CHARACTERIZING THE STRUCTURE OF THE TYPE III-B CRISPR-CAS CMR COMPLEX USING CRYO-EM

by

Brynn Durecki

A thesis submitted to Johns Hopkins University in conformity with the requirements for the degree of Master of Science

Baltimore, Maryland
August 2023
Abstract

CRISPR-Cas systems are adaptive immune systems found in prokaryotes. Since their discovery, many CRISPR-Cas systems have been characterized biochemically and structurally, and they are divided into two main classes based on their components. Class 1 CRISPR-Cas systems consist of multi-subunit effector complexes (Makarova et al., 2015). Within Class 1, Type III systems are among the most ancient CRISPR-Cas complexes found in bacteria and archaea. Consisting of multiple protein subunits and a CRISPR RNA (crRNA), type III effectors specifically bind complementary target RNA, which activates both RNA and DNA cleavage activity to protect against invading nucleic acids (Hale et al., 2009; Samai et al., 2015). The subtype III-B complex structure has been solved in many organisms, but little is known about the structural and mechanistic details of type III-B CRISPR-Cas mediated immunity in the system Thermotoga maritima. Previous work in the lab has established the biochemical basis for RNA-activated DNA cleavage by a particular type III-B complex known as the Cas RAMP module (Cmr) in T. maritima, as well as the tolerance of mismatches between the crRNA and the target RNA to elicit this cleavage (Estrella et al., 2016; Johnson et al., 2019). The objective of this thesis was to characterize the structure of the Cmr complex using cryo-EM. The complex was purified, frozen at cryogenic temperature, and imaged, and the resulting images were averaged and processed to obtain a 3D density map of the target unbound complex.

We hoped to elucidate the precise catalytic mechanism of target RNA cleavage by the Cmr complex; however, due to low resolution, particle-limited maps and the formation of a sub-stoichiometric subcomplex, we were unable to see the molecular basis for the Cmr complex’s interference activity. Future experiments will focus on optimizing purification protocols and obtaining higher resolution target unbound and target bound structures to investigate specific
interactions between the target RNA and key amino acid residues of the complex’s catalytic subunit, Cmr4. This work will not only provide the structural basis for CRISPR-Cas interference in an understudied organism, but it may also give insights that can inform the development of genetic engineering tools.
Preface

First and foremost, I would like to thank Dr. Scott Bailey for his guidance, patience, and support over the course of my thesis. This work has opened my eyes to the complexity and challenge of science at the microscale. I would also like to thank Dr. Jennifer Kavran, my secondary thesis reader, for her time, patience, and input to bring this work to completion.

I would also like to thank Dr. Brian Learn for tolerating my endless questions and leading me through this challenging protocol. Without his support, kindness, and sense of humor, I would not have made it as far as I did in the lab. To Charlie, who took on the immense task of training me in the lab, thank you for being a great sport in the face of my many mistakes and growing pains. To Morgan, Zoë, Haobo, Katelyn, Elvar, Coco, and Max, thank you for welcoming me with open arms into the lab and for creating such an amazing environment to learn and grow as a scientist and person.

To Domenica, I could not have survived Immunology I without our study sessions. You are a queen and I am so grateful to have had you by my side throughout this program. To Selime, the best roommate a girl could ask for, the kindest confidante, and my best friend – I couldn’t have made it without you. I am so proud of both of you and I cannot wait to see how you thrive.

To Dr. Eddie Hall, for first giving me a chance at a laboratory bench. You inspired a lifelong love for the nitty-gritty details of experimenting, through the highs and lows and everything in between. Thank you for letting me in your organic chemistry lab (despite my innate clumsiness).

To Dr. Michael Duffy, for spiritual guidance and constant mentorship on Tuesday mornings, thank you for listening to me rant about everything from obscure science and ethical
conundrums. Graduate school (and life in general) is a long-haul, but our weekly chats were the greatest source of solace in times of mental and emotional strife. I am forever grateful to you.

To my family, the most important people in my life. To my Mom, for always answering my phone calls and being there even when far away. To my Dad, for inspiring my love of science and always encouraging and entertaining my endless questions. To Blake and Kallie, my siblings by blood and by choice who have been there for me through thick and thin. I love you all.

Lastly, I would like to thank Nelson, who has been a rock and a role model throughout this process. I am so happy that Hopkins brought us together. Thank you for everything.
# Table of Contents

Abstract .......................................................................................................................... ii
Preface ............................................................................................................................. iv
Table of Contents ........................................................................................................ vi

**Chapter 1**

Introduction .................................................................................................................. 1

**Chapter 2**

Background .................................................................................................................. 19

**Chapter 3**

Methods ....................................................................................................................... 28

**Chapter 4**

Results and Discussion ............................................................................................. 34

**Chapter 5**

Conclusion and Future Directions ............................................................................. 38

References ................................................................................................................... 41
Chapter 1

Introduction
Introduction

CRISPR-Cas systems are adaptive immune systems found in bacteria and archaea that provide defense against viruses or foreign plasmids, widely referred to as mobile genetic elements or MGEs (Jinek et al., 2012; Makarova et al., 2015; Frost et al., 2005; Faure et al., 2019). CRISPR-Cas systems derive unique sequences known as spacers from MGEs and incorporate them into the CRISPR array through a process known as adaptation (Sternberg et al., 2016; Makarova et al., 2020). Transcription of the CRISPR array and subsequent processing produces CRISPR RNAs (crRNAs) that assemble with protein subunits to form the CRISPR-Cas effector complex (Marraffini, 2015). The CRISPR-Cas effector complex identifies, binds, and cleaves nucleic acids to activate an immune response (Barrangou, 2013). CRISPR-Cas systems are divided into two classes depending upon the composition of their effector complexes. Class 1 systems utilize the combined functions of multiple subunits to carry out an immune response, whereas Class 2 systems depend upon a single protein for effector complex function (Makarova et al., 2015).

Within Class 1 are types I, III, and IV, each of which are further classified into a variety of subtypes. Type III CRISPR-Cas systems are multi-subunit complexes with subtypes III-A through III-F, each of which display differences in subunit composition and complex organization (Makarova et al., 2020). Most type III systems contain the signature gene cas10, which is translated into a large protein subunit that contains a histidine-aspartate (HD) domain and a Palm domain, similar to those of nucleic acid polymerases (Makarova et al., 2015; Estrella et al., 2016). These domains are important for activating ssDNA cleavage and for producing signaling molecules known as cyclic oligoadenylates (cOAs), respectively (Jinek et al., 2012; Kazlauskiene et al., 2017; Niewohner et al., 2017). Type III systems are unique in that they can
degrade both RNA and DNA derived from invaders to elicit an immune response (Samai et al., 2015; Peng et al., 2015). When the crRNA binds a complementary strand of target RNA, the complex undergoes conformational changes that result in cleavage of the target RNA strand and activation of Cas10 to cleave single-stranded DNA (ssDNA) throughout the cell environment (Estrella et al., 2016). Cas10 works in tandem with other subunits in the type III CRISPR-Cas effector complex to provide an adaptive immune response (Table 1).

**Table 1.** Subunits of the various subtypes of type III CRISPR-Cas systems. Based on classification of Makarova et al., 2020.

