Investigations of Prokaryotic Defense Systems

Hannah Domgaard
Utah State University

Follow this and additional works at: https://digitalcommons.usu.edu/etd

Part of the Biochemical and Biomolecular Engineering Commons, and the Other Chemistry Commons

Recommended Citation
https://digitalcommons.usu.edu/etd/8639

This Thesis is brought to you for free and open access by the Graduate Studies at DigitalCommons@USU. It has been accepted for inclusion in All Graduate Theses and Dissertations by an authorized administrator of DigitalCommons@USU. For more information, please contact digitalcommons@usu.edu.
INVESTIGATIONS OF PROKARYOTIC DEFENSE SYSTEMS

by

Hannah Domgaard

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

of

MASTER OF SCIENCE

in

Biochemistry

Approved:

Ryan N. Jackson, Ph.D.                Joan Hevel, Ph.D.
Major Professor                      Committee Member

Sean J. Johnson, Ph.D.               D. Richard Cutler, Ph.D.
Committee Member                    Vice Provost of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2022
ABSTRACT

Investigations of Prokaryotic Defense Systems

by

Hannah Domgaard, Master of Science

Utah State University, 2022

Major Professor: Ryan Jackson, Ph.D.
Department: Chemistry and Biochemistry

Prokaryotic defense systems utilize protein-mediated mechanisms and chemical signaling to protect against mobile genetic elements (MGEs) such as phages and plasmids. This thesis contains studies of two defense systems, CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repeats - CRISPR Associated) and Wadjet systems. CRISPR-Cas systems defend against MGEs with programmable complexes composed of protein and an RNA guide. The general defense mechanisms of five of the six CRISPR system types are known. However, many CRISPR systems, such as type IV and variant systems such as type V-A2, remain uncharacterized. Type IV systems lack nuclease-associated genes, making it difficult to predict how immune function is delivered. To better understand type IV function, we identified a type IV-A CRISPR operon in Pseudomonas aeruginosa strain 83 adjacent to a damage-inducible helicase (dinG-like) gene. This thesis describes some of the first biochemical and structural studies performed on a CRISPR-associated DinG protein (CasDinG), showing that CasDinG is an ATP-dependent helicase that plays a pivotal role in type IV-A function. Type V systems house many potent genome editing tools, such as Cas9 and Cas12a.
Bioinformatic analysis identified new variant subtypes of Cas12a called Cas12a2. Cas12a2 was recently used to create insertion-deletion events in eukaryotic cell lines. However, how these variant proteins generated double-strand DNA (dsDNA) breaks is unknown. This thesis describes preliminary studies performed on Cas12a2 variants showing that they can process and bind nucleic acids.

Wadjet systems share homology with *Escherichia coli* chromosome condensin complexes, and recent research has shown that they limit the transformation efficiency of plasmids. However, the mechanism behind this action remains unknown. This thesis describes preliminary studies performed on *Azotobacter vinelandii* JetB and the cloning of Wadjet genes. This work is essential as conducting basic research on prokaryotic immune systems has continuously led to the development of innovative tools used in genome editing and medical and scientific research.

(180 pages)
PUBLIC ABSTRACT

Investigations of Prokaryotic Defense Systems

Hannah Domgaard

Bacteria are constantly threatened with infection by mobile genetic elements (MGE) such as bacteriophage and plasmids. Bacteriophage and plasmids require the bacteria's cellular infrastructure to replicate their genomes. Rampant replication can lead to cell death which is one reason why bacteria have developed a diverse array of immune systems to prevent or limit infection. This thesis studies three types of bacterial immune systems, type IV-A CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repeat – CRISPR associated), type V-A2 CRISPR-Cas systems, and Wadjet systems.

The type IV-A system lies adjacent to a dinG-like helicase gene. Research has shown that this system can target plasmids preventing their spread throughout a microbial population. This system is reliant upon the dinG-like gene, but how this system mechanistically prevents plasmid sharing is not understood.

The type V-A2 system has been shown to be capable of editing the genome of rice by generating breaks in the DNA. How this editing takes place, and the other biochemical mechanisms of this protein are not understood either. This thesis provides the preliminary framework for studying these putative genome editing tools.

Wadjet immune systems prevent the sharing of plasmids between bacteria. These systems share structural similarities with proteins responsible for separating bacterial chromosomes during cellular replication. This thesis contains the foundation for the characterization of the system by cloning genes from native host systems and then recombinantly expressing and purifying proteins.
ACKNOWLEDGMENTS

I am grateful for the past five years as I have learned much about the rigor and exacting standards of scientists in the field of CRISPR. These past years have taught me much about my passions, and I am happy to have had the opportunity to pursue a degree here at Utah State University. I thank all the following people for helping me along the way.

My husband, for the continual love and support he has given me.

My child, for providing me with the motivation to finish this degree.

My dogs, for providing snuggles and unconditional love.

My parents, for instilling in me a love of learning.

My siblings, for the laughter and support over the years.

My advisor, for allowing me to work on these projects and for his mentorship.

My lab associates, for the help on the bench.

My friends, for the tears and laughter shared over coffee.

Collaborators, for their hard work and effort put forth on manuscripts.

Hannah Domgaard
# CONTENTS

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
</tr>
<tr>
<td>Public Abstract</td>
</tr>
<tr>
<td>Acknowledgments</td>
</tr>
<tr>
<td>List of Tables</td>
</tr>
<tr>
<td>List of Figures</td>
</tr>
<tr>
<td>List of Abbreviations</td>
</tr>
<tr>
<td>Chapter 1</td>
</tr>
<tr>
<td>Bacterial Immune Systems</td>
</tr>
<tr>
<td>CRISPR-Cas Prokaryotic Adaptive Immune Systems</td>
</tr>
<tr>
<td>Wadjet Defense Systems</td>
</tr>
<tr>
<td>References</td>
</tr>
<tr>
<td>Chapter 2</td>
</tr>
<tr>
<td>Abstract</td>
</tr>
<tr>
<td>Introduction</td>
</tr>
<tr>
<td>Type IV CRISPR-Cas Systems</td>
</tr>
<tr>
<td><em>Pseudomonas aeruginosa</em> type IV-A CRISPR-Cas</td>
</tr>
<tr>
<td>CasDinG and Non CasDinG Helicases</td>
</tr>
<tr>
<td>Methods</td>
</tr>
<tr>
<td>Cloning &amp; Gene Alignments</td>
</tr>
<tr>
<td>Expression &amp; Purification of CasDinG</td>
</tr>
<tr>
<td>Westerns</td>
</tr>
<tr>
<td>Crystallization and Structure determination</td>
</tr>
<tr>
<td>Nucleic Acid Substrate Preparation</td>
</tr>
<tr>
<td>Nucleic acid-binding assays</td>
</tr>
<tr>
<td>Malachite Green ATPase Assays</td>
</tr>
<tr>
<td>Preparation of chemically competent type IV-A CRISPR Cas cells</td>
</tr>
<tr>
<td><em>In Vivo</em> Plasmid Competition Assay</td>
</tr>
<tr>
<td>Helicase assays</td>
</tr>
<tr>
<td>Type IV-A Complex (+/- CasDinG) Expression &amp; Purification</td>
</tr>
<tr>
<td>Far-UV Circular Dichroism (CD)</td>
</tr>
</tbody>
</table>
Chapter 3

Results

Methods

Introduction

Abstract

Type V CRISPR-Cas Systems

Wadjet Defense System

Acknowledgments

References

Discussion

In Vivo Plasmid Curing Assay

Type IV-A Complex and CasDinG

Plasmid Cleavage Assay

Protein Purification

Crystallization and structure determination

Nucleic acid-binding assays

ATPase Assays

Helicase assays

Mutational Analysis

In Vivo Plasmid Curing Assay

Type IV-A Complex and CasDinG

Discussion

Acknowledgments

References

Chapter 3

Abstract

Introduction

Type V CRISPR-Cas Systems

Wadjet Defense System

Methods

Nucleic Acid Substrate Preparation

Cloning of Cas12a2

Protein Expression & Purification

Pre-crRNA Processing Assay

DNA Binding Assays

DNA Cleavage Assays

Negative Stain & Cryo-EM of SuCas12a2

SAXS-SEC of Cas12a2

Negative Stain SmCas12a2

Cloning of Wadjet

Expression & Purification of 2ST Av JetB

Far-UV Circular Dichroism (CD)

Results

Protein Expression & Purification

SAXS of Cas12a2

Negative Stain SmCas12a2

Cryo-EM of SuCms1

Pre-crRNA Processing

Far- UV Circular Dichroism (CD) of SuCas12a2
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA Binding Assays</td>
<td>110</td>
</tr>
<tr>
<td>DNA Cleavage Assays</td>
<td>111</td>
</tr>
<tr>
<td>Wadjet Cloning, Protein Purification, &amp; Structure Prediction</td>
<td>114</td>
</tr>
<tr>
<td>Discussion</td>
<td>114</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>119</td>
</tr>
<tr>
<td>References</td>
<td>120</td>
</tr>
<tr>
<td><strong>Chapter 4</strong></td>
<td>124</td>
</tr>
<tr>
<td>Introduction</td>
<td>124</td>
</tr>
<tr>
<td>Type IV-A Summary &amp; Future Directions</td>
<td>125</td>
</tr>
<tr>
<td>Cas12a2 Summary &amp; Future Directions</td>
<td>131</td>
</tr>
<tr>
<td>Wadjet Summary &amp; Future Directions</td>
<td>134</td>
</tr>
<tr>
<td>References</td>
<td>137</td>
</tr>
<tr>
<td><strong>Appendices</strong></td>
<td>140</td>
</tr>
<tr>
<td>APPENDIX A</td>
<td>141</td>
</tr>
<tr>
<td>Chapter 2 Supplementary Information</td>
<td>141</td>
</tr>
<tr>
<td>APPENDIX B</td>
<td>157</td>
</tr>
<tr>
<td>Chapter 3 Supplementary Information</td>
<td>157</td>
</tr>
</tbody>
</table>
### LIST OF TABLES

| Table 2-1 CasDinG Data Collection and Refinement Statistics | 45 |
| Table 2-2 Apparent K_d for Nucleic Acid Substrates | 51 |
| Table 2-3 Michaelis-Menten Kinetic Parameters for ATP hydrolysis | 55 |
| Table 2-4 Apparent K_d of ssDNA Alternative PAM Substrates | 68 |
| Table A-1 Mutant(s) CasDinG Binding to ssDNA | 151 |
| Table A-2 CasDinG Binding to ssDNA and Blunt Duplex DNA | 151 |
| Table A-3 Constructs Generated of CasDinG by Hannah Domgaard | 151 |
| Table A-4 Substrates Used in CasDinG Assays | 152 |
| Table A-5 Primers to Mutate and Sequence CasDinG | 153 |
| Table A-6 Constructs of CasDinG Molecular Weights and Extinction Coefficients | 154 |
| Table A-7 ApoCasDinG Crystal Hits | 155 |
| Table B-1 Chapter 2 Construct Molecular Weights and Extinction Coefficients | 160 |
| Table B-2 Substrates Used in Cas12a2 Assays | 160 |
| Table B-3 Primers to Sequence and Mutate SuCas12a2 (Cms1) | 160 |
| Table B-4 Primers to Sequence and Mutate SmCas12a2 (Cms1) | 162 |
| Table B-5 Primers to Clone and Sequence Wadjet | 164 |
| Table B-6 Chapter 2 Constructs Cloned by Hannah Domgaard | 165 |
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1-1</td>
<td>CRISPR-Cas Systems are Adaptive Immune Systems</td>
<td>3</td>
</tr>
<tr>
<td>Figure 1-2</td>
<td>CRISPR-Cas Immune Systems are Diverse</td>
<td>6</td>
</tr>
<tr>
<td>Figure 2-1</td>
<td>CasDinG is Distinct from Non-CasDinG</td>
<td>29</td>
</tr>
<tr>
<td>Figure 2-2</td>
<td>Purification of CasDinG</td>
<td>43</td>
</tr>
<tr>
<td>Figure 2-3</td>
<td>X-Ray Structure of Strep-Tagged Apo CasDinG</td>
<td>46</td>
</tr>
<tr>
<td>Figure 2-4</td>
<td>CasDinG Accessory Domain Comparisons</td>
<td>47</td>
</tr>
<tr>
<td>Figure 2-5</td>
<td>Domain Alignments Comparing PaCasDinG to <em>E. coli</em> DinG</td>
<td>48</td>
</tr>
<tr>
<td>Figure 2-6</td>
<td>Alignment of Apo CasDinG to 6FWS</td>
<td>50</td>
</tr>
<tr>
<td>Figure 2-7</td>
<td>CasDinG Binding to Nucleic Acid Substrates</td>
<td>52</td>
</tr>
<tr>
<td>Figure 2-8</td>
<td>$V_0$ Plots of ATP Hydrolysis by CasDinG</td>
<td>54</td>
</tr>
<tr>
<td>Figure 2-9</td>
<td>Michaelis-Menten Curves of ATP Hydrolysis by CasDinG</td>
<td>56</td>
</tr>
<tr>
<td>Figure 2-10</td>
<td>CasDinG is a 5’- 3’ ATP-Dependent Helicase</td>
<td>59</td>
</tr>
<tr>
<td>Figure 2-11</td>
<td>Unwinding is Stimulated by ATP and Divalent Ions</td>
<td>60</td>
</tr>
<tr>
<td>Figure 2-12</td>
<td>Functional Analysis of CasDinG Domains <em>via</em> Mutagenesis</td>
<td>63</td>
</tr>
<tr>
<td>Figure 2-13</td>
<td><em>In Vivo</em> Plasmid Curing Assay</td>
<td>65</td>
</tr>
<tr>
<td>Figure 2-14</td>
<td>Type IV-A Protein Purification</td>
<td>66</td>
</tr>
<tr>
<td>Figure 2-15</td>
<td>Type IV-A Preliminary Binding and Cleavage Data</td>
<td>69</td>
</tr>
<tr>
<td>Figure 2-16</td>
<td>Hypothesized Role of CasDinG in CRISPR System Function</td>
<td>76</td>
</tr>
<tr>
<td>Figure 3-1</td>
<td>Organization of Type V-A2 Systems</td>
<td>87</td>
</tr>
<tr>
<td>Figure 3-2</td>
<td><em>Azotobacter vinelandii</em> Wadjet Overview</td>
<td>90</td>
</tr>
<tr>
<td>Figure 3-3</td>
<td>Protein Purification of Type V-A2 Proteins</td>
<td>103</td>
</tr>
</tbody>
</table>
Figure 3-4 SmCas12a2 RNP Negative Stain ................................................................. 105
Figure 3-5 SuCas12a2 Low Resolution Map and Model........................................... 106
Figure 3-6 SuCas12a2 crRNA Processing and Secondary Structure Validation........... 108
Figure 3-7 Model Used for the Development of RNA Processing Mutants............... 109
Figure 3-8 SuCas12a2 Binds DNA Substrates ......................................................... 112
Figure 3-9 SuCas12a2 Cleaves and/or Degrades DNA ........................................... 113
Figure 3-10 Recombinantly Expressed and Purified AzJetB is a Dimer...................... 115
Figure A-1 CasDinG Binding to Single-Stranded DNA of Variable Length (+ATP).... 141
Figure A-2 Mutant(s) CasDinG Binding to ssDNA (17 nt.) ....................................... 142
Figure A-3 Far UV-Circular Dichroism of CasDinG and Mutants ......................... 142
Figure A-4 Oligonucleotides Used in CasDinG Assays ........................................... 143
Figure A-5 Binding of CasDinG to ssDNA (40nt.) and Blunt DNA Duplex ............... 144
Figure A-6 Type IV RNP + CasDinG Binding to Target ssDNA TTTC PAM ............... 144
Figure A-7 Robetta Model of CasDinG with Mutations Marked.............................. 145
Figure A-8 CasDinG Helicase Motifs Highlighted in the RecA domains.................... 146
Figure A-9 EMBOSS-NEEDLE alignment of PaCasDinG to E. coli DinG ................. 147
Figure A-10 EMBOSS-NEEDLE alignment of PaCasDinG to S. aureus DinG ......... 149
Figure B-1 SAXS-SEC of SuCas12a2 ......................................................................... 158
Figure B-2 I-TASSER Homology Model of SuCas12a2 .......................................... 159
Figure B-3 DALI Top 5 Server Results for AvJet Alphafold Models ....................... 159
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRISPR</td>
<td>Clustered Regularly Interspaced Short Palindromic Repeats</td>
</tr>
<tr>
<td>Cas</td>
<td>CRISPR associated</td>
</tr>
<tr>
<td>crRNA</td>
<td>CRISPR derived RNA</td>
</tr>
<tr>
<td>gRNA</td>
<td>guide RNA</td>
</tr>
<tr>
<td>SEC</td>
<td>Size Exclusion Chromatography</td>
</tr>
<tr>
<td>SAXS</td>
<td>Small Angle X-Ray Scattering</td>
</tr>
<tr>
<td>dsDNA/ssDNA</td>
<td>Double-stranded or single-stranded deoxyribonucleic acid</td>
</tr>
<tr>
<td>dsRNA/ssRNA</td>
<td>double stranded or single stranded ribonucleic acid</td>
</tr>
<tr>
<td>Csf</td>
<td>CRISPR-Cas subtype as in <em>A. ferrooxidans</em></td>
</tr>
<tr>
<td>DinG</td>
<td>damage inducible gene G</td>
</tr>
<tr>
<td>RuvC</td>
<td>Recombination UV C - an endonuclease domain</td>
</tr>
<tr>
<td>Nuc</td>
<td>nuclease domain</td>
</tr>
<tr>
<td>PAM</td>
<td>protospacer adjacent motif</td>
</tr>
<tr>
<td>PFS</td>
<td>protospacer flanking sequence</td>
</tr>
<tr>
<td>RNP</td>
<td>ribonucleoprotein complex</td>
</tr>
<tr>
<td>MGE</td>
<td>mobile genetic element</td>
</tr>
<tr>
<td>RecA</td>
<td>Recombination A – nucleic acid and NTP binding domain</td>
</tr>
<tr>
<td>vFeS</td>
<td>vestigial iron-sulfur coordinating domain</td>
</tr>
<tr>
<td>P_i</td>
<td>inorganic phosphate</td>
</tr>
<tr>
<td>SF2</td>
<td>Super Family 2, refers to superfamily 2 helicases</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>PAGE</td>
<td>polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>NTP</td>
<td>nucleotide triphosphates</td>
</tr>
<tr>
<td>CD</td>
<td>circular dichroism</td>
</tr>
<tr>
<td>Cryo-EM</td>
<td>cryo-electron microscopy</td>
</tr>
</tbody>
</table>
Bacterial Immune Systems

Mobile genetic elements (MGE), such as bacteriophages and plasmids, rely upon prokaryotic hosts to replicate their genetic material (del Solar et al., 1998; Rampersad & Tennant, 2018). This reliance comes at a high cost for the microbes, often resulting in cell death (Suttle, 2007). The evolutionary pressure resulting from prokaryotic and MGE interactions has resulted in an arms race, with the development of microbial defense systems on the one hand and the generation of viral counter defense pathways that prevent microbial defense on the other (Stern & Sorek, 2010, Makarova et al., 2011; Fortier & Sekulovic, 2013; Pawluk et al., 2017). These microbial defenses or immune systems utilize protein machinery, nucleic acid, and chemical signaling systems to identify and defend against invasive genetic elements (Stern & Sorek, 2010; Hall et al., 2011; Bernheim & Sorek, 2019). For example, restriction-modification systems use restriction enzymes that bind and cleave specific nucleic acid sequences, and abortive infections utilize chemical signaling pathways which result in host cell death rather than the death of the bacterial colony.

New microbial immune systems and variants within those systems are being discovered yearly and are prime targets for new research projects. For example, toxin-antitoxin systems, bacteriophage exclusion (BREX) systems, CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repeats – CRISPR associated) systems, Wadjet, Thoeris, Septu, and others require further biochemical and structural characterization
The study of prokaryotic immune systems is vital, as basic research identifies and promotes the development of tools that can be repurposed for scientific and medical research and or diagnostics. For instance, the study of CRISPR-Cas systems led to the research and development of viral diagnostic tests used to detect the presence of SARS-CoV-2, a virus associated with a recent pandemic (Broughton et al., 2020; Rahimi et al., 2021).

Furthermore, research on CRISPR systems is leading to the development of treatments for HIV-1, sickle cell, cancer, and blindness (Ledford, 2020; Frangoul et al., 2021; Herskovitz et al., 2021; Katti et al., 2022).

This thesis will focus primarily on CRISPR-Cas and Wadjet systems, described in detail in chapters two and three. Characterization of these systems is performed using both in vivo and in vitro biochemical assays and structural analysis using cryo-electron microscopy and x-ray crystallography.

**CRISPR-Cas Prokaryotic Adaptive Immune Systems**

Bacteria and archaea utilize CRISPR-Cas systems to defend against MGE. CRISPR-Cas systems act as adaptive immune systems by generating and maintaining an immunological memory of infection (Figure 1-1) (Barrangou et al., 2007). The CRISPR operon is located on the host chromosome or a host plasmid and is composed of a CRISPR array and CRISPR-associated genes. CRISPR arrays are composed of direct repeat sequences, ~30 base pairs long, separated by spacer sequences, which are also roughly 30 base pairs long (Boilotin et al., 2005). The spacer sequences serve as the genetic memory of infection (Makarova et al., 2006).
Note. These systems defend against mobile genetic elements such as phage through adaptation, biogenesis, and interference stages.
CRISPR systems generate immunity in stages (Carter & Wiedenheft, 2015). The first stage, adaptation, occurs when the CRISPR adaptation complex, Cas1-Cas2, recognizes, processes, and integrates foreign nucleic acid sequences into the CRISPR array adjacent to the leader sequence. The result of integration is a new spacer sequence and thus, a new memory of infection (Yosef et al., 2012; Núñez et al., 2015; Xiao et al., 2017). Interestingly, integration of new spacer sequences occurs near the leading direct repeat sequence. This means a more recent infection can be identified from an older infection (Held et al., 2010; Yosef et al., 2012). This information allows for a more thorough understanding of contemporary threats.

CRISPR RNA (crRNA) biogenesis defines the second stage. The CRISPR operon is transcribed, resulting in Cas proteins and a long pre-crRNA sequence. The long pre-crRNA is processed into shorter guide RNA (gRNA) sequences by either Cas or host nucleases (Haurwitz et al., 2010; Charpentier et al., 2015). Next, Cas proteins assemble on the crRNA, comprised of a truncated direct repeat and a spacer sequence, to form a ribonucleoprotein complex (RNP) capable of surveilling the cell for complementary nucleic acid (Brouns et al., 2008; Charpentier et al., 2015).

After RNP complex assembly, the CRISPR-Cas system enters the interference stage of immunity. RNP complexes will probe the cell matrix for foreign nucleic acid targets by examining the complementarity of DNA or RNA sequences to the gRNA. This surveillance process is enhanced through complex recognition of a protospacer adjacent motif (PAM) on the foreign nucleic acid. PAM sequences are usually only 3-5 nucleotides in length and act as a self vs. non-self label (Marraffini & Sontheimer, 2010). If the RNP complex recognizes the PAM, complementarity between the gRNA and the
bound nucleic acid will be probed (Mojica et al., 2009; Shah et al., 2013; Sternberg et al.,
2014; Gleditzsch et al., 2018). If sufficient complementarity exists, targeting will occur via a cleavage event by either the RNP complex or a recruited nuclease. This cleavage event will destroy the bound target sequence, inhibiting a given mobile genetic element from propagating throughout a microbial community (Brouns et al., 2008; Marraffini & Sontheimer, 2008; Hille et al., 2018).

CRISPR-Cas systems utilize a diverse array of enzymes and differ in their crRNA processing, target substrate preference (DNA/RNA), and the mechanism of target substrate interference. The classification of systems is based on the structure and gene contents of the system. Two classes and six sub-types are used to classify and categorize these systems. Class 1 systems (types I, III, and IV) are composed of multi-subunit RNP complexes. In contrast, class 2 systems (types II, V, VI) are composed of single subunit RNP complexes (Figure 1-2). The six types of CRISPR-Cas systems are divided into ~50 subtypes, with new subtypes being identified yearly (Makarova et al., 2015; Makarova et al., 2019; Koonin & Makarova, 2022).

Class 1 CRISPR-Cas systems describe large multi-subunit RNP complexes. There are three class 1 CRISPR systems (I, III, and IV). Type I and type III systems are composed of a multi-protein complex called Cascade (CRISPR-associated complex for antiviral defense). These RNP complexes contain multiple Cas7 backbone proteins, a Cas6/Csf5 subunit, a large subunit Cas8/Cas10 and then Cas11, which act as small subunits (Jackson et al., 2014; McMahon et al., 2020). Cas6/Csf5 is used by type I and type III systems to process the long pre-crRNA into a shorter crRNA during biogenesis (Carte et al., 2010; Barrangou, 2013; Charpentier et al., 2015). Type I systems bind
**Figure 1-2**

*CRISPR-Cas Immune Systems are Diverse*

<table>
<thead>
<tr>
<th>Type</th>
<th>Binds</th>
<th>Cleaves</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>dsDNA</td>
<td>dsDNA</td>
</tr>
<tr>
<td>III</td>
<td>RNA</td>
<td>RNA, ssDNA</td>
</tr>
<tr>
<td>IV</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>II</td>
<td>dsDNA</td>
<td>dsDNA</td>
</tr>
<tr>
<td>V</td>
<td>dsDNA, RNA</td>
<td>DNA, RNA</td>
</tr>
<tr>
<td>VI</td>
<td>RNA</td>
<td>RNA</td>
</tr>
</tbody>
</table>

*Note.* Class 1 Systems form multiprotein complexes. The CRISPR Cascade complex from the Type I systems is represented in the diagram. Class 2 systems form a single protein complex. The CRISPR Cas9 complex is represented in the diagram.
double-stranded DNA (dsDNA) and recruit the cas3 helicase/nuclease to unwind and degrade dsDNA (Sinkunas et al., 2011; Mulepati & Bailey, 2013; Jackson et al., 2014; Mulepati et al., 2014; van Erp et al., 2015; Xiao et al., 2018). Type III systems bind RNA, while the cas10 subunit in these systems will bind and cleave ssDNA and RNA (Hale et al., 2009; Staals et al., 2013; Tamulaitis et al., 2014). Type IV systems contain CRISPR-cas subtype as in A. ferrooxidans genes (csf) csf1, csf2, csf3, and csf5. They are adjacent to type IV accessory genes (dinG, cysH, and cas10-like) and form RNP complexes (Taylor, 2021; Özcan et al., 2018; Zhou et al., 2021). These systems have been shown to disrupt plasmid transformation efficiencies and interfere with gene expression without DNA nuclease activity (Crowley et al., 2019; Guo et al., 2022). However, little is understood about the type IV systems and the mechanisms they use to interfere with MGE.

Class 2 CRISPR-Cas Systems describe single subunit RNP complexes. There are three class 2 CRISPR systems (II, V, and VI). Type II systems utilize Cas9-like proteins, recognize guanine-rich PAM sequences, and cleave DNA adjacent to the PAM using an HNH and RuvC-like domain, producing blunt double-stranded DNA breaks (Garneau et al., 2010; Gasiunas et al., 2012; Jinek et al., 2012). Type V systems utilize Cas12a-like proteins, recognize thymine-rich PAM sequences, and cleave distally to the PAM using a single RuvC-like domain resulting in staggered double-stranded DNA breaks (Zetsche et al., 2015). Type VI systems utilize Cas13-like enzymes, do not have a set PAM but may use a protospacer flanking sequence (PFS) for initial recognition and then bind and cleave single-stranded RNA (Abudayyeh et al., 2016).

