In particular, determining the structure and function of the archaeal type IV-C system from *T. onnurineus* would dramatically

Table 7-1. *cas* genes I cloned, expressed, and purified in the Jackson lab.
Expression/purification trials are recorded in my lab notebooks (#1-#4) starting at the indicated page number.

Cas Type	Organism	Cas gene	Expression vector archive number	Expression Trials	Purification trials
IV-A	***Acidithiobacillus ferrooxidans***	*csf1*	203, 204, 237, 241, 242, 348, 438, 439	#1-7, #1-45, #1-83, #2-55	#1-57
		csf2	202, 239, 349, 501-506	#1-7	
		csf3	238, 240, 350, 498-500	#1-7, #1-67	
		cas6	201, 347, 443-445, 536, 879	#1-7	*
		dinG	231, 351, 352, 440-442, 497, 528		
		RNP complex	223+239+680	#2-44	
		RNP complex + dinG	243+239+680+351	#2-49	
IV-B	***Mycobacterium sp. MCS***	*csf1*	458-463, 670-675	#2-55, #2-75	#2-67
		csf2	464-466, 521-525, 676-677	#2-57, #2-75	
		csf3	467-471	#2-57, #2-75	
		csf4/cas11	472-476	#2-57, #2-75	#2-61
		RNP complex	471+476+522+463	#2-72	#2-72
IV-A	***Rhodoferax ferrireducens***	*csf1*	507-512, 981, 1041	#2-55	
		csf2	513-515, 517		
		csf3	448		
		cas6	452-454, 519, 520	#3-40	
		csf5	449-451	#3-40	
		dinG			
		cas1	455-457		
		cas3	446, 447		
IV-B	***Mycobacterium sp. JS623***	*csf1*	477-479, 735-740	#2-55, #3-134	#2-88, #3-27, #3-155
		csf2	480, 527, 741-746		
		csf3	481-483, 489-491	#3-134	#4-27
		csf4/cas11	484-488, 492-496		
		RNP complex	1211-1214, 1242, 1286, 1288	#3-33, #3-51	#3-36, #3-63, #3-102
IV-A	***Mahella australiensis***	*csf1*	994-996		
		csf2	997, 998		
		csf3			
		cas6	876-878, 880-882, 1294	**	**
		RNP complex	1062, 1285	#3-43	
IV-A	***Pseudomonas aeruginosa PA83***	*csf1*	1340-1346	#3-131	#3-155, #4-10
		csf2	1347-1350, 1358	#3-131	#4-23
		csf3	1215-1221	#3-131	#4-23
		csf5 (cas6)	1351-1356, 1359	#3-131	#4-2, #4-10
		dinG	1222-1227, 1313		
		RNP complex	1284, 1287, 1293, 1306, 1314-1315, 1496, 1651-1652	#3-122, #3-136, #4-57	#4-23, #4-122, #4-133
		RNP complex + dinG	1290-1292, 1316, 1357	#3-100, #3-112	#4-83

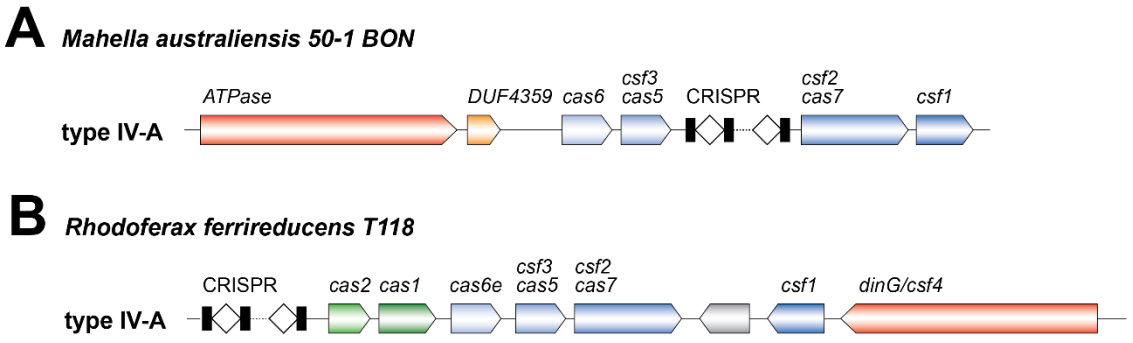

Figure 7-5. Operons of interesting type IV systems. (**A**) The *M. australiensis* type IV-A operon encodes an ATPase distinct from CasDinG. (**B**) The *R. ferrireducens* type IV-A operon encodes the adaptation proteins Cas1 and Cas2.

advance our understanding of type IV CRISPR-Cas systems, as little information is known about type IV-C systems.

One major open question for type IV systems is how the CRISPR is maintained and whether new spacers can be acquired. Bioinformatics studies suggest that some type IV systems can engage in crosstalk with type I CRISPR-Cas systems, potentially co-opting their adaptation proteins to introduce new spacers into the type IV CRISPR (Pinilla-Redondo et al., 2019; Bernheim et al., 2020). A few type IV systems encode Cas1, Cas2, and/or Cas4 proteins (Makarova et al., 2015, 2020; Pinilla-Redondo et al., 2019; Crowley et al., 2019). For example, a type IV-A system in *Rhodoferax ferrireducens* encodes the *cas1* and *cas2* genes (**Figure 7-5**). I have cloned most of the *R. ferrireducens* type IV genes into expression vectors, including *cas1* (**Table 7-1**). Future work in the Jackson lab should attempt to determine the structure and biochemistry of the type IV Cas1 and determine whether it has a specialized function in type IV systems.

CONCLUSION

As we better understand the structures, functions, and mechanisms of the type IV CRISPR-Cas systems, we will gain a more complete understanding of prokaryotic biology and potentially uncover novel activities that may lead to biotechnological tools that can advance science and medicine.

REFERENCES

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. (2016). C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science*, *353*(6299). https://doi.org/10.1126/science.aaf5573

Afonine, P. V., Poon, B. K., Read, R. J., Sobolev, O. V., Terwilliger, T. C., Urzhumtsev, A., & Adams, P. D. (2018). Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallographica Section D: Structural Biology*, *74*(6), 531–544. https://doi.org/10.1107/S2059798318006551

Almendros, C., Guzmán, N. M., Díez-Villaseñor, C., García-Martínez, J., & Mojica, F. J. M. (2012). Target Motifs Affecting Natural Immunity by a Constitutive CRISPR-Cas System in Escherichia coli. *PLoS ONE*, *7*(11), e50797. https://doi.org/10.1371/journal.pone.0050797

Al-Shayeb, B., Sachdeva, R., Chen, L.-X., Ward, F., Munk, P., Devoto, A., Castelle, C. J., Olm, M. R., Bouma-Gregson, K., Amano, Y., He, C., Méheust, R., Brooks, B., Thomas, A., Lavy, A., Matheus-Carnevali, P., Sun, C., Goltsman, D. S. A., Borton, M. A., … Banfield, J. F. (2020). Clades of huge phages from across Earth's ecosystems. *Nature*, *578*(7795), 425–431. https://doi.org/10.1038/s41586-020-2007-4

Altschul, S. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research*, *25*(17), 3389–3402. https://doi.org/10.1093/nar/25.17.3389

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Aravind, L., & Koonin, E. V. (1998). The HD domain defines a new superfamily of metal-dependent phosphohydrolases. *Trends in Biochemical Sciences*, *23*(12), 469–472. https://doi.org/10.1016/S0968-0004(98)01293-6

Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y., & Wishart, D. S. (2016). PHASTER: A better, faster version of the PHAST phage search tool. *Nucleic Acids Research*, *44*(W1), W16–W21. https://doi.org/10.1093/nar/gkw387

Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., & Ben-Tal, N. (2016). ConSurf 2016: An improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Research*, *44*(W1), W344–W350. https://doi.org/10.1093/nar/gkw408

Baker, N. A., Sept, D., Joseph, S., Holst, M. J., & McCammon, J. A. (2001). Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(18), 10037–10041. https://doi.org/10.1073/pnas.181342398

Beloglazova, N., Petit, P., Flick, R., Brown, G., Savchenko, A., & Yakunin, A. F. (2011). Structure and activity of the Cas3 HD nuclease MJ0384, an effector enzyme of the CRISPR interference. *The EMBO Journal*, *30*(22), 4616–4627. https://doi.org/10.1038/emboj.2011.377

Bernheim, A., Bikard, D., Touchon, M., & Rocha, E. P. C. (2020). Atypical organizations and epistatic interactions of CRISPRs and cas clusters in genomes and their

mobile genetic elements. *Nucleic Acids Research*, *48*(2), 748–760. https://doi.org/10.1093/nar/gkz1091

Blum, M., Chang, H.-Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., Nuka, G., Paysan-Lafosse, T., Qureshi, M., Raj, S., Richardson, L., Salazar, G. A., Williams, L., Bork, P., Bridge, A., Gough, J., Haft, D. H., Letunic, I., Marchler-Bauer, A., … Finn, R. D. (2020). The InterPro protein families and domains database: 20 years on. *Nucleic Acids Research*, *49*(D1), D344–D354. https://doi.org/10.1093/nar/gkaa977

Borges, A. L., Davidson, A. R., & Bondy-Denomy, J. (2017). The Discovery, Mechanisms, and Evolutionary Impact of Anti-CRISPRs. *Annual Review of Virology*, *4*(1), 37–59. https://doi.org/10.1146/annurev-virology-101416-041616

Bork, P., & Koonin, E. V. (1994). A P-loop-like motif in a widespread ATP pyrophosphatase domain: Implications for the evolution of sequence motifs and enzyme activity. *Proteins*, *20*(4), 347–355. https://doi.org/10.1002/prot.340200407

Bravo, J. P. K., Bartnik, K., Venditti, L., Gail, E. H., Davidovich, C., Lamb, D. C., Tuma, R., Calabrese, A. N., & Borodavka, A. (2020). Structural basis of rotavirus RNA chaperone displacement and RNA annealing. *BioRxiv*, 2020.10.26.354233.