<table>
<thead>
<tr>
<th>Cas Genes</th>
<th>Csm Nomenclature</th>
<th>Cmr Nomenclature</th>
<th>General Role in Complex</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>cas10</em></td>
<td>Csm1 (III-A)</td>
<td>Cmr2 (III-B)</td>
<td>Characteristic subunit of Type III systems. Located at 3’ end of crRNA in complex; has ssDNase and cOA synthetase activity</td>
</tr>
<tr>
<td></td>
<td>Csm3 (III-A/F)</td>
<td>Cmr1 (III-B/C)</td>
<td>Subunit that filaments along the crRNA in complex; number of subunits can vary depending on subtype</td>
</tr>
<tr>
<td></td>
<td>Csm5 (III-A/D/E)</td>
<td>Cmr4 (III-B/C)</td>
<td>Subunit located at 5’ end of crRNA in complex; interacts directly with Cas10 and is linked to preventing autoimmunity</td>
</tr>
<tr>
<td><em>cas5</em></td>
<td>Csm4 (III-A/E)</td>
<td>Cmr3 (III-B/C)</td>
<td>Subunit located at 5’ end of crRNA in complex; interacts directly with Cas10 and is linked to preventing autoimmunity</td>
</tr>
<tr>
<td></td>
<td>Csx10 (III-D)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>cas11</em></td>
<td>Csm2 (III-A/D/E)</td>
<td>Cmr5 (III-B/C)</td>
<td>Small subunit responsible for scaffolding of the filament</td>
</tr>
<tr>
<td><em>cas6</em></td>
<td>-</td>
<td>-</td>
<td>Found in subtypes III-A and III-B; not part of the effector complex, but necessary for crRNA processing and maturation</td>
</tr>
<tr>
<td>-</td>
<td>Csm6 (III-A)</td>
<td>Csx1 (III-B)</td>
<td>CARF domain-containing RNases that are not associated with complex, but are activated by cOA signaling as a result of target RNA binding complex</td>
</tr>
</tbody>
</table>
Due to their widespread nature throughout archaea and bacteria, type III systems can vary greatly between subtypes. The two most widely studied type III systems are the Cas subtype Mtube (Csm) complex and the Cas Module RAMP (Cmr) complex. Both complexes are recognizable by their seahorse-shaped architecture (Rouillon et al., 2013; Zhang et al., 2012). The Csm complex is found in subtypes III-A and III-D and is characterized by a five-subunit structure. The Cmr complex of subtypes III-B and III-C have six subunits. Interestingly, the Cas10 of subtype III-C has inactive Palm domains, while Cas10 of subtype III-D is missing the HD domain (Makarova et al., 2015). Though little was previously known about subtype III-E, recent studies have elucidated its unique architecture with several Cas7 proteins fused to a small Cas11 (Makarova et al., 2020; Özcan et al., 2021; van Beljouw et al., 2021). Type III-F is unique from other type III subtypes in that it only encodes one Cas7 in its genetic locus (Makarova et al., 2020). The sequence similarity of genes encoding the Cas7, Cas5, and Cas10 of subtype III-F is significant, yet this subtype’s small subunit bears no resemblance to the Cas11 proteins of other type III systems (Makarova et al., 2020). Consequently, there is much diversity and heterogeneity amongst type III CRISPR-Cas systems. The aim of this review is to highlight structural and mechanistic studies of type III CRISPR-Cas-mediated interference, as well as to explore future directions for type III CRISPR-Cas system research and applications.

Mechanisms of nucleic acid cleavage in Type III systems

Type III CRISPR-Cas systems are able to cleave RNA and ssDNA. Upon assembly of various subunits and a mature crRNA, type III CRISPR-Cas complexes survey the cellular environment for target RNA. The crRNA recognizes and binds complementary target RNA sequences. The complex cleaves target RNA at six nucleotide intervals in what is known as the 5’ ruler endonuclease mechanism (Hatoum-Aslan et al., 2013). This ruler mechanism is
conserved amongst all type III CRISPR-Cas systems that cleave target RNA. The Cas7 subunits that filament along the backbone of the complex contain thumb-like β-hairpin protrusions that cause every sixth nucleotide of the crRNA and target RNA duplex to become displaced (Hatoum-Aslan et al., 2013; Hale et al., 2009; Barrangou, 2013). These flipped out bases allow for the scissile bonds of the target RNA to be in range of conserved catalytic residues in certain Cas7 (Csm3/Cmr4) subunits. Cleavage of target RNA is metal-dependent, and magnesium and manganese ions must be present for RNA cleavage to occur (Hale et al., 2009; Zhang et al., 2012; Staals et al., 2013; Samai et al., 2015).

Upon cleavage of the target RNA, the histidine-aspartate (HD) domain of Cas10 activates cleavage of transcriptionally active ssDNA (Staals et al., 2013; Jinek et al., 2012; Estrella et al., 2016; Samai et al., 2015), and promotes the Palm domains of Cas10 to synthesize cyclic oligoadenylates (Niewoehner et al., 2017; Kazlauskiene et al., 2017). The ssDNase and cOA synthetase activities of Cas10 are deactivated in a time-dependent manner, such that when target RNA is cleaved by the complex via the 5’ ruler mechanism, the activities of Cas10 stop (Samai et al., 2015; Hatoum-Aslan et al., 2013). This allows the complex to disengage from the cleaved RNA transcript and resume surveillance for foreign nucleic acids.

High-resolution structures of the Csm complex found in S. thermophilus with target RNA substrates, as well as ATP or a non-activating analogue of ATP, AMPPnP, gave insights into the mechanistic details of Csm complex function (Guo et al., 2019a). It was found that without Csm5 to cap the 3’ end of the crRNA, an added Csm3 subunit can fill the space and continue the filament of Csm2 and Csm3 subunits along the crRNA. Thus, the stoichiometry of type III CRISPR-Cas complexes can vary based on the functional length of the crRNA spacer sequence. Structural studies of the Csm complex T. onnurineus suggest that crRNA-directed cleavage
patterns of target RNA can differ with variations in complex stoichiometry (Jia et al., 2019b). A Csm complex with only one Csm2 and two Csm3 subunits robustly cleaves up to one site of the target RNA, while a complex with two Csm2 and three Csm3 subunits cleaves up to another site further along the target RNA (Jia et al., 2019b). Both types of complexes are able to cleave both RNA and DNA efficiently. Thus, it is suggested that the stoichiometry of the Csm complex can be written as Csm1\(_2\)n\(_3\)+14\(_1\)5\(_1\) (You et al., 2019; Jia et al., 2019b). The stoichiometry of Cmr is similar to Csm in that it can vary depending upon the length of the crRNA spacer sequence; as such, it can be written as Cmr1\(_1\)2\(_1\)3\(_1\)4\(_n\)+15\(_n\)6\(_1\) (Sofos et al., 2020). Variations in complex stoichiometry allow for enhanced adaptive immunity and flexibility to respond to various invader-derived spacer sequences.