In chapter 2 of this thesis, I discuss my work on the least understood CRISPR
type, the type IV-A CRISPR system and the CRISPR associated protein, CasDinG, and its role in CRISPR immunity. I show evidence that CasDinG acts as a 5’-3’ ATP-dependent helicase and present the first crystal structure of a type IV-associated helicase. Furthermore, mutational analysis of this protein highlights accessory domains that are essential for helicase, ATPase, and in vivo function. This work is significant as it provides key insight into the interference mechanism of type IV-A immune systems which have been presumed to utilize helicase activity, but this work is the first to prove CasDinG helicase function. Knowledge of how type IV-A systems utilize helicase activities could provide new insight into another aspect of how immune systems use helicases as means of thwarting mobile genetic element infections.

In chapter 3 of this thesis, I discuss the foundational work I performed on the class 2 single-subunit type V-A2 Cas12a2 proteins. I provide evidence showing that Cas12a2 processes its own crRNA guide through the use of lysine and arginine residues which were identified using a structure prediction model. I show preliminary evidence that Cas12a2 can bind and cleave nucleic acid. Furthermore, my work on these systems resulted in a low-resolution cryo-electron microscopy map from which some structural analyses could be made. This work was significant as it provided the first proof that Cas12a2 could process a crRNA for use as a guide RNA, fulfilling the second essential step for CRISPR immunity. Additionally, my work provided the foundation for collaborators and fellow lab members to identify the interference mechanism of Cas12a2. Overall, the research performed on Cas12a2 may lead to the development of a new CRISPR-Cas tool capable of acting as a diagnostic or as tool in gene therapeutics or cancer treatment (Dymetrenko et al., 2022). Some of my work on SuCas12a2 was
included in a manuscript currently under revision in the journal Nature, of which I am listed as a co-author.

**Wadjet Defense Systems**

Bioinformatic analysis of defense islands and known prokaryotic immune systems resulted in the identification of 28 phylogenetically distinct immune systems. One of these systems, Wadjet, was found to limit plasmid transformation efficiency in *B. subtilis* (Doron et al., 2018). How this system disrupted plasmid transformation efficiency is unknown. Furthermore, there is no other biochemical or structural data in the current literature which describes Wadjet systems.

Structure prediction software showed that Wadjet systems share structural similarities with the *E. coli* MukBEF chromosome condensin complex proteins (Doron et al., 2018). Condensin complexes bind, condense, and segregate nucleic acid as it is replicated in the cell during mitosis (Hirano, 2012). Knowledge of a condensin-like system, such as Wadjet, could provide powerful insight into how bacteria regulate horizontal gene transfer of plasmids or manage the replication of plasmids. Additionally, it is unknown whether Wadjet systems can target other nucleic acids, such as single-stranded phage as this was not tested in the initial *in vivo* studies involving Wadjet systems. Furthermore, *B. Subtilis* utilizes a plasmid transformation strategy in which the plasmid is integrated into the host as single-stranded DNA, thus whether Wadjet systems recognize ssDNA or dsDNA (plasmid) is unknown (Yadav et al., 2012).

In chapter 3 of this thesis, I discuss cloning of the Wadjet genes, *jetABC*, from two prokaryotic species, *Azotobacter vinelandii* DJ and *Mycobacterium* MCS, into
ligation-independent expression vectors and show evidence that AvJetB is recombinantly expressed and purified as a dimer. My work on these Wadjet systems provides the preliminary work necessary to study these systems in vitro and is significant as the dimerization of JetB supports the hypothesis that Wadjet systems share similarities with condensin systems which form dimeric complexes. Studies of condensins and condensin-like systems can provide powerful insight into nucleic acid segregation, without which cells die (Hirano, 2016).

Chapter 4 of this thesis discusses a summary of the Wadjet and type IV-A and type V-A2 CRISPR-Cas immune system projects. This summary includes a brief discussion of the biochemical research performed as well as future directions for each project.

Basic research on novel prokaryotic immune systems allows for the identification of tools used by microbes to manipulate nucleic acids and defend against infectious events. Determining biochemical parameters surrounding protein-protein and protein-nucleic acid interactions results in a foundation of knowledge that promotes genome editing research, medical research, and recombinant nucleic acid technology which can lead to treatments for genetic disease and repurposed biotechnological tools (Cho et al., 2013; Cong et al., 2013; Mali et al., 2013).

References
https://doi.org/10.1126/science.1138140


https://doi.org/10.1038/s41579-019-0278-2


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


Dmytrenko, O., Neumann, G., Hallmark, T., Keiser, D., Crowley, V., & Vialetto, E. et al. (2022). Cas12a2 elicits abortive infection via RNA-triggered destruction of double-stranded DNA. https://doi.org/10.1101/2022.06.13.495973


https://doi.org/10.1093/nar/gks173


CHAPTER 2

CASDING IS A 5’- 3’ ATP-DEPENDENT HELICASE

Abstract

*Pseudomonas aeruginosa* strain 83 encodes a type IV-A CRISPR-Cas system that relies on five genes (*csf1, csf2, csf3, cas6*, and *dinG*) and a CRISPR array to protect against invasive plasmids. Phylogenetic analysis shows that type IV-A CRISPR-associated DinG (*CasDinG*) is distinct from chromosomally associated DinG proteins which are helicases involved in recombination and repair. This distinction suggested that *CasDinG* may play a CRISPR-specific role in type IV-A immunity. This thesis contains biochemical and structural analyses essential for understanding *CasDinG* function. We provide the first structure of a type IV *CasDinG* helicase using x-ray crystallography. Fluorescence anisotropy techniques were used to determine binding affinities to nucleic acid substrates. Colorimetric ATPase assays monitored the hydrolysis of adenosine triphosphate in the presence of nucleic acids and helicase assays determined the preferred directionality of strand displacement. Furthermore, *in vitro* and *in vivo* mutational analysis of domains and motifs provided functional insight into the role of *CasDinG* as a CRISPR-associated helicase.

Introduction

Type IV CRISPR-Cas Systems

Type IV biochemical mechanisms and structures remain largely uncharacterized,
as large complexes can be challenging to study via recombinant expression and purification techniques. Work in recent years has shown that type IV systems form multi-subunit RNP complexes, process guide RNAs, and can reduce the transformation efficiencies of target plasmids (Özcan et al., 2018; Crowley et al., 2019; Taylor et al., 2019; Zhou et al., 2021). Type IV systems are found almost exclusively on plasmids and are not typically associated with adaptation machinery or identifiable nuclease (Makarova et al., 2017, 2019; Pinilla-Redondo et al., 2019). The mechanisms behind how these systems acquire new spacers, enact interference, and whether they recruit or assemble with associated proteins is unknown.

There are three subtypes of type IV CRISPR-Cas systems (A, B, & C) (Makarova et al., 2020; Taylor et al., 2021). Type IV-A systems lie adjacent to dinG-like genes and are commonly adjacent to CRISPR arrays. Type IV-B systems lie adjacent to cysH-like genes and are not found adjacent to CRISPR arrays and type IV-C systems lie adjacent to cas10-like genes and may or may not lie adjacent to CRISPR arrays (Taylor et al., 2021). Many type IV systems are found on plasmids and contain spacer sequences that target plasmids, suggesting that these systems may play a role in inter-plasmid warfare (Pinilla-Redondo et al., 2019). Because some type IV systems lack the adaptation genes cas1 and cas2, it is theorized that these systems rely upon other CRISPR-Cas systems found in their hosts, such as the type I-E system, to generate new spacer sequences (Pinilla-Redondo et al., 2019). Due to the lack of identifiable nuclease genes close to or within type IV-A and type IV-B systems, there is speculation on how interference occurs. Perhaps endogenous host nucleases are recruited, there are unidentified nuclease motifs, or interference occurs on a transcriptional level.
Out of the type IV systems, the most studied are the type IV-A systems which have been shown to process a long pre-crRNA into shorter guide RNA sequences and inhibit plasmid transformation recombinantly (Crowley et al., 2019; Taylor et al., 2019). It was recently reported that there is inhibition of gene expression without the cleavage of DNA. This indicates that type IV-A systems may utilize gene repression or transcription interference techniques to combat phage or plasmid (Guo et al., 2022). Whether the type IV-A systems house unidentified nuclease motifs or recruit nucleases to cleave sequences bound by the RNP has yet to be discovered. Furthermore, there are no solved structures of the type IV-A complex or the associated CasDinG protein.

*Pseudomonas aeruginosa* type IV-A CRISPR-Cas

The type IV-A CRISPR-Cas system in *Pseudomonas Aeruginosa* strain 83 is located on a plasmid and is composed of four cas/csf genes (*csf1, csf2, csf3, cas6*), the *dinG*-like gene (*casdinG*), and a CRISPR array (Figure 2-1). Hannah Taylor, a graduate student of the Jackson lab, has shown via recombinant expression and purification that the type IV-A RNP complex is composed of *csf1, csf2, csf3*, and *cas6* in a predominantly 1:2:1:1 ratio. This ratio was confirmed by mass spectrometry; however, alternative ratios were also identified (Taylor et al., 2021). Initial *in vivo* characterization was performed by Dr. Val Crowley using both plasmid competition and plasmid curing assays. Omission of any type IV-A RNP genes leads to a non-functional CRISPR immune system incapable of targeting complementary nucleic acid. Omission of the *casdinG* gene or mutation of the Walker B DEAH motif also leads to a non-functional immune system, showing that the binding or hydrolysis of ATP is essential for CasDinG function.
(Crowley et al., 2019). Herein lies a fundamental question which is, what is the role of CasDinG in type IV-A immunity and what does ATP have to do with it?

**CasDinG and Non CasDinG Helicases**

Helicases are enzymes that remodel nucleic acid through the hydrolysis of nucleotide triphosphates (NTPs) and are categorized based on shared sequence motifs. There are six superfamilies of helicases, superfamilies I and II are typically monomeric, families III-VI are typically hexameric and helicases from all families contain RecA-like folds which house motifs important for NTP and nucleic acid binding (Pyle, 2008). Superfamily II (SF2) proteins, house subfamilies of helicases which contain superfamily II (SF2) helicase motifs: Q, I, Ia, Ib, Ic, II, III, IV, V, VI (Singleton et al., 2007; Pyle, 2008; Umate et al., 2011). These motifs are essential for ATP binding/hydrolysis, nucleic acid binding/unwinding, and are typically located at the interface of the RecA1 and RecA2 lobes (Tuteja & Tuteja, 2004; Pyle, 2008). In many cases, the hydrolysis of ATP provides the chemical energy needed to generate the mechanical energy used to drive protein conformational changes. These protein conformational changes result in the remodeling of nucleic acid (Ye et al., 2004). These protein conformational changes can result in translocation along single-stranded or duplexed nucleic acids with or without unwinding. However, helicases can also remodel nucleic acid without translocation by generating sufficient distortion in a nucleic acid strand to displace duplexes (Pyle, 2008). Not all proteins categorized within helicase families are active helicases but have adopted alternative roles requiring DNA binding, such as nuclease activity (McRobbie et al., 2012). Characterization of helicases is typically performed by determining ATP
hydrolysis rates, step size, binding requirements, whether accessory proteins are used or needed, and whether there are directionality or polarity restraints (Pyle, 2008).

Helicases associated with CRISPR-Cas systems have been limited to the type I systems that utilize the helicase/nuclease Cas3 during adaptation and interference (Sinkunas et al., 2011). During interference, Cas3 is recruited to target bound Cascade complex where it will nick the R-loop formed between the RNP complex and duplexed nucleic acid and then degrade the DNA as it unwinds in a 3’-5’ direction (Sinkunas et al., 2011; Westra et al., 2012; Mulepati & Bailey, 2013). The degradation of DNA by Cas3 will then enhance primed adaptation allowing for increased uptake of new spacer sequences (Datsenko et al., 2012). Phylogenetic analysis of Cas3 proteins from archaea and bacteria shows conservation of SF2 family helicase motifs and inclusion of a non-SF2 conserved asparagine in motif IV (Jackson et al., 2014).

DinG stands for damage-inducible gene and was identified in connection with DNA-damage response pathways. *DinG* genes belong to the XPD family of SF2 helicases (Lewis & Mount, 1992; Koonin, 1993). XPD is a 5’-3’ ATP-dependent eukaryotic helicase component in the transcription factor IIH complex and relies on an iron-sulfur (FeS) cluster for helicase activity (Liu et al., 2008). FeS clusters found within helicases, nucleases, and primases are essential for a variety of purposes, including structural stability, DNA damage sensing, DNA binding, helicase activity, exonuclease activity, and the formation of active protein complexes (Ren et al., 2009; White, 2009; Yeeles et al., 2009; Netz et al., 2011; Stiban et al., 2014; Baranovskiy et al., 2018; Mariotti et al., 2020). Characterization of XPD, Rad3, and multiple DinG homologs show a dependence on iron-sulfur clusters for helicase activity (Rudolf et al., 2006; Pugh et al.,
2008; Ren et al., 2009; Thakur et al., 2014).

For the purposes of this thesis, there are two classes of DinG enzyme: CRISPR-associated DinG (CasDinG) and non-CRISPR-associated DinG (non-CasDinG). Bioinformatic analysis has shown that CasDinG are phylogenetically distinct from chromosomally associated DinG helicases or non-CasDinG (Taylor et al., 2021). Meaning, CasDinG amino acid sequences are more similar to one another than they are to non-CasDinG DinG sequences, suggesting that CasDinG proteins may play a role unique to CRISPR-Cas immune system function.

To date there are two types of non-CasDinG enzymes that have been characterized. The first is the FeS domain-containing species, the most well studied from *Escherichia coli*. *E. coli* DinG is a 5’-3’ ATP-dependent helicase implicated in DNA repair and replication pathways (Voloshin et al., 2003; Voloshin & Camerini-Otero, 2007). *E. coli* DinG contains a Walker A and B motif housed between the RecA1 and RecA2 domains. The structurally essential FeS domain is positioned between the Walker A and B motifs, from N to C termini, as an accessory to the RecA1 domain, similar to the XPD helicase (Cheng & Wigley, 2018). The second type of non-CasDinG is the nuclease domain-containing DinG from *Staphylococcus aureus*. *S. aureus* DinG is a 3’-5’ ATP tunable exonuclease and contains a large N-terminal exonuclease domain followed by a RecA body containing Walker A and B motifs. This enzyme does not contain a FeS cluster, or FeS containing domain and does not possess identifiable helicase activity (McRobbie et al., 2012).

Comparison of the amino acid sequences of CasDinG to Non-CasDinG shows no conservation of an exonuclease domain and no conservation of the cysteine residues
necessary for coordination of an FeS cluster but does show conservation of the Walker A and Walker B motifs (Figure 2-1; A-9; A-10). The lack of cysteine residues and structural knowledge of fold similarities led to the renaming of FeS domain in CasDinG to the vestigial FeS domain. A closer examination of Pa83 CasDinG amino acid sequences to the S. aureus and E. coli non-Cas DinG homologs shows low identity and similarity scores using EMBOSS needle pair-wise alignments. The highest alignment scores are 18.9% identity and 30.3% similarity with E. coli DinG (Figure A-9). These low similarity and identity scores are common among SF2 helicases as these types of enzymes are incredibly diverse in function. Typically, the similarity and identity scores reflect the maintenance of highly conserved helicase motifs, whereas the regions between motifs are variable, allowing for diversity in protein function (Fairman-Williams et al., 2010). With this knowledge, the lack of a typical FeS domain and an N-terminal exonuclease domain alongside the knowledge that CasDinG is a necessary component of type IV-A targeting suggest that CasDinG may perform a function specific to CRISPR immunity (Crowley et al., 2019). Furthermore, Alphafold and Robetta structure predictions of CasDinG show a model with an N-terminal fold and a vestigial FeS domain (vFeS) that lacks the cysteines needed to coordinate Fe atoms but maintains an ordered alpha-helical conformation (Figure A-7) (Baek et al., 2021; Mirdita et al., 2022).

This thesis contains both in vitro and in vivo biochemical studies assessing the biological role of CasDinG. Knowing that CasDinG contains nucleic acid and ATP binding motifs. Nucleic acid binding assays, ATPase assays, helicase assays, and plasmid curing assays were performed to elucidate functional roles of CasDinG. To understand the function of CasDinG, the molecular structure was solved using x-ray crystallography,
Figure 2-1

*CasDinG is Distinct from Non-CasDinG*

**A.** Schematic comparing CasDinG domain organization from *P. aeruginosa* 83 to Non-CasDinG from both *E. coli* and *S. aureus*. **B.** Operon schematic of the type IV-A CRISPR-Cas system genes, composed of a CRISPR array, *csf5, csf1, csf2, csf3*, and *casdinG*.

*Note.* A. Schematic comparing CasDinG domain organization from *P. aeruginosa* 83 to Non-CasDinG from both *E. coli* and *S. aureus*. B. Operon schematic of the type IV-A CRISPR-Cas system genes, composed of a CRISPR array, *csf5, csf1, csf2, csf3*, and *casdinG*. 
allowing us to gather structural insight into how a non-FeS coordinating enzyme can unwind nucleic acids.

**Methods**

**Cloning & Gene Alignments**

As described in a previous publication, the *dinG* gene from *P. aeruginosa* 83 was synthesized by TWIST bioscience and cloned into a pET StrepII TEV ligation independent cloning (LIC) vector (2R-T) vector (Crowley et al., 2019). Mutants were generated using the Q5 Mutagenesis kit (NEB Biolabs). Decisions on where to mutate domains were based on a predicted model of the CasDinG structure generated by Robetta prediction software. It should be noted that mutations were first generated as a broad means of identifying regions of activity (Figure A-7). For a list of constructs generated, see Table A-3; for primers, see Table A-5.

Clustal Omega Alignments and EMBOSS NEEDLE pair-wise alignments *E. coli* (P27296) and *S. aureus* (Q2FGY5) DinG sequences were obtained through the Uniprot database server and aligned to *Pa*83 CasDinG using the Clustal Omega multiple sequence alignment tool (McWilliam et al., 2013). Pairwise sequence alignments were performed between *Pa*83 CasDinG and *E. coli/S. aureus* sequences using an EMBOSS Needle alignment (Madeira et al., 2022).

**Expression & Purification of CasDinG**

The cloned vector containing the *dinG* gene from *Pseudomonas aeruginosa* 83 NCBI ref: WP_088922490.1 was transformed into *Escherichia coli* BL21 HMS174
(DE3) chemically competent cells (Novagen). This cell line provides a RecA mutation in
an E. coli K-12 background, supposedly this strain can genes whose protein products may
cause loss of the DE3 prophage. From the transformation plate, a colony was picked and
placed into an overnight outgrowth. 1 L of Luria-Bertani (LB) medium in a 2.8 L flask
was inoculated with 1 mL of both 1000 x metals mix (0.1 M FeCl$_3$-6H$_2$O, 1 M CaCl$_2$, 1
M MnCl$_2$-4H$_2$O, 1 M ZnSO$_4$-7H$_2$O, 0.2M CoCl$_2$-6H$_2$O, 0.1 M CuCl$_2$-2H$_2$O, 0.2M NiCl$_2$
6H$_2$O, 0.1 M Na$_2$MoO$_4$-2H$_2$O, 0.1 M Na$_2$SeO$_3$-5H$_2$O, 0.1 M H$_3$BO$_3$) and 1000 x 1 M
MgSO$_4$ (Studier, 2015). The supplemented medium was then inoculated with the
overnight starter, and cells were grown to an optical density between 1.0 - 1.3 OD$_{600}$ at
37 °C. When OD was reached, the cells were induced with 0.5 mM IPTG (isopropyl B-
D-1-thiogalactopyranoside), and the temperature was dropped to 20 °C for 5 hours. The
cells were then harvested via high-speed centrifugation and stored at -80 °C. All mutants
were expressed in the same manner.

Cells were homogenized on ice with lysis buffer (100 mM Tris Base pH 8.0, 150
mM NaCl, 1 mM TCEP (tris(2-carboxyethyl) phosphine)). The protease inhibitors
Aprotinin 1000 x (0.5 mg/mL), Leupeptin 1000 x (0.5 mg/mL), Pepstatin A 1000 x (0.7
mg/mL), & PMSF 150 x (25 mg/mL) were added prior to cell lysis. Probe sonication for
cell lysis was performed at settings of 4/60. The lysate was clarified by high-speed
centrifugation at 16 K RPM for 30 minutes.

All purification steps were performed at 4 °C. The supernatant was loaded onto
strep resin (Strep-Tactin®XT 4Flow®, IBA). The resin was washed with lysis buffer and
then eluted with elution buffer (100 mM Tris Base, 150 mM NaCl, 50 mM Biotin, 1 mM
TCEP, pH 8.0). Fractions with DinG were pooled. Protein was loaded onto a desalting
column (HiPrep 26/10 Desalting, GE Healthcare) that had been pre-equilibrated with low
salt buffer (100 mM Tris Base pH 8.0, 10 mM NaCl, 1 mM TCEP), and then DinG
fractions were run over the column. Protein elution was collected and then run over a
heparin column (HiTrap Heparin HP, GE Healthcare). The column was washed with 47.5
mM NaCl buffer. The protein was eluted into a high salt buffer (100 mM Tris Base pH
8.0, 500 mM NaCl, 1 mM TCEP). Samples were spun down using a spin concentrator
(Corning® Spin-X® UF 50 MWCO) before placement on a size exclusion column
(HiLoad 26/600 Superdex 200 pg., GE Healthcare). Protein was eluted off of the size
exclusion column into the high salt buffer (100 mM Tris Base pH 8.0, 500 mM NaCl, 1
mM TCEP). Protein samples were concentrated and stored at 4 °C, as freezing at -80 °C
resulted in loss of activity. All mutants were purified in the same manner. Protein was
assessed for purification after each step via 12% SDS-PAGE.

Protein concentration was determined by UV-Vis spectroscopy (Thermofisher
UV-vis Nanodrop), using the Beer-lambert law to correct absorbance values for
extinction coefficient (assuming all cysteines are reduced) and molecular weight as
determined by Expasy Protparam (Table A-6).

**Western Blots**

Western blots were performed using the Pierce Fast Western Blot kit from
ThermoFisher Scientific (catalog #35055). The standard protocol was altered by
increasing the time the primary antibody working dilution was incubated with the blot.
The length of time was increased from 30 minutes at room temperature to 60 minutes at
room temperature. The Precision Plus Protein™ WesternC™ Blotting Standards (Biorad,
#1610376) were also utilized. The Strep-Tactin conjugated HRP antibodies were sufficient for both the ladder and the strep-tagged protein samples.

**Crystallization and Structure determination**

The purified recombinant strep-tagged CasDinG protein was concentrated at 5 mg/ml and crystallized using 0.225 M Imidazole pH 8.0, 3.5 % PEG8000, and 30 % sucrose using hanging-drop vapor diffusion at room temperature. The crystal used for structure determination was comprised of 1 µL (5mg/ml) protein solution to 2.6 µL mother liquor and 0.4 µL 30 % sucrose. The crystal was then soaked in a cryoprotecting solution composed of 30 % ethylene glycol and mother liquor, then mounted on a loop and cooled to 100 K. Diffraction data were collected at the SSRL beamline 9-2. The data were indexed, integrated, and scaled using HKL3000. A data set to 2.95 Å was recorded with the space group determined as P65. Alphafold structure predictions were generated using the google colab Alpha 2.0 servers to solve the phases through molecular replacement (Mirdita et al., 2022). The motor domains of these structure predictions were then used as the starting models in Phenix for phasing, for a more in-depth explanation of molecular replacement using an Alphafold model see Appendix A, Supplemental Methodology. Model building was performed in COOT, the structures were refined using PHENIX, and validation was performed using Molprobity within PHENIX and the PDB deposition servers (Emsley et al., 2010; Liebschner et al., 2019).

**Nucleic Acid Substrate Preparation**

Nucleic acids were synthesized by Integrated DNA Technologies (IDT). Nucleic acids were labeled with a fluorescein (FAM) label on the 5’ and 3’ end by IDT. To make
duplexed nucleic acids, complementary oligonucleotides were mixed in an equimolar ratio in the presence of NEB buffer 2.1 and heated to 95°C. These oligonucleotides were slowly cooled to room temperature before being run on 12-15% native PAGE gels. Duplex bands were then gel extracted, ethanol precipitated, and reconstituted in water.

**Nucleic acid-binding assays**

Nucleic acid-binding activities of strep-tagged CasDinG were monitored using a fluorescence polarization-based assay. Anisotropy data were collected using a BioTek Synergy H4 Hybrid Multi-Mode Microplate Reader equipped with polarizers and bandpass filters. The polarizers and bandpass filters provided 485 ±20 nm excitation and detection of fluorescence emission at 528 ± 20 nm. Each reaction (80 µL) contained a limiting concentration (10 nM) of 5’ FAM-labeled nucleic acid substrate. CasDinG and 5’ FAM nucleic acid were assayed at room temperature with increasing concentrations of CasDinG (0 – 2.5 µM) in a binding buffer (100 mM Tris pH 8.0, 1 mM TCEP, and 5 mM MgCl₂). Change in anisotropy relative to FAM-nucleic acid was plotted as a function of CasDinG concentration. The apparent dissociation constant (Kₐ) for each nucleic acid substrate was determined by fitting the raw data to a single site saturation binding model in GraphPad Prism for Windows version 9.3.0. For a list of oligonucleotides used in these assays, see supplemental Figure A-4 and Table A-4.

**Malachite Green ATPase Assays**

Concentrations of inorganic phosphate (Pᵢ) were determined with a Malachite Green Phosphate Assay kit (BioAssay Systems, Haward, CA, USA). Activated Malachite Green reagent was added to wells of a 384 well plate (Corning Assay Plate, 384 wells,
Black with clear bottom, non-binding surface, Low flange, no lid, polystyrene, 3766).

Before the reaction, assay components were incubated at 37 °C for 15 minutes. Reactions were started with the addition of ATP run at 37 °C. The reaction was quenched in the activated Malachite Green reagent at designated time points, between 30 - 150 seconds for nucleic acid containing reactions and 0 - 4 minutes for basal reactions. The quenched reactions were developed for 30 minutes before sample measurement. The absorbance values of the samples were obtained using a Synergy H4 Hybrid Multi-Mode Microplate Reader measuring absorbance at 620 nm.

The initial velocities of CasDinG nucleotide triphosphate hydrolysis were determined in the presence of ssDNA (40 nt.) with 10 nM CasDinG, 100 nM nucleic acid, and 600 µM nucleotide triphosphate (NEB) in the buffer (50 mM Tris pH 7.5, 0.1 mg/ml Recombinant Albumin, 1 mM MgCl₂, 0.4 mM TCEP). Substrate comparison initial velocities utilized 10 nM CasDinG, 100 nM nucleic acid, and 600 µM ATP, which was reconstituted in lab and spectroscopically verified, in the buffer (50 mM Tris pH 7.5, 0.1 mg/ml Recombinant Albumin, 1 mM MgCl₂, 1 mM TCEP). Structural mutants were assayed for ATPase activity in the presence of single stranded DNA (40 nt.) using 20 nM CasDinG and 200 nM ssDNA (40 nt.). Michaelis-Menten curves were generated for CasDinG (20 nM) + single stranded phosphorothioated RNA (200 nM) pH 7.5, CasDinG (10 nM) + single stranded DNA (100 nM) pH 7.5, CasDinG (18.5 nM) + single stranded DNA (150 nM) pH 6.8 and ApoCasDinG pH 6.8 (500 nM). Michaelis-Menten reactions at pH 7.5 utilized the buffer (50 mM Tris pH 7.5, 0.1 mg/ml Recombinant Albumin, 1 mM MgCl₂, 1 mM TCEP) whereas reactions at pH 6.8 used the buffer (20 mM PIPES pH 6.8, 0.1 mg/ml bovine serum albumin (BSA), 50 mM NaCl, 1 mM TCEP, 5 mM MgCl₂).
For a list of oligonucleotides used in these assays, see supplemental Figure A-4 and Table A-4. GraphPad Prism for windows version 9.3.0 fits the data set to the Michaelis-Menten equation (Equation 1). Where v is the initial reaction velocity of the reaction, [S] is the ATP concentration, K_M is the Michaelis constant, and V_max is the maximum velocity of the enzyme.