Bravo, J. P. K., Borodavka, A., Barth, A., Calabrese, A. N., Mojzes, P., Cockburn, J. J. B., Lamb, D. C., & Tuma, R. (2018). Stability of local secondary structure determines selectivity of viral RNA chaperones. *Nucleic Acids Research*, *May*, 293191. https://doi.org/10.1101/293191

Bravo, J. P. K., Dangerfield, T. L., Taylor, D. W., & Johnson, K. A. (2021). Remdesivir is a delayed translocation inhibitor of SARS CoV-2 replication. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2021.01.035

Brouns, S. J. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J. H., Snijders, A. P. L., Dickman, M. J., Makarova, K. S., Koonin, E. V., & van der Oost, J. (2008a). Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. *Science*, *321*(5891), 960–964. https://doi.org/10.1126/science.1159689

Brouns, S. J. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J. H., Snijders, A. P. L., Dickman, M. J., Makarova, K. S., Koonin, E. V., & van der Oost, J. (2008b). Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. *Science*, *321*(5891), 960–964. https://doi.org/10.1126/science.1159689

Byrd, A. K., & Raney, K. D. (2012). Superfamily 2 helicases. *Frontiers in Bioscience (Landmark Edition)*, *17*, 2070–2088.

Carroll, K. S., Gao, H., Chen, H., Stout, C. D., Leary, J. A., & Bertozzi, C. R. (2005). A Conserved Mechanism for Sulfonucleotide Reduction. *PLoS Biology*, *3*(8). https://doi.org/10.1371/journal.pbio.0030250

Carte, J., Wang, R., Li, H., Terns, R. M., & Terns, M. P. (2008). Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes & Development*, *22*(24), 3489–3496. https://doi.org/10.1101/gad.1742908

Chartron, J., Shiau, C., Stout, C. D., & Carroll, K. S. (2007). 3'-Phosphoadenosine-5'-phosphosulfate reductase in complex with thioredoxin: A structural snapshot in the catalytic cycle. *Biochemistry*, *46*(13), 3942–3951. https://doi.org/10.1021/bi700130e

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., & Richardson, D. C. (2010). MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography*, *66*(1), 12–21. https://doi.org/10.1107/S0907444909042073

Cheng, K., & Wigley, D. B. (2018). DNA translocation mechanism of an XPD family helicase. *ELife*. https://doi.org/10.7554/elife.42400.001

Chowdhury, S., Carter, J., Rollins, M. C. F., Golden, S. M., Jackson, R. N., Hoffmann, C., Nosaka, L., Bondy-Denomy, J., Maxwell, K. L., Davidson, A. R., Fischer, E. R., Lander, G. C., & Wiedenheft, B. (2017). Structure Reveals Mechanisms of Viral Suppressors that Intercept a CRISPR RNA-Guided Surveillance Complex. *Cell*, *169*(1), 47-57.e11. https://doi.org/10.1016/j.cell.2017.03.012

Crooks, G. E. (2004). WebLogo: A Sequence Logo Generator. *Genome Research*, *14*(6), 1188–1190. https://doi.org/10.1101/gr.849004

Crooks, G., Hon, G., Chandonia, J., & Brenner, S. (2004). WebLogo: A sequence logo generator. *Genome Res*, *14*, 1188–1190. https://doi.org/10.1101/gr.849004.1

Crowley, V. M., Catching, A., Taylor, H. N., Borges, A. L., Metcalf, J., Bondy-Denomy, J., & Jackson, R. N. (2019). A Type IV-A CRISPR-Cas System in *Pseudomonas aeruginosa* Mediates RNA-Guided Plasmid Interference *In Vivo*. *The CRISPR Journal*, *2*(6), 434–440. https://doi.org/10.1089/crispr.2019.0048

Datsenko, K. A., Pougach, K., Tikhonov, A., Wanner, B. L., Severinov, K., & Semenova, E. (2012). Molecular memory of prior infections activates the CRISPR/Cas

adaptive bacterial immunity system. *Nature Communications*, *3*(1), 945. https://doi.org/10.1038/ncomms1937

Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A., Eckert, M. R., Vogel, J., & Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, *471*(7340), 602–607. https://doi.org/10.1038/nature09886

Dwarakanath, S., Brenzinger, S., Gleditzsch, D., Plagens, A., Klingl, A., Thormann, K., & Randau, L. (2015). Interference activity of a minimal Type I CRISPR–Cas system from Shewanella putrefaciens. *Nucleic Acids Research*, *43*(18), 8913–8923. https://doi.org/10.1093/nar/gkv882

Elmore, J., Deighan, T., Westpheling, J., Terns, R. M., & Terns, M. P. (2015). DNA targeting by the type I-G and type I-A CRISPR-Cas systems of Pyrococcus furiosus. *Nucleic Acids Research*, *43*(21), 10353–10363. https://doi.org/10.1093/nar/gkv1140

Elmore, J. R., Sheppard, N. F., Ramia, N., Deighan, T., Li, H., Terns, R. M., & Terns, M. P. (2016). Bipartite recognition of target RNAs activates DNA cleavage by the Type III-B CRISPR–Cas system. *Genes & Development*, *30*(4), 447–459. https://doi.org/10.1101/gad.272153.115

Estrella, M. A., Kuo, F.-T., & Bailey, S. (2016). RNA-activated DNA cleavage by the Type III-B CRISPR-Cas effector complex. *Genes & Development*, *30*(4), 460–470. https://doi.org/10.1101/gad.273722.115

Fairman-Williams, M. E., Guenther, U.-P., & Jankowsky, E. (2010). SF1 and SF2 helicases: Family matters. *Current Opinion in Structural Biology*, *20*(3), 313–324. https://doi.org/10.1016/j.sbi.2010.03.011

Faure, G., Shmakov, S. A., Yan, W. X., Cheng, D. R., Scott, D. A., Peters, J. E., Makarova, K. S., & Koonin, E. V. (2019). CRISPR–Cas in mobile genetic elements: Counter-defence and beyond. *Nature Reviews Microbiology*. https://doi.org/10.1038/s41579-019-0204-7

Garneau, J. E., Dupuis, M.-È., Villion, M., Romero, D. A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadán, A. H., & Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*, *468*(7320), 67–71. https://doi.org/10.1038/nature09523

Gasiunas, G., Sinkunas, T., & Siksnys, V. (2014). Molecular mechanisms of CRISPR-mediated microbial immunity. *Cellular and Molecular Life Sciences*, *71*(3), 449–465. https://doi.org/10.1007/s00018-013-1438-6

Goddard, T. D., Huang, C. C., Meng, E. C., Pettersen, E. F., Couch, G. S., Morris, J. H., & Ferrin, T. E. (2018). UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Science*, *27*(1), 14–25. https://doi.org/10.1002/pro.3235

Grodick, M. A., Segal, H. M., Zwang, T. J., & Barton, J. K. (2014a). DNA-Mediated Signaling by Proteins with 4Fe–4S Clusters Is Necessary for Genomic Integrity. *Journal of the American Chemical Society*, *136*(17), 6470–6478. https://doi.org/10.1021/ja501973c

Grodick, M. A., Segal, H. M., Zwang, T. J., & Barton, J. K. (2014b). DNA-Mediated Signaling by Proteins with 4Fe–4S Clusters Is Necessary for Genomic Integrity. *Journal of the American Chemical Society*, *136*(17), 6470–6478. https://doi.org/10.1021/ja501973c