**Avoiding autoimmunity**

The nucleic acid cleavage mechanism of type III CRISPR-Cas systems is a powerful and effective tool against invading nucleic acids. Unlike other CRISPR-Cas systems, type III systems do not recognize a protospacer adjacent motif (PAM) to avoid cleavage of its own nucleic acids. In order to avoid cleavage of nucleic acids derived from the host organism itself, which would be detrimental to its survival, type III systems depend upon sequence complementarity between the 5’ tag of the crRNA and the 3’ anti-tag of the target RNA. When the 5’ tag of crRNA is fully complementary to the 3’ anti-tag of target RNA, the effector complex will cleave target RNA, but ssDNA cleavage activity of Cas10 will not be activated (Johnson et al., 2019; Estrella et al., 2016). Conversely, when the 5’ tag of crRNA is non-complementary to the 3’ anti-tag of target RNA, Cas10 will be activated and ssDNA cleavage will occur. The precise mechanism by which type III effector complexes avoid autoimmunity depends upon the composition of the complex. In the Csm complex of *S. thermophilus*, an important arginine residue of a Cas5 subunit (Csm4)
prevents base-pairing between crRNA and the fully complementary target RNA strand at nucleotide -6 (You et al., 2019). This occurs after a conformational change is induced in complex to accommodate the target RNA while preventing duplex formation beyond that key nucleotide to prevent an autoimmune response. The duplex formed between the 5’ tag and 3’ anti-tag is lodged in a cleft between the Cas10 (Csm1) and a Cas5 subunit (Csm4), which eventually displaces Cas10 and rotates the complex toward the target RNA binding channel of the Csm complex (You et al., 2019).

Conversely, the 3’ anti-tag of non-complementary target RNA interacts with a linker and loop L1 in Cas10 (Csm1), which is thought to be involved in activating ssDNA cleavage (You et al., 2019). In the presence of the 3’ anti-tag of the non-complementary target RNA strand, the HD and Palm1 domains of Cas10 (Csm1) shift and the linker becomes more ordered, perhaps allosterically activating the ssDNase and cyclic oligoadenylate synthesis activities of each domain, respectively. This interaction is involved in discriminating self from non-self RNA in Csm complexes, but the region that mitigates this mechanism is not present in the type III-B Cas10 (Cmr2) (You et al., 2019). Instead, the unique stalk loop of the Cas7 (Cmr3) is implicated in the Cmr complex’s self-recognition mechanism (Guo et al., 2019b; Sofos et al., 2020). When the 5’ tag base-pairs with the 3’ anti-tag of the self target RNA, Cas7 (Cmr3) undergoes a conformational change that rotates the stalk loop 90 degrees inward into a retracted state. The rotation is coordinated with the movement of Cas10 (Cmr2) away from the target RNA binding channel. Ultimately, the retracted state of the stalk loop results in conformational changes throughout the complex that favor the Cmr2-inactive state, which halts ssDNA cleavage and cyclic oligoadenylate synthesis (Sofos et al., 2020).

**Cyclic oligoadenylate signaling pathway**
A unique feature of type III CRISPR-Cas systems is their ability to mount a robust immune response to invading nucleic acids not only through target RNA and ssDNA cleavage, but also through cyclic oligonucleotide (cOA) signaling (Kazlauskiene et al., 2017; Niewoehner et al., 2017). The structural basis for cOA signaling has been reviewed extensively (Molina et al., 2020; Huang and Zhu, 2021). Here we will give a brief overview of this pathway and discuss recent findings. As aforementioned, the HD nuclease domain of Cas10 is responsible for degrading DNA (Makarova et al., 2020; Estrella et al., 2016). Cas10 also possesses two Palm domains, one of which contains a GGDD motif (Makarova et al., 2002; Koonin and Makarova, 2017). Early structural studies of Cas10 show a potential binding site for adenyl nucleotides created by the GGDD motif and the amino P-loop motif, the residues of which are conserved amongst many polymerases and cyclases with Palm domains (Cocozaki et al., 2012; Kazlauskiene et al., 2017). It was eventually shown that in addition to degrading ssDNA, Cas10 generates cOAs upon target RNA binding (Kazlauskiene et al., 2017; Niewoehner et al., 2017). The cOAs produced by Cas10 bind and activate RNases with an N terminal CRISPR-associated Rossmann fold (CARF) domain and a C-terminal higher eukaryotes and prokaryotes nucleotide-binding (HEPN) domain (Makarova et al., 2014; Jia et al., 2019a). Csm6, found in type III-A, and Csx1, found in type III-B, III-C, and III-D are considered ancillary RNases as they are not a part of the effector complex but elicit downstream effects as part of the overall immune response (Jia et al., 2019a, 6; Molina et al., 2019). Csm6 is activated by the second messenger cyclic tetra-adenylate (cA₄) and Csx1 is activated by cyclic hexa-adenylate (cA₆), and crystal structures of these RNases in complex with their respective cOAs shows that one cOA binds each CARF domain in a symmetrical homodimer that induces a conformational change and activates the HEPN domain (Jia et al., 2019a; Molina et al., 2019; Garcia-Doval et al., 2020). It is predicted
that the CARF domain of Csm6 autoregulates RNase activity by degrading its cOA activators, which is a specialized mechanism of CARF family proteins.

Accessory nucleases provide an additional layer of defense in type III systems. The cA₄-activated CRISPR ancillary nuclease 1 (Can1) has two CARF domains and two DNA nuclease-like domains, which are involved in nicking of supercoiled DNA that leads to the collapse of viral DNA replication forks (McMahon et al., 2020). In the type III-A CRISPR-Cas system of *Staphylococcus epidermidis*, two nucleases associated with the degradosome—a complex formed from multiple protein subunits and responsible for processing and degrading RNA in bacteria—were linked to the system’s overall immune response (Chou-Zheng and Hatoum-Aslan, 2019). A study of the type III-B system of *Vibrio metoecus* (VmeCmr) showed that upon target RNA binding, cyclic triadenylate (cA₃) is generated and activates an endonuclease encoded by the VmeCmr genetic locus, NucC (Grüschow et al., 2021). NucC has non-specific dsDNA degradation activity, and a recent study of jumbo phages that infect type III CRISPR-Cas containing *Serratia* suggests that NucC activation triggers an abortive infection mechanism (Mayo-Muñoz et al., 2022). Jumbo phages attempt to subvert their host by forming nucleus-like structures around their genetic material. NucC combats this by degrading all dsDNA within the cell, including the host chromosome, which sets off cell death pathways to extinguish viral replication and propagation throughout the population. Interestingly, *Serratia* encodes type I-E, I-F, and III-A CRISPR-Cas systems, but only the interference and cyclic oligoadenylate signaling capabilities of type III were able to mount a defense against jumbo phages (Malone et al., 2020; Mayo-Muñoz et al., 2022). The collective action of type III systems to initiate antiviral responses in bacteria has been termed the cyclic oligonucleotide-based anti-phage signaling system (CBASS), and systematic studies have drawn comparisons between CBASS and the
mammalian antiviral defense pathway, cGAS-STING (Cai et al., 2014; Lowey et al., 2020). Both pathways utilize cyclic adenylate molecules as second messengers in response to nucleic acid detection, underscoring a conserved mechanism for defense against viral threats. This gives evidence to the elegance and complexity of immunity mediated by type III CRISPR-Cas systems.