Equation 1
\[ v = \frac{V_{\text{max}} [S]}{K_M + [S]} \]

ATP concentrations were spectroscopically verified at 280 nm with crystal cuvettes and a spectrophotometer. Concentrations were confirmed using the Beer-Lambert law (Equation 2). Where A is the absorption, E is the extinction coefficient, l is the path length, and c is the concentration. The molar extinction coefficient for ATP was 15,400 M^{-1}cm^{-1}. ATP was frozen at -80°C, and aliquots were used and then discarded after a single thaw.

Equation 2
\[ A = \varepsilon lc \]

**Preparation of chemically competent type IV-A CRISPR Cas cells**

Transformations were prepared using an open flame and standard aseptic technique. BL21 HMS174 (DE3) E. coli cells (Novagen), thawed on ice and 50-100 ng of both plasmid #1284(pCDF_Pa_csf1_csf2_cas6) and plasmid #1290(pACYC-PaCR83-csf3-dinG), 2367 (PACYC-PA83CR-csf3-dinG-NTerm-truncation), 2368 (PACYC-PA83CR-csf3-dinG-ArchKO), 2369 (PACYC-PA83CR-csf3-dinG-FeS-KO) or 2370 (PACYC-PA83CR-csf3-dinG-DEAH_AAAH-KO), were added to the cells. Cells were
then heat shocked in a 42 °C water bath for 45 seconds, followed by a cold shock on ice for 5 minutes. 450 μL of antibiotic-free LB broth was added to the cells, followed by a 1 hour incubation at 37 °C in a shaking incubator, 200 RPM. Cells were then plated on LB agar containing streptomycin and chloramphenicol at 50 and 25 μg/mL concentrations, respectively. Plates were then incubated at 37 °C for 16-24 hrs. LB broth containing streptomycin and chloramphenicol (50 and 25 μg/mL, respectively) was incubated at 37 °C for 16-24 hrs. in a shaking incubator, 200 RPM in preparation for the next day.

Colonies from the overnight plates were used to inoculate 25 mL of the prepared LB broth in a 50 mL falcon tube. These cells were then incubated at 37 °C until an OD$_{600}$ between 0.2-0.3 when they were induced with 0.1 mM IPTG (100 μM IPTG). Cells were allowed to grow for an additional 45 minutes at 37 °C before being cold-shocked on ice for 20 minutes. Cells were then spun down at 2700 x G for 15 minutes at 4 °C. The supernatant was decanted, and the cells were resuspended in 12 mL RF1 Buffer (100 mM RbCl, 50 mM MnCl$_2$4H$_2$O, 30 mM Potassium acetate, 15% m/v glycerol). Cells were then allowed to rest on ice for 15 minutes before being spun down at 870 x G for 15 minutes. The supernatant was decanted, and cells were resuspended in 1 mL of RF2 Buffer (10 mM MOPS, 10 mM RbCl, 75 mM CaCl$_2$2H$_2$O 15% m/v glycerol). Cell solution was incubated on ice for 15 minutes prior to aliquoting cells in 100 μL volumes and flash freezing at -80 °C.

**In Vivo Plasmid Competition Assay**

Using type IV-A CRISPR-Cas containing chemically competent cells, 10 ng of target or non-target plasmid, #2380 (pET27b_CA01_GGAAA) and 1095 (pET27b(+)-
non-target_TTTC) respectively, was added to a 100 μL cell aliquot. Cells were then heat shocked at 42 °C for 30-40 seconds, followed by a cold shock on ice for 1-3 minutes. 400 μL of LB containing chloramphenicol 25 μg/mL, streptomycin 50 μg/mL, and 0.1 mM IPTG were then added to the cold shocked cells. Cells were then incubated at 37 °C for 16-24 hrs. in a shaking incubator, 200 RPM, followed by plating of cells onto a triple antibiotic LB agar selection plate (Chloramphenicol 25 μg/mL, streptomycin and kanamycin 50 μg/mL, and IPTG 0.1 mM IPTG). The cells were then plated in rows of 6 with 10 μL volumes. A dilution series was performed and plated from left to right in dilutions of 1, 10, 100, 1000, 10,000, and 100,000-fold. The inoculating drops were allowed to dry and then placed in a 37 °C incubator for 24 hrs. Plates were imaged using a Bio-Rad Imager with a non-stained blot setting. Colonies were then counted manually.

**Helicase assays**

**Direct Substrate and Mutant Comparisons**

15 nM 5' fluorescein (FAM) labeled nucleic acid was incubated in the presence of 100 nM WT CasDinG and 1 mM ATP in the helicase buffer (25 mM Tris pH 7.5, 1 mM MgCl₂, 1 mM TCEP, 0.1 mg/ml recombinant albumin) for approximately 20 minutes at 37 °C before being quenched in 2 x STOP buffer (10 mM EDTA (Ethylenediaminetetraacetic acid), 1% SDS (Sodium Dodecyl Sulfate), 20% glycerol) with 500 nM unlabeled displaced strand. Samples were run on 15% TBE native PAGE gels. The mutant analysis utilized the same nucleic acid substrate as WT CasDinG and an equivalent amount of protein.
**Direct NTP Comparisons**

15 nM 5' FAM-labeled nucleic acid was incubated in the presence of 100 nM WT CasDinG and 1 mM ATP analogue (ATP, ADP, ATPγS, and AMP-PNP) in the helicase buffer (25 mM Tris pH 7.5, 1 mM MgCl₂, 1 mM TCEP, 0.1 mg/ml recombinant albumin) for approximately 20 minutes at 37 °C before being quenched in 2 x STOP buffer with 500 nM unlabeled displaced strand. Samples were run on 15% TBE native PAGE gels.

**Metal Comparisons**

15 nM 5' FAM labeled nucleic acid was incubated in the presence of 100 nM WT CasDinG and 1 mM ATP and 1 mM divalent salt (MgCl₂, MnCl₂, ZnCl₂, CaCl₂, NiCl₂, CoCl₂, CuCl₂) in the helicase buffer (25 mM Tris pH 7.5, 1 mM TCEP, 0.5 mg/ml BSA) for approximately 20 minutes at 37 °C before being quenched in 2 x STOP buffer with 500 nM unlabeled displaced strand. Samples were run on 15% TBE native PAGE gels.

**Time Courses**

15 nM FAM-labeled nucleic acid substrate was exposed to 25 nM CasDinG over the course of 10 minutes in the presence of 1 mM MgCl₂ and 1 mM ATP in the helicase buffer (25 mM Tris pH 7.5, 0.1 mg/mL recombinant albumin, 1 mM TCEP) at 37 °C. Samples were quenched in 2 x STOP buffer at times between 0-10 minutes and run on 15% TBE native PAGE gels. For a list of oligonucleotides used in these assays, see supplemental Figure A-4 and Table A-4.

**Gel Analysis**
All PAGE gels for helicase assays were imaged using a BioRad Imaging system and analyzed using BioRad ImageLab software. Percent unwound were quantified using ImageLab software, and reported data is the average of three experiments, with error bars representing the standard deviation from the mean. Graphs were made in GraphPad Prism for Windows version 9.3.0.

**Type IV-A Complex (+/- CasDinG) Expression & Purification**

The cloned vectors containing the type IV-A complex (csf1, csf2, csf3, cas6, and the CRISPR array) and casdinG (No tag sequence) were transformed into *Escherichia coli* BL21 HMS174 (DE3) chemically competent cells (Novagen). The NCBI reference codes are WP_088922490.1 (CasDinG protein sequence) and NZ_CP017294.1 (Pa83 plasmid sequence) To obtain the genes for the type IV-A complex refer to the plasmid sequence. This protocol can be used either for just RNP complex purification or RNP + CasDinG purification. From the transformation plate, a colony was picked and placed into a 25 mL overnight outgrowth. 1 L of LB medium in a 2.8 L flask was supplemented with 1000 x Metals Mix and 1000 x MgSO₄ (Studier, 2015). The media was then inoculated with a 20 mL overnight starter. Cells were grown to an optical density between 1.0-1.3 OD₆₀₀ at 37 °C 200 RPM in a shaking incubator. When OD₆₀₀ was reached, the cells were induced with 0.5 mM IPTG, and the temperature was dropped to 20 °C for 5 hours. The cells were then harvested via high-speed centrifugation and stored at -80 °C.

Cells were homogenized on ice with 35 mL lysis buffer (100 mM Tris base pH 8.0, 150 mM NaCl, 1 mM TCEP). The protease inhibitors 1000 x Aprotinin (0.5mg/mL),
1000 x Leupeptin (0.5 mg/mL), and 1000 x Pepstatin A (0.7 mg/mL) were added prior to cell lysis. Probe sonication for cell lysis was performed at settings of 4/60 for eight minutes. The lysate was then clarified by high-speed centrifugation at 16 K RPM for 35 minutes.

All purification steps were performed at 4 °C. The supernatant was loaded onto a nickel column (HisTrap HP, GE Healthcare) that had been pre-equilibrated with lysis buffer. The column was washed with wash buffer (100 mM Tris pH 8.0, 500 mM NaCl, 150 mM Imidazole, 1 mM TCEP) and then with lysis buffer. The protein sample was then eluted with Ni elution buffer (100 mM Tris pH 8.0, 150 mM NaCl, 1 M Imidazole, 1 mM TCEP). Fractions with complex were pooled. The sample was then run over strep resin (Strep-Tactin XT 4 Flow, IBA). The column was washed with lysis buffer followed by protein elution using the strep elution buffer (100 mM Tris base pH 8.0, 150 mM NaCl, 50 mM Biotin, 1 mM TCEP), which was allowed to incubate with the resin for 15 minutes prior to sample collection. Eluted samples containing protein were spun down using a spin concentrator (Corning® Spin-X® UF 100 MWCO) for placement on a size exclusion column (Superdex 200 10/300 GL, GE Healthcare). Protein was eluted into a final buffer (100 mM Tris pH 8.0, 150 mM NaCl, 1 mM TCEP) using the sizing column. Protein samples were then concentrated and stored at 4 °C for use or flash frozen at -80 °C.

Protein concentration was determined by UV-Vis spectroscopy (Thermofisher UV-vis Nanodrop), and the appropriate extinction coefficient was applied as determined by Expasy Protparam, assuming a 1:2:1:1, Csf1:Csf2: Csf3:Csf5/Cas6 ratio of purified type IV-A RNP complex (Table A-3).
Far-UV Circular Dichroism (CD)

Far-UV CD spectra were collected for strep-tagged WT CasDinG, ΔDEAH_AAAH, and ΔN. terminal truncation in a low salt buffer (100 mM Tris pH 8.0, 10 mM NaCl, 1 mM TCEP). All data were collected using a JASCO model J-1500 spectropolarimeter. CD spectra were collected from 190 to 260 nm at 10°C using 0.1 cm quartz cuvettes, 0.3-1 nm data sampling, a 50 nm/min scan rate, and a 1 second data integration time. Measurements were performed at 0.5 mg/mL protein concentration, and CD signals were normalized by converting to mean residue molar ellipticity. WT and ΔDEAH_AAAH were collected in triplicate with the average values plotted. ΔN-terminal truncation was collected once.

Plasmid Cleavage Assay

Target (GGAAA PAM sequence) and non-target plasmid were held at 120 ng concentration in a 10 µL reaction containing 5.37 µM type IV-A RNP complex, 100 nM DinG (WT), and 1 mM ATP in a reaction buffer (25 mM Tris pH 7.5, 1 mM MgCl₂, 1 mM TCEP, 0.1 mg/ml recombinant albumin) for 50 minutes at 37 °C before being quenched in a 2 x STOP buffer. Samples were run on a 1 % agarose gel.

Results

Protein Purification

To assess the function of CasDinG, we recombinantly expressed the Pa83 CRISPR associated dinG gene (NCBI ref: WP_088922490.1) using E. coli HMS174 (DE3) cells and confirmed the expression of the strep tagged 81.1 kDa protein using
Figure 2-2

Purification of CasDinG

*Note.* A. Western blot showing expressed strep-tagged CasDinG from cell pellets using strep-tactin antibodies. B. Size exclusion chromatogram showing strep-tagged WT CasDinG (81.1 kDa) with UV absorption at 280 and 253 nM respectively off a Superdex 200 pg. 26/600 column. A representative 12 % SDS PAGE gel, stained with Coomassie blue, showing purified protein is below the curve.
western blots. Analysis of size exclusion chromatograms were consistent with the elution of a monomeric CasDinG species (Figure 2-2).

**Crystallization and structure determination**

To truly understand the mechanism of CasDinG, the first crystal structure of CasDinG was solved at a resolution of 2.95Å (Table 2-1; Figure 2-3). The phases of the x-ray crystal data were solved by using an Alphafold structure prediction model for molecular replacement. The solved model shows a core composed of a RecA-like helicase core and two accessory domains.

Electron density became apparent at amino acid position 104, indicating that the N-terminus is flexible. Alphafold structure predictions of CasDinG showed variability in the positioning of the N-terminus, supporting the hypothesis that the N-terminus is flexible. Two of the accessory domains, the arch and vestigial FeS (vFeS), were resolved and are located as inserts into the RecA1 lobe. There are 11 helicase motifs that can be identified in the RecA body, which allow us to infer from the structure that this is a nucleic acid and NTP binding enzyme, similar to other DinG proteins (Figure A-8).

Alignment of the CasDinG model with the solved structure of *E. coli* DinG bound to single-stranded DNA, PDB 6FWR, shows conservation of a nucleic acid binding pocket within the RecA-like cleft (Figure 2-3). Comparison of the arch domains between homologs shows that the arch from CasDinG is composed of fewer amino acids but the overall architecture of the domain is conserved, although the CasDinG structure lacks a β-loop between alpha helices two and three (Figure 2-4; Figure 2-5). Analysis of the
Table 2-1

*CasDinG Data Collection and Refinement Statistics*

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Native</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data collection</strong></td>
<td></td>
</tr>
<tr>
<td>Beamline</td>
<td>SSRL 9-2</td>
</tr>
<tr>
<td>Space group</td>
<td>P65</td>
</tr>
<tr>
<td>Cell dimensions</td>
<td></td>
</tr>
<tr>
<td>a, b, c (Å)</td>
<td>123.3, 123.3, 136.9</td>
</tr>
<tr>
<td>α, β, γ (deg)</td>
<td>90.0, 90.0, 120</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.88684</td>
</tr>
<tr>
<td>Resolution** (Å)</td>
<td>50-2.95 (3.06-2.95)*</td>
</tr>
<tr>
<td>R&lt;sub&gt;merge&lt;/sub&gt; (%)</td>
<td>18.8 (135.6)</td>
</tr>
<tr>
<td>CC&lt;sub&gt;1/2&lt;/sub&gt;</td>
<td>0.994 (0.766)</td>
</tr>
<tr>
<td>I/σI</td>
<td>20.0 (2.0)</td>
</tr>
<tr>
<td>Observations</td>
<td>49009 (4849)</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>24916 (2453)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>20.5 (16)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>100.0 (99.8)</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>50-2.95 (3.06-2.95)</td>
</tr>
<tr>
<td>No. reflections</td>
<td>24916</td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
</tr>
<tr>
<td>R&lt;sub&gt;work&lt;/sub&gt;/R&lt;sub&gt;free&lt;/sub&gt;</td>
<td>18.5/21.5</td>
</tr>
<tr>
<td>No. atoms</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>4578</td>
</tr>
<tr>
<td>Water</td>
<td>0</td>
</tr>
<tr>
<td>Ligands</td>
<td>0</td>
</tr>
<tr>
<td>B-factors</td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td>76.2</td>
</tr>
<tr>
<td>R.m.s. deviations</td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.002</td>
</tr>
<tr>
<td>Bond angles (deg)</td>
<td>0.52</td>
</tr>
<tr>
<td>Ramachandran</td>
<td></td>
</tr>
<tr>
<td>Favored (%)</td>
<td>96.33</td>
</tr>
<tr>
<td>Allowed (%)</td>
<td>3.77</td>
</tr>
<tr>
<td>Outliers (%)</td>
<td>0</td>
</tr>
<tr>
<td>Clashscore</td>
<td>3.4</td>
</tr>
</tbody>
</table>

*Note.* *Values in parentheses are for highest-resolution shell. **Resolution limit used the criterion of I/σI > 2.0.*
Figure 2-3

X-Ray Structure of Strep-Tagged Apo CasDinG

A. The solved structure of Apo CasDinG is composed of the large RecA body (light blue), the vestigial FeS domain (dark blue) and the Arch domain (red). B. Alignment of Apo CasDinG (light blue) to 6FWR E. coli DinG bound to a poly T 12 single stranded DNA substrate (pink), highlights a binding pocket for nucleic acid.

*Note.* A. The solved structure of Apo CasDinG is composed of the large RecA body (light blue), the vestigial FeS domain (dark blue) and the Arch domain (red). B. Alignment of Apo CasDinG (light blue) to 6FWR *E. coli* DinG bound to a poly T 12 single stranded DNA substrate (pink), highlights a binding pocket for nucleic acid.
Figure 2-4

CasDinG Accessory Domain Comparisons

Note. A. Arch domain comparison between ApoCasDinG and *E. coli* DinG. B. vFeS CasDinG domain comparison to *E. coli* DinG FeS domain. C. Zoomed in view of acidic aspartate glutamate pair in the vFeS which reaches into the nucleic acid binding pocket.
Figure 2-5

Domain Alignments Comparing PaCasDinG to E. coli DinG.

**A**. Secondary structure matching of the FeS domains of PaCasDinG to *E. coli* DinG. 
**B**. DALI alignment of arch domains.

*Note.* A. Secondary structure matching of the FeS domains of PaCasDinG to *E. coli* DinG. B. DALI alignment of arch domains.
FeS domains from the two homologs confirm the lack of a coordinated Fe atom in the CasDinG structure. Examination of the vFeS domain reveals the lack of a loop-helix loop between alpha helices two and three and the presence of an extended loop between alpha helices three and four (Figure 2-4, Figure 2-5).

A secondary structure matching alignment of CasDinG to *E. coli* DinG reveal an RMSD of 2.93 Å. Further alignment with the solved structure of *E. coli* DinG bound to an ADP BeF$_3$ Mg complex, PDB 6FWS, highlights the binding pocket for ATP and analogs in the CasDinG structure (Figure 2-6). Measurements between the conserved aspartate residue in the Walker B motif and the magnesium complex show a 6.7Å distance, whereas the 6FWS structure shows a 4.7Å distance length, indicating that conformational changes will occur when CasDinG binds an NTP and ssDNA substrate (the 6FWS structure is *E. coli* DinG bound to both ssDNA and an ADP complex).

**Nucleic acid-binding assays**

Knowing that CasDinG contained SF2 helicase motifs and that the structure of CasDinG highlighted a nucleic acid binding pocket, it was presumed that CasDinG may bind nucleic acids such as DNA or RNA. CasDinG nucleic acid-binding preferences were examined using fluorescence anisotropy technology on a Synergy H4 plate reader, and anisotropy values were analyzed using GraphPad Prism software. The binding constants were derived using a one-site saturated binding curve and are presented in Table 2-2. Binding data show that CasDinG can bind both DNA and RNA with apparent affinities of 81 +/- 5.4 nM and 61 +/- 8.1 nM affinities (Table 2-2; Figure 2-6). As these affinities are
Figure 2-6

Alignment of Apo CasDinG to 6FWS

Note. Image highlights the NTP binding site with the Walker B motifs highlighted in salmon (CasDinG) and green (6FWS E. coli DinG).
within a similar range, we could not determine if there was a true preference for one nucleic acid substrate over the other. Knowing that ATP binding and hydrolysis can lead to protein conformational changes, and therefore affinity changes, we hypothesized that CasDinG affinity for ssDNA may alter in the presence of ATP or analogues. We found that in the presence of ATP the apparent $K_d$ was $140 \pm 16$. This was at most a 2-fold decrease in affinity. In comparison, the apparent binding affinity of CasDinG for ssDNA in the presence of the non-hydrolysable analog AMP-PNP was $47 \pm 4.9$. Which was within a 2-fold increase in binding affinities. From this data we concluded that there were no substantial differences in nucleic acid binding affinities in the presence of analogs (Figure 2-7).

CasDinG was also able to bind both 17 nucleotide (nt.) and 40 nt. single-stranded DNA substrates with similar binding affinities, even in the presence of ATP (Figure A-1; Table 2-1). DinG binding to a blunt-ended duplexed showed no binding in direct comparison to single-stranded DNA, suggesting that single stranded DNA is required for loading (Figure A-2).

### ATPase Assays

<table>
<thead>
<tr>
<th>Substrate</th>
<th>$K_d$ (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ssDNA (17 nt.)</td>
<td>81 +/- 5.4</td>
</tr>
<tr>
<td>ssRNA-PT</td>
<td>61 +/- 8.1</td>
</tr>
<tr>
<td>ssDNA (17 nt.) ATP</td>
<td>140 +/- 16</td>
</tr>
<tr>
<td>ssDNA (17 nt.) ADP</td>
<td>37 +/- 3.2</td>
</tr>
<tr>
<td>ssDNA (17 nt.) AMP-PNP</td>
<td>47 +/- 4.9</td>
</tr>
</tbody>
</table>

**Table 2-2**

*Apparent $K_d$ for Nucleic Acid Substrates*
Figure 2-7

*CasDinG Binding to Nucleic Acid Substrates*

**A.** WT CasDinG binding to 5’ FAM labeled 17 nt. Single stranded DNA substrate.

**B.** WT CasDinG binding to a 5’ FAM labeled 17 nt. single stranded DNA substrate in the presence of ATP, ADP, or AMP-PNP.

**C.** WT CasDinG binding to a 5’ FAM labeled phosphorothioated RNA substrate.

*Note.* A. WT CasDinG binding to 5’ FAM labeled 17 nt. Single stranded DNA substrate. B. WT CasDinG binding to a 5’ FAM labeled 17 nt. single stranded DNA substrate in the presence of ATP, ADP, or AMP-PNP. C. WT CasDinG binding to a 5’ FAM labeled phosphorothioated RNA substrate.
As helicases are defined by their ability to remodel nucleic acid using the hydrolysis of NTPs, colorimetric malachite green assays were used to measure the liberation of free phosphate from the provided NTP. SF2 family helicases typically have nucleic acid stimulated ATPase activity (Christiansen, 2003). Thus, hydrolysis of ATP was assessed in the presence of single-stranded DNA (ssDNA) and single-stranded RNA (ssRNA), each substrate forty nucleotides in length. Rates of inorganic phosphate production were higher for ssDNA than ssRNA, with close to zero rates of hydrolysis associated with no nucleic acid. DNA substrates of differing lengths were assessed, and the highest hydrolysis rates were seen for ssDNA (40 nt.), followed by double-stranded DNA (dsDNA) with a 5’ DNA overhang (34 nt.). Duplexed DNA substrates with a 3’ overhang also showed ATP hydrolysis, but rates were not as high as ssDNA (40 nt.) or the 5’ overhang duplexed DNA. Blunt dsDNA substrates and ssDNA (17 nt.) showed rates of ATP hydrolysis equivalent to basal CasDinG (Figure 2-8). This data highlighted that a nucleic acid substrate needs to be greater in length than 17 nt. to see ATP hydrolysis.

Helicases can sometimes hydrolyze other NTPs, so we assessed hydrolysis rates for ATP, UTP, CTP, and GTP in the presence of ssDNA (40 nt.). The highest rates of NTP hydrolysis were achieved with ATP, with minimal hydrolysis rates for GTP and UTP. There were no measurable hydrolysis rates for CTP in the conditions tested (Figure 2-7). This data showed that CasDinG will preferentially hydrolyze ATP in the presence of ssDNA.

To further understand the ATPase activities of CasDinG, Michaelis-Menten kinetics were performed. Curves assessing ATP hydrolysis rates for CasDinG without
Figure 2-8

*Vo Plots of ATP Hydrolysis by CasDinG*

A. ATP hydrolysis by WT CasDinG is stimulated in the presence of both single stranded DNA and single stranded phosphorothioated RNA. B. ATP hydrolysis is stimulated by single stranded nucleic acids with sufficient length. C. WT CasDinG primarily hydrolyzes ATP. Error bars represent standard deviation from the mean of three replicates.

*Note.*
nucleic acid (basal) pH 6.8, with ssDNA pH 6.8, with ssDNA pH 7.5 and with ssRNA-PT pH 7.5 were generated (Figure 2-9). Analysis of the Michaelis-Menten CasDinG curves reveal meager rates of ATP hydrolysis in the absence of nucleic acid at pH 6.8, with a k_{cat} of 0.28 +/- 0.037 molecules of ATP hydrolyzed per second, whereas in the presence of ssDNA at pH 6.8, we report a k_{cat} of 11 +/- 0.74 ATP sec^{-1}. At pH 7.5 we saw k_{cat} values of 32 +/- 0.99 ATP sec^{-1} with ssDNA and 23 +/- 0.71 ATP sec^{-1} for ssRNA-PT (Table 2-3). This data showed that a pH shift from 6.8 - 7.5 resulted in a 2.9-fold difference in the turnover number of ATP hydrolysis when CasDinG is in the presence of both ATP and ssDNA (Table 2-3).

**Table 2-3**

*Michaelis-Menten Kinetic Parameters for ATP hydrolysis*

<table>
<thead>
<tr>
<th>Substrate</th>
<th>[CasDinG] nM</th>
<th>[nucleic acid] nM</th>
<th>pH</th>
<th>V_{max} (µM sec^{-1})</th>
<th>K_{m} (µM)</th>
<th>k_{cat} (ATP sec^{-1})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basal</td>
<td>500 nM</td>
<td>0 nM</td>
<td>6.8</td>
<td>0.14 +/- 0.019</td>
<td>2.0 x10^2 +/- 77</td>
<td>0.28 +/- 0.037</td>
</tr>
<tr>
<td>ssDNA (40 nt.)</td>
<td>18.5 nM</td>
<td>150 nM</td>
<td>6.8</td>
<td>0.20 +/- 0.014</td>
<td>80 +/- 24</td>
<td>11 +/- 0.74</td>
</tr>
<tr>
<td>ssDNA (40 nt.)</td>
<td>10 nM</td>
<td>100 nM</td>
<td>6.8</td>
<td>0.014</td>
<td>0.0099</td>
<td>160 +/- 15</td>
</tr>
<tr>
<td>ssRNA-PT (17 nt.)</td>
<td>20 nM</td>
<td>200 nM</td>
<td>7.5</td>
<td>0.32 +/- 0.014</td>
<td>180 +/- 17</td>
<td>32 +/- 0.99</td>
</tr>
</tbody>
</table>

The differences between ssDNA and ssRNA-PT at pH 7.5 are within one order of magnitude of each other. *E. coli* DinG has been reported to hydrolyze 1.59 molecules of ATP sec^{-1} without nucleic acid and 24.1 molecules of ATP sec^{-1} in the presence of ssDNA (Voloshin et al., 2003).
Figure 2-9

Michaelis-Menten Curves of ATP Hydrolysis by CasDinG

Note. A. Rates of ATP hydrolysis at various ATP concentrations using 500 nM WT CasDinG with no nucleic acid substrates, pH 6.8. B. Rates of ATP hydrolysis at various ATP concentrations using WT CasDinG and a 40-nucleotide single stranded DNA substrate at pH 6.8 C. Rates of ATP hydrolysis at various ATP concentrations using WT CasDinG and a 40-nucleotide single stranded DNA substrate at pH 7.5 D. Rates of ATP hydrolysis at various ATP concentrations using WT CasDinG and phosphorothioated single stranded RNA substrate at pH 7.5.
Therefore, we concluded from these Michaelis-Menten curves that CasDinG ATPase rates were within a similar range as that of Non-Cas DinG enzymes. To make accurate comparisons between all curves reported here, we would need to repeat the experiments in same buffers/pHs for each substrate or lack thereof. To determine whether the phosphorothioated RNA backbone impacts ATPase rates, another curve will need to be generated with a native RNA backbone in RNAse-free conditions.