Guo, T. W., Bartesaghi, A., Yang, H., Falconieri, V., Rao, P., Merk, A., Eng, E. T., Raczkowski, A. M., Fox, T., Earl, L. A., Patel, D., & Subramaniam, S. (2017). Cryo-EM Structures Reveal Mechanism and Inhibition of DNA Targeting by a CRISPR-Cas Surveillance Complex. *Cell*, *171*(2), 414-426.e12. https://doi.org/10.1016/j.cell.2017.09.006

Hale, C., Kleppe, K., Terns, R. M., & Terns, M. P. (2008). Prokaryotic silencing (psi)RNAs in Pyrococcus furiosus. *RNA*, *14*(12), 2572–2579. https://doi.org/10.1261/rna.1246808

Hale, C. R., Zhao, P., Olson, S., Duff, M. O., Graveley, B. R., Wells, L., Terns, R. M., & Terns, M. P. (2009). RNA-Guided RNA Cleavage by a CRISPR RNA-Cas Protein Complex. *Cell*, *139*(5), 945–956. https://doi.org/10.1016/j.cell.2009.07.040

Han, W., Li, Y., Deng, L., Feng, M., Peng, W., Hallstrøm, S., Zhang, J., Peng, N., Liang, Y. X., White, M. F., & She, Q. (2017). A type III-B CRISPR-Cas effector complex mediating massive target DNA destruction. *Nucleic Acids Research*, *45*(4), 1983–1993. https://doi.org/10.1093/nar/gkw1274

Haurwitz, R. E., Jinek, M., Wiedenheft, B., Zhou, K., & Doudna, J. A. (2010). Sequence- and Structure-Specific RNA Processing by a CRISPR Endonuclease. *Science*, *329*(5997), 1355–1358. https://doi.org/10.1126/science.1192272

Hayes, R. P., Xiao, Y., Ding, F., van Erp, P. B. G., Rajashankar, K., Bailey, S., Wiedenheft, B., & Ke, A. (2016). Structural basis for promiscuous PAM recognition in Type I-E Cascade from E. coli. *Nature*, *530*(7591), 499–503. https://doi.org/10.1038/nature16995

Heler, R., Samai, P., Modell, J. W., Weiner, C., Goldberg, G. W., Bikard, D., & Marraffini, L. A. (2015). Cas9 specifies functional viral targets during CRISPR–Cas adaptation. *Nature*, *519*(7542), 199–202. https://doi.org/10.1038/nature14245

Hille, F., Richter, H., Wong, S. P., Bratovič, M., Ressel, S., & Charpentier, E. (2018). The Biology of CRISPR-Cas: Backward and Forward. *Cell*, *172*(6), 1239–1259. https://doi.org/10.1016/j.cell.2017.11.032

Jackson, R. N., Golden, S. M., Erp, P. B. G. van, Carter, J., Westra, E. R., Brouns, S. J. J., Oost, J. van der, Terwilliger, T. C., Read, R. J., & Wiedenheft, B. (2014). Crystal structure of the CRISPR RNA–guided surveillance complex from Escherichia coli. *Science*, *345*(6203), 1473–1479. https://doi.org/10.1126/science.1256328

Jackson, R. N., Golden, S. M., van Erp, P. B. G., Carter, J., Westra, E. R., Brouns, S. J. J., van der Oost, J., Terwilliger, T. C., Read, R. J., & Wiedenheft, B. (2014a). Structural biology. Crystal structure of the CRISPR RNA-guided surveillance complex from Escherichia coli. *Science (New York, N.Y.)*, *345*(6203), 1473–1479. https://doi.org/10.1126/science.1256328

Jackson, R. N., Golden, S. M., van Erp, P. B. G., Carter, J., Westra, E. R., Brouns, S. J. J., van der Oost, J., Terwilliger, T. C., Read, R. J., & Wiedenheft, B. (2014b). Crystal structure of the CRISPR RNA-guided surveillance complex from

Escherichia coli. *Science*, *345*(6203), 1473–1479. https://doi.org/10.1126/science.1256328

Jackson, R. N., van Erp, P. B., Sternberg, S. H., & Wiedenheft, B. (2017). Conformational regulation of CRISPR-associated nucleases. *Current Opinion in Microbiology*, *37*, 110–119. https://doi.org/10.1016/j.mib.2017.05.010

Jackson, S. A., McKenzie, R. E., Fagerlund, R. D., Kieper, S. N., Fineran, P. C., & Brouns, S. J. J. (2017). CRISPR-Cas: Adapting to change. *Science*, *356*(6333). https://doi.org/10.1126/science.aal5056

Jia, N., Jones, R., Sukenick, G., & Patel, D. J. (2019). Second Messenger cA4 Formation within the Composite Csm1 Palm Pocket of Type III-A CRISPR-Cas Csm Complex and Its Release Path. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2019.06.013

Jia, N., Mo, C. Y., Wang, C., Eng, E. T., Marraffini, L. A., & Patel, D. J. (2019). Type III-A CRISPR-Cas Csm Complexes: Assembly, Periodic RNA Cleavage, DNase Activity Regulation, and Autoimmunity. *Molecular Cell*, *73*(2), 264-277.e5. https://doi.org/10.1016/j.molcel.2018.11.007

Jore, M. M., Lundgren, M., van Duijn, E., Bultema, J. B., Westra, E. R., Waghmare, S. P., Wiedenheft, B., Pul, Ü., Wurm, R., Wagner, R., Beijer, M. R., Barendregt, A., Zhou, K., Snijders, A. P. L., Dickman, M. J., Doudna, J. A., Boekema, E. J., Heck, A. J. R., van der Oost, J., & Brouns, S. J. J. (2011). Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nature Structural & Molecular Biology*, *18*(5), 529–536. https://doi.org/10.1038/nsmb.2019

Jung, T.-Y., An, Y., Park, K.-H., Lee, M.-H., Oh, B.-H., & Woo, E. (2015). Crystal Structure of the Csm1 Subunit of the Csm Complex and Its Single-Stranded DNA-Specific Nuclease Activity. *Structure*, *23*(4), 782–790. https://doi.org/10.1016/j.str.2015.01.021

Kamruzzaman, M., & Iredell, J. R. (2020). CRISPR-Cas System in Antibiotic Resistance Plasmids in Klebsiella pneumoniae. *Frontiers in Microbiology*, *10*. https://doi.org/10.3389/fmicb.2019.02934

Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Kazlauskiene, M., Kostiuk, G., Venclovas, Č., Tamulaitis, G., & Siksnys, V. (2017). A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems. *Science*, *357*(6351), 605–609. https://doi.org/10.1126/science.aao0100

Kazlauskiene, M., Tamulaitis, G., Kostiuk, G., Venclovas, Č., & Siksnys, V. (2016). Spatiotemporal Control of Type III-A CRISPR-Cas Immunity: Coupling DNA Degradation with the Target RNA Recognition. *Molecular Cell*, *62*(2), 295–306. https://doi.org/10.1016/j.molcel.2016.03.024

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. (2018). Cas4 Facilitates PAM-Compatible Spacer Selection during CRISPR Adaptation. *Cell Reports*, *22*(13), 3377–3384. https://doi.org/10.1016/j.celrep.2018.02.103

Klompe, S. E., Vo, P. L. H., Halpin-Healy, T. S., & Sternberg, S. H. (2019). Transposon-encoded CRISPR–Cas systems direct RNA-guided DNA integration. *Nature*. https://doi.org/10.1038/s41586-019-1323-z

Koonin, E. V., & Krupovic, M. (2015). Evolution of adaptive immunity from transposable elements combined with innate immune systems. *Nature Reviews Genetics*, *16*(3), 184–192. https://doi.org/10.1038/nrg3859

Koonin, E. V., & Makarova, K. S. (2017). Mobile Genetic Elements and Evolution of CRISPR-Cas Systems: All the Way There and Back. *Genome Biology and Evolution*, *9*(10), 2812–2825. https://doi.org/10.1093/gbe/evx192

Koonin, E. V., & Makarova, K. S. (2019). Origins and evolution of CRISPR-Cas systems. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*(1772). https://doi.org/10.1098/rstb.2018.0087

Koonin, E. V., Makarova, K. S., & Zhang, F. (2017). Diversity, classification and evolution of CRISPR-Cas systems. *Current Opinion in Microbiology*, *37*, 67–78. https://doi.org/10.1016/j.mib.2017.05.008

Lander, G. C., Stagg, S. M., Voss, N. R., Cheng, A., Fellmann, D., Pulokas, J., Yoshioka, C., Irving, C., Mulder, A., Lau, P. W., Lyumkis, D., Potter, C. S., & Carragher, B. (2009). Appion: An integrated, database-driven pipeline to facilitate EM image processing. *Journal of Structural Biology*, *166*(1), 95–102. https://doi.org/10.1016/j.jsb.2009.01.002

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, *10*(3). https://doi.org/10.1186/gb-2009-10-3-r25

Lee, H., Zhou, Y., Taylor, D. W., & Sashital, D. G. (2018). Cas4-Dependent Prespacer Processing Ensures High-Fidelity Programming of CRISPR Arrays. *Molecular Cell*, *70*(1), 48-59.e5. https://doi.org/10.1016/j.molcel.2018.03.003

Lewis, L. K., Jenkins, M. E., & Mount, D. W. (1992). Isolation of DNA damage-inducible promoters in Escherichia coli: Regulation of polB (dinA), dinG, and dinH by LexA repressor. *Journal of Bacteriology*, *174*(10), 3377–3385.