Not only do nucleases directly protect against invading nucleic acids, but some also regulate the cyclic oligoadenylate pathway via cOA degradation. Ring nucleases activated by cA₄, known as Crn1 and Crn2, degrade cA₄ in a time-dependent manner to terminate the cOA signal (Samolygo et al., 2020). Csx3/Crn3, another enzyme found in many type III systems, has been characterized as an RNA deadenylase and cA₄-specific ring nuclease, highlighting the conserved necessity for these nucleases in cyclic oligoadenylate signaling pathways (Shmakov et al., 2018; Shah et al., 2019; Athukoralage et al., 2020b). Degradation of cyclic oligoadenylates acts as an important regulatory step in the type III CRISPR-Cas immune response, as unbridled or inappropriate signaling could have detrimental effects to the host. These include targeting the host’s own genetic material or triggering cell dormancy and programmed cell death pathways, which are considered to be the most severe and, thus, final course of action against invading nucleic acids. In Sulfolobus islandicus, a membrane-associated DHH-DHHA1 family nuclease (MAD) was found to rapidly degrade cOA in a metal-dependent manner to deactivate Csx1 (Zhao et al., 2020). It is predicted that MAD acts in the late stage of the type III CRISPR-Cas immune response. The current model suggests that ring nucleases provide the first layer of control to degrade cOAs, while MAD exhibits non-specific degradation of cOAs, RNAs, and even cytoplasmic DNA as a secondary regulator. Many recent studies have shown that nucleases
are vital not only for mediating immunity against invading nucleic acids, but also for preventing the toxic effects of the type III CRISPR-Cas immune response.

**Anti-CRISPRs**

Viruses have evolved their own mechanisms to subvert the type III CRISPR-Cas immune response. Anti-CRISPRs (Acrs) are small proteins of around 150 amino acids or less encoded or carried by viruses (van Houte et al., 2016). Most Acrs are specific to subtype, but some have cross-activity within classes. Mechanisms of deactivation by Acrs mainly act to disrupt DNA binding and substrate cleavage. Some viruses have adapted to target the cyclic oligoadenylate pathway, such as AcrIIIB1, the first Acr found to inhibit type III systems (Bhoobalan-Chitty et al., 2019). Although the exact mechanism of action of AcrIIIB1 is unknown, it is thought that it binds type III-B complexes directly to inhibit it. Other mechanisms of anti-CRISPR action involve inhibiting COA synthesis, binding and sequestering COA, or degrading COA to prevent RNase activation. In type III CRISPR-Cas systems, the DNase function of Cas10 is constitutively turned on when target RNA binds crRNA in effector complex (Citation). The cOA synthetase activity of Cas10 is necessary for activation of RNase activity of Csx1/Csm6, which targets middle or late-stage transcription of viral genes (Citation). Viruses that encode a DNase inhibitor alone cannot survive this aspect of type III immunity. Thus, viruses that survive to middle or late-stage viral gene transcription tend to target functions that inhibit the activation of Csx1/Csm6 RNase (Bhoobalan-Chitty et al., 2019). Preventing RNase activation prohibits the cell from entering dormancy or cell death pathways, allowing virus to continue to replicate and mature. AcrIII-1 is a viral ring nuclease that targets a cOA used widely amongst type III systems, cyclic tetra-adenylate (cA₄). Mimicking the action of endogenous ring nucleases, AcrIII-1
rapidly degrades cA4 molecules at a catalytic rate that is 60 times more active than Crn1 derived from *Sulfolobus islandicus* (Athukoralage et al., 2020a).

In order to survive viral infection, CRISPR-Cas systems must find ways to adapt to Acrs. As aforementioned, a common mechanism of Acrs is to target CARF-domain containing RNases of type III systems (Bhoobalan-Chitty et al., 2019; Athukoralage et al., 2020a). The Csm complex of *S. thermophilus* may combat this by encoding two Csm6 homologs, termed StCsm6 and StCsm6’ (Kazlauskiene et al., 2017). The two homologs have only 34% amino acid identity, suggesting that StCsm may have evolved to avoid detection by Acrs that target Csm6. Another mechanism of CRISPR-Cas inhibition is to target specific cOAs, such as AcrIII-1 and its substrate, cA4 (Athukoralage et al., 2020a). Intriguingly, cA4 activates almost all accessory CRISPR nucleases discovered to date. Only the CARF-domain containing nucleases of Firmicutes and Actinobacteria are activated by cA6, indicating that they may have adapted to using a different cOA in response to Acrs (Athukoralage et al., 2020a). Still other Acrs bind the effector protein or complex to halt the impending immune response (Thavalingam et al., 2019; Bhoobalan-Chitty et al., 2019). The previously mentioned SisCmr-β complex may have an interesting solution to this dilemma. One prediction for the function of the multiple unique Cmr7 subunits that surround SisCmr-β is that they create steric hinderance to protect the complex (Sofos et al., 2020).

**Cas7-11 (gRAMP)**

Until recently, only the genetic locus that encodes subtype III-E CRISPR-Cas systems was known (Makarova et al., 2020). As we have seen in subtypes III-A through III-D, type III systems are classified as Class 1 multi-subunit effector complexes. Subtype III-E, however, is characterized by a fusion between Cas7 and Cas11 subunits, what is referred to as giant repeat-
associated mysterious protein (gRAMP) or Cas7-11 (Makarova et al., 2020; Özcan et al., 2021; van Beljouw et al., 2021). Recent studies show four Cas7 subunits fused to one Cas11 subunit. The genes that encode this single polypeptide also encode a Cas1 adaptation protein that typically has a reverse transcriptase domain, which suggests that type III-E systems acquire spacers for its CRISPR array from RNA invaders (van Beljouw et al., 2021). Interestingly, Cas7-11 does not need a Cas6 or additional accessory proteins to process crRNA transcripts into a mature crRNA. This functional capability of Cas7-11 is similar to that of other effector proteins, such as type V Cas12 and type VI Cas13 (Yan et al., 2019). The resulting mature crRNA contains a guide sequence of around 20-nt in length and a 5’ CRISPR repeat of 14-28 nt, the final 14 nucleotides of which are highly conserved amongst type III-E systems. A putative hairpin loop formed by the repeat crRNA 5’ tag is thought to form a bulge that Cas7-11 recognizes for cleavage (Özcan et al., 2021).

While crRNA cleavage is metal-independent, target RNA cleavage is metal-dependent and occurs according to the conserved 6-nt ruler mechanism previously mentioned for other type III CRISPR-Cas systems. The spacing of the catalytic Cas7 subunits along the crRNA and target RNA duplex is responsible for this pattern, which allows cleavage at two distinct sites. Biochemical assays validate genetic analyses that a conserved aspartate residue is necessary for cleavage at one site. The mechanism of cleavage at the second site is unknown, though is still thought to be carried out by another Cas7 domain (Özcan et al., 2021). The functions of the remaining Cas7 domains, as well as that of the large insert in one Cas7 subunit, are also undefined.