Helicase assays

Non-Cas DinG enzymes if they are helicases, have 5’-3’ polarity, suggesting that this enzyme might be the same given its ability to bind nucleic acid and hydrolyze ATP in the presence of both RNA and DNA (Voloshin et al., 2003; Thakur et al., 2014). Here we investigated the substrate preferences of CasDinG for different DNA and RNA duplexes with 5’ and 3’ overhangs (Figure A-4). CasDinG was first assessed for helicase activity against 5’ FAM-labeled DNA duplexes with either a 5’ overhang, a 3’ overhang, or a blunt end (Figure 2-10). Out of these substrates, CasDinG only unwound the 5’ overhang dsDNA substrate. This is the first data showing a CRISPR associated helicase to have 5’-3’ polarity. Given that DinG homologs can unwind RNA/DNA hybrids, we next investigated whether CasDinG would be able to unwind 5’ FAM-labeled DNA/RNA hybrids and RNA duplexes (Figure 2-10) (Voloshin & Camerini-Otero, 2007). CasDinG was found capable of unwinding a 5’ overhang duplexed-hybrid only when the overhang was single-stranded DNA. There was no unwinding observed for a duplexed RNA substrate or for a hybrid duplex in which the overhang was single-stranded RNA.

The no ATP control on the native PAGE gels shows a small amount of intensity
associated with a single-stranded band, suggesting that the initial binding of CasDinG to duplex with single-stranded overhang(s) results in some displacement of duplex due to thermal fraying. However, direct comparison of the ATP-containing reactions shows a clear and consistent increase in substrate displacement, consistent with CasDinG being a 5’-3’ ATP-dependent helicase (Figure 2-10).

Time trial comparisons of 5’ FAM-labeled nucleic acid substrates with 25 nM protein concentrations show that active helicase substrates are more than 70% unwound within the first 60 seconds of the reaction. The highest percentages of unwound duplex were found with the 5’ DNA overhang hybrid duplex, indicating there may be a preference for hybrid duplexes, however given that helicases can perform annealing reactions as well as unwinding reactions, rates of unwinding would have to be determined at multiple protein concentrations to determine true nucleic acid unwinding preferences.

To further assess whether duplex displacement was dependent upon the hydrolysis of ATP, CasDinG was assayed for helicase activity with a 5’ FAM labeled 5’ overhang duplexed DNA substrate in the presence of ATP, ADP, AMP-PNP, or ATPγS. It was only in the presence of ATP that substantial unwinding of duplex substrates was observed, indicating that hydrolysis of ATP is essential for proper helicase function (Figure 2-11).

Non-Cas DinG enzymes, such as E. coli DinG, rely upon divalent metals to unwind duplexed nucleic acids (Voloshin et al., 2003). Divalent metals can be used to coordinate the phosphate backbone of ATP and nucleic acid substrates. Here we
Figure 2-10

*CasDinG is a 5’-3’ ATP-Dependent Helicase*

A. Native PAGE analysis of DNA helicase substrates.

B. Native PAGE analysis of RNA and DNA hybrid helicase substrates.

C. Graph displaying the percent of helicase substrate unwound as a function of time. Error bars represent standard deviation from the mean of three replicates.

*Note.* A. Native PAGE analysis of DNA helicase substrates. B. Native PAGE analysis of RNA and DNA hybrid helicase substrates. C. Graph displaying the percent of helicase substrate unwound as a function of time. Error bars represent standard deviation from the mean of three replicates.
Figure 2-11

Unwinding is Stimulated by ATP and Divalent Ions

Note. A. Native PAGE showing that unwinding of a duplexed DNA substrate is dependent upon ATP hydrolysis. B. Bar graph depicting the % of duplexed DNA substrate unwound in the presence of ATP analogues. Error bars represent standard deviation from the mean of three replicates with the no ATP control acting as the control for 0% unwound. C. Native PAGE showing unwinding of duplexed DNA substrate in the presence of various divalent ions.
investigated the divalent metal requirements for CasDinG in the presence of a 5’ FAM-labeled 5’ overhang DNA duplex (Figure 2-11). Data show that CasDinG can displace DNA in the presence of Mg$^{2+}$, Mn$^{2+}$, Ca$^{2+}$, Ni$^{2+}$, and Co$^{2+}$ but not Zn$^{2+}$ or Cu$^{2+}$.

**Mutational Analysis**

CasDinG mutants were initially generated to determine whether DNAse/RNase activity visualized on helicase assays was due to CasDinG or contaminating nucleases (nuclease data not shown). These mutations were made using a structural model generated by the Robetta server (Figure A-7). Although the nuclease data did not result in a successful avenue of study, we utilized these mutations to understand the function of accessory domains.

The Walker B motif, DEAH, located at position 337, was mutated to AAAH. The ΔN-terminus truncation eliminated the amino acids 18-128 and included deletion of the Q motif. The ΔArch (350-469) and ΔvFeS (195-287) with a x5 glycine linker, were domain knockouts generated at the junction between the specified domain and Domain1 of the RecA core (Figure 2-12). All mutants were assayed for their ability to unwind a 5’ FAM-labeled DNA with a 5’ overhang. ΔArch and ΔDEAH both abolished DNA unwinding activity, whereas ΔN-terminus truncation and ΔvFeS showed appreciable amounts of displaced nucleic acid (Figure 2-12). However, it should be mentioned, that the ΔN-terminus truncation helicase activity has not been consistent. Alternative protein preparations have resulted in no duplex unwinding. Whether this is due to protein preparation errors, or other human errors has yet to be elucidated but could be remedied through more biological replicates deciphering ΔN-terminus truncation function.
Next the ΔDEAH, ΔN-terminal truncation and ΔArch were assessed for their ability to bind ssDNA (17 nt.). All mutants tested maintained the ability to bind and had similar binding affinities ranging from 42 - 94 nM (Figure A-2; Table A-1). There were not sufficient concentrations of the ΔvFeS to assess binding affinities. Next rates of ATP hydrolysis were assessed for all four mutants in the presence of ssDNA (40 nt.). ΔArch and ΔN-terminus truncation maintained similar rates of ATP hydrolysis to WT, whereas ΔvFeS and ΔDEAH resulted in diminished ATP hydrolysis (Figure 2-12). Given that the ΔvFeS mutant showed helicase activity in a 20-minute reaction and little to no ATPase activity in a 3 minute reaction, it is hypothesized that the ΔvFeS is capable of minimal rates of ATP hydrolysis needed to unwind duplexes. To fully understand the mechanism of ATP hydrolysis, Michaelis-Menten kinetics should be performed using the ΔvFeS construct.

To ensure that the removal of domains did not disrupt the folding of CasDinG, far UV circular dichroism was used to assess the structural integrity of purified proteins. This data was collected for WT CasDinG, ΔDEAH, and ΔN-terminal truncation. ΔArch and ΔvFeS have not had CD data collected as protein concentrations were too low for the volumes needed to assay. Analysis of the CD data shows that the protein is mostly alpha-helical, and a comparison of the mutants with WT shows no substantial loss of structural integrity when purifying these mutants (Figure A-3).

Comparison of the x-ray crystal generated model and the predicted model, along with increased knowledge of helicase motifs, revealed that the initial mutants may have been overly harsh in their construction, meaning that more of a domain was deleted than necessary. New mutants utilizing the x-ray crystal structure as a reference with fewer
**Figure 2-12**

*Functional Analysis of CasDinG Domains via Mutagenesis*

A. Gene schematics comparing WT CasDinG to each of the studied mutants. B. Rates of ATP hydrolysis by CasDinG and mutants as measured by concentration of inorganic phosphate as a function of time. C. Native PAGE analysis of a duplexed DNA helicase substrate in the presence of CasDinG (mutants) and ATP. D. Bar graph analysis of % duplexed DNA unwound as a function of CasDinG construct. Error bars represent standard deviation from the mean of three replicates.

*Note.* A. Gene schematics comparing WT CasDinG to each of the studied mutants. B. Rates of ATP hydrolysis by CasDinG and mutants as measured by concentration of inorganic phosphate as a function of time. C. Native PAGE analysis of a duplexed DNA helicase substrate in the presence of CasDinG (mutants) and ATP. D. Bar graph analysis of % duplexed DNA unwound as a function of CasDinG construct. Error bars represent standard deviation from the mean of three replicates.
amino acid deletions, may be more accurate in determining the function of accessory domains in CasDinG function.

**In Vivo Plasmid Curing Assay**

Previous research by Dr. Crowley showed that all type IV CRISPR-Cas genes are necessary for type IV-A system targeting, including the *ding*-like gene (Crowley et al., 2019). Using a similar plasmid curing assay, we assessed the necessity of the accessory domains (ΔArch, ΔN. terminal truncation, and ΔvFeS) for type IV-A immune system function (Figure 2-13). All four mutants (ΔDEAH_AAAH, ΔN-terminal truncation, ΔArch, and ΔvFeS) inhibited the targeting ability of the type IV-A system *in vivo*. Although noticeably, there is variability in the ratio of target to non-target colonies in both the ΔN-terminal truncation and ΔvFeS data sets. Additional biological replicates could allow for greater understanding of whether the data skews closer towards a knockout of immune system function or not. However, in comparison to the WT positive control, all mutants did not provide proper type IV CRISPR immune system function.

**Type IV-A Complex and CasDinG**

Protein expression conditions were identified which allowed for a high yield of type IV-A RNP complex to be purified. Furthermore, conditions were identified that allowed for the co-expression and co-purification of CasDinG with the type IV-A RNP complex (Figure 2-14). Protein purification occurred over two affinity columns (nickel & streptavidin) followed by size exclusion. Affinity tags were placed on Csf1 and Csf2 with no tags on the CasDinG. SDS-PAGE after the affinity columns showed that CasDinG
**Figure 2-13**

*In Vivo Plasmid Curing Assay.*

**Note.** A. Schematic of the protocol used to count colonies for the *in vivo* plasmid curing assay. B. Bar graph showing the ratio of target/non target plasmid containing cells on the y axis as a function of CasDinG construct used in the assay, in addition to the rest of the type IV-A immune system. A ratio of ~1 or higher implies a dysfunctional immune system.
Figure 2-14

_Type IV-A Protein Purification_

**A.** Size exclusion chromatogram of the Type IV-A RNP complex eluting off a Superdex 200 Increase 10/300 GL column. **B.** SDS PAGE of recombinantly expressed and purified Type IV-A ribonucleoprotein complex (Csf2, Csf5/Cas6, Csf1, and Csf3). **C.** Size exclusion chromatogram of the Type IV-A RNP + CasDinG complex eluting off a Superdex 200 Increase 10/300 GL column. **D.** SDS PAGE of recombinantly expressed and purified Type IV-A ribonucleoprotein complex and CasDinG stained with Coomassie blue.

**Note.** A. Size exclusion chromatogram of the Type IV-A RNP complex eluting off a Superdex 200 Increase 10/300 GL column. B. SDS PAGE of recombinantly expressed and purified Type IV-A ribonucleoprotein complex (Csf2, Csf5/Cas6, Csf1, and Csf3). C. Size exclusion chromatogram of the Type IV-A RNP + CasDinG complex eluting off a Superdex 200 Increase 10/300 GL column. D. SDS PAGE of recombinantly expressed and purified Type IV-A ribonucleoprotein complex and CasDinG stained with Coomassie blue.
was copurifying with the complex. However, size exclusion chromatography shows that CasDinG copurifies with the type IV-A RNP complex in a later elution than the type IV-A RNP complex without CasDinG, indicating that CasDinG does not form a stable interaction with the complex. If a stable interaction had formed, we would have expected that the CasDinG + RNP complex would have resulted in complex that eluted prior to RNP complex alone.

Additionally, the estimated molecular weights for type IV-A RNP complex resemble a 1:5-6:1:1 ratio rather than the previously described 1:2:1:1 ratio of Csf1:Csf2:Csf3:Csf5/Cas6, respectively (Taylor, 2021). Later elutions are associated with smaller molecules, leading to the hypothesis that the interaction between CasDinG and type IV-A RNP complex is transient, dependent upon nucleic acid substrates, or a result of column preferences. It is unknown what the conditions for the interaction are.

Knowing that the type IV-A CRISPR-cas system can target plasmids in vivo led to the development of in vitro binding assays using fluorescence anisotropy. There is currently no peer-reviewed published work reporting in vitro binding affinities for type IV-A RNP complexes to nucleic acid substrates. The type IV-A RNP complex was assessed for binding affinities to single-stranded DNA targets with differing PAM sequences (Figure 2-15). The estimated $K_d$ values for these substrates were all below 10 nM, suggesting high affinities between the complex and target substrates (Table 2-4).

This data indicated that all three PAMs would be suitable for use in in vitro studies. The type IV-A RNP complex that had been co-expressed with CasDinG was assayed for single-stranded DNA (TTTC PAM sequence) binding once. This data showed a preliminary $K_d$ of 5.523 nM, suggesting that CasDinG does not interfere with RNP
Table 2-4

Apparent $K_d$ of ssDNA Alternative PAM Substrates

<table>
<thead>
<tr>
<th>Substrate</th>
<th>$K_d$ (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ssDNA Target GGAAA</td>
<td>9.3 +/- 0.81</td>
</tr>
<tr>
<td>ssDNA Target GCGAT</td>
<td>9.0 +/- 1.4</td>
</tr>
<tr>
<td>ssDNA Target AAAAA</td>
<td>8.4 +/- 1.3</td>
</tr>
</tbody>
</table>

Complex binding target sequences (Figure A-6).

Current literature suggests that type IV-A systems can reduce gene expression without DNase activity (Guo et al., 2022), however in vitro plasmid cleavage data shows that non-target and target-containing plasmids can be nicked in the presence of the type IV-A complex (Figure 2-15). The nicking occurs whether CasDinG is present or not. To understand whether this nicking activity is real or due to contaminating nucleases, point mutations eliminating this activity are necessary. If nicking occurs, it could allow for CasDinG to load on an open ssDNA end.

It should be noted that all fluorescence anisotropy reactions with estimated $K_d$ values below 100 nM should be viewed cautiously, as substrate concentrations of nucleic acid were held constant at 10 nM. Therefore, we can only reasonably state that nM binding is occurring with these substrates. The probe's concentration should be at least nine times lower than the estimated $K_d$ to obtain accurate values using the traditional fit of $\%$ bound = $[S]/([S] + K_d)$. With the values generated here, pico or femto molar probe
Figure 2-15

Type IV-A Preliminary Binding and Cleavage Data

Note. A. Binding of Type IV-A RNP complex to single stranded nucleic acid sequences with alternative PAM sequences. Measured using fluorescence anisotropy. B. Agarose gel showing cleavage a GGAAA PAM containing plasmid.
levels would be needed for accurate $K_d$ estimation. Data collected with FAM-labeled substrates at those lower concentrations suffers from higher background noise. Thus, an alternative method of labeling substrates, such as radiolabeling may be helpful.

**Discussion**

The role of the *dinG*-like gene associated with type IV-A CRISPR-Cas systems was unknown prior to the work performed in this thesis. The biochemical characterizations performed on CasDinG establish this enzyme as a 5’-3’ ATP-dependent helicase with an essential role in the *in vivo* function of the type IV-A CRISPR-Cas system from *Pseudomonas aeruginosa strain 83*. Initial binding assessments revealed a preference for single-stranded nucleic acid substrates. Furthermore, it was shown that single-stranded nucleic acid was required for binding to duplexed substrates as there was no binding to a blunt duplexed nucleic acid substrate. Binding affinities for ssRNA were within the same range as ssDNA. From this we can infer that there is no substantial affinity difference between the two substrates. It is known that helicases can translocate along nucleic acid by changing the affinity of one nucleic acid-binding domain from another by utilizing ATP hydrolysis (Pyle, 2008; Cheng & Wigley, 2018). Assessment of binding affinities of ssDNA in the presence of ATP analogs did not show data indicative of higher or lower binding affinities. It may be that the use of fluorescence anisotropy to measure binding affinities is not sensitive enough to detect the changes in domain affinity for nucleic acid in the presence of ATP with single-stranded DNA.

Subsequent analysis of the NTPase activities of CasDinG suggests that ATP is the preferred NTP substrate in the conditions tested. ATPase velocity plots showed higher
rates of ATP hydrolysis using ssDNA (40 nt.) over ssRNA substrates (40 nt.). Interestingly no ATP hydrolysis was visualized for a ssDNA (17 nt.) which CasDinG was shown to bind. This data shows that there is a required nucleic acid substrate length for ATP hydrolysis. The lack of ATP hydrolysis in the presence of ssDNA (17 nt.) could be the reason why there were no discernable binding differences of ssDNA (17 nt.) in the presence of ATP analogues. However, this does not explain the lack of binding affinity differences between ssDNA (40 nt.) in and out of the presence of ATP. The binding data combined with ATPase data suggest that CasDinG surveys its environment for single-stranded nucleic acids on which it can bind. Following nucleic acid binding, CasDinG may then translocate on ssDNA. As no helicase activity was visualized for RNA overhang substrates, the ATP hydrolysis rates differences seen between CasDinG bound to DNA versus RNA-PT may be due to either functional activity differences or limitations imposed by the phosphorothioated backbone of the RNA substrate. It should be noted that helicase assays did not use RNA-PT.

Helicase data showed unwinding of duplexed nucleic acid with a 5’ overhang and little to no displacement of 3’ overhang substrates. This data shows that CasDinG has a preferred polarity of 5’-3’ which differs from that of Cas3 (3’-5’), the only other CRISPR associated helicase to be characterized thus far (Sinkunas et al., 2011). CasDinG does not bind and cannot unwind blunt dsDNA duplexes suggesting that a single-stranded nucleic acid is required for loading CasDinG onto a given substrate. Furthermore, substantial displacement of the labeled strand is dependent upon the presence of ATP and a forked substrate is not required to see displacement. Furthermore, RNA-containing substrates did not show unwinding when the loading strand was RNA, despite data that shows
CasDinG binding ssRNA-PT and hydrolyzing ATP in the presence of ssRNA-PT. This result could be due to the A-form nature of RNA and may coincide with the lower rates of ATP hydrolysis seen with CasDinG in the presence of ssRNA-PT. CasDinG could perhaps displace transcription machinery which would support the hypothesis presented by the Guo group, who show that type IV-A systems can inhibit gene expression without cleavage of the gene (Guo et al., 2022). However, experiments showing the ability of CasDinG to displace proteins are still needed. Helicase assays showed duplex unwinding dependence on divalent ions. CasDinG was able to unwind duplexed DNA in the presence of Mg$^{2+}$, Mn$^{2+}$, Ca$^{2+}$, and Ni$^{2+}$ but not Cu$^{2+}$ or Zn$^{2+}$.

This thesis presents the first crystal structure of a CRISPR-associated DinG helicase. The 2.95 Å structure of CasDinG reveals a model that maintains a core composed of two RecA domains with two accessory domains inserted into the RecA1 lobe. The vFeS domain appears as an accessory to the RecA domain, branching off the RecA1 lobe, and does not contain any density indicative of bound metal ions. The vFeS domain adopts a closed alpha-helical domain held close to the RecA body beneath the arch domain. One of the alpha-helices in the vFeS domain is near the binding pocket for nucleic acid with an aspartate glutamate pair at positions 271 and 272. These residues may play a role in coordinating metal ions, but that has yet to be studied. When CasDinG is aligned with the polyT substrate from the 6FWR E. coli DinG structure, we see that the entire vFeS domain clashes with the ssDNA substrate, indicating that domain rearrangements may occur upon binding nucleic acid substrates.

Interestingly, deletion of the vFeS domain resulted in reduced rates of ATP hydrolysis and a non-targeting type IV-A system in vivo, suggesting a structural role in
forming an ATP binding/hydrolysis pocket. Deletion of the vFeS domain did not abolish DNA duplex unwinding, suggesting that ATP hydrolysis still occurs at a lower rate than WT. Alternatively, the vFeS domain deletion construct could adopt a conformation adapted to displacing or kinking nucleic acid duplexes. This data set is significant, as non-CasDinG homologs require the presence of a FeS cluster to unwind duplexed nucleic acid, whereas we see that CasDinG does not require a FeS cluster nor even the vFeS domain to unwind duplexes. Although we know that unwinding still occurs, we do not know whether the rates of unwinding are altered. It would be prudent to determine \( k_{\text{unwind}} \) for domain mutants in future studies and confirm results with biological replicates.

The arch domain appears as a second insert in the RecA1 lobe. Deletion of the arch domain results in an enzyme incapable of unwinding duplexed nucleic but capable of binding ssDNA and hydrolyzing ATP. This data suggests that the arch may play a role in strand separation. The N-terminus was not observed in the electron density, and there is no structural data on which to comment. However, the Alphafold structure prediction showed that there are structured alpha helical regions and unstructured loops in the N-terminus. The absence of the N-terminus in the x-ray structure suggests that perhaps the N-terminus is flexible. The N-terminus may play a role in CasDinG immune function as deletion results in a dysfunctional type IV-A immune system incapable of targeting plasmid, however given that the Q motif is removed with this mutation, lack of targeting should have been anticipated. Further experimentation using an N-terminal truncation that does not remove the Q motif should be pursued. Arch domains in other SF2 family proteins may function as interaction sites for other proteins, suggesting that this may be a site of interaction between CasDinG and the type IV-A complex; this is an avenue of
research for future studies (Abdulrahman et al., 2013). Additionally, future structures of CasDinG bound to DNA substrates and ATP analogs will provide further insight into the accessory domains' structural, functional and regulatory roles.

Early experiments conducted by Dr. Val Crowley revealed that a WT DEAH motif is essential for the targeting ability of the type IV-A system in vivo. This thesis' in vivo analysis of domain mutations suggests that a CasDinG enzyme with both ATPase and helicase activity is essential for proper CRISPR system function. These findings are significant because we have determined a biochemical mechanism for a type IV system protein directly involved in CRISPR plasmid interference. This allows us to make an educated hypothesis on the interference mechanisms utilized by these systems.

SDS-PAGE gels show that CasDinG and the type IV-A RNP complex co-purify over a size exclusion column, suggesting that there may be interactions mediated by nucleic acid, between CasDinG and complex. The size exclusion chromatogram showed RNP complexes with CasDinG eluting at later elution volumes than isolated type IV-A RNP. This data was not anticipated as size exclusion columns work by sorting larger or higher molecular weight molecules from smaller ones. We would have expected that a stable CasDinG-type IV-A RNP complex would have eluted at a volume expected for a complex of 255 kDa, assuming a 1:2:1:1 ratio of RNP complex proteins (Taylor, 2021). Instead, we observe an RNP complex eluting at a volume equivalent to a complex of 297 kDa. This size of complex indicates an alternative ratio of RNP enzymes closer to 1:5-6:1:1 (Csf1:Csf2:Csf3:Csf5/Cas6). The CasDinG-type IV RNP complex eluted at a volume equivalent to a complex of 185 kDa, significantly less than the predicted 255 kDa, suggesting that the interaction between CasDinG and RNP may be transient or be
dependent on the binding of nucleic acid substrates. More studies evaluating the relationship between CasDinG and RNP complex are needed before definitive statements are made regarding RNP complex composition.

Fluorescence anisotropy binding assays revealed that the type IV-A RNP complex binds at least three ssDNA PAM sequences with nM affinity. Binding assays also showed that the CasDinG + type IV-A RNP complex binds ssDNA with similar affinities to RNP complex alone. Knowing that CasDinG is a helicase, it would be prudent to evaluate the CasDinG + type IV-A complex affinity for duplexed nucleic acid and the type IV-A RNP complex (297 kDa) affinity for duplexed nucleic acids. This information would be helpful for understanding when and if CasDinG is for type IV-A complex binding. CasDinG may be necessary during RNP complex formation and interference of MGE.

Preliminary plasmid cleavage data indicates that small amounts of nickase activity may be associated with the RNP complex. The single-stranded DNA breaks resulting from nickase activity could provide a loading zone for CasDinG.

Chapter 2 of this thesis reveals a hypothetical model in which CasDinG could promote alternative RNP complex compositions and act as a 5’-3’ helicase that runs on DNA, knocking off RNA transcripts (Figure 2-16). Alternatively, CasDinG could displace transcription machinery, thus blocking transcription of RNA and thereby inhibiting translation. Whether CasDinG can displace the type IV-A RNP complex when bound to a target sequence should be assessed in the future using translocase assays. Future work on type IV systems will require in-depth knowledge of CasDinG functionality and a more in-depth analysis of RNP complex composition. Functional knowledge of a type IV-A helicase mechanism provides powerful insight into the breadth
Figure 2-16

Hypothesized Role of CasDinG in CRISPR System Function

*P. aeruginosa* 83 Type IV-A RNP Complex

1. Type IV-A RNP complex loads onto dsDNA (plasmid)
2. CasDinG loading onto a nicked substrate? Then runs in a 5’ - 3’ direction
3. CasDinG knocked off RNA transcripts adjacent to R-loop? RNAs are degraded by host?
and depth of microbial immune systems and enhances our understanding of the versatility of SF2 helicases. As helicases are utilized in nearly every facet of nucleic acid biology, the continued study of CasDinG and the type IV-A RNP complex could result in a new tool to be used in gene therapeutics by acting as a gene expression regulator.

Acknowledgments

I would like to acknowledge the following individuals for their contributions:

Dylan Keiser for cloning the Pa83 dinG gene into the 2RT expression vector when rotating through the Jackson lab. Hannah Taylor for the foundational work performed on type IV CRISPR-Cas systems and for cloning the type IV-A Pa83 CRISPR-Cas genes into expression vectors. Matt Armbrust for his help developing a high throughput fluorescence anisotropy assay, assistance with protein purifications which led to the discovery of protein crystallization conditions, intellectual contributions, and his help with collecting data via PAGE analysis. Christian Cahoon for his help with ATPase data collection, the intellectual conception of domain mutations, and his work on the in vivo analysis of domain mutations. Val Crowley for developing the initial plasmid curing/competition assay used to analyze CRISPR-Cas Systems in the Jackson lab and Olive Redman for continuing to develop in vivo assays. A special thanks to Sam Barker for his help using the SSRL beamline. Figures were created with BioRender.com

References

Abdulrahman, W., Ilit, I., Radu, L., Braun, C., Maglott-Roth, A., & Giraudon, C. et al. (2013). ARCH domain of XPD, an anchoring platform for CAK that conditions
TFIIH DNA repair and transcription activities. *Proceedings Of The National Academy Of Sciences, 110*(8). https://doi.org/10.1073/pnas.1213981110


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

https://doi.org/10.1038/ncomms1937


https://doi.org/10.1107/s0907444910007493


https://doi.org/10.1016/j.sbi.2010.03.011


Ren, B., Duan, X., & Ding, H. (2009). Redox Control of the DNA Damage-inducible Protein DinG Helicase Activity via Its Iron-Sulfur Cluster. *Journal Of Biological Chemistry, 284*(8), 4829-4835. https://doi.org/10.1074/jbc.m807943200


CHAPTER 3

IN VITRO CHARACTERIZATIONS OF BACTERIAL IMMUNE SYSTEM PROTEINS CAS12A2 AND JETB

Abstract

*Sulfuricurvum* sp. PC08-66 encodes a type V-A2 CRISPR-Cas system that consists of a *cas12a2* gene and a CRISPR array to protect against invasive nucleic acids. Phylogenetic analysis shows that type V-A2 Cas12a2 is distinct from Cas12a, a DNA-specific reprogrammable nuclease. Cas12a2 proteins consistently have a lower molecular weight than Cas12a, contain a domain foreign to Cas12a, maintain RuvC nuclease motifs, and are sometimes found adjacent to *cas12a* genes. These distinctions suggest that Cas12a2 may play a unique role in type V immunity. This thesis contains biochemical and structural analyses that aid in understanding Cas12a2 function. Pre-crRNA processing assays monitored the cleavage of RNA in the presence of SuCas12a, an essential step in CRISPR immune activity. DNA binding and cleavage assays monitored nuclease activity. Circular dichroism verified the structural integrity of mutants, and a preliminary low-resolution structure of SuCas12a2 bound to a guide RNA was determined using cryo-electron microscopy. Wadjet is a novel immune system that inhibits plasmid transformation efficiencies in *B. subtilis*. To better understand this system, this thesis contains methods to recombinantly express and purify JetB and circular dichroism data to verify structural integrity.
Introduction

Type V CRISPR-Cas Systems

Type V systems are Class 2 CRISPR-Cas systems comprised of single proteins which complex with guide RNA (Koonin & Makarova, 2017). These systems have been used in genome editing applications as they only require a single gRNA, can process their pre-crRNA into a gRNA, and contain a single RuvC nuclease domain (Shmakov et al., 2017; Świat et al., 2017; Alok et al., 2020). The most studied of the Type V enzymes, Cas12a, comes from the type V-A subtype and uses a single RNA guide to cleave double-strand DNA, resulting in a cut with staggered sticky ends (Zetsche et al., 2015).