Lintner, N. G., Kerou, M., Brumfield, S. K., Graham, S., Liu, H., Naismith, J. H., Sdano, M., Peng, N., She, Q., Copié, V., Young, M. J., White, M. F., & Lawrence, C. M. (2011). Structural and Functional Characterization of an Archaeal Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-associated Complex for Antiviral Defense (CASCADE). *Journal of Biological Chemistry*, *286*(24), 21643–21656. https://doi.org/10.1074/jbc.M111.238485

Liu, T. Y., Iavarone, A. T., & Doudna, J. A. (2017). RNA and DNA Targeting by a Reconstituted Thermus thermophilus Type III-A CRISPR-Cas System. *PLOS ONE*, *12*(1), e0170552. https://doi.org/10.1371/journal.pone.0170552

Madeira, F., Park, Y. mi, Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A. R. N., Potter, S. C., Finn, R. D., & Lopez, R. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research*, *47*(W1), W636–W641. https://doi.org/10.1093/nar/gkz268

Makarova, K. S., Aravind, L., Wolf, Y. I., & Koonin, E. V. (2011). Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. *Biology Direct*, *6*(1), 38. https://doi.org/10.1186/1745-6150-6-38

Makarova, K. S., & Koonin, E. V. (2015). Annotation and Classification of CRISPR-Cas Systems. In M. Lundgren, E. Charpentier, & P. C. Fineran (Eds.), *CRISPR* (Vol. 1311, pp. 47–75). Springer New York. https://doi.org/10.1007/978-1-4939-2687-9_4

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., … Koonin, E. V. (2015). An updated evolutionary classification of CRISPR–Cas systems. *Nature Reviews Microbiology*, *13*(11), 722–736. https://doi.org/10.1038/nrmicro3569

Makarova, K. S., Wolf, Y. I., Iranzo, J., Shmakov, S. A., Alkhnbashi, O. S., Brouns, S. J. J., Charpentier, E., Cheng, D., Haft, D. H., Horvath, P., Moineau, S., Mojica, F. J. M., Scott, D., Shah, S. A., Siksnys, V., Terns, M. P., Venclovas, Č., White, M. F., Yakunin, A. F., … Koonin, E. V. (2020). Evolutionary classification of CRISPR–Cas systems: A burst of class 2 and derived variants. *Nature Reviews Microbiology*. https://doi.org/10.1038/s41579-019-0299-x

Makarova, K. S., Wolf, Y. I., Iranzo, J., Shmakov, S. A., Alkhnbashi, O. S., Brouns, S. J. J., Charpentier, E., Cheng, D., Haft, D. H., Horvath, P., Moineau, S., Mojica, F. J. M., Scott, D., Shah, S. A., Siksnys, V., Terns, M. P., Venclovas, Č., White, M. F., Yakunin, A. F., … Koonin, E. V. (2020). Evolutionary classification of CRISPR–Cas systems: A burst of class 2 and derived variants. *Nature Reviews Microbiology*, *18*(2), 67–83. https://doi.org/10.1038/s41579-019-0299-x

Marraffini, L. A., & Sontheimer, E. J. (2008). CRISPR Interference Limits Horizontal Gene Transfer in Staphylococci by Targeting DNA. *Science*, *322*(5909), 1843–1845. https://doi.org/10.1126/science.1165771

Marraffini, L. A., & Sontheimer, E. J. (2010). Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature*, *463*(7280), 568–571. https://doi.org/10.1038/nature08703

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal [Online]*, *17*(1), 10–12.

McKitterick, A. C., LeGault, K. N., Angermeyer, A., Alam, M., & Seed, K. D. (2019). Competition between mobile genetic elements drives optimization of a phage-encoded CRISPR-Cas system: Insights from a natural arms race. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*(1772), 20180089. https://doi.org/10.1098/rstb.2018.0089

McRobbie, A.-M., Meyer, B., Rouillon, C., Petrovic-Stojanovska, B., Liu, H., & White, M. F. (2012). *Staphylococcus aureus* DinG, a helicase that has evolved into a nuclease. *Biochemical Journal*, *442*(1), 77–84. https://doi.org/10.1042/BJ20111903

Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J., & Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology*, *155*(3), 733–740. https://doi.org/10.1099/mic.0.023960-0

Mulepati, S., & Bailey, S. (2011). Structural and Biochemical Analysis of Nuclease Domain of Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-

associated Protein 3 (Cas3). *Journal of Biological Chemistry*, *286*(36), 31896–31903. https://doi.org/10.1074/jbc.M111.270017

Mulepati, S., Héroux, A., & Bailey, S. (2014). Crystal structure of a CRISPR RNA–guided surveillance complex bound to a ssDNA target. *Science*, *345*(6203), 1479–1484. https://doi.org/10.1126/science.1256996

Naser, I. B., Hoque, M. M., Nahid, M. A., Tareq, T. M., Rocky, M. K., & Faruque, S. M. (2017). Analysis of the CRISPR-Cas system in bacteriophages active on epidemic strains of Vibrio cholerae in Bangladesh. *Scientific Reports*, *7*(1), 14880. https://doi.org/10.1038/s41598-017-14839-2

Newire, E., Aydin, A., Juma, S., Enne, V. I., & Roberts, A. P. (2020). Identification of a Type IV-A CRISPR-Cas System Located Exclusively on IncHI1B/IncFIB Plasmids in Enterobacteriaceae. *Frontiers in Microbiology*, *11*. https://doi.org/10.3389/fmicb.2020.01937

Niewoehner, O., Garcia-Doval, C., Rostøl, J. T., Berk, C., Schwede, F., Bigler, L., Hall, J., Marraffini, L. A., & Jinek, M. (2017). Type III CRISPR–Cas systems produce cyclic oligoadenylate second messengers. *Nature*, *548*(7669), 543–548. https://doi.org/10.1038/nature23467

Niewoehner, O., Jinek, M., & Doudna, J. A. (2014). Evolution of CRISPR RNA recognition and processing by Cas6 endonucleases. *Nucleic Acids Research*, *42*(2), 1341–1353. https://doi.org/10.1093/nar/gkt922

Nobrega, F. L., Walinga, H., Dutilh, B. E., & Brouns, S. J. J. (2020). Prophages are associated with extensive CRISPR–Cas auto-immunity. *Nucleic Acids Research*, gkaa1071. https://doi.org/10.1093/nar/gkaa1071

Nuñez, J. K., Bai, L., Harrington, L. B., Hinder, T. L., & Doudna, J. A. (2016). CRISPR Immunological Memory Requires a Host Factor for Specificity. *Molecular Cell*, *62*(6), 824–833. https://doi.org/10.1016/j.molcel.2016.04.027

Nuñez, J. K., Kranzusch, P. J., Noeske, J., Wright, A. V., Davies, C. W., & Doudna, J. A. (2014). Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. *Nature Structural & Molecular Biology*, *21*(6), 528–534. https://doi.org/10.1038/nsmb.2820

O'Brien, R. E., Santos, I. C., Wrapp, D., Bravo, J. P. K., Schwartz, E. A., Brodbelt, J. S., & Taylor, D. W. (2020). Structural basis for assembly of non-canonical small subunits into type I-C Cascade. *Nature Communications*, *11*. https://doi.org/10.1038/s41467-020-19785-8

O'Brien, R. E., Santos, I. C., Wrapp, D., Taylor, D. W., Bravo, J. P. K., Schwartz, E. A., & Brodbelt, J. S. (2020). Structural basis for assembly of non-canonical small subunits into type I-C Cascade. *Nature Communications*, 1–6. https://doi.org/10.1038/s41467-020-19785-8

Osawa, T., Inanaga, H., Sato, C., & Numata, T. (2015a). Crystal structure of the crispr-cas RNA silencing cmr complex bound to a target analog. *Molecular Cell*, *58*(3), 418–430. https://doi.org/10.1016/j.molcel.2015.03.018