Unlike other type III systems, Cas7-11 does not have the ability to produce cyclic oligoadenylates. It has not been linked to affect downstream activities associated with cOA
signaling. Interestingly, most known type III-E systems have acquired an accessory protein known as Csx29 instead. Csx29 has a tetratricopeptide repeat domain (TPR) that is fused to a caspase-like protease (CHAT), which implies that Csx29 may have a role in inducing caspase-like functions in protein degradation or cell death pathways (van Beljouw et al., 2021). A recent paper validated this finding, using cryo-EM to characterize the structure of the type III-E effector complex associated with the TPR-CHAT/Csx29 (Hu et al., 2022). The complex, termed Craspase, is able to cleave target RNA and initiate specific protease activity. As with Cas10 inactivation in other Type III systems, Craspase is not activated when the target RNA contains a protospacer flanking sequence that matches the crRNA to avoid autoimmunity (Hu et al., 2022). These results prove promising for the development of novel and highly specific RNA and protein knockdown techniques, as Cas7-11 has been shown to have significantly less off-target effects than comparable technologies such as short hairpin RNAs (shRNAs) and Cas13. Fusion of an adenosine deaminase (ADAR2) to Cas7-11 allowed for highly site-specific RNA modification of adenosine to inosine in vivo (Özcan et al., 2021). This high specificity came at a cost to cleavage efficiency, but it nonetheless provides a launchpad for future innovation in RNA targeting technology. The ability of Craspase to target specific sequences of RNA and to regulate itself from cleaving host nucleic acids and proteins has great potential for the development of diagnostics, tool, and potential treatments (Hu et al., 2022).

**Overview**

The adaptive immune response elicited by type III CRISPR-Cas systems is powerful. As the most ancient and widely used CRISPR-Cas system in prokaryotes, type III systems have evolved a sophisticated series of mechanisms to combat invaders. After adapting invader-derived spacer sequences into the CRISPR array, transcription of the array and subsequent processing
allows for the generation of mature crRNAs (Sternberg et al., 2016; Ye et al., 2020). Mature crRNAs bind to multi-subunit type III CRISPR complexes and survey the cellular environment for RNA transcripts (Hatoum-Aslan et al., 2011; Hale et al., 2012). Mismatches between the 5’ tag of the crRNA and the 3’ anti-tag of the target RNA activate the ssDNA cleavage and cOA synthesis activities of the signature Cas10 protein. Autoimmunity is avoided by base-pairing interactions between the 5’ tag of crRNA and the 3’ anti-tag of self RNA transcripts (Hale et al., 2012; Johnson et al., 2019). Cleavage of target RNA bound in the complex is metal-dependent and acts as a timer mechanism, as the activities of Cas10 are inhibited when target RNA dissociates from the complex (Hale et al., 2009; Staals et al., 2013; Hatoum-Aslan et al., 2013; Jia et al., 2019b). cOAs synthesized from Cas10 activate CARF domain proteins, such as Csm6/Csx1, which have non-specific RNA cleavage activity and cut mRNAs throughout the cellular environment (Jia et al., 2019a; Molina et al., 2019). The activity of these RNases can be indirectly inhibited by endogenous regulatory factors known as ring nucleases, which degrade cOAs. Anti-CRISPRs (Acrs) have co-opted the mechanisms of ring nucleases to avoid destruction by the immune response (Athukoralage et al., 2020a; Thavalingam et al., 2019). Recent structures of the gRAMP or Cas7-11 complex is translated as a singular protein and can process its own crRNA, but it still cleaves RNA via the ruler mechanism conserved amongst other type III systems (Özcan et al., 2021). As its genetic locus does not contain the signature cas10 gene, the gRAMP has not been shown to have cOA signaling activity (van Beljouw et al., 2021). Craspase is an effector complex associated with a caspase-like protein that has RNA-targeting protease activity (Hu et al., 2022). Structural studies using x-ray crystallography and cryo-EM have given many insights into the mechanisms of the type III CRISPR-Cas adaptive immune response, as well as the mechanisms that viral invaders use to subvert host immunity.
**Future Directions**

Since the discovery of CRISPR-Cas systems, an explosion of information about prokaryotic adaptive immune systems has brought type III systems into light. Studies of type III-A/D Csm complexes and type III-B/C Cmr complexes have dominated the field for years, yet the recent characterization of the Cas7-11 complex has incited interest in understanding type III-E systems (Özcan et al., 2021; van Beljouw et al., 2021). Type III systems have long been overlooked for use in genetic engineering tools due to their complicated assembly and narrow field of applications, but other systems such as Cas9 and Cas12 can be toxic to some bacterial strains (Liu et al., 2020). As type III systems are some of the most widely dispersed CRISPR-Cas systems in nature, their use as endogenous tools is attractive. The specificity and simplicity of the Cas7-11 CRISPR-Cas system provides an interesting opportunity for RNA knockdown technology and microbial engineering. Structural studies of Cas7-11 could provide insights in modifying and fine-tuning the complex for innovative uses.

Some research groups have sought opportunities for diagnostic development in the inherent signal amplification of the type III cyclic oligoadenylate pathway. In this pathway, Csm Cas10 synthesizes cOAs, protons, and pyrophosphates, each of which can be detected by colorimetric or fluorometric outputs (Santiago-Frangos et al., 2021). Santiago-Frangos et al. designed guide RNAs that target highly conserved segments of the SARS-CoV-2 genome and coupled *Thermus thermophilus* Csm complex (TtCsm) to reverse transcription loop-mediated isothermal amplification (RT-LAMP). In doing so, they developed an assay that could detect viral RNA at approximately 200 copies/μL in under 30 minutes, repurposing the type III cOA signaling pathway for a rapid, sensitive, and effective diagnostic tool. Similarly, Steens et al. utilized a LAMP-CRISPR based assay to detect a synthetic copy of the SARS-CoV-2 E-gene in
the atto-molar \((10^{-18}\) moles per liter) range (Steens et al., 2021). These proof-of-concept studies provide evidence that type III CRISPR-Cas systems can serve as programmable diagnostic tools.

While much has been discovered about the factors that act in the cOA pathway, little is known about the downstream effectors of cOA signaling and how they directly lead to cell dormancy and cell death. It has been suggested that various transcription factors and other proteins could be involved in the prokaryotic immune response. One such transcription factor, Csa3b, binds cOAs and the promoter of \(cas\) genes involved in the adaptation or acquisition stage of the type III CRISPR-Cas immune response (Ye et al., 2020). Csa3b has been implicated in feedback repression and points to a sophisticated method of regulation in type III systems. Additionally, newly discovered anti-CRISPRs (Acrs) can inform our understanding of CRISPR-Cas systems. Many Acrs target common enzymes or substrates along the cOA pathway, such as AcrIII-1 and cA4 (Athukoralage et al., 2020a). Others, like AcrIIIB1, bind and inhibit entire type III complexes (Bhoobalan-Chitty et al., 2019). Acrs that are yet to be identified may target downstream effectors in the cOA signaling pathway; as such, further study of Acrs may help illuminate unrevealed mechanisms of type III CRISPR-Cas mediated immunity.

Despite the surge of discovery surrounding CRISPR-Cas systems over the last decade, there are still gaps in the field. Type III-F systems, for example, were only recently distinguished as a separate subtype (Makarova et al., 2020). It is unique amongst type III systems in that its complex contains only one Cas7-like subunit. In the HD domain of its Cas10-like large subunit, conserved catalytic residues remain intact, indicating that it has ssDNA cleavage activity. However, the Cas10-like subunit does not have an active Palm domain, nor does the III-F loci encode for any CARF domain containing proteins. Thus, type III-F systems do not use functional
cOA signaling pathways to elicit immune responses. Further studies of type III-F systems could elucidate currently unknown mechanisms of ssDNA cleavage.