Bioinformatic studies surrounding Cas12a led to the discovery of Cas12a variant nucleases termed Cas12a2. These Cas12a2 genes are sometimes directly adjacent to Cas12a genes and share the same CRISPR array (Begemann et al., 2017). These genes can differ in length and domain organization from Cas12a, suggesting a difference in function (Figure 3-1). In particular, the Cas12a2 genes possess a novel domain of unknown function between RuvC motifs 1 and 2 and lack the Nuc domain housed by Cas12a.

Cas12a2 variants were shown to edit the genome of Oryza sativa, producing insertion-deletion events in targeted genes, suggesting that Cas12a2 targets dsDNA, similar to Cas12a (Begemann et al., 2017). However, the mechanism and structure of the Cas12a2 variants remain unknown and provide an opportunity to discover novel genome-editing tools.

In this thesis, PAGE analysis shows that SuCas12a2 processes a long pre-crRNA
Figure 3-1

Organization of Type V-A2 Systems

**A.** Organization of type V-A2 CRISPR systems in the Jackson lab. Cas12a is colored gold and Cas12a2 is colored pink. **B.** Protein comparison of type V-A2 SuCas12a2 in comparison to type V-A LbCas12a. The RuvC nuclease motifs and notable domains and structural features are labeled. Figure adapted from Begemann et al. 2017

*Note.* A. Organization of type V-A2 CRISPR systems in the Jackson lab. Cas12a is colored gold and Cas12a2 is colored pink. B. Protein comparison of type V-A2 SuCas12a2 in comparison to type V-A LbCas12a. The RuvC nuclease motifs and notable domains and structural features are labeled. Figure adapted from Begemann et al. 2017
into gRNA, binds single-stranded and double-stranded DNA with a TTTC PAM and may possess DNase activity. Cryo-electron microscopy shows a substantial conformational shift upon binding gRNA. This work is significant as this was the first time bioprocessing was shown for the Cas12a2 variant proteins and the amino acid residues essential for RNA processing were identified. This provided a foundation for the discovery of the interference mechanism by the Jackson and Biesel labs (Dymetrenko et al., 2022).

Overall, this work has allowed for the characterization of a microbial protein which can be repurposed for use as a biotechnology tool as it’s requirements for nucleic acid targeting differ from other type V proteins.

**Wadjet Defense System**

Bioinformatic searches using genome databases in combination with secondary structure predictions using the PHYRE server identified a putative immune system called Wadjet, near known defense genes in defense islands (Doron et al., 2018). Wadjet systems are a four-gene system composed of jetABCD, however there are systems which contain only the jetABC genes. Homology searches suggest that these genes are distantly related to condensin systems such as the MukBEF condensin complex from *E. coli* (Doron et al., 2018). MukBEF complexes form protein rings composed of homo and heterodimeric complexes and are critical for segregating genomic material during cell replication (Rybenkov et al., 2014). HHPRD analysis of Pfams contained by these proteins indicate that jetC houses an ATPase domain and jetD shares similarities with topoisomerase IV (Doron et al.; 2018). Recent work has shown that these systems reduced the transformation efficiency of the plasmid pHCMC05 in *B. Subtilis* (Doron et
al., 2018). However, how these systems interact with plasmids or other nucleic acids, such as ssDNA phage, are unknown. Knowing that all cells require genomic segregation and compaction, both biochemical and structural characterization of these systems is needed. By studying a putative condensin system, we stand to learn more about genetic compaction and segregation methods that could alter transfection or transformation rates of nucleic acid into various cell types.

Wadjet systems may form protein complexes capable of binding nucleic acid, like distantly related condensin systems, and may loop DNA or supercoil DNA to disrupt the ability of foreign plasmids to segregate correctly during cellular replication. This thesis provides the framework for studying Wadjet systems from *Azotobacter vinelandii* DJ and *Mycobacterium* MCS. Alphafold prediction models of *A. vinelandii jet* genes in tandem with DALI server searches support the findings by Doron and show that Jet proteins share structural similarities to condensin systems, such as structural maintenance of chromosome proteins. However, all DALI server matches show less than 20% similarity (Figure 3-2; Figure B-3) (Holm, 2022; Mirdita et al., 2022). It is interesting to note that the Wadjet systems from both *A. vinelandii* DJ and *Mycobacterium* MCS do not contain a *jetD* gene, suggesting that either these systems contain non-functional immune systems or recruit endogenous proteins.

With the genes from these systems transplanted to recombinant expression vectors, one can take steps towards identifying this novel condensin-interference system's structure and determine the biochemical methods it employs to limit plasmid replication. This thesis provides the first biochemical data showing that AvJetB purifies as a dimer and maintains its structural integrity throughout the purification process. This work lays
Figure 3-2

Azotobacter vinelandii Wadjet Overview

A. Schematic of the Wadjet defense system, highlighting proximity of jet genes to transposase genes. B. Alphafold generated models of Wadjet proteins (JetA, JetB, and JetC) alongside a theoretical condensin complex model. The theoretical condensin complex model was Adapted from “DNA looping in Transcriptional Regulation”, by BioRender.com (2020). Retrieved from https://app.biorender.com/biorender-templates.

Note. A. Schematic of the Wadjet defense system, highlighting proximity of jet genes to transposase genes. B. Alphafold generated models of Wadjet proteins (JetA, JetB, and JetC) alongside a theoretical condensin complex model. The theoretical condensin complex model was Adapted from “DNA looping in Transcriptional Regulation”, by BioRender.com (2020). Retrieved from https://app.biorender.com/biorender-templates.
the foundation for the future expression and purification of other Wadjet genes which allow for in depth biochemical characterization of Wadjet complexes. Overall, the study of nucleic acid condensin complexes will provide enhanced understanding of chromatin compaction and its effects on transcription and translation as well as an enhanced understanding of plasmid sharing within microbial communities.

Methods

Nucleic Acid Substrate Preparation

IDT and Eurofins synthesized nucleic acids. Nucleic acids, if labeled in-house, were labeled with a fluorescein label on the 5' end or were radiolabeled on the 5' end with (γ-32P) -ATP (Perkin Elmer) and T4 polynucleotide kinase (NEB). Radiolabeled nucleic acids were separated from excess 32P ATP with a MicroSpin G-25 column (GE Healthcare).

Single-stranded nucleic acids were then gel purified using 12% 7 M UREA PAGE gels, gel extracted, ethanol precipitated, and then reconstituted in water. Duplexed nucleic acids were generated by adding complementary oligonucleotides in an equimolar ratio in the presence of NEB buffer 2.1 and heated to 95°C. These oligonucleotides were slowly cooled to room temperature before running on 12-15% native PAGE gels. Duplex bands were then gel extracted, ethanol precipitated, and reconstituted in water.

Cloning of Cas12a2

SmCas12a2 was obtained from Benson Hill Biosystems and was codon-optimized for plant cell lines. SmCas12a2 mutants were created using the Q5 Site-Directed
Mutagenesis kit (New England BioLabs (NEB)). The N-terminal HIS tagged SuCas12a2 gene was obtained from Benson Hill Biosystems in a plant codon-optimized pet vector containing the CRISPR array that codes for a guide RNA. SuCas12a2 mutants were generated using the Q5 Site-Directed Mutagenesis kit (NEB). See Table B-3 for a list of primers and Table B-4 for constructs generated.

**Protein Expression & Purification**

**SmCas12a2 Expression & Purification**

The cloned vector containing the SmCas12a2 gene was transformed into *Escherichia coli* BL21 (DE3) chemically competent cells (Novagen). From the transformation plate, a colony was picked and placed into an overnight outgrowth with 0.1 mg/ml ampicillin. 500 mL of Luria-Bertani medium in a 2.8 L flask is inoculated with a 5 mL overnight starter. Cells are grown to an optical density between 0.4-0.5 OD₆₀₀ at 30°C 200 RPM in a shaking incubator. When OD₆₀₀ is reached, the cells were induced with 0.05 mM IPTG, and the temperature was dropped to 12°C for 16 hours. The cells were then harvested via high-speed centrifugation and stored at -80°C. All mutants were expressed in the same manner.

Cells were homogenized on ice with 8 mL/gram lysis buffer (25 mM NaPO₄ pH 7.4, 0.5 M NaCl, 10 % glycerol, 2 mM MgCl₂, 20 mM imidazole, 5 mM Tris [2-carboxyethyl] phosphine hydrochloride (TCEP), 0.1 % TWEEN 20). The protease inhibitors 1000 x Aprotinin (0.5mg/mL), 1000 x Leupeptin (0.5 mg/mL), 1000 x Pepstatin A (0.7 mg/mL), & 150x phenylmethylsulphonyl fluoride (PMSF) (25 mg/mL) and 0.25 mg/mL lysozyme were added prior to cell lysis. Cells were incubated on ice for
20 minutes. Probe sonication for cell lysis was performed at settings of 5/50. The lysate was then clarified by high-speed centrifugation at 15 K RPM for 15 minutes.

All purification steps were performed at 4°C. The supernatant was loaded onto a nickel affinity column (Nickel His trap FF crude, GE Healthcare) that had been pre-equilibrated with lysis buffer. The column was washed with lysis buffer and then eluted with elution buffer (25 mM NaPO₄ pH 7.4, 0.5 M NaCl, 10 % glycerol, 2 mM MgCl₂, 500 mM imidazole, 5 mM TCEP, 0.1 % TWEEN 20). Fractions with SmCas12a2 were pooled. Protein was loaded onto a desalting column (HiPrep 26/10 Desalting, GE Healthcare) that had been pre-equilibrated into a low salt buffer (25 mM NaPO₄ pH 7.4, 50 mM NaCl, 10 % glycerol, 2 mM MgCl₂, 5 mM TCEP, 0.1% TWEEN 20), and then SmCas12a2 fractions were run over the column. Protein elution was collected and then run over an IEX column (HiTrap SP HP cation exchange, GE Healthcare). The column was washed with low salt buffer. The protein was eluted into a high salt buffer (25 mM NaPO₄ pH 7.4, 1 M NaCl, 10% glycerol, 2 mM MgCl₂, and 5 mM TCEP, 0.1% TWEEN 20. Samples were spun down using a spin concentrator (Corning® Spin-X® UF 100 MWCO) for placement on a size exclusion column (HiLoad 26/600 Superdex 200 pg., GE Healthcare). Protein was eluted into a final buffer (25 mM HEPES pH 7.4, 100 mM NaCl, 2 mM MgCl₂, 1 mM DTT, 10% glycerol). Protein samples were concentrated and stored at 4°C.

Protein concentration was determined by UV-Vis spectroscopy (Thermofisher UV-vis Nanodrop), and the appropriate extinction coefficient was applied (Table B-1).

ApoSuCas12a2 Expression and Purification

The Sulfuricurvum sp. PC08-66 cas12a2 gene was obtained from Benson Hill
Biosystems in a plant codon-optimized pet vector with an N-terminal HIS tag.

SuCas12a2 mutants were created using the Q5 Site-Directed Mutagenesis kit (New England BioLabs (NEB)).

The cloned vector containing the sucas12a2 gene was transformed into Escherichia coli BL21 (DE3) chemically competent cells (Novagen). From the transformation plate, a colony was picked and placed into 1 L of prepared ZY Autoinduction media in a 2.8 L baffled flask (Studier, 2015). Cells are grown for 5 hours at 37°C 200 RPM in a shaking incubator. After 5 hours or when cells were turbid, the temperature was dropped to 24°C for 24 hours. The cells were then harvested via high-speed centrifugation and stored at -80°C. All mutants were expressed in the same manner.

Cells were homogenized on ice with 8 mL/gram lysis buffer (25 mM Tris pH 7.2, 0.5 M NaCl, 10% glycerol, 2 mM MgCl2, 10 mM imidazole, 1 mM Tris [2-carboxyethyl] phosphine hydrochloride (TCEP)). The protease inhibitors 1000 x Aprotinin (0.5 mg/mL), 1000 x Leupeptin (0.5 mg/mL), 1000 x Pepstatin A (0.7 mg/mL), & 150 x phenylmethylsulphonyl fluoride (PMSF) (25 mg/mL) and 0.25 mg/mL lysozyme were added prior to cell lysis. Cells were incubated on ice for 20 minutes. Probe sonication for cell lysis was performed at settings of 5/50. The lysate was then clarified by high-speed centrifugation at 16 K RPM for 35 minutes.

All purification steps were performed at 4°C. The supernatant was loaded onto a nickel affinity column (Nickel His trap FF crude, GE Healthcare) that had been pre-equilibrated with lysis buffer. The column was washed with 50-fold dilution of nickel elution buffer (25 mM Tris pH 7.2, 0.5 M NaCl, 10 % glycerol, 2 mM MgCl2, 500 mM
imidazole, 1 mM TCEP) and then eluted. Fractions with SuCas12a2 were pooled. Protein was loaded onto a desalting column (HiPrep 26/10 Desalting, GE Healthcare) that had been pre-equilibrated into a low salt buffer (25 mM Tris pH 7.2, 50 mM NaCl, 10 % glycerol, 2 mM MgCl₂, 1 mM dithiothreitol (DTT)), and then SuCas12a2 fractions were loaded onto the column. Protein elution was collected and run over an IEX column (HiTrap SP HP cation exchange, GE Healthcare). The column was washed with a 20-fold dilution of the high salt buffer. The protein was eluted into a high salt buffer (25 mM Tris pH 7.2, 1 M NaCl, 10 % glycerol, 2 mM MgCl₂, and 1 mM DTT) using a linear gradient of 5-100 % high salt buffer. Samples were spun down using a spin concentrator (Corning® Spin-X® UF 100 MWCO) for placement on a size exclusion column (HiLoad 26/600 Superdex 200 pg., GE Healthcare). Protein was eluted into a final buffer (100 mM HEPES pH 7.2, 150 mM KCl, 2 mM MgCl₂, 1 mM TCEP, 10 % glycerol). Protein samples were concentrated and stored at 4°C.

Protein concentration was determined by UV-Vis spectroscopy (Thermofisher UV-vis Nanodrop), and the appropriate extinction coefficient was applied (Table B-1).

**SuCas12a2+gRNA Expression and Purification**

The cloned vector containing the sucas12a2 gene and the CRISPR array was transformed into *Escherichia coli* BL21 NiCo (DE3) chemically competent cells (Novagen). From the transformation plate, a colony was picked and placed into an overnight outgrowth. 1 L of Terrific Broth pH 7.2, supplemented with 2 mL 1000 x Metals Mix and 1 mL MgSO₄, in a 2.8 L flask was inoculated with 10 mL overnight starter (Studier, 2015). Cells were grown to an optical density of 0.6 OD₆₀₀ at 37°C 200
RPM in a shaking incubator. When OD$_{600}$ is reached, the cells were induced with 0.1 mM IPTG (isopropyl B-D-1-thiogalactopyranoside), and the temperature was dropped to 18°C for 15 hours. The cells were then harvested via high-speed centrifugation and stored at -80°C. All mutants were expressed in the same manner.

Cells were homogenized on ice with 8 mL/gram lysis buffer (25 mM Tris pH 7.2, 0.5 M NaCl, 10 % glycerol, 2 mM MgCl$_2$, 10 mM imidazole, 1 mM Tris [2-carboxyethyl] phosphine hydrochloride (TCEP)). The protease inhibitors 1000 x Aprotinin (0.5mg/ml), 1000 x Leupeptin (0.5 mg/ml), 1000 x Pepstatin A (0.7 mg/ml), & 150 x phenylmethylsulphonyl fluoride (PMSF) (25 mg/ml) and 0.25 mg/ml lysozyme were added prior to cell lysis. Cells were incubated on ice for 20 minutes. Probe sonication for cell lysis was performed at settings of 5/50. The lysate was then clarified by high-speed centrifugation at 16 K RPM for 35 minutes.

All purification steps were performed at 4°C. The supernatant was loaded onto a nickel affinity column (Nickel His trap FF crude, GE Healthcare) that had been pre-equilibrated with lysis buffer. The column was washed with 50-fold dilution of nickel elution buffer (25 mM Tris pH 7.2, 0.5 M NaCl, 10 % glycerol, 2 mM MgCl$_2$, 500 mM imidazole, 1 mM TCEP) and then eluted. Fractions with SuCas12a2 were pooled. Protein was loaded onto a desalting column (HiPrep 26/10 Desalting, GE Healthcare) that had been pre-equilibrated into a low salt buffer (25 mM Tris pH 7.2, 50 mM NaCl, 10 % glycerol, 2 mM MgCl$_2$, 1 mM dithiothreitol (DTT), and then SuCas12a2 fractions were loaded onto the column. Protein elution was collected and run over an IEX column (HiTrap Q HP cation exchange, GE Healthcare). The column was washed with a 20-fold dilution of the high salt buffer. The protein was eluted into a high salt buffer (25 mM Tris
pH 7.2, 1 M NaCl, 10 % glycerol, 2 mM MgCl₂, and 1 mM DTT) using a linear gradient of 5-100 % high salt buffer. Samples were spun down using a spin concentrator (Corning® Spin-X® UF 100 MWCO) for placement on a size exclusion column (HiLoad 26/600 Superdex 200 pg., GE Healthcare). Protein was eluted into a final buffer (100 mM HEPES pH 7.2, 150 mM KCl, 2 mM MgCl₂, 1 mM TCEP, 10 % glycerol). Protein samples were concentrated and stored at 4°C.

Protein concentration was determined by UV-Vis spectroscopy (Thermofisher UV-vis Nanodrop), and the appropriate extinction coefficient was applied (Table B-1).

**Pre-crRNA Processing Assay**

692 nM SuCas12a2 and 25 picomoles or 300 CPM of pre-crRNA (80 nt.) were incubated at 37°C in NEB buffer 3.1. 10 µL aliquots were removed at time points between 0-60 minutes, quenched with 10 µL phenol-chloroform, mixed, added 5 µL of water followed by more mixing, and then centrifuged. The aqueous layer was removed from the phenol, mixed with RNA loading buffer, and resolved on a 12% Denaturing (7 M urea) PAGE gel with a 30 bp. marker. The gel was dried and then exposed to a phosphoscreen before being imaged on a Typhoon imager. Cleaved and uncleaved fractions were quantified using ImageLab software as a function of % band cleaved. The triplicate gels for analysis had samples that remained in the well, and this band was not included in gel analysis. The data is the average of at least three experiments, and the error bars represent the standard deviation from the mean. The ΔKRR_AAA mutant was assessed using the same methodology as wildtype.
DNA Binding Assays

5' $^{32}$P labeled DNA with a TTTC protospacer adjacent motif was exposed to increasing concentrations of SuCas12a2 RNP complex (0-10 µM) in NEB Buffer 3.1 and allowed to incubate for 30 minutes at 37 °C. Samples were then run on a 12% native PAGE gel. Bound and unbound bands were quantified using ImageLab software as a function of % band bound. Reported data is as follows: dsDNA binding was analyzed for two replicates, and the ssDNA was analyzed with only one replicate. Error bars, when present, represent the standard deviation from the mean. See Table B-2 for substrates.

DNA Cleavage Assays

5' FAM-labeled ssDNA with a TTTC PAM was exposed to 500 nM SuCas12a2 RNP complex in NEB Buffer 3.1. Time points were quenched in basic phenol-chloroform and then treated with RNAse A. Samples were removed from the top layer of the phenol-chloroform mixture, mixed with 2x formamide loading dye, and run on a 15% denaturing PAGE gel. See Table B-2 for substrates.

Negative Stain & Cryo-EM of SuCas12a2

Methodology for the negative stain and cryo-electron microscopy of SuCas12a2 are available through the David Taylor lab.

SAXS-SEC of Cas12a2

Dr. Robert Rambo performed small-angle x-ray scattering size exclusion chromatography (SAXS-SEC) at the Diamond Light Source. The complete methodology can be obtained from him. In brief, Cas12a2 was purified in the presence of 0.1%
TWEEN 20 and then desalted into a storage buffer (100 mM HEPES pH 7.2, 150 mM KCl, 1 mM TCEP, 1% glycerol, and 2 mM MgCl₂) and was placed over a Shodex KW-403 column for SAXS-SEC analysis.

**Negative Stain SmCas12a2**

Negative stain images of SmCas12a2 were generated through a collaboration between Hannah Domgaard and Dr. David Timm at the University of Utah’s Cryo-EM facilities. Hannah Domgaard prepared the protein samples. SmCas12a2 was desalted into NEB Buffer 3.1 prior to grid plating. The method below was written and developed by Dr. Timm.

Grids were glow discharged for 30 seconds. A 3.5 µL drop of the sample was applied to the grid for each sample. Samples were incubated on the grid at room temperature for 30 seconds. Grids were sequentially washed with 20 µL of NaCl & 20 µL of 150 mM NaCl (2-5 sec wash times), with filter paper blotting after each wash. Grids were stained with 10 µL 0.75% uranyl formate (UF) for 2-5 seconds, blotted, stained with another 10 µL 0.75% UF for 2-5 seconds, blotted, then allowed to dry in air. Imaging was on a T12 microscope. Search images were taken at 2,700 x magnification; exposure images were taken at 30,000 x magnification.

**Cloning of Wadjet**

Developed primers to PCR amplify Wadjet-associated genes out of organisms housed within the Department of Chemistry & Biochemistry at Utah State University. These organisms are *Mycobacterium MCS* and *Azotobacter vinelandii* DJ995. 5 µL of glycerol stock cell culture with 100 µL water were added together and then heat-treated
for 10 minutes at 95°C for use as a DNA substrate in the Q5 PCR kit from New England Biolabs. Ligation Independent Cloning (LIC) was used to transfer Wadjet genes into expression vectors. Reference Table B-5 for primers and Table B-6 for constructs generated.

**Expression & Purification of 2ST Av JetB**

The cloned vector containing the Sumo-HIS AvJetB gene was transformed into *Escherichia coli* BL21 Tuner (DE3) chemically competent cells (Novagen). From the transformation plate, a colony was picked and placed into an overnight outgrowth. 500 mL of Luria-Bertani medium in a 2.8 L flask was inoculated with 10 mL overnight starter. Cells were grown to an optical density between 0.6-0.7 OD$_{600}$ at 37°C 200 RPM in a shaking incubator. When OD$_{600}$ was reached, the cells were induced with 1 mM IPTG, and the temperature was dropped to 18°C for 18 hours. The cells were then harvested via high-speed centrifugation and stored at -80°C.

Cells were homogenized on ice with 5-8 mL/gram lysis buffer (25 mM Tris pH 7.4, 0.5 M NaCl, 10 % glycerol, 2 mM MgCl$_2$, 10 mM Imidazole, 1 mM DTT). The protease inhibitors 1000 x Aprotinin (0.5 mg/ml), 1000 x Leupeptin (0.5 mg/ml), 1000 x Pepstatin A (0.7 mg/ml), & 150 x PMSF (25 mg/ml) and 0.25 mg/ml lysozyme were added prior to cell lysis. Probe sonication for cell lysis was performed at settings of 5/50 for five minutes. The lysate was then clarified by high-speed centrifugation at 16 K RPM for 25 minutes.

All purification steps were performed at 4°C. The supernatant was loaded onto nickel resin in a gravity flow column that had been pre-equilibrated with lysis buffer. The
lysate was allowed to batch bind with the resin for 1 hour in the cold room on a shaking platform. The column was washed with wash buffer (25 mM Tris pH 7.4, 2 M NaCl, 10 % glycerol, 2 mM MgCl₂, 25 mM Imidazole, 1 mM DTT) and then washed with lysis buffer. Protein sample was then eluted with Ni elution buffer (25 mM Tris pH 7.4, 0.5 M NaCl, 10 % glycerol, 2 mM MgCl₂, 0.5 M Imidazole, 1 mM DTT). Fractions with AvJetB were pooled. Protein was loaded onto a desalting column (HiPrep 26/10 Desalting, GE Healthcare) that had been pre-equilibrated into a low salt buffer (25 mM Tris pH 7.4, 50 mM NaCl, 10 % glycerol, 2 mM MgCl₂, 1 mM DTT), and then AvJetB fractions were run over the column. Protein elution was collected and run over an IEX column (HiTrap Q HP anion exchange, GE Healthcare). The column was washed with 15% high salt buffer (25 mM Tris pH 7.4, 1 M NaCl, 10 % glycerol, 2 mM MgCl₂, 1 mM DTT). The protein was eluted into 100 % high salt buffer. Samples were spun down using a spin concentrator (Corning® Spin-X® UF 100 MWCO) for placement on a size exclusion column (HiLoad 26/600 Superdex 200 pg., GE Healthcare). Protein was eluted into a final buffer (50 mM HEPES pH 7.4, 75 mM KCl, 1 mM MgCl₂, 1 mM DTT, 5 % glycerol). Protein samples were concentrated and stored at 4°C.

Protein concentration was determined by UV-Vis spectroscopy (Thermofisher UV-vis Nanodrop), and the appropriate extinction coefficient was applied (Table B-1)

**Far-UV Circular Dichroism (CD)**

Far-UV CD spectra were collected for ApoSuCas12a2, SuCas12a2 RNP, SuCas12a2 ΔKRR_AAA, Sumo-HIS AvJetB, and AvJetB. All data were collected using a JASCO model J-1500 spectropolarimeter. CD spectra were collected from 190 to 260
nm at 10 °C using 0.1 cm quartz cuvettes, 0.5 nm data sampling, a 50 nm/min scan rate, and a 1-second data integration time. Measurements were performed at 0.5 (suCas12a2) and 0.79 (JetB) mg/mL protein concentration, and CD signals were normalized by converting to mean residue molar ellipticity. All proteins were placed in PBS for analysis. Results for SuCas12a2, N-Sumo-HIS AvJetB, were collected once, and cleaved tag AvJetB data were collected in triplicate.

Results

Protein Expression & Purification

To biochemically and structurally characterize Cas12a2 proteins, Apo and gRNA bound SmCas12a2 and SuCas12a2 were recombinantly expressed and purified from *E. coli* DE3 cells (Figure 3-3). Initially, concentrations of purified protein were sufficient for biochemical assays but not for x-ray crystallography techniques. This finding led to the pursual of alternative techniques that would provide structural data, such as size exclusion small-angle x-ray scattering (SAXS-SEC) and cryo-electron microscopy.