Osawa, T., Inanaga, H., Sato, C., & Numata, T. (2015b). Crystal Structure of the CRISPR-Cas RNA Silencing Cmr Complex Bound to a Target Analog. *Molecular Cell*, *58*(3), 418–430. https://doi.org/10.1016/j.molcel.2015.03.018

Özcan, A., Pausch, P., Linden, A., Wulf, A., Schühle, K., Heider, J., Urlaub, H., Heimerl, T., Bange, G., & Randau, L. (2018). Type IV CRISPR RNA processing and

effector complex formation in Aromatoleum aromaticum. *Nature Microbiology*. https://doi.org/10.1038/s41564-018-0274-8

Özcan, A., Pausch, P., Linden, A., Wulf, A., Schühle, K., Heider, J., Urlaub, H., Heimerl, T., Bange, G., & Randau, L. (2019). Type IV CRISPR RNA processing and effector complex formation in Aromatoleum aromaticum. *Nature Microbiology*, *4*(1), 89–96. https://doi.org/10.1038/s41564-018-0274-8

Peissert, S., Sauer, F., Grabarczyk, D. B., Braun, C., Sander, G., Poterszman, A., Egly, J.-M., Kuper, J., & Kisker, C. (2020). In TFIIH the Arch domain of XPD is mechanistically essential for transcription and DNA repair. *Nature Communications*, *11*(1), 1667. https://doi.org/10.1038/s41467-020-15241-9

Peters, J. E., Makarova, K. S., Shmakov, S., & Koonin, E. V. (2017). Recruitment of CRISPR-Cas systems by Tn7-like transposons. *Proceedings of the National Academy of Sciences*, *114*(35), E7358–E7366. https://doi.org/10.1073/pnas.1709035114

Pinilla-Redondo, R., Mayo-Muñoz, D., Russel, J., Garrett, R. A., Randau, L., Sørensen, S. J., & Shah, S. A. (2019). Type IV CRISPR–Cas systems are highly diverse and involved in competition between plasmids. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkz1197

Poranen, M. M., & Tuma, R. (2004). Self-assembly of double-stranded RNA bacteriophages. *Virus Research*, *101*(1), 93–100. https://doi.org/10.1016/j.virusres.2003.12.009

Punjani, A., Rubinstein, J. L., Fleet, D. J., & Brubaker, M. A. (2017). CryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods*, *14*(3), 290–296. https://doi.org/10.1038/nmeth.4169

Redding, S., Sternberg, S. H., Marshall, M., Gibb, B., Bhat, P., Guegler, C. K., Wiedenheft, B., Doudna, J. A., & Greene, E. C. (2015). Surveillance and Processing of Foreign DNA by the Escherichia coli CRISPR-Cas System. *Cell*, *163*(4), 854–865. https://doi.org/10.1016/j.cell.2015.10.003

Ren, B., Duan, X., & Ding, H. (2009). Redox control of the DNA damage-inducible protein DinG helicase activity via its iron-sulfur cluster. *The Journal of Biological Chemistry*, *284*(8), 4829–4835. https://doi.org/10.1074/jbc.M807943200

Rohou, A., & Grigorieff, N. (2015). CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *Journal of Structural Biology*, *192*(2), 216–221. https://doi.org/10.1016/j.jsb.2015.08.008

Rollie, C., Schneider, S., Brinkmann, A. S., Bolt, E. L., & White, M. F. (2015). Intrinsic sequence specificity of the Cas1 integrase directs new spacer acquisition. *ELife*, *4*. https://doi.org/10.7554/eLife.08716

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A., & Huelsenbeck, J. P. (2012). MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Systematic Biology*, *61*(3), 539–542. https://doi.org/10.1093/sysbio/sys029

Samai, P., Pyenson, N., Jiang, W., Goldberg, G. W., Hatoum-Aslan, A., & Marraffini, L. A. (2015). Co-transcriptional DNA and RNA Cleavage during Type III CRISPR-

Cas Immunity. *Cell*, *161*(5), 1164–1174. https://doi.org/10.1016/j.cell.2015.04.027

Sashital, D. G., Jinek, M., & Doudna, J. A. (2011). An RNA-induced conformational change required for CRISPR RNA cleavage by the endoribonuclease Cse3. *Nature Structural & Molecular Biology*, *18*(6), 680–687. https://doi.org/10.1038/nsmb.2043

Savage, H., Montoya, G., Svensson, C., Schwenn, J. D., & Sinning, I. (1997). Crystal structure of phosphoadenylyl sulphate (PAPS) reductase: A new family of adenine nucleotide α hydrolases. *Structure*, *5*(7), 895–906. https://doi.org/10.1016/S0969-2126(97)00244-X

Scheres, S. H. W. (2012). RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *Journal of Structural Biology*, *180*(3), 519–530. https://doi.org/10.1016/j.jsb.2012.09.006

Seed, K. D., Lazinski, D. W., Calderwood, S. B., & Camilli, A. (2013). A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. *Nature*, *494*(7438), 489–491. https://doi.org/10.1038/nature11927

Shmakov, S. A., Makarova, K. S., Wolf, Y. I., Severinov, K. V., & Koonin, E. V. (2018). Systematic prediction of genes functionally linked to CRISPR-Cas systems by gene neighborhood analysis. *Proceedings of the National Academy of Sciences*, *115*(23), E5307–E5316. https://doi.org/10.1073/pnas.1803440115

Shmakov, S. A., Sitnik, V., Makarova, K. S., Wolf, Y. I., Severinov, K. V., & Koonin, E. V. (2017). The CRISPR Spacer Space Is Dominated by Sequences from Species-Specific Mobilomes. *MBio*, *8*(5). https://doi.org/10.1128/mBio.01397-17

Sinkunas, T., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., & Siksnys, V. (2011). Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system: Cas3 nuclease/helicase. *The EMBO Journal*, *30*(7), 1335–1342. https://doi.org/10.1038/emboj.2011.41

Staals, R. H. J., Agari, Y., Maki-Yonekura, S., Zhu, Y., Taylor, D. W., van Duijn, E., Barendregt, A., Vlot, M., Koehorst, J. J., Sakamoto, K., Masuda, A., Dohmae, N., Schaap, P. J., Doudna, J. A., Heck, A. J. R., Yonekura, K., van der Oost, J., & Shinkai, A. (2013). Structure and Activity of the RNA-Targeting Type III-B CRISPR-Cas Complex of Thermus thermophilus. *Molecular Cell*, *52*(1), 135–145. https://doi.org/10.1016/j.molcel.2013.09.013

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., … van der Oost, J. (2014). RNA targeting by the type III-A CRISPR-Cas Csm complex of Thermus thermophilus. *Molecular Cell*, *56*(4), 518–530. https://doi.org/10.1016/j.molcel.2014.10.005

Steinegger, M., & Söding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotechnology*, *35*(11), 1026–1028. https://doi.org/10.1038/nbt.3988

Sternberg, S. H., LaFrance, B., Kaplan, M., & Doudna, J. A. (2015). Conformational control of DNA target cleavage by CRISPR–Cas9. *Nature*, *527*(7576), 110–113. https://doi.org/10.1038/nature15544

Sternberg, S. H., Richter, H., Charpentier, E., & Qimron, U. (2016). Adaptation in CRISPR-Cas Systems. *Molecular Cell*, *61*(6), 797–808. https://doi.org/10.1016/j.molcel.2016.01.030

Suloway, C., Pulokas, J., Fellmann, D., Cheng, A., Guerra, F., Quispe, J., Stagg, S., Potter, C. S., & Carragher, B. (2005). Automated molecular microscopy: The new Leginon system. *Journal of Structural Biology*, *151*(1), 41–60. https://doi.org/10.1016/j.jsb.2005.03.010

Tamulaitis, G., Kazlauskiene, M., Manakova, E., Venclovas, Č., Nwokeoji, A. O., Dickman, M. J., Horvath, P., & Siksnys, V. (2014). Programmable RNA Shredding by the Type III-A CRISPR-Cas System of Streptococcus thermophilus. *Molecular Cell*, *56*(4), 506–517. https://doi.org/10.1016/j.molcel.2014.09.027

Taylor, H. N., Warner, E. E., Armbrust, M. J., Crowley, V. M., Olsen, K. J., & Jackson, R. N. (2019). Structural basis of Type IV CRISPR RNA biogenesis by a Cas6 endoribonuclease. *RNA Biology*. https://doi.org/10.1080/15476286.2019.1634965