Recent structural and biochemical studies have provided key insights into how prokaryotes protect themselves from invading threats. These studies not only inform the design of genetic engineering and diagnostic tools, but they also highlight intriguing similarities between prokaryotic and mammalian adaptive immune systems. Type III CRISPR-Cas systems are ancient and found across many phyla of bacteria and archaea. The immune response they elicit is elaborate, consisting of cyclic oligoadenylate second messengers, which been likened to those of the mammalian cGAS-STING pathway, and transcription factors, which have monumental effects on the fate of the infected cell. Thus, the complexity and sophistication of the type III CRISPR-Cas adaptive immune system should not be overlooked. Future structural and biochemical analysis can further elucidate important mechanisms of immunity in type III CRISPR-Cas systems.
Chapter 2

Background
**Background**

CRISPR-Cas systems are adaptive immune systems in bacteria. Though these systems come in a variety of types based on their genetic loci as aforementioned, the focus of this thesis will be centered on type III CRISPR-Cas systems. Type III systems consist of multiple protein subunits and a crRNA, which can specifically recognize and bind complementary segments of RNA derived from viral transcripts, known as the RNA protospacer (Samai et al., 2015). When the crRNA and RNA protospacer base-pair, multiple activities are induced, including cleavage of the RNA protospacer and activation of various other enzymes (Hale et al., 2009; Jinek et al., 2012; Cocozaki et al., 2012). The ability of type III CRISPR-Cas systems to specifically identify and elicit a series of reactions that drive immunity in the cell. The three main activities of type III CRISPR-Cas that mediate immunity are DNA cleavage, specific RNA cleavage, and non-specific RNA cleavage (Jinek et al., 2012; Samai et al., 2015; Estrella et al., 2016).

**Figure 1.** Activities of the Type III-B Cmr complex. Made in BioRender.

*Mechanism of nucleic acid cleavage by Type III-B CRISPR-Cas complex*

In order to combat invading and foreign mobile genetic elements, Type III-B CRISPR-Cas systems utilize three different cleavage activities. One of these activities is the cleavage of
transcriptionally active single-stranded DNA (ssDNA). The signature subunit of the Type III CRISPR-Cas complexes is the Cas10, which is also known as Cmr2 in Type III-B systems (Cocozaki et al., 2012). It is suspected to have a nuclease domain that is responsible for dsDNA cleavage. Known as the histidine-aspartate (HD) domain, it is suspected to cleave phosphodiester bonds, but previous research has shown that the HD domain of Cmr2 is not required for target RNA cleavage (Aravind and Koonin, 1998).

*T. maritima* MSB8 contains three distinct Cas modules within its genome, which include Types I-B, III-A, and III-B (Makarova et al., 2015). Six proteins are encoded in the Type III-B module, Cmr1-6, which form the bulk of the effector complex. CRISPR RNAs (crRNAs) are short sequences that are transcribed from CRISPR loci and processed (Brouns et al., 2008; Mojica et al., 2005; Barrangou, 2013). The crRNA is then incorporated into the effector complex in order to identify and cleave nucleic acid sequences complementary to the crRNA’s guide region (Brouns et al., 2008; Hale et al., 2009; Jinek et al., 2012).

**Figure 2.** Schematic of genetic loci of Type III-B Cmr subunits and the assembled Cmr complex without a target RNA bound. The complex consists of six unique protein subunits, with four copies of Cmr4 and three copies of Cmr5, and the CRISPR RNA (crRNA). Made in BioRender.
The rationale for the Cmr complex cleaving transcriptionally active ssDNA stems from work on related CRISPR systems. The structurally and genetically similar Type III-A CRISPR-Cas complex, known as Csm, has been shown to cleave DNA coupled to transcription (Samai et al., 2015). Previous work in our lab intended to establish that Cmr complexes also target transcriptionally active DNA through biochemical assays. In order to understand the mechanism of DNA cleavage by the Cmr complex, the purified Cmr complex was incubated with noncomplementary ssRNA, complementary ssRNA, complementary ssDNA, and dsRNA (Estrella et al., 2016). The goal of this experiment was to determine which type of nucleic acid activated DNA cleavage activity under biologically relevant conditions. Only the ssRNA complementary to the crRNA was found to be cleaved at a site 14 nucleotides from the 3’ end of the crRNA with which it was base-paired. Cleavage was dependent upon the 5’ handle of the crRNA being intact and upon the presence of magnesium or manganese ions (Hale et al., 2012; Staals et al., 2013; Samai et al., 2015). RNA targets in other Cmr systems are cleaved at 6-nucleotide intervals in a manner known as the ruler mechanism, and biochemical analysis of the T. maritima Cmr complex supported these results (Hatoum-Aslan et al., 2011; Estrella et al., 2016). The complex cleaved its RNA target at four sites, with the first site resulting in minimal cleavage and the other three sites showing significant cleavage activity (Estrella et al., 2016). Time-course experiments suggested that the ssRNA target was cleaved rapidly in a sequential manner, starting with the second site and moving toward the fourth (Hale et al., 2009; Staals et al., 2013; Estrella et al., 2016). This work suggests that there are four catalytic subunits in the complex, each responsible for cleaving at sites 6 nucleotides from one another. Further experiments showed that ssRNA cleavage by the Cmr complex did not strictly require perfect complementarity between the crRNA and RNA target; in fact, cleavage at site four seemed to be
enhanced when mismatches were introduced into the RNA target strand (Estrella et al., 2016). The 2’-hydroxyl group adjacent to the scissile phosphate in the catalytic subunit is key in the mechanism of Cmr complex cleavage.

Thus, DNA cleavage in vitro by the Cmr complex requires the presence of a complementary ssRNA transcript and manganese ions. It was found that while the presence of the ssRNA transcript is required for DNA cleavage, cleavage of target ssRNA inhibits ssDNA cleavage activity by the complex (Estrella et al., 2016). It is suspected that this acts as a timer mechanism, as turning the complex’s activity off after the target transcript is cleaved maintains efficiency and prevents detrimental effects to the host cell (Estrella et al., 2016). RNA target binding is what allows for and triggers ssDNA cleavage. Cmr2 (Cas10) is suspected to enact DNA cleavage, and deactivating mutations in the two highly conserved residues in the HD motif (His32 and Asp33) of Cmr2 obliterated DNA activity while preserving ssRNA cleavage activity. This work suggests that the conserved HD motif in Cmr2 is necessary for the complex’s ssDNA cleavage activity.

As aforementioned, other work in the type III-A system has shown that the HD domain of Csm1 (Cas10) is not required for DNA cleavage. Rather, the GGDD motif in the Palm domain is (Ramia et al., 2014; Samai et al., 2015; Hatoum-Aslan et al., 2014). The GGDD motif is a conserved site in the Palm domain of the signature Cas10 protein, and it is similar in structure to catalytic domains of DNA polymerases and nucleotide cyclases (Jia et al., 2019b; Kazlauskiene et al., 2017; Niewoehner et al., 2017; Mönttinen et al., 2016). To assess the role of the GGDD motif in DNA cleavage by the Cmr complex, a mutant that replaced the two aspartate residues in the motif with alanine residues was generated. This mutant was unable to cleave both ssDNA and the ssRNA target, and further work showed that the mutant did not even bind the complementary
RNA target (Estrella et al., 2016). In fact, these mutations resulted in a dysfunctional complex that could not accurately bind or cleave RNA or DNA. Thus, the GGDD motif is not implicated in the DNA cleavage activity of the *T. maritima* Cmr complex.