SAXS of Cas12a2

Protein samples were sent for small-angle x-ray scattering to provide low-resolution data on the conformational arrangement of protein samples, for instance to determine conformational differences between ApoCas12a2 and gRNA bound Cas12a2. SAXS-SEC analysis revealed that all Cas12a2 proteins were soluble but trapped in a detergent micelle, presumably TWEEN 20 (Figure B-1). This detergent was initially included in our protein preparations in an attempt to purify more soluble protein, as most
Figure 3-3

Protein Purification of Type V-A2 Proteins

Note. SDS-PAGE gels showing purified Apo SuCas12a2, SuCas12a2 bound to a guide RNA, and Apo SmCas12a2.
of our expressed protein at the time remained in the cell pellet after lysis. It was unanticipated for our purified protein to be trapped in a micelle. This information led to the development of alternative Cas12a2 protein purification techniques.

**Negative Stain SmCas12a2**

SmCas12a2 was prepared for negative stain analysis to assess whether cryo-electron microscopy could be used to solve the molecular structure. Negative stain data collected at the University of Utah show that the SmCas12a2 RNP complex is a monomer in solution, with particles ranging from 7 nm to larger. The monodisperse particles show that SmCas12a2 is a viable candidate for cryo-electron microscopy (Figure 3-4).

**Cryo-EM of SuCms1**

Collaboration with the David Taylor lab at UT-Austin led to the generation of negative stain data, 2-D forward projections, and a low resolution 4 Å electron density map of the SuCas12a2 RNP complex (Figure 3-5). The 2-D forward projections allowed for comparison of ApoSuCas12a2 and SuCas12a2 RNP, showing a global conformational rearrangement when bound to the guide RNA. Model building into the 4 Å map resulted in the placement of identifiable secondary structure elements such as alpha helices and an A-form RNA into density representative of where the direct repeat portion of the guide RNA is bound. The model building of the RNP complex halted when it became apparent that the connectivity of the electron density would be insufficient to place a backbone of the structure through the density. Attempts to dock Cas12a models into the density did not provide enough structural insight to continue modeling the RNP complex.
**Figure 3-4**

*SmCas12a2 RNP Negative Stain*

*Note.* Purified SmCas12a2 was prepared for negative staining using uranyl formate. Using a T12 microscope, search images were taken at 2,700x magnification, exposure images were taken at 30,000x magnification.
Figure 3-5

**SuCas12a2 Low Resolution Map and Model**

Note. A. Negative stain and 2D particle classification of Apo and RNP SuCas12a2 generated by the Taylor lab. B. 2D forward projections of Apo and RNP SuCas12a2 generated by the Taylor lab. C. Cryo-electron density generated maps with modeled secondary structure and A-form RNA.
Pre-crRNA Processing

One of the critical stages for CRISPR immunity is processing a long pre-crRNA into smaller guide RNAs. Without this step, no ribonucleoprotein complexes will assemble which means that no targeting of foreign nucleic acid will occur. Using denaturing PAGE gels and a radiolabeled substrate, we show that ApoSuCas12a2 will process an 80 nt. pre-crRNA into a 30 bp gRNA (Figure 3-6). In-gel densitometry analysis shows that SuCas12a2 processed ~30% of the pre-crRNA. It is possible that the substrate utilized was non-optimal, as it was not a full pre-crRNA transcript composed of multiple direct repeat and spacer sequences. Alternatively, it is possible that the number of active SuCas12a2 molecules was low given that we utilized upwards of 500 nM protein. Dr. Crowley performed analysis of the processed gRNA and identified the exact cleavage site via a reverse transcriptase PCR reaction. She found that cleavage occurs upstream of the 2nd predicted RNA stem-loop in the direct repeat (Dmytrenko et al., 2022).

To further understand the mechanism of biogenesis used by SuCas12a2, we identified the biogenesis active site in the N-terminal region of the protein. Cas12a2 proteins share sequence similarities in the N-terminus, knowing this we generated an I-TASSER homology model, which referenced the PDB deposition 5NG6 of FnCas12a bound to RNA (Figure B-2) (Yang et al., 2014; Swarts et al., 2017). FnCas12a utilizes histidine and lysine residues to process a guide RNA. Knowing that basic amino acids form a biogenesis active site in Cas12a, we looked at the N-terminal region of the I-TASSER Cas12a2 homology model and identified lysine and arginine residues that could play a role in SuCas12a2 biogenesis (Figure 3-7). The residues K791, R792, and R795
**Figure 3-6**

*SuCas12a2 crRNA Processing and Secondary Structure Validation*

**Note.** A. Denaturing PAGE analysis shows that SuCas12a2 processes a guide RNA. B. Graph depicts % guide RNA cleaved as a function of time. C. Denaturing PAGE analysis shows that mutation of residues KRR into AAA abolishes processing activity. D. Bar graph analysis depicts % guide RNA cleaved between WT SuCas12a2 and ΔKRR_AAA SuCas12a2. E. Far UV Circular Dichroism analysis shows that ΔKRR_AAA SuCas12a2 maintains the same level of structural integrity as SuCas12a2.
**Figure 3-7**

*Model Used for the Development of RNA Processing Mutants*

A. FnCas12a + gRNA 5NG6  
B. SuCas12a2 + gRNA 5NG6

*Note.* A. 5NG6 FnCas12a bound to an RNA guide with the RNA processing residues highlighted in green. B. ITASSER homology model of SuCas12a2 modeled with the RNA from 5NG6 with the residues presumed to be important for RNA processing in green.
were selected and mutated to alanines based on their position in the active site as well as their charge. The secondary structure of this mutant was assessed via circular dichroism. RNP. The triple mutant was assessed for activity alongside WT SuCas12a2 and was deficient in processing the long pre-crRNA, with less than 20% of the pre-crRNA being cleaved, confirming the role of the KRR residues in biogenesis. This work is significant as the residues responsible for biogenesis, a key step in CRISPR immunity, were identified.

**Far-UV Circular Dichroism (CD) of SuCas12a2**

Far-UV circular dichroism (CD) was performed on ApoSuCas12a2, SuCas12a2 RNP complex, and SuCas12a2 ΔKRR_AAA to assess structural integrity. All showed similar CD profiles, which indicated a largely alpha-helical structure. Furthermore, the similarities between spectra suggest that the proteins remain similarly folded and that the lack of activity we see for the ΔKRR_AAA in the crRNA processing assays is a result of the amino acid mutations and not a result of disrupting the secondary structure of the protein (Figure 3-6). It is from these results that we can have confidence in our bioprocessing residue selections and purification methods.

**DNA Binding Assays**

To understand how Cas12a2 could generate dsDNA breaks in *Oryza sativa*, we needed to understand what the substrates for the ribonucleoprotein complex were (Begemann et al., 2017). This involved *in vitro* studies of the interference aspect of CRISPR immunity. For interference to occur, binding to a nucleic acid substrate must
occur, so DNA binding assays using 5’ P$^{32}$ labeled dsDNA and 5’ FAM-labeled ssDNA on native PAGE gels were conducted. Data show that Cas12a2 can bind both single stranded and double-stranded DNA with a TTTC PAM. The data collection and analysis were not performed in triplicate as DNA nuclease activity sporadically appeared; thus, a $K_d$ analysis could not be performed. However, visual analysis indicates that high concentrations of Cas12a2 in the upper nanomolar to lower micromolar are needed for dsDNA binding (Figure 3-8).

**DNA Cleavage Assays**

After substrate binding, interference is typically achieved via cleavage event. It is for this reason that in vitro DNA cleavage assays were utilized to determine the molecular mechanism of SuCas12a2 interference. Native PAGE gels intended to show binding of SuCas12a2 to 5’P$^{32}$ labeled duplexed nucleic acid showed both binding and cleavage (Figure 3-9). However, this cleavage activity was sporadic and protein preparation dependent. Denaturing PAGE gels assaying cleavage of single-stranded DNA with a TTTC PAM also showed cleavage; yet again, this cleavage activity was sporadic and protein preparation dependent (Figure 3-9). Attempts to determine exact cleavage conditions were unsuccessful and included the use of a variety of buffers, DNA substrates, concentrations of substrates and proteins, times, reductants, and RuvC mutant analysis (data not shown). Under no condition tested did were consistent cleavage conditions identified. It was later revealed in work performed by Dylan Keiser, that the first interference substrate of Cas12a2 is RNA and that after target RNA is bound by the ribonucleoprotein complex, indiscriminate cleavage of ssRNA, ssDNA, and dsDNA can
Figure 3-8

*SuCas12a2 Binds DNA Substrates*

**A.** Native electromobility shift assays show that SuCas12a2 RNP can bind both 5’ P32 single stranded and duplexed DNA substrates. B. Graph shows % bound double-stranded DNA as a function of SuCas12a2 concentration. Error bars represent standard deviation between two data sets. C. Graph shows % bound single stranded DNA as a function of SuCas12a2 concentration. There are no error bars as only one gel was sampled. 

*Note.* A. Native electromobility shift assays show that SuCas12a2 RNP can bind both 5’ P32 single stranded and duplexed DNA substrates. B. Graph shows % bound double-stranded DNA as a function of SuCas12a2 concentration. Error bars represent standard deviation between two data sets. C. Graph shows % bound single stranded DNA as a function of SuCas12a2 concentration. There are no error bars as only one gel was sampled.
Figure 3.9

*SuCas12a2 Cleaves and/or Degrades DNA*

**Note.** A. Native electromobility PAGE analysis intended to show binding to double stranded 5’ P32 DNA. Increasing concentrations of SuCas12a2 RNP correlate with increasing band intensity of a cleaved/degraded product. B. Denaturing PAGE gel showing cleavage of a 5’ FAM labeled single stranded DNA substrate.
occur (Dymetrenko et al., 2022).

**Wadjet Cloning, Protein Purification, & Structure Prediction**

Prokaryotic species containing Wadjet defense systems were identified in the department using the extensive bioinformatic search performed by the Sorek lab (Doron et al., 2018). All three genes from *Azotobacter vinelandii DJ* and *Mycobacterium MCS* were successfully cloned into LIC expression vectors using LIC cloning (Table B-5; B-6). The Sumo-HIS AvJetB was the only Wadjet gene for which expression and purification conditions were established. SDS-PAGE analysis shows that Sumo-HIS AvJetB purifies at a molecular weight of 48 kDa, and after TEV tag cleavage, AvJetB runs at a molecular weight close to 25 kDa. Size exclusion chromatography shows that cleaved tag AvJetB elutes with an elution volume equivalent to a protein near 48 kDa, showing that AvJetB purifies as a dimer. This is the first *in vitro* data showing dimerization of a Wadjet associated protein. Circular dichroism of tagged and cleaved tag AvJetB shows no loss of secondary structure after TEV cleavage of the large SUMO-HIS tag (Figure 3-10).

**Discussion**

The type V-A2 Cas12a2 proteins had previously been shown to cause insertion/deletion events in *Oryza Sativa*. However, the mechanism behind the targeted double stranded DNA breaks was unknown. The work performed in this thesis provided some of the first attempts to characterize Cas12a2 proteins *in vitro*. Initial attempts to structurally analyze Cas12a2 proteins from *Smithella* and *Sulfuricurvum* sp. PC08-66 resulted in SAXS data confirming the presence of micelles in protein purifications. This
**Figure 3-10**

*Recombinantly Expressed and Purified AvJetB is a Dimer*

**A.** SDS PAGE gel of N-terminal Sumo-HIS tagged AvJetB (~48 kDa.) off a size exclusion column. **B.** SDS PAGE of TEV treated N-term. Sumo-HIS AvJetB (Cleaved Tag; ~25 kDa.) off a size exclusion column. **C.** Far-UV Circular dichroism showing the maintenance of structural integrity between tagged and cleaved tag JetB. **D.** Size exclusion chromatogram of cleaved tag AvJetB eluting with an elution volume equivalent to a dimer ~49 kDa.

*Note.* A. SDS PAGE gel of N-terminal Sumo-HIS tagged AvJetB (~48 kDa.) off a size exclusion column. B. SDS PAGE of TEV treated N-term. Sumo-HIS AvJetB (Cleaved Tag; ~25 kDa.) off a size exclusion column. C. Far-UV Circular dichroism showing the maintenance of structural integrity between tagged and cleaved tag JetB. D. Size exclusion chromatogram of cleaved tag AvJetB eluting with an elution volume equivalent to a dimer ~49 kDa.
led to the removal of detergents from future protein protocols. Negative stain analysis of SmCas12a2 showed a monodisperse particle optimal for cryo-electron microscopy techniques. As expression and purification conditions for SmCas12a2 have been identified, this protein should be selected for future cryo-electron microscopy pursuits as it may harbor differences from SuCas12a2.

SuCas12a2 structural analysis led to the development of negative stain data, 2D forward projections showing global conformational rearrangements between Apo and RNP complex, and a low-resolution map in which some secondary structures could be modeled. Although the attempts to build a model of SuCas12a2 were insufficient here, the work discussed in this thesis provided the foundation for others in the Jackson and Taylor labs to generate a high-resolution map of SuCas12a2 bound to nucleic acid substrates (Bravo et al., 2022).

In vitro analysis geared towards understanding the bioprocessing stage of CRISPR immunity showed that SuCas12a2 processed a long pre-crRNA into a short 30 bp guide RNA. Although in-gel densitometry revealed that only ~30% of substrate was cleaved. This low cleavage rate could be because the pre-crRNA substrate used in these assays was a truncated pre-crRNA composed of only one direct repeat sequence and spacer, rather than the biologically relevant pre-crRNA composed of multiple direct repeats and spacers. Alternatively, protein preparation, storage, and reaction concentrations could have played a role in the low percent cleavage. Perhaps using non-frozen protein or higher concentrations of protein would have resulted in higher percentages of cleaved nucleic acid. Successful visualization of bioprocessing led to the identification of the bioprocessing active site.
The amino acids KRR (751, 752, and 755 respectively) were identified as catalytic candidates in the biogenesis active site and were found to be vital for processing activity as mutating these residues into alanines substantially lowered RNA cleavage. Dr. Val Crowley demonstrated later through reverse transcriptase PCR that the pre-crRNA is processed between two adenosine residues on the 5’ end of the second hairpin of the direct repeat (Dmytrenko et al., 2022). The work performed here is significant, as it has elucidated the mechanism of bioprocessing for SuCas12a2. Additionally, if Cas12a2 is to be repurposed as a tool for gene therapeutics or genome engineering, knowledge of gRNA requirements is needed in order to correctly assemble the tool in vitro.

After determining the parameters for gRNA processing, the requirements for CRISPR interference were explored through DNA binding and cleavage experiments. In vitro analysis showed that SuCas12a2 could bind both duplexed and single-stranded DNA, although binding affinity appeared to be in the hundreds of nM, suggesting that the PAM sequence used in our assays was non-optimal, or that an alternative binding substrate was preferred. Additionally, binding activity of SuCas12a2 to DNA was intermittently accompanied by nuclease activity, resembling cleavage products of CasX, a fellow member of the type V family of proteins (Liu et al., 2019). This nuclease activity sometimes resembled DNA degradation when witnessed with dsDNA and a single cut site when with ssDNA. However, none of the cleavage activity was consistently reproducible between protein purifications leading us to believe that we had either purified contaminating nucleases or had not identified the appropriate conditions to determine the biochemical parameters for DNA cleavage.

The SuCas12a2 project was passed from my hands to Dylan Keiser two years ago.
Since then, Dylan Keiser was able to identify the biologically relevant RNA substrate and show RNA-guided RNAse activity (Dmytrenko et al., 2022). He utilized RuvC nuclease motif mutants to break RNAse activity and demonstrated that after SuCas12a2 binds and cleaves the target RNA substrate, it will begin cleaving DNA and RNA indiscriminately. From his work, the data behind the indiscriminate DNA cleavage in the Jackson lab began to make sense. Perhaps DNA would get caught in a suitable conformation for cleavage without a target RNA first being bound. Perhaps there was prep-dependent activity in which proteins would be purified with contaminating RNAs that SuCas12a2 could bind and use to turn on the indiscriminate DNAse activity. The study of SuCas12a2 is vital as it provides a new tool for medical scientists, genome engineers, and researchers to harness the power of a reprogrammable nuclease capable of targeting specific RNA sequences.

Wadjet systems have been found to inhibit plasmid transformation in B. subtilis, however the mechanism behind that inhibition was unknown, additionally structures of Wadjet system proteins are unknown as well. Initial homology searches of these proteins suggested secondary structure similarities with condensin complexes, suggesting roles in DNA binding and segregation. It is with this knowledge that the Wadjet systems from A. vinelandii DJ and Mycobacterium MCS were cloned into LIC expression vectors. The AvJetB protein was successfully expressed and purified through the use of affinity and size exclusion chromatography. Size exclusion chromatography revealed that AvJetB was purified as a dimer. Dimerization was not unanticipated, as condensin complexes typically form hetero and homo dimeric complexes (Rybenkov et al., 2014). This work supports the hypothesis that Wadjet systems will form similar dimeric complexes. The
cloning performed on the *jet* genes set the groundwork for a successful project studying a plasmid inhibition system as rapid testing of expression conditions could result in the purification of the JetA and JetC proteins. With all protein components assembled, proper study of the Wadjet complex could be pursued through ATPase and DNA binding assays.

In conclusion, the study of Cas12a2 has laid down important foundational work on variant CRISPR immune systems. If these systems are studied further, they could provide insight into microbial immune systems and provide new tools for nucleic acid recognition and manipulation. Specifically, there are many diseases whose roots are due to genetic defects. The use of a reprogrammable nuclease, such as Cas12a2, could provide alleviation from genetic disease, such as cystic fibrosis, through gene therapy (Maule et al., 2020). Although there are many CRISPR-Cas proteins that are already being pursued for such ventures, Cas12a2 may provide a more suitable tool in certain conditions. The continued study of JetB and the other *jet* genes, could result in knowledge regarding plasmid control and sharing within microbial communities. This information can provide context for why some plasmids are successfully shared within a community rather than others.

**Acknowledgments**

I want to acknowledge Dr. Val Crowley, who determined the location of the pre-crRNA processing site, generated mutants, and provided critical experimental feedback and assistance. Dr. Crowley also assisted in the generation of mutants for various Cas12a2 constructs. I would like to thank Dylan Keiser for contributing to the development of mutants used in various constructs and for his later work on determining
the actual nucleic acid target of Cas12a2, as it provided explanations for the data I collected. Thanks to David Timm for his assistance in generating negative stain data for SmCas12a2, to the Taylor lab for their work on the preliminary SuCas12a2 cryo-electron structure and current work on SuCas12a2 models, and to Robert Rambo for his collection of SAXS data. Thank you to Benson Hill Biosystems, who provided gene expression constructs and valuable insight into the Cas12a2 systems. I thank the Seefeldt lab for generously contributing a sample of *Azotobacter vinelandii* DJ for use in our Wadjet studies. Figures were created with BioRender.com

References


Dmytrenko, O., Neumann, G., Hallmark, T., Keiser, D., Crowley, V., & Vialletto, E. et al. (2022). Cas12a2 elicits abortive infection via RNA-triggered destruction of double-stranded DNA. https://doi.org/10.1101/2022.06.13.495973


CHAPTER 4
SUMMARY & FUTURE DIRECTIONS

Introduction

The work presented in this thesis provides a basic characterization of the type IV-A CasDinG enzyme, presents preliminary data on the type IV-A RNP complex, and lays a foundation for understanding the type V-A2 protein SuCas12a2 and the Wadjet complex from Azotobacter vinelandii.

In Chapter 1, three microbial immune systems were introduced, the type IV-A CRISPR-Cas system, the type V-A2 CRISPR-Cas system, and the Wadjet system. CRISPR-Cas systems are adaptive and heritable prokaryotic immune systems that defend against mobile genetic elements, while Wadjet systems have been shown to defend against plasmids. The motivation for studying these systems was to discover their biochemical mechanisms and structures through in vitro analyses, which would provide key insight into how prokaryotes defend against infectious viruses and proteins. The work presented in this thesis provides pivotal information on three prokaryotic immune systems which could result in new genome editing tools and an enhanced understanding of the diversity of microbial immune systems.

In Chapter 2, a detailed investigation of the CasDinG protein was conducted, resulting in a thorough biochemical characterization of the first 5’-3’ ATP dependent CRISPR-associated helicase as well as details the first structure of a CRISPR-associated DinG helicase. Furthermore, size exclusion chromatograms and SDS PAGE showed the formation of a transient complex between CasDinG and the type IV-A RNP complex.
This is the first data showing that CasDinG is a functional helicase, providing key information on a protein which is essential for type IV-A CRISPR system functionality.

In Chapter 3, the initial biochemical and structural characterizations performed on type V-A2 Cas12a2 proteins were presented. These data revealed that SuCas12a2 can process a gRNA from a long pre-crRNA utilizing lysine and arginine residues in the bioprocessing active site. Furthermore, cryo-EM data showed that SuCas12a2 undergoes a global conformational rearrangement upon gRNA binding. This work laid the foundation for others to decipher the biochemical basis of CRISPR interference in type V-A2 systems. With this knowledge, SuCas12a2 may be repurposed for use as an RNA-guided nuclease, adding another Cas protein to the genome editing toolbox. Work on the type V systems was followed by the presentation of preliminary data on the Wadjet system *A. vinelandii*. Size exclusion chromatogram analysis showed evidence of AvJetB dimer formation which supports the hypothesis the Wadjet systems are related to chromatin condensin complexes. Additionally, the cloning performed on these Wadjet systems provides a path for others to characterize Wadjet proteins from either *Azotobacter vinelandii* or *Mycobacterium* MCS, which could provide novel information on plasmid defense systems.

**Type IV-A Summary & Future Directions**

Assays which assessed nucleic acid binding preferences, ATPase hydrolysis rates, helicase polarity, and *in vivo* functionality were utilized in order to provide basic characterization of this enzyme.

To understand CasDinG nucleic acid preferences, binding assays were performed
which showed that CasDinG could bind both DNA and RNA with similar affinities. DNA binding assays with ATP and analogs showed that the binding affinity of CasDinG for ssDNA did not change substantially, indicating that ATP binding does not enhance affinity for nucleic acid substrates.

Binding assays were followed by ATPase assays to determine whether CasDinG ATPase activity was stimulated by nucleic acid. Michaelis-Menten kinetics revealed that CasDinG hydrolyzes 32 ATP sec$^{-1}$ with ssDNA and 22 ATP sec$^{-1}$ with PT-RNA substrates. These DNA ATPase rates are similar to values reported for *E. coli* DinG (Voloshin et al., 2003). The phosphorothioated RNA (PT-RNA) substrate was utilized rather than an RNA substrate when we encountered RNAse activity in our assays. The phosphorothioated backbone protects against contaminating nuclease activity (Putney et al., 1981). However, with the added protection of the RNA backbone, there is increased steric strain from the addition of the sulfur group on the phosphate backbone resulting in a less flexible backbone (Zhang et al., 2012). Whether or not the phosphorothioated backbone inhibits hydrolysis of ATP will have to be assessed in the future through either an RNA substrate in RNAse-free conditions or by incorporating fewer phosphorothioates into the RNA substrate used as it is only with additional experimentation that a true comparison of RNA to DNA ATPase rates could be evaluated.

Future ATPase assays could be performed in the presence of type IV-A RNP and target nucleic acid sequences to assess whether there are increased rates of ATP hydrolysis by CasDinG. Cas3 helicase/nuclease proteins have shown increased ATP hydrolysis rates when introduced to the R-loop formed between the Cascade RNP complex and the target sequence (Mulepati & Bailey, 2013).
CasDinG can bind both ssDNA and ssRNA and hydrolyze ATP in their presence but will only unwind duplexed nucleic acids, which contain a DNA overhang, despite binding and hydrolyzing ATP in the presence of RNA. This data suggests that a more thorough analysis of CasDinG function in regard to ssRNA may be prudent. Perhaps, in addition to being a DNA helicase, CasDinG may be performing another functional role with small RNAs.

CasDinG was identified as a 5'-3' ATP-dependent helicase capable of utilizing an array of divalent ions to displace duplexed nucleic acid, similar to other DinG enzymes (Voloshin et al., 2003). Mg\(^{2+}\) is typically used in helicase reactions as it can be bound and coordinated by the aspartate residue in the conserved Walker B (DEAH) helicase motif, while the adjacent glutamate can function as a catalytic residue acting on the phosphate backbone of ATP, which is being coordinated by the Mg\(^{2+}\) (Frick et al., 2007). The research presented here shows that CasDinG can unwind nucleic acid using Mg\(^{2+}\), Mn\(^{2+}\), Ca\(^{2+}\), Ni\(^{2+}\), and Co\(^{2+}\) but not Zn\(^{2+}\) or Cu\(^{2+}\). ATPase assays should be performed using zinc and copper to rule out the lack of ATP hydrolysis as the reason for no unwinding. If ATP can be hydrolyzed in the presence of Zn\(^{2+}\) or Cu\(^{2+}\), then the lack of displacement may be protein destabilization or, alternatively, hyper-stabilization of duplexed nucleic acid (Voloshin et al., 2003).

Time-course analysis of helicase activity showed quick unwinding of substrates using 15 nM substrate and 25 nM protein. The quick unwinding suggests that a stop-flow assay may be prudent in order to obtain more data points with various percentages of substrate unwound. Alternatively, one could generate longer duplexes and perform native PAGE analysis with the hope that longer duplexes would require longer incubation.
periods with CasDinG to be fully unwound.

In depth characterization of CasDinG as a helicase would provide information on parameters such as a real $k_{\text{unwind}}$. To perform a real $k_{\text{unwind}}$ analysis, one would need to set up time constraints and then perform helicase reactions in triplicate for multiple concentrations of CasDinG. This data would provide information on the rate of unwinding for a given substrate, allowing the comparison of unwinding rates between DinG enzymes and DinG substrates. This would allow for statements to be made in regard to whether CasDinG preferentially unwinds duplex DNA or hybrid duplexes.

Some DinG enzymes have been found capable of unwinding G quadruplexes and are recruited to DNA replication forks (Thakur et al., 2014). With this in mind, complex nucleic acid structures such as Holliday junctions could be assessed as substrates for CasDinG helicase activity. Additionally, the minimum length of single-stranded nucleic acid needed as a loading zone for helicase activity should be tested, perhaps a longer loading or shorter loading zone increases or decreases helicase function. This data would provide crucial information on the requirements for CasDinG function in regard to type IV-A CRISPR-Cas system immune function.

Mutational analysis of the accessory domains and Walker B motif revealed that some domains are essential for robust ATPase activity (vFeS & Walker B motif) and helicase activity (Walker B motif, Arch). The vFeS and Walker B mutants resulted in inhibited ATPase rates in comparison to WT. Interestingly the vFeS mutant displayed fairly robust helicase activity despite inhibited rates of ATP hydrolysis. The arch and walker B mutations box displayed less unwinding in comparison to WT. Data regarding the N-terminal truncation was ambiguous as different preparations of this mutant showed
both unwinding and the lack thereof. Further biological replicates of each mutant are advised, perhaps with adjustments to the size of the domain knock outs.

To provide enhanced understanding of the function of each accessory domain, studies should migrate from whole domain mutations to single amino acid mutations. Point mutations would require the identification of residues within each of the accessory domains which could contribute to the ATPase, DNA binding, or DNA unwinding activities of CasDinG. For instance, there is a glutamate aspartate pair at positions 271 and 272, which reaches into the nucleic acid binding pocket of CasDinG. These acidic residues could be mutated to alanines and then assayed for their potential role in nucleic acid remodeling, whether through direct interactions with nucleic acid or stabilization of the ATP hydrolysis active site.