Thakur, R. S., Desingu, A., Basavaraju, S., Subramanya, S., Rao, D. N., & Nagaraju, G. (2014). Mycobacterium tuberculosis DinG Is a Structure-specific Helicase That Unwinds G4 DNA Implications for Targeting G4 DNA as a Novel Therapeutic Approach. *Journal of Biological Chemistry*, *289*(36), 25112–25136. https://doi.org/10.1074/jbc.M114.563569

van Erp, P. B. G., Jackson, R. N., Carter, J., Golden, S. M., Bailey, S., & Wiedenheft, B. (2015). Mechanism of CRISPR-RNA guided recognition of DNA targets in Escherichia coli. *Nucleic Acids Research*, *43*(17), 8381–8391. https://doi.org/10.1093/nar/gkv793

Voloshin, O. N., & Camerini-Otero, R. D. (2007a). The DinG Protein from *Escherichia coli* Is a Structure-specific Helicase. *Journal of Biological Chemistry*, *282*(25), 18437–18447. https://doi.org/10.1074/jbc.M700376200

Voloshin, O. N., & Camerini-Otero, R. D. (2007b). The DinG Protein from *Escherichia coli* Is a Structure-specific Helicase. *Journal of Biological Chemistry*, *282*(25), 18437–18447. https://doi.org/10.1074/jbc.M700376200

Voloshin, O. N., Vanevski, F., Khil, P. P., & Camerini-Otero, R. D. (2003a). Characterization of the DNA Damage-inducible Helicase DinG from *Escherichia coli*. *Journal of Biological Chemistry*, *278*(30), 28284–28293. https://doi.org/10.1074/jbc.M301188200

Voloshin, O. N., Vanevski, F., Khil, P. P., & Camerini-Otero, R. D. (2003b). Characterization of the DNA Damage-inducible Helicase DinG from *Escherichia coli*. *Journal of Biological Chemistry*, *278*(30), 28284–28293. https://doi.org/10.1074/jbc.M301188200

Wang, J., Li, J., Zhao, H., Sheng, G., Wang, M., Yin, M., & Wang, Y. (2015). Structural and Mechanistic Basis of PAM-Dependent Spacer Acquisition in CRISPR-Cas Systems. *Cell*, *163*(4), 840–853. https://doi.org/10.1016/j.cell.2015.10.008

Wang, L., Jiang, S., Deng, Z., Dedon, P. C., & Chen, S. (2018). DNA phosphorothioate modification—A new multi-functional epigenetic system in bacteria. *FEMS Microbiology Reviews*, *43*(2), 109–122. https://doi.org/10.1093/femsre/fuy036

Wei, Y., Terns, R. M., & Terns, M. P. (2015). Cas9 function and host genome sampling in Type II-A CRISPR–Cas adaptation. *Genes & Development*, *29*(4), 356–361. https://doi.org/10.1101/gad.257550.114

Westra, E. R., Semenova, E., Datsenko, K. A., Jackson, R. N., Wiedenheft, B., Severinov, K., & Brouns, S. J. J. (2013). Type I-E CRISPR-Cas Systems Discriminate Target from Non-Target DNA through Base Pairing-Independent PAM Recognition. *PLOS Genetics*, *9*(9), e1003742. https://doi.org/10.1371/journal.pgen.1003742

Westra, E. R., van Erp, P. B. G., Künne, T., Wong, S. P., Staals, R. H. J., Seegers, C. L. C., Bollen, S., Jore, M. M., Semenova, E., Severinov, K., de Vos, W. M., Dame, R. T., de Vries, R., Brouns, S. J. J., & van der Oost, J. (2012). CRISPR Immunity Relies on the Consecutive Binding and Degradation of Negatively Supercoiled Invader DNA by Cascade and Cas3. *Molecular Cell*, *46*(5), 595–605. https://doi.org/10.1016/j.molcel.2012.03.018

Wu, Y., & Brosh, R. M., Jr. (2012). DNA helicase and helicase–nuclease enzymes with a conserved iron–sulfur cluster. *Nucleic Acids Research*, *40*(10), 4247–4260. https://doi.org/10.1093/nar/gks039

Yan, W. X., Hunnewell, P., Alfonse, L. E., Carte, J. M., Keston-Smith, E., Sothiselvam, S., Garrity, A. J., Chong, S., Makarova, K. S., Koonin, E. V., Cheng, D. R., & Scott, D. A. (2019). Functionally diverse type V CRISPR-Cas systems. *Science*, *363*(6422), 88–91. https://doi.org/10.1126/science.aav7271

Yin, Y., Yang, B., & Entwistle, S. (2019). Bioinformatics Identification of Anti-CRISPR Loci by Using Homology, Guilt-by-Association, and CRISPR Self-Targeting Spacer Approaches. *MSystems*, *4*(5), e00455-19, /msystems/4/5/msys.00455-19.atom. https://doi.org/10.1128/mSystems.00455-19

Yosef, I., Goren, M. G., & Qimron, U. (2012). Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. *Nucleic Acids Research*, *40*(12), 5569–5576. https://doi.org/10.1093/nar/gks216

You, D., Wang, L., Yao, F., Zhou, X., & Deng, Z. (2007). A Novel DNA Modification by Sulfur: DndA Is a NifS-like Cysteine Desulfurase Capable of Assembling DndC as an Iron−Sulfur Cluster Protein in *Streptomyces lividans* †. *Biochemistry*, *46*(20), 6126–6133. https://doi.org/10.1021/bi602615k

You, L., Ma, J., Wang, J., Artamonova, D., Wang, M., Liu, L., Xiang, H., Severinov, K., Zhang, X., & Wang, Y. (2019). Structure Studies of the CRISPR-Csm Complex Reveal Mechanism of Co-transcriptional Interference. *Cell*, *176*(1–2), 239-253.e16. https://doi.org/10.1016/j.cell.2018.10.052

Zhang, J., Graham, S., Tello, A., Liu, H., & White, M. F. (2016). Multiple nucleic acid cleavage modes in divergent type III CRISPR systems. *Nucleic Acids Research*, *44*(4), 1789–1799. https://doi.org/10.1093/nar/gkw020

Zheng, S. Q., Palovcak, E., Armache, J. P., Verba, K. A., Cheng, Y., & Agard, D. A. (2017). MotionCor2: Anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nature Methods*, *14*(4), 331–332. https://doi.org/10.1038/nmeth.4193

Zhou, C., Wu, Q., Wang, B., Lin, P., Wu, M., & Yu, X. (2021). Calcium-responsive kinase LadS modulates type I–F CRISPR-Cas adaptive immunity. *Biochemical and Biophysical Research Communications*, *546*, 155–161. https://doi.org/10.1016/j.bbrc.2021.01.100

Zhou, Y., Bravo, J. P. K., Taylor, H. N., Steens, J. A., Jackson, R. N., Staals, R. H. J., & Taylor, D. W. (2021). Structure of a type IV CRISPR-Cas ribonucleoprotein complex. *IScience*, 102201. https://doi.org/10.1016/j.isci.2021.102201

Zimmermann, L., Stephens, A., Nam, S.-Z., Rau, D., Kübler, J., Lozajic, M., Gabler, F., Söding, J., Lupas, A. N., & Alva, V. (2018). A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *Journal of Molecular Biology*, *430*(15), 2237–2243. https://doi.org/10.1016/j.jmb.2017.12.007

APPENDIX

PERMISSIONS LETTERS

Hannah Taylor <hannahwagner12@aggiemail.usu.edu> Wed, May 19, 2021 at 3:35 PM

Hi all,

I will be submitting my dissertation soon and it will be archived at the Utah State University library. You are a co-author on one or more publications (listed below) that I have included in my dissertation. If you have any concerns about the inclusion of these articles in my dissertation, please reach out to me.

It has been a pleasure working and collaborating with you.

Thanks!

Zhou, Y., Bravo, J. P. K., Taylor, H. N., Steens, J. A., Jackson, R. N., Staals, R. H. J., & Taylor, D. W. (2021). Structure of a type IV CRISPR-Cas ribonucleoprotein complex. IScience, 24(3). https://doi.org/10.1016/j.isci.2021.102201

Taylor, H. N., Warner, E. E., Armbrust, M. J., Crowley, V. M., Olsen, K. J., & Jackson, R. N. (2019). Structural basis of Type IV CRISPR RNA biogenesis by a Cas6 endoribonuclease. RNA Biology. https://doi.org/10.1080/15476286.2019.1634965

Crowley, V. M., Catching, A., Taylor, H. N., Borges, A. L., Metcalf, J., Bondy-Denomy, J., & Jackson, R. N. (2019). A Type IV-A CRISPR-Cas System in Pseudomonas

aeruginosa Mediates RNA-Guided Plasmid Interference In Vivo. The CRISPR Journal, 2(6), 434–440. https://doi.org/10.1089/crispr.2019.0048

Taylor, H. N., Laderman, E., Armbrust, M. J., Hallmark, T., Keiser, D., Bondy-Denomy, J., & Jackson, R. N. (2021). Positioning diverse type IV structures and functions within class 1 CRISPR-Cas systems. Frontiers in Microbiology.