The model of DNA and RNA cleavage enacted by the Cmr complex, known collectively as interference, is proposed to begin with RNA protospacer transcription. When the host RNA polymerase transcribes the protospacer, the Cmr complex’s crRNA base-pairs with the newly formed transcript. Binding of the target transcript activates the Cmr complex, and the HD domain of Cmr2 initiates cleavage of transcriptionally active ssDNA. While cleavage of the displaced ssDNA of the non-template strand occurs, the complex cleaves target RNA at four distinct sites 6-nucleotides apart. Once the RNA target is cleaved, DNA nuclease activity is deactivated, indicating that a timer mechanism regulates nucleic acid cleavage by the complex.

*Avoiding autoimmunity in Type III-B CRISPR-Cas interference*

Type III CRISPR-Cas systems require target transcription to confer immunity. When the transcript binds, DNA cleavage by the type III effector complex is activated. However, unregulated degradation of nucleic acids can lead to severe consequences for the host (Jia et al., 2019b; Johnson et al., 2019). Other types of CRISPR-Cas systems recognize short DNA sequences known as protospacer adjacent motifs (PAMs) to distinguish between foreign and host-derived sequences (Westra et al., 2013; Mojica et al., 2005; Semenova et al., 2011). Type III systems do not recognize PAMs. In order to avoid autoimmunity, type III systems recognize a distinct sequence of nucleotides that flanks the 3’ end of the RNA target site known as the protospacer flanking site (PFS) (Johnson et al., 2019). Mechanisms of PFS recognition can vary between different subtypes of type III CRISPR-Cas systems. In type III-A systems, an 8-nucleotide crRNA tag base pairs with a complementary PFS known as an anti-tag to deactivate
the Csm complex (You et al., 2019; Jia et al., 2019b). Previous work in our lab characterized the mechanism of PFS recognition and autoimmunity avoidance in type III-B Cmr complex of *T. maritima*. As alluded to earlier, type III CRISPR-Cas systems are notoriously tolerant of mismatches. RNA target strands containing numerous mismatches can still activate RNA and DNA cleavage (Estrella et al., 2016; Maniv et al., 2016; Silas et al., 2018). This robust immune response prevents viral escape in type III systems more so than other CRISPR-Cas systems (Silas et al., 2018). The majority of data regarding type III systems comes from studies of the subtype III-A, but little was understood about how mismatches between the crRNA and its target mediate the mechanism of type III-B DNA cleavage. Previous colleagues in our lab performed experiments to test how mismatches between the crRNA tag and the PFS of the target RNA strand affect the activities of the Cmr complex (Johnson et al., 2019).

Non-complementarity between the PFS of the target RNA and the 5’ tag of the crRNA activates the complex. Positions -1 and -3—that is, the sites that are up to three nucleotides from the start of the RNA protospacer—are key to the activation of the Cmr complex’s DNA cleavage activity (Johnson et al., 2019). Experiments showed that a target containing a fully noncomplementary PFS to the crRNA elicited the same amount of DNA cleavage activity as a target with anti-tag sequence at all positions except for positions -1 to -3. These positions of the PFS are vital for the regulation of DNA cleavage (Johnson et al., 2019). Mismatches between the crRNA and the rest of the RNA target are tolerated by the Cmr complex, but mismatches in the first 5 nucleotides that follow the PFS inhibit DNA cleavage activity (Johnson et al., 2019). Significant mismatches between the RNA target and the crRNA result in the complex poorly binding the target, which in turn results in a failure to activate DNA cleavage. This research established that although the *T. maritima* Cmr complex follows lenient rules to activate DNA
cleavage, it uses strict rules to deactivate DNA cleavage by the complex. Thus, the *T. maritima* Cmr complex errs on the side of caution, prioritizing the destruction of invading nucleic acids over the potential for autoimmunity (Johnson et al., 2019).

*Overview of structural characterization of biomolecular complexes via cryo-EM*

Though these findings are significant, there is a lack of structural data for the *T. maritima* Cmr complex to support them. The objective of this thesis is to validate these biochemical findings by analyzing the structure of the *T. maritima* Cmr complex. Of the three major techniques used to determine biomolecular structures—nuclear magnetic resonance, X-ray crystallography, and cryo-electron microscopy—cryo-EM offers many key advantages (Murata and Wolf, 2018). Cryo-EM requires a smaller amount of sample for structure determination and the sample preparation is relatively easier than that necessary for NMR and X-ray crystallography (Thompson et al., 2016). Sample preparation for cryo-EM involves rapidly freezing the protein or complex of interest at cryogenic temperatures (-193 °C or 80 K); whereas techniques such as x-ray crystallography involve complicated crystallization procedures that lock proteins in static states, this rapid freezing technique allows for target structure determination at near-native and dynamic conformations (Murata and Wolf, 2018). Cryo-EM is a robust technique for studying biological macromolecules that are sensitive to radiation, allowing samples to remain intact in fully or partially hydrated forms and reducing overall radiation damage as compared to X-ray crystallography (Henderson, 1995). Another benefit of cryo-EM is the wide range of sample sizes that it is able to accommodate, from small protein complexes to entire tissue sections (Murata and Wolf, 2018). As such, cryo-EM is remarkably suited for studying biomolecular complexes such as CRISPR-Cas complexes.
Thesis rationale

As structural databases continue to grow with information regarding CRISPR-Cas systems, there continues to be a gap in the field regarding structural information about the *T. maritima* Cmr complex. Our lab has characterized the biochemical mechanisms of the Cmr complex’s activities; namely, its role in targeting transcriptionally active DNA (Estrella et al., 2016) and in specific cleavage of the RNA protospacer (Johnson et al., 2019). Although the structures of individual subunits and entire Cmr complexes of other organisms have been solved, there is little structural information regarding the *T. maritima* Cmr complex in particular. We therefore intended to characterize the structure of the *T. maritima* Cmr complex using cryo-EM. The objective was to complete the story of the Cmr complex and gain insight into how the complex cleaves nucleic acids from a structural standpoint.
Chapter 3

Methods
Methods

A schematic of the experimental design is outlined below:

**Figure 3.** Project workflow. Cmr2-6, His-tagged Cmr1 (His-Cmr1), and crRNA are purified individually. The subunits are combined, incubated at 80 °C for 20 minutes, and purified over analytical size-exclusion chromatography. The purified Cmr complex is frozen on surface of grids are imaged on a cryo-electron microscope (Titan Krios Cryo-TEM). The resulting images are processed using software such as RELION and cryoSPARC. Maps are refined to create a three-dimensional model of the protein and nucleic acid complex. Made in BioRender.