The structure of CasDinG was solved using x-ray crystallography and molecular replacement. The structure revealed two accessory domains, the arch and vFeS and there was no electron density to describe the N-terminus of the protein. Overall, structural comparison of CasDinG to E. coli DinG revealed structural conservation of the helicase core with slight modifications to the arch domain and confirmation of the lack of coordinated Fe atoms in the vFeS domain. In the future, crystallization of a truncated CasDinG structure containing only the N termini, or CasDinG bound to substrate allowing for the N terminus to be resolved, would provide structural information on a domain critical for immune system function. Expression conditions would have to be evaluated as attempts to purify just the N-terminus using a WT CasDinG protein expression profile were previously unsuccessful.

CasDinG will co-express and co-purify with the type IV-A RNP complex. The
exact nature of their association is unknown as co-purification shows a transient complex. Whether this transient complex is based on protein-protein contacts or through shared nucleic acid substrates is unknown. *In vitro* assembly of purified type IV-A RNP complex and CasDinG in the presence of various nucleic acid substrates followed by size exclusion chromatography could provide information on the recruitment of CasDinG to complex; alternatively, chemical cross-linking could be used to determine if there are protein contacts in the absence of nucleic acid substrates. This has been done with other type IV-A RNP complexes and DinG proteins successfully and may reveal the true nature of the Pa83 CasDinG and type IV-A RNP association (Özcan et al., 2018).

Future structural analysis of CasDinG should include the crystallization of CasDinG in the presence of substrates such as single-stranded DNA, single-stranded RNA, ADP-BeF$_3$, and perhaps even a stalled unwinding complex. Perhaps these substate CasDinG structures could be assisted by the ApoCasDinG crystallization conditions that provided crystal hits, Table A-7. The next step would be to obtain a type IV-A RNP complex structure and/or a structure that included CasDinG with the RNP complex. It may be beneficial to work towards performing cryo-electron microscopy with these large complexes as that has previously been successful with other CRISPR-Cas complexes isolated in the Jackson lab (Taylor, 2021; Zhou et al., 2021).

In order to more fully understand the *P. aeruginosa strain 83* type IV-A RNP complex, binding affinities for ssDNA with various PAM sequences were identified. All ssDNA PAM sequences tested resulted in low nM binding affinities. As the *in vivo* assays test for plasmid targeting, the assessment of binding affinities for duplexed nucleic acid targets should be a top priority in future studies. Additionally, the type IV-A RNP
complex showed indiscriminate plasmid nicking/linearization regardless of the presence of CasDinG. Further in vitro assays are needed to determine whether this activity is the result of contaminating nucleases or whether there are catalytic residues within the type IV-A complex capable of causing breaks in nicked or duplexed nucleic acid. If the type IV-A complex is indeed causing cleavage of nucleic acid, that would contrast with a recent publication showing type IV-A immune function without DNAse activity (Guo et al., 2022). Alternatively, if the type IV-A system houses nuclease activity, it may do so as a requirement for CasDinG loading onto ssDNA. Knowledge of any nuclease activity associated with the type IV-A immune system would provide critical data on the interference phase of CRISPR immunity.

Overall, the studies performed on the type IV-A CRISPR-Cas system from P. aeruginosa strain 83 in this thesis reveal a model in which the CasDinG enzyme acts a 5’ 3’ ATP dependent helicase which could possibly be recruited to an R-loop generated by the type IV-A RNP complex binding to a target plasmid. CasDinG could then perhaps knock off RNA transcripts or proteins bound near the R-loop resulting in CRISPR interference.

**Cas12a2 Summary & Future Directions**

The purification of multiple Cas12a2 proteins was discussed in this thesis and SEC-SAXS data prompted changes to the purification protocols so that there would be no contaminating detergents, trapping Cas12a2 proteins in micelles. This protocol change allowed future Cas12a2 protein purifications to be more accurately assessed for activity in vitro.
Denaturing PAGE and radiolabeled substrates were used to show that SuCas12a2 processes a guide RNA and relies upon the KRR basic amino acids (positions 791, 792, and 795) in the biogenesis active site for RNA processing. This information is important as this information describes a pivotal step in CRISPR immunity, without guide RNA processing there will be no targeted CRISPR interference.

We believed that Cas12a2 targeted DNA as previous work by collaborators showed DNA editing in *Oryza sativa* (Begemann et al., 2017). It was with this knowledge that PAGE analysis with radiolabeled DNA substrates were used to show that SuCas12a2 can bind ssDNA and dsDNA. Sometimes these assays coincided with DNAse activity and other times no DNAse activity was visualized. Because of the variability of protein preparations, no real binding affinities were able to be assessed.

To determine whether SuCas12a2 was responsible for the DNAse activity seen in the binding assays, nuclease assays were performed. The breadth of the nuclease assays performed with target DNA was not discussed in detail in this thesis, however it should suffice to say that temperatures, buffers, protein concentrations, and protein preparation methods were all tested. It was thought that there was either a contaminating nuclease responsible for the DNA cleavage seen in the binding assays or that we had not identified the correct *in vitro* conditions for optimal target DNA cleavage. Later work performed by Dylan Keiser explained the DNA nuclease activity. He found that RNA was the primary Cas12a2 substrate, which, when bound, would activate indiscriminate DNA and RNA cleavage using the catalytic RuvC motifs (Dmytrenko et al., 2022).

In the future a thorough biochemical analysis of SuCas12a2 determining binding affinities and cleavage rates should be performed. This information would be pivotal in
order to utilize SuCas12a2 as a tool for cell editing applications. For this to be accomplished, binding affinities using nuclease deficient RuvC motif mutants will need to be established for the target RNA sequence, and then DNA and RNA in trans. All mutants generated by me, and others should be analyzed via circular dichroism to ensure that the activity loss is due to the mutation itself and not the widespread loss of secondary structure.

This thesis described preliminary structural analyses of Cas12a2 proteins. Negative stain data and 2D classifications of ApoSuCas12a2 and SuCas12a2 RNP showed a global conformation rearrangement of gRNA bound protein. Additionally, a low-resolution map (4 Å) was generated for the RNP structure. This preliminary map and structural work laid a foundation for the successful determination of a higher resolution map and model prepared by Thom Hallmark and the Taylor lab (Bravo et al. 2022). This high-resolution structure is currently being prepared for publication. Future attempts to solve alternative structures of SuCas12a2 in apo form and as a ternary complex with either DNA or RNA should be pursued.

The bioinformatics analysis performed by Benson Hill Biosystems highlighted other Cas12a2 variants that may be of interest (Begemann et al., 2017). The biochemical characteristics of ObCas12a2 may differ from that of SuCas12a2 and may provide an alternative tool to the genome editing toolbox by highlighting an enzyme that is more or less specific for particular nucleic acid substrates. Furthermore, the Smithella sp. SCADC and Microgenomates sp. Cas12a2 genes are of interest as they are directly adjacent to Cas12a genes and share a CRISPR array. An avenue of study that would be beneficial in a broader sense would be to determine the biochemical basis of adaptation for the Type
V-A2 systems, as there is no current data on adaptation from Type V systems.

Purification of SmCas12a2 has already been performed and resulted in the collection of negative stain data showing monodisperse particles, implying that it is a great candidate for future cryo-EM experiments. SmCas12a2 should be followed up on with both structural and biochemical analysis. Experiments assessing whether both Cas12a and Cas12a2 can process a pre-crRNA and perform interference would provide clues as to whether these proteins interact directly or function as systems that complement one another. For instance, if one system fails to target an infecting sequence, the other system may be able to target and eliminate the infection.

In conclusion, my work on the Cas12a2 proteins resulted in the identification of the bioprocessing active site in SuCas12a2. Furthermore, my work laid the foundation by which others to discovered that SuCas12a2 is a reprogrammable RNA nuclease with trans cleavage of both DNA and RNA, which results in widespread cell death due to the indiscriminate DNA and RNA cleavage (Dmytrenko et al., 2022; Bravo et al., 2022). The Jackson lab is currently trying to determine whether SuCas12a2 can be utilized in eukaryotic cells as an RNA guided nuclease. This work is important as it could provide a potential therapeutic tool, which targets cancer cells.

**Wadjet Summary & Future Directions**

The Wadjet genes from *A. vinelandii* DJ and *Mycobacterium* MCS have been cloned into expression vectors. Recombinant expression and purification parameters for AvJetB were identified, and size exclusion chromatography indicated that AvJetB purifies as a dimer. This data is significant as it supports the hypothesis that Wadjet
systems form dimeric complexes.

Expression tests of AvJetC and AvJetA are needed in order to obtain protein for single protein experiments such as DNA binding and ATPase assays. Once JetC is purified, an immediate assessment of ATPase activity would be prudent, as this protein shares structural similarity with MukB, an active ATPase (Doron et al., 2018; Zawadzka et al., 2018). When expression/purification conditions are established for all Wadjet proteins, experiments can be performed to determine whether the Wadjet system acts similar to chromatin condensin complexes. Furthermore, the generation of mutants along the dimer interfaces of all Jet proteins would allow us to determine where protein interactions are occurring, and which residues are essential for DNA recognition and binding.

The easiest method to move studies forward would be to finish the development of a destination vector, which would house all three Av Wadjet genes allowing for co-expression of the Wadjet complex. Alternatively, generating tagless versions of the genes and attempting to co-express utilizing multiple plasmids may yield results as this has been successful with other Jackson lab projects (CasDinG + Type IV-A RNP Complex purification). If the complex is successfully purified, crystal trays should be pursued as there are no Wadjet structures in the PDB database. Analytical ultracentrifugation could provide sedimentation coefficients that can provide insight into the oligomeric state of the potential Wadjet complex. Distantly related condensin systems vary in their oligomeric state based on ATP binding and ATP hydrolysis states; thus, these conditions would also need to be assessed (Keenholtz et al., 2017).

Experiments that determine whether the *A. vinelandii* DJ Wadjet proteins inhibit
plasmid transformation are needed, considering this particular set of Wadjet genes is only composed of jetABC and no jetD. The absence of jetD in A. vinelandii DJ and many other species poses an interesting question, as researchers found that jetD was required for plasmid interference in Bacillus subtilis (Doron et al., 2018). Although it is tempting to say that Wadjet systems in species without jetD may be inactive, it could be that these species contain an endogenous protein that could perform the same function as jetD. This may be an exciting story to follow where we could perform pull-down assays using naturally expressed WADJET in the A. vinelandii DJ cell lines. This would require learning homologous recombination techniques to add tags to endogenous jet genes in A. vinelandii DJ.

A significant question to pursue would be how Wadjet systems can distinguish between endogenous genomic material and foreign material such as a plasmid. It may be that Wadjet can distinguish plasmid material through recognition of DNA supercoiling state, the origin of replication recognition, or by being expressed when non-chromosomal material such as plasmids are replicated. This hypothesis is based on the knowledge that plasmids are self-replicating and can replicate independently of chromosomal DNA as they have their own ORI sequence (Hayes, 2003). This then begs the question of whether Wadjet is continuously expressed or whether specific cellular signals are needed to induce the expression. This could be assessed by performing real-time PCR in A. vinelandii DJ.

In conclusion, the work performed on Wadjet systems has resulted in the successful cloning of A. vinelandii and M. MCS jet genes into expression vectors. Furthermore, the first data showing AvJetB purifying as a dimer was presented. This data
is significant as it supports the claim that Wadjet systems are similar to chromatin condensin complexes.

In conclusion, the work presented in this thesis provides biochemical and structural details on the enzyme CasDinG, which enhances the knowledge of both SF2 helicases and type IV-A CRISPR- Cas systems. The work on CasDinG fills a crucial gap in the mechanism of type IV-A systems. Furthermore, work in this thesis revealed the bioprocessing active site of SuCas12a2, laid a foundation for the study of Cas12a2 nuclease, and provided a foundation for the study of Wadjet plasmid inhibition systems. Overall, this work has shown that basic research on novel prokaryotic immune systems can result in the identification of potential DNA/RNA tools.

References


Dmytrenko, O., Neumann, G., Hallmark, T., Keiser, D., Crowley, V., & Vialletto, E. et al. (2022). Cas12a2 elicits abortive infection via RNA-triggered destruction of double-stranded DNA. https://doi.org/10.1101/2022.06.13.495973


APPENDICES
Chapter 2 Supplementary Information

Supplementary Figures

Figure A-1

*CasDinG Binding to Single-Stranded DNA of Variable Length (+ATP)*
Figure A-2

*Mutant(s) CasDinG Binding to ssDNA (17 nt.)*

![Graph showing anisotropy vs. CasDinG concentration for different mutants.]

Figure A-3

*Far UV-Circular Dichroism of CasDinG and Mutants*

![Graph showing mean residue ellipticity vs. wavelength for different mutants.]

**Figure A-4**

*Oligonucleotides Used in CasDinG Assays*

<table>
<thead>
<tr>
<th>Construct</th>
<th>Oligo #(#s)</th>
<th>Length</th>
<th>Visual Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blunt dsDNA</td>
<td>14-17</td>
<td>17 nt.</td>
<td></td>
</tr>
<tr>
<td>5' Overhang DNA Duplex</td>
<td>14-20</td>
<td>17 nt. duplex, overhang 17 nt.</td>
<td></td>
</tr>
<tr>
<td>3' Overhang DNA Duplex</td>
<td>14-21</td>
<td>17 nt. duplex, overhang 17 nt.</td>
<td></td>
</tr>
<tr>
<td>5' DNA Overhang Hybrid Duplex</td>
<td>20-36</td>
<td>17 nt. duplex, overhang 17 nt.</td>
<td></td>
</tr>
<tr>
<td>5' RNA Overhang Hybrid Duplex</td>
<td>14-37</td>
<td>17 nt. duplex, overhang 17 nt.</td>
<td></td>
</tr>
<tr>
<td>5' Overhang RNA Duplex</td>
<td>36-37</td>
<td>17 nt. duplex, overhang 17 nt.</td>
<td></td>
</tr>
<tr>
<td>ssRNA-PT</td>
<td>42</td>
<td>17 nt.</td>
<td></td>
</tr>
<tr>
<td>ssDNA (40 nt.)</td>
<td>82</td>
<td>40 nt.</td>
<td></td>
</tr>
<tr>
<td>ssDNA (17 nt.)</td>
<td>14</td>
<td>17 nt.</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* Related to Table A-5

*no FAM label in ATPase assays*
Figure A-5

Binding of CasDinG to ssDNA (40 nt.) and Blunt DNA Duplex

![Graph showing binding of CasDinG to ssDNA and Blunt DNA Duplex](image)

Figure A-6

Type IV-A RNP + CasDinG Binding to Target ssDNA TTTC PAM

![Graph showing binding of Type IV-A RNP + CasDinG to target ssDNA](image)

$K_d = 5.23 \pm 1.3$ nM
Figure A-7

Robetta Model of CasDinG with Mutations Marked
Figure A-8

*CasDinG Helicase Motifs Highlighted in the RecA Domains*
### EMBOSS-NEEDLE Alignment of PaCasDinG to E. coli DinG

```plaintext
# Program: needle
# Rundate: Thu 28 Jul 2022 19:33:16
# Commandline: needle
# -auto
# -stdout
# -asequence emboss_needle-I20220728-193315-0439-45218992-p2m.asequence
# -bsequence emboss_needle-I20220728-193315-0439-45218992-p2m.bsequence
# -datafile EBLOSUM62
# -gapopen 10.0
# -gapextend 0.5
# -endopen 10.0
# -endextend 0.5
# -aformat3 pair
# -sprotein1
# -sprotein2
# Align_format: pair
# Report_file: stdout

#=======================================
# Aligned_sequences: 2
# 1: CasDinG
# 2: E. coli
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
# Length: 887
# Identity: 168/887 (18.9%)
# Similarity: 269/887 (30.3%)
# Gaps: 332/887 (37.4%)
# Score: 147.0
#=======================================

<table>
<thead>
<tr>
<th>CasDinG</th>
<th>E. coli</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 MKLAQGAFVDVRGALSPTQDSLRWARTLLSEAVDQGLDSLFPVLQDV</td>
<td>0</td>
</tr>
<tr>
<td>51 STFTVTLQPALAVRLKALADQHNTPSVVYAGLIEAMRRRSEGSVVAEAP</td>
<td>100</td>
</tr>
<tr>
<td>101 MELPADLPGEGAVRPLQLKQAEEKTAA--GKIVFEDAAATGK--G</td>
<td>146</td>
</tr>
<tr>
<td>147 RMIASLAAAIAIKDGTVVVSAPLATWQLVNDMKDIPEVRRV----GLTL</td>
<td>192</td>
</tr>
<tr>
<td>193 SLGRPNFISPQRTLEWAIDNERADLAHWIEGKFLSLRSMETSKVISH</td>
<td>241</td>
</tr>
</tbody>
</table>
```
E. coli
112 AGMGGRYVCFRNMLTALASTPTEPTQDLLALFLDELPFNNVEEQKRLAKLRG 161
CasDinG
242 LECWILLEADALLAEEDLPADSL--LTSEDPADECAPQQLY-------VAMR 282
E. coli
162 DLLTYKWDGLRHOHTIAIDDDLWRRLSTDKASCLNRNCYYRECIPFVAR 211
CasDinG
283 SNYTEAGILLCSTLMWWAHHTRMMQVNRGMDKDEDEAPTGLSLPHFIDT 332
E. coli
212 REIQEAELVANHALV----------MAAMEAEVLDP--PKN-------------L 244
CasDinG
333 LIVDEAHLLEAFAVSYVTHTTLRPLMRTIEGLGSGRKPHDALK------- 378
E. coli
245 LVLDEGHHLDV-----------------------------ARDALEMSAE 266
CasDinG
379 -------ELFTQMQVASARSTNTSLNVPSDVP----------ELI 407
E. coli
267 ITAPWYRQLDFTKL-VATCMQFPRKPIPIAIPELRNAHCEELYLI 315
CasDinG
408 PALKDTVK---------------------TGALPTKGSMDRASVIRATRA--ND 442
E. coli
316 ASLNNIINLMPQAEARHFPAMELPDé----------VLEICQRALKTE 357
CasDinG
433 ALSCHSRLRI---------------EVTPVHSYPMILLSGRSLQARALLGNATG 482
E. coli
358 MLRGLAEFLNDLSEKTGSHDVRLRIILIQMNQALGMFEAQSQLKLRLAS 407
CasDinG
483 GATLVSA--TLLETTGD--NGLSLTRW-------KLEVPTERAALFPLPHFPAN 522
E. coli
408 LAQSGAPVTWATREEQGLHLFWCVGIRVSQDRLLLWRISPhtIV 457
CasDinG
523 TTAP-------------VLHGKEF--CA 536
E. coli
458 TSATLRLSNFSLQEMGLKEKADGFVRDVALDSPNFHCEFQKIVIPRMV 507
CasDinG
537 EP--DSSPERWATECGFTI--GQVASTAQQGTLVFLCTSYQNTELLAGRGA 583
E. coli
508 EPSIDNESEQHIAEFAFRKQVESKHLGMLVLpFAS---------GRAMQR 549
CasDinG
584 LGD-------RLIQSIPKSTASATCLAQFKKAK-GIRPWVGLGAAWTTGID 627
E. coli
550 FLDYVTDLRLMLLVQGQDQPRYKLVELHRKRVANGERSVVLGQSFAGLD 599
CasDinG
628 LSDHSLPD------NPEDLRLLSDLVITRIPVQ------NRS alternatives 665
E. coli
600 LQGDSLQVHIKIAFPID------SPVVITE---GEWLKSLNRPGFPEVQLS 643
CasDinG
666 AIGGFRISSEQAEAWHRQGLGRRLVRPVPHTKNLW------VLDARI------Y 707
E. coli
644 PSASFNLIQ--------VGRLIRSHG-------CGEVVIYDKRLLTKY 679
CasDinG
708 G-----GAANVAPFRQ--------ILDRYKKA---------- 726
E. coli
680 GRKLLDALPVFPEQPEVEPSEIVKKKKGKTSRPPP 716

#---------------------------------------
#---------------------------------------
Figure A-10

EMBOSS-NEEDLE alignment of PaCasDinG to S. aureus DinG

```
# Program: needle
# Rundate: Thu 28 Jul 2022 19:34:53
# Commandline: needle
# -auto
# -stdout
# -asequence emboss_needle-I20220728-193533-0953-75622219-01m.asequence
# -bsequence emboss_needle-I20220728-193533-0953-75622219-01m.bsequence
# -datafile EBLOSUM62
# -gapopen 10.0
# -gapextend 0.5
# -endopen 10.0
# -endextend 0.5
# -aformat3 pair
# -sprotein1
# -sprotein2
# Align_format: pair
# Report_file: stdout
```

```
# Aligned_sequences: 2
# 1: CasDinG
# 2: S.
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
# Length: 1029
# Identity: 152/1029 (14.8%)
# Similarity: 292/1029 (28.4%)
# Gaps: 435/1029 (42.3%)
# Score: 110.0

CasDinG    1 ------------------------------- 0
S.         1 MGMATYAVVDLETGQQLDFDIIIGITFVRNQIIDTYHSINTLNEI  50
CasDinG    1 ------------------------------- 0
S.         51 PPFIQALTSSIEENMLQQAPYFNQVAEYDKIKIFVFHVHFDLNFIK 100
CasDinG    1 -------MKILQGAFDVIRIGALSPPTDQRTLSARTKALSEAVDQGLDSL 43
S.         101 KAFKDCINQVYRPKVDLTLIFKIAFPTDK------------------ 130
CasDinG    44 PVPLQDVSTFTVLQPALAVRLKADQHNTPVSVYAGLIEAMRRRRSES 93
S.         131 ----------------SYQLSELAERH--------------GITLANAHRADE 153
CasDinG    94 GSVAEAPM--------------ELPADALPGEAVREVL--------- 118
```
Supplementary Tables

Table A-1

*Mutant(s) CasDinG Binding to ssDNA*

<table>
<thead>
<tr>
<th>Construct</th>
<th>$K_d$ (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>∆Arch</td>
<td>42 +/- 4.9</td>
</tr>
<tr>
<td>∆DEAH</td>
<td>94 +/- 18</td>
</tr>
<tr>
<td>∆N. Terminus Trunc.</td>
<td>65 +/- 9.6</td>
</tr>
</tbody>
</table>

*Note.* Related to Figure A-2.

Table A-2

*CasDinG Binding to ssDNA and Blunt Duplex DNA*

<table>
<thead>
<tr>
<th>Substrate</th>
<th>$K_d$ (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ssDNA (17 nt.)</td>
<td>61 +/- 4.6</td>
</tr>
<tr>
<td>Blunt Duplex DNA</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Table A-3

*Constructs Generated of CasDinG by Hannah Domgaard*

<table>
<thead>
<tr>
<th>Construct</th>
<th>Antibiotic</th>
<th>Tag</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>2RTPadING Arch K/O and</td>
<td>AMP</td>
<td>N-strep</td>
<td>knocks out amino acids 418-537 of 2rtpading and mutates DEAH_AAAH</td>
</tr>
<tr>
<td>DEAH.AAAH</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2RTPaDING ED_AA Set 2 324_325 AMP N-strep Mutates ED_AA set in vFeS positions 324 and 325
2RTPaDING ED-AA Set 1 317_318 AMP N-strep Mutates ED_AA set1 in vFeS positions 317 and 318
2RTPaDING FeS K/O Glycine Link AMP N-strep knocks out amino acids 264-355, and adds in a 5 glycine link
2RTPaDING N-Term Truncation DEAH-AAAH AMP N-strep knocks out amino acids 1-195 of 2rtpading
2RTPaDING Arch K/O AMP N-strep knocks out amino acids 418-537 of 2rtpading
2RTPaDING N-Term solo mutation AMP N-strep keeps only the N terminus 1-225, known as N terminus isolation
2RTPaDING N-Term Short. Truncation AMP N-strep knocks out amino acids 1-195 of 2rtpading
2RTPaDING Notag (#2) AMP N-strep same pet vector as 2rtpading but the strep tag has been removed
2RTPaDING N-TERMINAL TRUNCATION DEAH_AAAAH AMP N-strep knocks out amino acids 1-195 of 2rtpading and then mutates DEAH_AAAAH

Table A-4

Substrates Used in CasDinG Assays

<table>
<thead>
<tr>
<th>Oligo Name &quot;DinG&quot;</th>
<th>DNA/RNA</th>
<th>Sequence</th>
<th>Length (nt.)</th>
<th>Complementary Partner</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>DNA</td>
<td>TCGTCACCAGTACAAAC</td>
<td>17</td>
<td>20,21,17,37</td>
<td>Helicase Assay, Anisotropy Assay</td>
</tr>
<tr>
<td>17</td>
<td>DNA</td>
<td>GTTTGTACTGGTGACGA</td>
<td>17</td>
<td>14</td>
<td>Helicase Assay, Anisotropy Assay</td>
</tr>
<tr>
<td>20</td>
<td>DNA</td>
<td>GTACTGGTGACGA</td>
<td>33</td>
<td>14</td>
<td>Helicase Assay, Anisotropy Assay</td>
</tr>
<tr>
<td>21</td>
<td>DNA</td>
<td>TTTTTTTTTTTTTTTT</td>
<td>33</td>
<td>14</td>
<td>Helicase Assay, Anisotropy Assay</td>
</tr>
<tr>
<td>36</td>
<td>RNA</td>
<td>UCGUCACCAGUAAAC</td>
<td>17</td>
<td>20</td>
<td>Helicase Assay</td>
</tr>
</tbody>
</table>
Table A-5

**Primers to Mutate and Sequence CasDinG**

<table>
<thead>
<tr>
<th>Name</th>
<th>Purpose</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>N-Terminus Solo F</td>
<td>To PCR amplify the N-terminal domain of PaDinG</td>
<td>CGC GGC GAT TTA GGG CGA CAC A</td>
</tr>
<tr>
<td>N-Terminus Solo R</td>
<td>To PCR amplify the N-terminal domain of PaDinG</td>
<td>GCG GCT GCC AGG GAA GCA</td>
</tr>
<tr>
<td>N-Term KO Long F</td>
<td>Mutates the N-terminal domain of PaDinG</td>
<td>CAA AGC TCC GAT CCT GAT TAC</td>
</tr>
<tr>
<td>N-Term KO Long R</td>
<td>Mutates the N-terminal domain of PaDinG</td>
<td>ACT GCA GCA GGC AAG ATC</td>
</tr>
<tr>
<td>N-TDm KO S F</td>
<td>Mutates the N-terminal domain of PaDinG</td>
<td>CAA AGC TCC GAT CCT GAT TAC</td>
</tr>
<tr>
<td>N-TDm KO S R</td>
<td>Mutates the Arch domain of PaDinG</td>
<td>GGA CGA TCC AAC CTG CAG</td>
</tr>
<tr>
<td>Arch Dm KO F</td>
<td>Mutates the Arch domain of PaDinG</td>
<td>GGT GTA GAC GGA CGC AAA</td>
</tr>
<tr>
<td>Arch Dm KO R</td>
<td>Mutates the Arch domain of PaDinG</td>
<td></td>
</tr>
<tr>
<td>HD_t7_F</td>
<td>T7 forward primer</td>
<td>TAATACGACTCACTATAGGG</td>
</tr>
<tr>
<td>HD_t7_term</td>
<td>T7 terminal primer</td>
<td>GCTAGTTATGCTAGCGG</td>
</tr>
<tr>
<td>PDinG_Ntrunc_F</td>
<td>Primer to truncate Pa83 DinG by taking off the N terminal domain</td>
<td>TAACGGCCACCGGTACCGGCA</td>
</tr>
<tr>
<td>PDinG_Ntrunc_R</td>
<td>Primer to truncate Pa83 DinG by taking off the N terminal domain</td>
<td>CCTCGGCAAAGACGATCTTGCCCTGC</td>
</tr>
</tbody>
</table>
off the N terminal domain
Mutagenesis primer to change D258 to a
and E259 to A
 Mutagenesis primer to change D258 to a
 and E259 to A
Mutagenesis primer to knock out the FeS domain using a 5 glycine linker
Mutagenesis primer to knock out the FeS domain using a 5 glycine linker
Mutagenesis primer to take the tag off of DinG
Mutagenesis primer to take the tag off of DinG