Bravo, Jack P <jack.bravo@austin.utexas.edu> Wed, May 19, 2021 at 3:38 PM

To: Hannah Taylor <hannahwagner12@aggiemail.usu.edu>

Hi Dr. Hannah,

I have no problem whatsoever, it's been great fun working with you. Congratulations again on passing your defence - best of luck in the future! Do you have any future post-PhD plans?

Cheers,

Jack

--

Jack Bravo, Ph.D.

Postdoctoral Research Scholar

Department of Molecular Biosciences

University of Texas at Austin

Staals, Raymond <raymond.staals@wur.nl> Thu, May 20, 2021 at 8:10 AM

To: Hannah Taylor <hannahwagner12@aggiemail.usu.edu>

Hey Hannah,

Really happy to hear this new, congratulations on getting your dissertation ready for submission. When is the big day?

Needless to say, I have no concerns whatsoever about including the papers where we are listed as co-author.

Cheers,

Raymond

Dylan Keiser <dylankeiser@aggiemail.usu.edu> Wed, Jul 7, 2021 at 9:29 AM

To: Hannah Taylor <hannahwagner12@aggiemail.usu.edu>

I, Dylan Keiser, hereby give permission to Hannah N. Taylor to reprint the following publication(s) in part or in full in her dissertation.

Taylor, H. N., Laderman, E., Armbrust, M. J., Hallmark, T., Keiser, D., Bondy-Denomy, J., & Jackson, R. N. (2021). Positioning diverse type IV structures and functions within class 1 CRISPR-Cas systems. Frontiers in Microbiology.

Joe Bondy-Denomy <joebondy@gmail.com> Wed, Jul 7, 2021 at 9:46 AM

To: Hannah Taylor <hannahwagner12@aggiemail.usu.edu>

I, Joseph Bondy-Denomy, hereby give permission to Hannah N. Taylor to reprint the following publication(s) in part or in full in her dissertation.

Crowley, V. M., Catching, A., Taylor, H. N., Borges, A. L., Metcalf, J., Bondy-Denomy, J., & Jackson, R. N. (2019). A Type IV-A CRISPR-Cas System in Pseudomonas aeruginosa Mediates RNA-Guided Plasmid Interference In Vivo. The CRISPR Journal, 2(6), 434–440.

Taylor, H. N., Laderman, E., Armbrust, M. J., Hallmark, T., Keiser, D., Bondy-Denomy, J., & Jackson, R. N. (2021). Positioning diverse type IV structures and functions within class 1 CRISPR-Cas systems. Frontiers in Microbiology.

David Taylor <dtaylor@utexas.edu> Wed, Jul 7, 2021 at 10:12 AM

To: Hannah Taylor <hannahwagner12@aggiemail.usu.edu>

Dear Hannah,

I, David W. Taylor, hereby give permission to Hannah N. Taylor to reprint the following publication(s) in part or in full in her dissertation.

Zhou, Y., Bravo, J. P. K., Taylor, H. N., Steens, J. A., Jackson, R. N., Staals, R. H. J., & Taylor, D. W. (2021). Structure of a type IV CRISPR-Cas ribonucleoprotein complex. IScience, 24(3).

Best,

David

--

David Taylor, Ph.D. (he/him)

Assistant Professor and

CPRIT Scholar

Department of Molecular Biosciences

Advisor, Biochemistry Graduate Studies

University of Texas at Austin

davidtaylorlab.com

Thom Hallmark <thom.hallmark@usu.edu> Wed, Jul 7, 2021 at 10:27 AM

To: "hannahwagner12@aggiemail.usu.edu" <hannahwagner12@aggiemail.usu.edu>

I, Thomson Hallmark, hereby give permission to Hannah N. Taylor to reprint the following publication(s) in part or in full in her dissertation.

Taylor, H. N., Laderman, E., Armbrust, M. J., Hallmark, T., Keiser, D., Bondy-Denomy, J., & Jackson, R. N. (2021). Positioning diverse type IV structures and functions within class 1 CRISPR-Cas systems. Frontiers in Microbiology.

Good luck!

Thom

Josie Metcalf <josiemetcalf2010@gmail.com> Wed, Jul 7, 2021 at 11:08 AM

To: Hannah Taylor <hannahwagner12@aggiemail.usu.edu>

I, Josie Metcalf, hereby give permission to Hannah N. Taylor to reprint the following publication(s) in part or in full in her dissertation.

Crowley, V. M., Catching, A., Taylor, H. N., Borges, A. L., Metcalf, J., Bondy-Denomy, J., & Jackson, R. N. (2019). A Type IV-A CRISPR-Cas System in Pseudomonas aeruginosa Mediates RNA-Guided Plasmid Interference In Vivo. The CRISPR Journal, 2(6), 434–440.

--

Thanks,

Josie Metcalf

Yi Zhou <yizhou@utexas.edu> Wed, Jul 7, 2021 at 11:29 AM

To: Hannah Taylor <hannahwagner12@aggiemail.usu.edu>

Hi Hannah,

Congratulations on graduation. Please see the statement below.

Thanks,

Yi

I, Yi Zhou, hereby give permission to Hannah N. Taylor to reprint the following publication(s) in part or in full in her dissertation.

Zhou, Y., Bravo, J. P. K., Taylor, H. N., Steens, J. A., Jackson, R. N., Staals, R. H. J., & Taylor, D. W. (2021). Structure of a type IV CRISPR-Cas ribonucleoprotein complex. IScience, 24(3).

Laderman, Eric <Eric.Laderman@ucsf.edu> Wed, Jul 7, 2021 at 12:12 PM

To: Hannah Taylor <hannahwagner12@aggiemail.usu.edu>

I, Eric Laderman, hereby give permission to Hannah N. Taylor to reprint the following publication(s) in part or in full in her dissertation.

Taylor, H. N., Laderman, E., Armbrust, M. J., Hallmark, T., Keiser, D., Bondy-Denomy, J., & Jackson, R. N. (2021). Positioning diverse type IV structures and functions within class 1 CRISPR-Cas systems. Frontiers in Microbiology.

Emily Warner <emilyecray@gmail.com> Wed, Jul 7, 2021 at 12:18 PM

To: Hannah Taylor <hannahwagner12@aggiemail.usu.edu>

I, Emily Warner, hereby give permission to Hannah N. Taylor to reprint the following publication(s) in part or in full in her dissertation.

Taylor, H. N., Warner, E. E., Armbrust, M. J., Crowley, V. M., Olsen, K. J., & Jackson, R. N. (2019). Structural basis of Type IV CRISPR RNA biogenesis by a Cas6 endoribonuclease. RNA Biology.

Sent from my iPhone

Matt Armbrust <matt.armbrust5@gmail.com> Wed, Jul 7, 2021 at 12:52 PM

To: Hannah Taylor <hannahwagner12@aggiemail.usu.edu>

I, Matt Armbrust, hereby give permission to Hannah N. Taylor to reprint the following publication(s) in part or in full in her dissertation.

Taylor, H. N., Warner, E. E., Armbrust, M. J., Crowley, V. M., Olsen, K. J., & Jackson, R. N. (2019). Structural basis of Type IV CRISPR RNA biogenesis by a Cas6 endoribonuclease. RNA Biology.

Taylor, H. N., Laderman, E., Armbrust, M. J., Hallmark, T., Keiser, D., Bondy-Denomy, J., & Jackson, R. N. (2021). Positioning diverse type IV structures and functions within class 1 CRISPR-Cas systems. Frontiers in Microbiology.

Valerie Crowley <valerie.crowley@gmail.com> Wed, Jul 7, 2021 at 1:16 PM

To: Hannah Taylor <hannahwagner12@aggiemail.usu.edu>

I, Valerie Crowley, hereby give permission to Hannah N. Taylor to reprint the following publication(s) in part or in full in her dissertation.

Taylor, H. N., Warner, E. E., Armbrust, M. J., Crowley, V. M., Olsen, K. J., & Jackson, R. N. (2019). Structural basis of Type IV CRISPR RNA biogenesis by a Cas6 endoribonuclease. RNA Biology.

Crowley, V. M., Catching, A., Taylor, H. N., Borges, A. L., Metcalf, J., Bondy-Denomy, J., & Jackson, R. N. (2019). A Type IV-A CRISPR-Cas System in Pseudomonas aeruginosa Mediates RNA-Guided Plasmid Interference In Vivo. The CRISPR Journal, 2(6), 434–440

Catching, Benjamin <Benjamin.Catching@ucsf.edu> Wed, Jul 7, 2021 at 4:21 PM

To: Hannah Taylor <hannahwagner12@aggiemail.usu.edu>

I, Benjamin Adam Catching, hereby give permission to Hannah N. Taylor to reprint the following publication(s) in part or in full in her dissertation.