<table>
<thead>
<tr>
<th>Construct</th>
<th>Molecular Weight (Daltons)</th>
<th>Extinction (M⁻¹ cm⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strep tag CasDinG ΔArch</td>
<td>68625.53</td>
<td>94420</td>
</tr>
<tr>
<td>Strep tag CasDinG ΔDEAH_AAAH</td>
<td>80915.93</td>
<td>95910</td>
</tr>
<tr>
<td>Strep tag CasDinG ΔN. Terminus Trunc.</td>
<td>69182.46</td>
<td>88920</td>
</tr>
<tr>
<td>Strep tag CasDinG ΔvFeS</td>
<td>72886.69</td>
<td>79410</td>
</tr>
<tr>
<td>Strep tag WT CasDinG</td>
<td>81131</td>
<td>95910</td>
</tr>
<tr>
<td>Type IV-A RNP</td>
<td>174612</td>
<td>463580</td>
</tr>
<tr>
<td>Type IV-A RNP + CasDinG</td>
<td>255743</td>
<td>559490</td>
</tr>
</tbody>
</table>
### Table A-7

**ApoCasDinG Crystal Hits**

<table>
<thead>
<tr>
<th>Screen- Temperature/ Well</th>
<th>Crystal type (- shower, + needle)</th>
<th>Salt</th>
<th>Buffer</th>
<th>pH</th>
<th>M</th>
<th>PEG</th>
<th>% PEG</th>
<th>Extra</th>
<th>% or</th>
<th>[ ]</th>
<th>Deep/Protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCSG II 4C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A8-1</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A11-2</td>
<td>-</td>
<td>Lithium</td>
<td>HCl</td>
<td>5</td>
<td>0.1</td>
<td>3350</td>
<td>20</td>
<td>0.3:0.1 uL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B5-1</td>
<td>-</td>
<td>NH4 Sulfate</td>
<td>pH 6.5</td>
<td>6.5</td>
<td>0.2</td>
<td>B000</td>
<td>30</td>
<td>0.3:0.3 uL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B6-2</td>
<td>-</td>
<td>Sodium</td>
<td>Formiate</td>
<td>0.2</td>
<td>0.2</td>
<td>3350</td>
<td>20</td>
<td>0.3:0.2 uL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C3-2</td>
<td>-</td>
<td>Ammonium</td>
<td>Acetate</td>
<td>0.2</td>
<td>7.5</td>
<td>0.1</td>
<td>3350</td>
<td>25</td>
<td>0.3:0.2 uL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCSG III 13C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B5-1</td>
<td>+</td>
<td>NH4 Sulfate</td>
<td>pH 6.5</td>
<td>6.5</td>
<td>0.2</td>
<td>B000</td>
<td>30</td>
<td>0.3:0.1 uL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E8-3 long skiny -</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sodium</td>
<td>Tartrate</td>
<td>CHES:NaOH</td>
<td>9.5</td>
<td>0.1</td>
<td>B000</td>
<td>20</td>
<td>0.3:0.1 uL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F11-1 long skiny -</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calcium</td>
<td>Acetate</td>
<td>Ethane</td>
<td>8</td>
<td>0.1</td>
<td>B000</td>
<td>10</td>
<td>0.3:0.3 uL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C5-3</td>
<td></td>
<td>Mg Acetate</td>
<td>pH 6.5</td>
<td>6.5</td>
<td>0.08</td>
<td>B000</td>
<td>16</td>
<td>0.3:0.3 uL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCSG III 8T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A7-3++</td>
<td></td>
<td>Calcium</td>
<td>Acetate</td>
<td>0.2</td>
<td>8</td>
<td>0.1</td>
<td>B000</td>
<td>10</td>
<td>0.3:0.2 uL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C5-3</td>
<td>- Mg Acetate</td>
<td>pH 6.5</td>
<td>0.08</td>
<td>B000</td>
<td>16</td>
<td>0.3:0.5 uL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D10-2++</td>
<td></td>
<td>Calcium</td>
<td>K Citrate</td>
<td>8.3</td>
<td>0.2</td>
<td>3350</td>
<td>20</td>
<td>0.3:0.2 uL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E8-3 long skiny -</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HEPESE:NaOH</td>
<td>Ethylene</td>
<td>7.5</td>
<td>0.1</td>
<td>B000</td>
<td>10</td>
<td>0.3:0.3 uL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sodium</td>
<td>Tartrate</td>
<td>Ethane</td>
<td>8</td>
<td>0.1</td>
<td>B000</td>
<td>10</td>
<td>0.3:0.3 uL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Supplementary Methodology

Alphafold Model Molecular Replacement

An Alphafold generated model was utilized to solve the phases of the x-ray diffraction data collected using the SSRL 9-2 beamline. First an Alphafold model was generated using the Alphafold 2.0 google colab server. Next, the Alphafold model was utilized in phaser (phenix program). The Alphafold model was trimmed of the arch domain and n terminal domain and was ultimately composed of the helicase core domain and the vestigial FeS domain (vFeS). Alphafold is able to generate a fairly reliable model based off of MSA (multiple sequence alignments) however domain placement of models may be biased if there is the presence of bound ligands such as nucleic acid or bound metals. The trimmed Alphafold model resulted in a solve (TFZ score > 8) and resulted in a low-resolution map. To improve the map, the Alphafold model was broken into helicase core domain and vFeS domains and then a rigid body fit was performed on each domain. Solvent flattening was then performed using the phenix program RESOLVE. This allowed for sufficient improvement of phases such that CasDinG model could be fit into the resulting electron density. The final model resulted in electron density for all domains, excluding the N-terminus which was not resolved.
APPENDIX B

Chapter 3 Supplementary Information

Supplementary Figures
Figure B-1

SAXS-SEC of SuCas12a2

Note. SEC chromatogram at A280 of SuCas12a2 (also known as SuCms1). Shoulder represents a larger species. B. SAXS SEC profile showing shoulder reversal confirming a larger mass species. C. P(r) distribution shows negative values around r=30 indicative of a micelle. D-max indicates something is embedded or inside the micelle. D. Log(q) plot is that of a typical detergent, such as TWEEN 20. E. UV absorption for sample, another indication of TWEEN 20.
**Figure B-2**

*I-TASSER Homology Model of SuCas12a2*

![I-TASSER Homology Model of SuCas12a2](image)

**Figure B-3**

*DALI Top 5 Server Results for AvJet Alphafold Models*

<table>
<thead>
<tr>
<th>A. Job: AvJetA</th>
<th>Query: s001A</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. Chain  Z</td>
<td>rmsd la1 nres</td>
</tr>
<tr>
<td>1: 3e9h-B</td>
<td>8.5 6.2 180 321</td>
</tr>
<tr>
<td>2: 5jv-A</td>
<td>8.1 8.9 131 108</td>
</tr>
<tr>
<td>3: 62hj-B</td>
<td>7.4 7.3 180 501</td>
</tr>
<tr>
<td>4: 3nhx-X</td>
<td>7.3 15.7 129 481</td>
</tr>
<tr>
<td>5: 1gj-A</td>
<td>7.2 19.4 158 863</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B. Job: AvJetB</th>
<th>Query: s001A</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. Chain  Z</td>
<td>rmsd la1 nres</td>
</tr>
<tr>
<td>1: 6epq-D</td>
<td>9.4 3.1 78 82</td>
</tr>
<tr>
<td>2: 7qwc-E</td>
<td>9.7 4.6 163 232</td>
</tr>
<tr>
<td>3: 6qfd-B</td>
<td>8.4 4.5 85 116</td>
</tr>
<tr>
<td>4: 5vyr-A</td>
<td>8.3 4.0 98 173</td>
</tr>
<tr>
<td>5: 5rr-B</td>
<td>8.3 5.9 122 374</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C. Job: AvJetC</th>
<th>Query: s001A</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. Chain  Z</td>
<td>rmsd la1 nres</td>
</tr>
<tr>
<td>1: 7hnx-A</td>
<td>11.2 27.8 146 217</td>
</tr>
<tr>
<td>2: 6g3-B</td>
<td>11.2 46.3 255 693</td>
</tr>
<tr>
<td>3: 5mgb-B</td>
<td>11.0 22.8 203 264</td>
</tr>
<tr>
<td>4: 5h69-A</td>
<td>10.5 7.3 217 252</td>
</tr>
</tbody>
</table>

**Supplementary Tables**
Table B-1

**Chapter 2 Construct Molecular Weights and Extinction Coefficients**

<table>
<thead>
<tr>
<th>Construct</th>
<th>Molecular Weight (Daltons)</th>
<th>Extinction (M⁻¹ cm⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SuCas12a2 Apo</td>
<td>143912.14</td>
<td>145890</td>
</tr>
<tr>
<td>SuCas12a2 RNP</td>
<td>161023.33</td>
<td>417590</td>
</tr>
<tr>
<td>C-terminal HIS SmCas12a2</td>
<td>128059.69</td>
<td>139930</td>
</tr>
<tr>
<td>HIS -SmCas12a2 RNP</td>
<td>145170.88</td>
<td>411630</td>
</tr>
<tr>
<td>Processed RNA guide</td>
<td>17111.19</td>
<td>271700</td>
</tr>
<tr>
<td>N term Sumo-HIS AvJetB</td>
<td>38912.78</td>
<td>19940</td>
</tr>
</tbody>
</table>

Table B-2

**Substrates Used in Cas12a2 Assays**

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>SuRNA</td>
<td>pre-crRNA used in SuCas12a2 processing assay</td>
<td>CACCUUAAUAUAUGUCUAACCGGCAACAAAUUUUCUAC</td>
</tr>
<tr>
<td>80nt</td>
<td>SuCas12a2</td>
<td>UGUUGUAGAUUGGAGCAACACCUGAAAGGCUUGAUG</td>
</tr>
<tr>
<td>TTTT_F_T</td>
<td>DNA used in binding and cleavage assay</td>
<td>ACAAACAGAAGAATCTACCTCTTCTGGAGACACACTGGAAGGA</td>
</tr>
<tr>
<td>TTTT_R_T</td>
<td>DNA used in cleavage assay</td>
<td>AGGCTTGATGAGCAAGTGGCGAGCAAGA</td>
</tr>
</tbody>
</table>

Table B-3

**Primers to Sequence and Mutate SuCas12a2 (Cms1)**

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>HDSunovel_licF</td>
<td>Primer to clone the novel domain out</td>
<td>TACTTCCAATCCAATGCAAAAACAGCT</td>
</tr>
<tr>
<td></td>
<td>of SuCms1 into LIC vectors</td>
<td>TGCCACACTC</td>
</tr>
<tr>
<td>HDSunovel_licR</td>
<td>Primer to clone the novel domain out</td>
<td>TTATCCACCTCAATGTTATTAGTAGAAG</td>
</tr>
<tr>
<td></td>
<td>of SuCms1 into LIC vectors</td>
<td>CCTTGTATGAATCTTACG</td>
</tr>
<tr>
<td>HD_t7_F</td>
<td>T7 forward primer</td>
<td>TAATACGACTCAGTAGGG</td>
</tr>
<tr>
<td>HD_t7_term</td>
<td>T7 terminal primer</td>
<td>GCTAGTTATGCTCAGGG</td>
</tr>
<tr>
<td>HDStrepRhino3</td>
<td>Primer to add a Strep Tag and Rhino</td>
<td>GGGTGCACCTGGAATGCCCTTCTCCAGG</td>
</tr>
<tr>
<td>182F</td>
<td>cleavage sequence to SuCms1 in the 3182 construct</td>
<td>GACCCCTTACAGCTTCTTCAACAT</td>
</tr>
<tr>
<td></td>
<td>Primer to add a Strep Tag and Rhino</td>
<td>TTTTCAACTGAAGGCTGGGCTCAGGCAGCT</td>
</tr>
<tr>
<td>HDStrepRhino3</td>
<td>Primer to add a Strep Tag and Rhino</td>
<td>TGCCATTGGATATCTCCTTTATTTAAG</td>
</tr>
<tr>
<td>182R</td>
<td>cleavage sequence to SuCms1 in the 3182 construct</td>
<td>GCTTAAAC</td>
</tr>
</tbody>
</table>
HDStrRhi3182seq

Primer used to sequence the Strep Tag and Rhino sequence of the 3182 construct
primer to sequence 2213 construct
Apo SuCms1

2213_seq6

Apo SuCms1

2213_seq5

Apo SuCms1

2213_seq4

Apo SuCms1

2213_seq3

Apo SuCms1

2213_seq2

Apo SuCms1

Sulf_Ruvc1_DA_F

primer to mutate SuCms1 ruvC1 aspartate to alanine

Sulf_Ruvc1_DA_R

primer to mutate SuCms1 ruvC1 aspartate to alanine

sulfRUVC2_ED_AA_F

primer to mutate SuCms1 ruvC2 aspartate/glutamate to alanines

sulfRUVC2_ED_AA_R

primer to mutate SuCms1 ruvC2 aspartate/glutamate to alanines

sulfRUVC3_DD_AA_F

primer to mutate SuCms1 ruvC3 glutamates to alanines

sulfRUVC3_DD_AA_R

primer to mutate SuCms1 ruvC3 glutamates to alanines

Sucms1_ruvc1s_eq_F

Primer to sequence SuCms1 ruvc1 domain

Sucms1_ruvc1s_eq_R

Primer to sequence SuCms1 ruvc1 domain

Sucms1_ruvc2s_eq_F

Primer to sequence SuCms1 ruvc2 domain

Sucms1_ruvc2s_eq_R

Primer to sequence SuCms1 ruvc2 domain

LIC_v1_suCMS1_F

Primer to clone SuCms1 into LIC vectors

LIC_v1_suCMS1_R

Primer to clone SuCms1 into LIC vectors

CMS1_2322_Ecoor1_F

Primer to add EcoR1 site to N terminus of SuCms1

CMS1_2322_Kpn1_R

Primer to add Kpn1 site to C terminus of SuCms1

RuvC3seqF

Primer to sequence SuCms1 RuvC3 domain
<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>HDStRhi3182n oCR_F</td>
<td>Primer used to eliminate the CRISPR from the strep-rhino 3182 construct</td>
<td>TAAACCGACCCAAGGTAC</td>
</tr>
<tr>
<td>HDStRhi3182n oCR_R</td>
<td>Primer used to eliminate the CRISPR from the strep-rhino 3182 construct</td>
<td>CCCTATAGTGAGTCGTATTTAG</td>
</tr>
</tbody>
</table>

### Table B-4

Primers to Sequence and Mutate SmCas12a2 (Cms1)

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>C_S_mut_234&amp;243_F</td>
<td>Cysteine to Serine mutant for SmCms1 position 234 and 243</td>
<td>CGGGCTGTGCGTGTTTCAAAAGACTATG</td>
</tr>
<tr>
<td>C_S_mut_234&amp;243_R</td>
<td>Cysteine to Serine mutant for SmCms1 position 234 and 243</td>
<td>AAGAGCTGTGTGCTGCTGTTTTCTTC</td>
</tr>
<tr>
<td>C_S_mut_291&amp;297_F</td>
<td>Cysteine to Serine mutant for SmCms1 position 291 and 297</td>
<td>TGCTGATGCGGGTGTGACAC</td>
</tr>
<tr>
<td>C_S_mut_291&amp;297_R</td>
<td>Cysteine to Serine mutant for SmCms1 position 291 and 297</td>
<td>AAAGCGGAAACCTTCTGCTGCTGTTTTATC</td>
</tr>
<tr>
<td>C_S_mut_303_F</td>
<td>Cysteine to Serine mutant for SmCms1 position 303</td>
<td>CAAATGCCAACGCGCTGTTTAC</td>
</tr>
<tr>
<td>C_S_mut_303_R</td>
<td>Cysteine to Serine mutant for SmCms1 position 303</td>
<td>CAGAACCTTGTGTGCTGACAC</td>
</tr>
<tr>
<td>C_A_mut_234&amp;243_F</td>
<td>Cysteine to Alanine mutant for SmCms1 position 234 and 243</td>
<td>TCGGGGCGCTGCGTTTCAAAAGAC</td>
</tr>
<tr>
<td>C_A_mut_234&amp;243_R</td>
<td>Cysteine to Alanine mutant for SmCms1 position 234 and 243</td>
<td>AGGCCCTGGTGCTGCTGTTTCTTC</td>
</tr>
<tr>
<td>C_A_mut_291&amp;297_F</td>
<td>Cysteine to Alanine mutant for SmCms1 position 291 and 297</td>
<td>TTGGCAGATGGCGGCTGACACC</td>
</tr>
<tr>
<td>C_A_mut_291&amp;297_R</td>
<td>Cysteine to Alanine mutant for SmCms1 position 291 and 297</td>
<td>AGCCGAAACCTTCTGCTGCTGTTTTATC</td>
</tr>
<tr>
<td>C_A_mut_303_F</td>
<td>Cysteine to Alanine mutant for SmCms1 position 303</td>
<td>CAAATGCAACCACGCGGCTGACAC</td>
</tr>
<tr>
<td>Name</td>
<td>Description</td>
<td>Sequence</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>C_A_mut_303_R</td>
<td>Cysteine to Alanine mutant for SmCms1 position 303 primer to sequence the first ruvc domain in smCms1</td>
<td>CAGAAACCTTCTTGCTTTG</td>
</tr>
<tr>
<td>Ruvc1_ncms1_seq_L</td>
<td>primer to sequence the first ruvc domain in smCms1</td>
<td>TGAACCGGAAGAAACGTAACC</td>
</tr>
<tr>
<td>Ruvc1_ncms1_seq_R</td>
<td>smCms1 primer to sequence CPKC mutant</td>
<td>TCGTTCAGAAAATAGCTCAGGT</td>
</tr>
<tr>
<td>NCMS1_CPKC_L_sq</td>
<td>Primer to sequence CPKC mutant</td>
<td>CTGGAATGGGCCGCTGTATC</td>
</tr>
<tr>
<td>NCMS1_CPKC_R_sq</td>
<td>Primer to mutate SmCms1 Ruvc1 domain from an aspartate to alanine</td>
<td>TCCGTTATCCACTTCCAAATGT</td>
</tr>
<tr>
<td>Smcms1Ruv1_D_A_F</td>
<td>Primer to mutate SmCms1 Ruvc1 domain from an aspartate to alanine</td>
<td>TATGGTATCGCCGGGCAGAA</td>
</tr>
<tr>
<td>Smcms1Ruv1_D_A_R</td>
<td>Primer to mutate SmCms1 Ruvc2 domain from glutamate/aspartate to alanines</td>
<td>AAACCACAGTTGTTCTTGTGAAC</td>
</tr>
<tr>
<td>Smcms1Ruv2ED_AA_F</td>
<td>primer to mutate SmCms1 Ruvc2 domain from glutamate/aspartate to alanines</td>
<td>ATTAGCATCGCAGCCCTGAACCAAAC</td>
</tr>
<tr>
<td>Smcms1Ruv2ED_AA_R</td>
<td>primer to mutate smcms1 ruvc3 aspartate to alanine</td>
<td>GATGCCACGATATTTCGT</td>
</tr>
<tr>
<td>Smcms1Ruv3_D_A_F</td>
<td>Primer to mutate smcms1 ruvc3 aspartate to alanine</td>
<td>AACGACCAGGCCAAGGTGGC</td>
</tr>
<tr>
<td>Smcms1Ruv3_D_A_R</td>
<td>ruvc3 aspartate to alanine SmCms1 Ruvc1 sequencing primer</td>
<td>CAGGCCCTAAACCCTTTCAG</td>
</tr>
<tr>
<td>SmRuvC1seq121318</td>
<td>Smcms1 Ruvc2 sequencing primer</td>
<td>GGGTTTTTCACGGATAACCAGCTC</td>
</tr>
<tr>
<td>SmRuvC2seq121318</td>
<td>SmCms1RuvC3 sequencing primer</td>
<td>CTCCAGTTCAGCGAGCTTAATC</td>
</tr>
<tr>
<td>SmRuvC3seq121318</td>
<td>SmCms1RuvC3 sequencing primer</td>
<td>CTTCCTTCGGGCTTTTAG</td>
</tr>
<tr>
<td>Cms1_RuvC3_D_A_F</td>
<td>Mutates the Ruvc3 domain from an aspartate to alanine. SmCms1</td>
<td>AACGACCAGGCCAAGGTGGC</td>
</tr>
<tr>
<td>Cms1_RuvC3_D_A_R</td>
<td>Mutates the Ruvc3 domain from an aspartate to alanine. SmCms1</td>
<td>CAGGCCCTCAAACCCTTCCAG</td>
</tr>
</tbody>
</table>
Cms1_RuvC1_D_A_F
Mutates the RuvC1 domain from an aspartate to alanine. SmCms1
TATGGTATCGCGCGGGCGAA

Cms1_RuvC1_D_A_R
Mutates the RuvC1 domain from an aspartate to alanine. SmCms1
AAACCAGTGGTTCTTGTTGAAC

Cms1_RuvC2_E_A_F
Mutates the RuvC2 domain from a glutamate to alanine. SmCms1
ATTAGCATGCAGACCTGAAAC

Cms1_RuvC2_E_A_R
Mutates the RuvC2 domain from a glutamate to alanine. SmCms1
GATGCCACGAGACCTGAAAC

smcms1_ruvc1_seqF
Primer for sequencing RuvC1 domain of Smcms1
AACGAAATTACCTTCGCGTTTG

smcms1_ruvc1_seqR
Primer for sequencing RuvC1 domain of Smcms1
GTATACTGTTCGTGCAGGTAACG

Table B-5

Primers to Clone and Sequencing Wadjet

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>HD_MMJetA_seq</td>
<td>Primer used to sequence jetA from</td>
<td>AGTCGCCGGATCTTGAAACTC</td>
</tr>
<tr>
<td>1</td>
<td>Mycobacterium MCS</td>
<td></td>
</tr>
<tr>
<td>HD_MMJetA_seq</td>
<td>Primer used to sequence jetA from</td>
<td>AGCAGATGGTCTGGTGCG</td>
</tr>
<tr>
<td>2</td>
<td>Mycobacterium MCS</td>
<td></td>
</tr>
<tr>
<td>HD_MMJetC_Seq</td>
<td>Primer used to sequence jetC from</td>
<td>ACCCGTTCGATCTGGCTCG</td>
</tr>
<tr>
<td>1</td>
<td>Mycobacterium MCS</td>
<td></td>
</tr>
<tr>
<td>HD_MMJetC_Seq</td>
<td>Primer used to sequence jetC from</td>
<td>ACGACATCGAGCGGCCTCG</td>
</tr>
<tr>
<td>2</td>
<td>Mycobacterium MCS</td>
<td></td>
</tr>
<tr>
<td>HD_MMJetCseq3</td>
<td>Primer used to sequence jetC from</td>
<td>AACTCGCGGAGCGAAGTTGTC</td>
</tr>
<tr>
<td>HD_AV_JetA_seq</td>
<td>Primer used to sequence jetA from</td>
<td>CAATCGATGTCCTTGAGTGTC</td>
</tr>
<tr>
<td>1</td>
<td>Azotobacter vinelandii</td>
<td></td>
</tr>
<tr>
<td>HD_AV_JetA_seq</td>
<td>Primer used to sequence jetA from</td>
<td>AAGCTCACGCATAGCTC</td>
</tr>
<tr>
<td>2</td>
<td>Azotobacter vinelandii</td>
<td></td>
</tr>
<tr>
<td>HD_AV_JetC_seq</td>
<td>Primer used to sequence jetC from</td>
<td>ATGCACTGGCCTTGAGTG</td>
</tr>
<tr>
<td>1</td>
<td>Azotobacter vinelandii</td>
<td></td>
</tr>
<tr>
<td>HD_AV_JetC_seq</td>
<td>Primer used to sequence jetC from</td>
<td>AGCTAGCAACTTGACTGTC</td>
</tr>
<tr>
<td>2</td>
<td>Azotobacter vinelandii</td>
<td></td>
</tr>
<tr>
<td>HD_AV_JetC_seq</td>
<td>Primer used to sequence jetC from</td>
<td>AAGCGTTCTTTGCCCCGAC</td>
</tr>
<tr>
<td>3</td>
<td>Azotobacter vinelandii</td>
<td></td>
</tr>
</tbody>
</table>
Table B-6

Chapter 2 Constructs Cloned By Hannah Domgaard

<table>
<thead>
<tr>
<th>Construct</th>
<th>Antibiotic</th>
<th>Tag</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>2ST_AVJetA</td>
<td>AMP</td>
<td>N-Sumo-HIS</td>
<td>Azotobacter Vinelandii</td>
</tr>
<tr>
<td>2ST_AVJetB</td>
<td>AMP</td>
<td>N-Sumo-HIS</td>
<td>Azotobacter Vinelandii</td>
</tr>
<tr>
<td>2ST_MMJetA</td>
<td>AMP</td>
<td>N-Sumo-HIS</td>
<td>Mycobacterium MCS</td>
</tr>
<tr>
<td>2ST_MMJetC</td>
<td>AMP</td>
<td>N-Sumo-HIS</td>
<td>Mycobacterium MCS</td>
</tr>
<tr>
<td>2RT_AVJetA</td>
<td>AMP</td>
<td>N-strep</td>
<td>Azotobacter Vinelandii</td>
</tr>
<tr>
<td>2RT_AVJetC</td>
<td>AMP</td>
<td>N-strep</td>
<td>Azotobacter Vinelandii</td>
</tr>
<tr>
<td>2RT_MMJetB</td>
<td>AMP</td>
<td>N-strep</td>
<td>Mycobacterium MCS</td>
</tr>
<tr>
<td>2RT_MMJetC</td>
<td>AMP</td>
<td>N-strep</td>
<td>Mycobacterium MCS</td>
</tr>
<tr>
<td>2ST_SuCas12a2_NovelDomain</td>
<td>AMP</td>
<td>N-Sumo-HIS</td>
<td>Sulfuricurvum Novel Domain</td>
</tr>
<tr>
<td>SuCas12a2_∆3182_strep_rhino</td>
<td>CAM</td>
<td>N Strep Rhino</td>
<td>Sulfuricurvum 3182 construct</td>
</tr>
<tr>
<td>strain</td>
<td>antibiotic</td>
<td>selection</td>
<td>host</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td>-----------</td>
<td>------</td>
</tr>
<tr>
<td>2BT_SmCas12a2 ∆RuvC1+2</td>
<td>AMP</td>
<td>N HIS</td>
<td>Smithella</td>
</tr>
<tr>
<td>SuCas12a2_∆RuvC3</td>
<td>CAM</td>
<td>N HIS</td>
<td>Sulfuricurvum</td>
</tr>
<tr>
<td>N-SmCas12a2_C-&gt;A_group1</td>
<td>AMP</td>
<td>N HIS</td>
<td>Smithella zinc finger mutant</td>
</tr>
<tr>
<td>N-SmCas12a2 ∆RuvC2</td>
<td>AMP</td>
<td>N HIS</td>
<td>Smithella</td>
</tr>
<tr>
<td>N-SmCas12a2 ∆RuvC3</td>
<td>AMP</td>
<td>N HIS</td>
<td>Smithella</td>
</tr>
<tr>
<td>SuCas12a2 ∆RuvC2 +3</td>
<td>AMP</td>
<td>N HIS</td>
<td>Sulfuricurvum</td>
</tr>
</tbody>
</table>