Crowley, V. M., Catching, A., Taylor, H. N., Borges, A. L., Metcalf, J., Bondy-Denomy, J., & Jackson, R. N. (2019). A Type IV-A CRISPR-Cas System in Pseudomonas aeruginosa Mediates RNA-Guided Plasmid Interference In Vivo. The CRISPR Journal, 2(6), 434–440.

Benjamin Adam Catching

Ph.D. Candidate, Biophysics

Andino Lab

University of California, San Francisco

Steens, Jurre <jurre.steens@wur.nl> Thu, Jul 8, 2021 at 10:01 AM

To: Hannah Taylor <hannahwagner12@aggiemail.usu.edu>

Hi Hannah,

I, Jurre, hereby give permission to Hannah N. Taylor to reprint the following publication(s) in part or in full in her dissertation.

Zhou, Y., Bravo, J. P. K., Taylor, H. N., Steens, J. A., Jackson, R. N., Staals, R. H. J., & Taylor, D. W. (2021). Structure of a type IV CRISPR-Cas ribonucleoprotein complex. IScience, 24(3).

All the best with finalizing everything!

Jurre

CURRICULUM VITAE

Hannah Nicole Taylor

LinkedIn: www.linkedin.com/in/hannah-nicole-taylor-21bb89159/
ORCid: https://orcid.org/0000-0002-5643-8998

EDUCATION

PhD, Utah State University, Logan, UT; 2021; Biochemistry.
BS, Utah State University, Logan, UT; 2016; Cellular/Molecular Biology, Chemistry minor; Magna Cum Laude.
AS, Utah Valley University, Orem, UT; 2013; General Studies; High Honors.

RESEARCH EXPERIENCE

Graduate Researcher:

Biochemical and structural analysis of CRISPR-Cas systems, 2016-2021.
PI: Ryan Jackson, Utah State University.

Undergraduate Researcher:

Determining the function of the Nodel gene, Summer 2015.
PI: Jan Christian, University of Utah.

Studying effects of nanoparticles on wheat, 2014-2015.
PI: Anne Anderson, Utah State University.

Lab Technician:

Studying multi-generational effects of diet, Spring 2016.
PI: Abby Benninghoff, Utah State University.

Determining evolutionary patterns of seed beetles, Summer 2014.
PI: Frank Messina, Utah State University.

PUBLICATIONS

1. Adams, J., Wright, M., Wagner, H., Valiente, J., Britt, D., Anderson, A. (2017). Cu from dissolution of CuO nanoparticles signals changes in root morphology. Plant Physiology and Biochemistry 110:103-117. DOI: 10.1016/j.plaphy.2016.08.005
*Undergraduate research publication

2. Taylor, H. N., Warner, E. E., Armbrust, M. J., Crowley, V. M., Olsen, K. J., Jackson, R. N. (2019). Structural basis of Type IV CRISPR RNA biogenesis by a Cas6 endoribonuclease. RNA Biology. DOI: 10.1080/15476286.2019.1634965

3. Crowley, V. M., Catching, A., Taylor, H. N., Borges, A. L., Metcalf, J., Bondy-Denomy, J., Jackson, R. N. (2019). A Type IV-A CRISPR-Cas System in Pseudomonas aeruginosa Mediates RNA-Guided Plasmid Interference In Vivo. The CRISPR Journal. DOI: 10.1089/crispr.2019.0048

4. Zhou, Y.*, Bravo, J. P. K.*, Taylor, H. N.*, Steens, J. A., Jackson, R. N., Staals, R. H. J., & Taylor, D. W. (2021). Structure of a type IV CRISPR-Cas ribonucleoprotein complex. IScience, 24(3). https://doi.org/10.1016/j.isci.2021.102201
*Co-1st authors

5. Taylor, H. N., Laderman, E., Armbrust, M. J., Hallmark, T., Keiser, D., Bondy-Denomy, J., & Jackson, R. N. (2021). Positioning diverse type IV structures and functions within class 1 CRISPR-Cas systems. Frontiers in Microbiology, 12. https://doi.org/10.3389/fmicb.2021.671522

POSTER PRESENTATIONS

World Microbe Forum
Online, international 2021

Hansen Life Sciences Retreat
Utah State University, Logan, UT 2019

Graduate Recruiting Weekend
Utah State University, Logan, UT 2019

American Association for the Advancement of Science Annual Meeting
Washington D. C. 2019

Hansen Life Sciences Retreat
Utah State University, Wellsville, UT 2018
*Awarded best poster

CRISPR Meeting
Vilnius, Lithuania 2018

Graduate Recruiting Weekend
Utah State University, Logan, UT 2018

Hansen Life Sciences Retreat
Utah State University, Logan, UT 2017

Graduate Recruiting Weekend
Utah State University, Logan, UT 2017

Student Research Symposium
Utah State University, Logan, UT 2016

Utah Conference on Undergraduate Research
University of Utah, Salt Lake City, UT 2016
Society for the Advancement of Chicanos/Hispanics and Native Americans in Science National Conference
Washington D. C. 2015
Fall Biology Undergraduate Research Symposium
Utah State University, Logan, UT 2014

ORAL PRESENTATIONS

Helicase-mediated mechanisms for defense against invasive plasmids by the type IV-A CRISPR-Cas bacterial immune system
American Society for Microbiology Intermountain Branch Meeting
Weber State University, Ogden, UT 2020
*Awarded best oral presentation in Session E
Structure of a Type IV-B CRISPR-Cas ribonucleoprotein complex
Hansen Life Sciences Retreat
Utah State University, Virtual Conference 2020
Guest lecture
Science Explorers
Leadership Education Academy, Virtual Classroom 2018
Type IV-A CRISPR RNA processing
Student Research Symposium
Utah State University, Logan, UT 2018
*Awarded best oral presentation by a graduate student in the Life Sciences
Jackson Lab overview
Hansen Life Sciences Retreat
Utah State University, Logan, UT 2017
Guest lecture
Science Explorers
Leadership Education Academy, Virtual Classroom 2017

TEACHING EXPERIENCE

Volunteer Tutor/Paid Tutor
General chemistry (4)** 2018
Organic chemistry (2) 2018
**Parentheses indicate number of students taught.

Teaching Assistant

Advanced Biochemistry Laboratory (20) 2017, (32) 2019
Organic Chemistry Laboratory (72) 2016, (72) 2017
General Chemistry Laboratory (72) 2020***
General Chemistry (~1000) 2020
***In addition to teaching, I also developed course material, facilitated a transition to online lab work, and produced online content.

Undergraduate Teaching Fellow

Principles of Genetics (60) 2015

DISTINCTIONS

Joseph Reuel Harris Graduate Scholarship
Utah State University 2020

Joshua E. Neimark Memorial Travel Assistance Award
American Association for the Advancement of Science 2019

Claude E. Zobell Scholarship Award
Utah State University 2018

RapiData Workshop Participant
Stanford Synchrotron Radiation Lightsource
SLAC National Accelerator Laboratory 2018

Society for the Advancement of Chicanos/Hispanics and Native Americans in Science National Conference Travel Grant
Utah State University 2017

Society for the Advancement of Chicanos/Hispanics and Native Americans in Science National Conference Travel Scholarship
SACNAS 2015

Summer Undergraduate Research Program Participant
University of Utah 2015

Christenson Memorial Scholarship
Utah State University 2015

Richard J and Marion A. Shaw Scholarship
Utah State University 2014

College of Science Minigrant
Utah State University 2014
*grant which funded my undergraduate research

Presidential Scholarship
Utah State University 2013-16
*Four year full-tuition academic scholarship

Dean's List Award
Utah State University 2013-15
Dean's List Award
Utah Valley University 2012-13

PROFESSIONAL ASSOCIATIONS

Society for the Advancement of Chicanos/Hispanics and Native Americans in Science (SACNAS)
American Association for the Advancement of Science (AAAS)
American Society for Microbiology (ASM) Intermountain Branch

SERVICE

GENIUS Olympiad Judge
Judged high-school oral poster presentations in an international competition
2021
American Association for the Advancement of Science Member Community Superhero
Facilitated and curated online discussions with scientists and the public
2019-21
Utah State University Native American Student Mentorship Program Mentor
Helped undergraduate students to complete a lab-based research project
2019
Utah State University Student Research Symposium Judge
Judged undergraduate and graduate oral and poster presentations
2019-20
Utah State University Graduate Enhancement Award Reviewer
Reviewed award applications
2019
Utah State University Society for the Advancement of Chicanos/Hispanics and Native Americans in Science Chapter President
Recruited students, organized events, and provided general leadership
2018-19
Utah State University Biotechnology Academy Mentor
Helped high school students complete a lab-based research project
2018