*Purification of Cmr2-6 Partial Complex and His-Cmr1*

Assembly of the *Thermotoga maritima* (*Tma*) Cmr complex begins with the purification of a partial complex consisting of subunits Cmr2-6. Plasmids encoding subunits *TmaCmr2-6* were transformed separately into NEB T7 Express competent *E. coli*. The plasmid encoding *TmaCmr6* contains a polyhistidine tag, known as a His-Tag. The plasmid encoding the catalytic subunit, *TmaCmr4*, has a disactivating mutation that prevents the assembled complex from degrading target RNA. Starter cultures were inoculated into 50 mL of sterilized Luria broth (LB) and left to grow overnight. After approximately 12-16 hours, 1 L flasks of LB were inoculated with 5 mL of each starter culture and left to grow at 36 °C until they reached an optical density of
0.300, measured by UV-Vis spectrophotometry of optical density at 600 nm. At this OD600, the cultures were moved to an incubator at 20 °C and induced at an OD600 of approximately 0.400. The cultures were allowed to grow overnight at 20 °C for 12-18 hours. These cultures were then centrifuged, and the cell pellets were collected and flash frozen in liquid nitrogen. On ice, 1L each of Cmr2 and Cmr3 and 2L each of Cmr4, Cmr5, and His-Cmr6 are thawed, combined, and resuspended in 200 mL of Cmr Lysis buffer (1 M KCl, 20 mM Tris at pH = 8.0, 10 mM imidazole at pH = 7.5, 1 mM TCEP). Protease inhibitors were added to the resuspended cell pellets to prevent potential degradation of the Cmr subunits during lysis. The resuspension was passed through cheesecloth and vacuum filtered before lysing with a microfluidizer. It is assumed that this method of lysis will allow for the Cmr subunits to interact and assemble into partial complexes during this procedure. The lysate was passed through the microfluidizer three times and then received a heat treatment at 65 °C for 10 minutes to denature non-Cmr proteins. The lysate was clarified via centrifugation and passed over a nickel-charged immobilized metal affinity chromatography (Bio-Rad Profinity IMAC) column equilibrated in the Cmr Lysis buffer. The His-tag of the Cmr6 subunit binds the nickel in the IMAC resin, allowing the Cmr2-6 complex to stick to the resin while other proteins flow off the column. The resin is washed with approximately 50 mL of Cmr Lysis buffer. The complex is then eluted from the column with 10 mL Cmr Elution buffer (500 mM KCl, 20 mM Tris at pH = 8.0, 250 mM imidazole at pH = 7.5, 1 mM TCEP). The fractions containing the complex are pooled and injected onto a size-exclusion chromatography (SEC) Superdex 200 pg column equilibrated in Cmr Gel Filtration buffer (200 mM KCl, 20 mM Tris at pH = 8.0, 1 mM TCEP). Samples were taken of the resulting fractions and analyzed via SDS-PAGE (Figure below). Those containing Cmr2-6 complex were pooled and concentrated using 30K MWCO spin concentrators. The final samples
were flash frozen and stored at -80 °C. His-Cmr1 was expressed and purified using the same protocol. Rather than mixing His-Cmr1 with the rest of the cell pellets, His-Cmr1 was purified separately because previous work has shown that it is unstable in complex without the crRNA (Estrella et al., 2016; Johnson et al., 2019).

**Assembly and purification of full T. maritima Cmr1-6 complex**

The crRNA was ordered from Sigma Aldrich. At room temperature, purified partial complex Cmr2-6 was combined with purified His-Cmr1 and crRNA in excess (1:2:2, respectively). The sample was incubated at 80 °C for 20 minutes. The sample was centrifuged at maximum speed at room temperature for 5 minutes before immediately injecting onto an analytical gel filtration Superose 6 column equilibrated in Cmr Gel Filtration buffer (200 mM KCl, 20 mM Tris at pH = 8.0, 1 mM TCEP). Samples were taken from the resulting fractions and analyzed via SDS-PAGE.

Figure 4. SDS-PAGE gels confirming the presence of His-Cmr1 and Cmr2-6 subunits post-purification.
Imaging complex using cryo-EM

The fully assembled Cmr1-6 and crRNA apo complex was given to Brian Learn to freeze on cryo-EM grids for imaging. The sample was frozen on two types of grids at varying concentrations to test freezing and ice conditions. Two AuFlat gold grids and two CF carbon/Cu grids were frozen with sample at 0.5 mg/mL and 1.0 mg/mL, respectively. The grids were screened and the CF carbon/Cu grids showed the most promise in terms of particles that looked like complex and were not aggregated. Images were collected on this grid on the Krios cryo-EM microscope at 40 electrons per A² and 8 shots per grid hole. A server error cut off imaging after only three hours of collecting data. The limited number of micrographs were processed for initial classification and map building.

Processing cryo-EM data

Initial motion correction, contrast transfer function (CTF) refinement, and image selection jobs were run in RELION-3.0 (Scheres, 2012). Particle picking was performed in crYOLo with a box size of 200 pixels or approximately 200.7 Å (Wagner et al., 2019). The rest
of the data processing, including 2D classification, ab-initio reconstruction, 3D classification, and homologous and non-uniform refinement, was carried out in cryoSPARC (Punjani et al., 2017). The general EM data processing workflow is outlined below.

**Figure 6.** Cryo-EM data processing workflow.
Chapter 4

Results and Discussion
Results and Discussion

Initial map of Cmr complex

Initial 2D classes were chosen for their resolution and general shape consistent with the expected structure. The particles selected from the most promising 2D classes (55,179 particles) were used to build an ab initio reconstruction. The ab initio reconstruction was performed with 2 classes to filter out any undesirable particles. Of the resulting classes, the one with the majority of particles (36,995 particles) was selected for homogenous refinement and subsequent non-uniform (Nu) refinement. This resulted in an ab initio map of the Cmr complex with a resolution of 3.94 Å, shown in Figure 7.

![Ab initio model of TmaCmr complex at a resolution of 3.94 Å.](image)

Figure 7. Ab initio model of TmaCmr complex at a resolution of 3.94 Å.

The density of the ab initio model was very weak at the lower end of the structure where the Cmr2 (Cas10) subunit is located. We then performed 3D classification on the 36,995 particles to attempt to tease out the classes with more robust Cmr2 density. We performed homogenous refinement on the particles from two classes (7,809 particles). The resulting density map with a resolution of 4.54 Å is shown in Figure 8.
Although the resolution of the map is low, we were able to observe the individual protein subunits of the complex. The resolution was too low to be able to visualize the crRNA. Cmr2 and Cmr3 are located at the 5’ end of the crRNA. Three Cmr4 subunits and two Cmr5 subunits form a helical filament around the crRNA. Cmr1 and Cmr6 cap the 3’ end of the crRNA. This structure does not have an RNA target bound and is catalytically inactive due to a mutation in Cmr4 (D26A). The stoichiometry of the catalytic Cmr4 subunits and the small Cmr5 subunits does not correlate with the biochemical data collected from this complex. Previous work in the lab shows the presence of four cleavage sites along the RNA protospacer, but the map from this structure suggests that there are only three catalytic subunits able to cleave the target RNA (Staals et al., 2013; Estrella et al., 2016). There is also a missing Cmr5 subunit according to work published by other labs in other organisms with Type III-B systems (Staals et al., 2013; Taylor et al., 2015; Zhang et al., 2012; Molina et al., 2020; Sofos et al., 2020). These results indicate that

Figure 8. Initial map after 3D classification and homogeneous refinement to a resolution of 4.54 Å.
the structure of the Cmr complex that was purified and obtained via this protocol was not the expected outcome stoichiometrically.

_Potential reasons for subcomplex formation_

There are a few potential explanations for the formation of this subcomplex. One is that the crRNA became degraded during assembly, so the complex formed with the incorrect stoichiometry around an incomplete piece of crRNA. Previous work on the Type III-A Csm complex in _Thermococcus onnurineus_ showed that the subunits that form helical filaments along the backbone of the complex can vary based on the length of the crRNA (Jia et al., 2019b). This suggests that a shorter Cmr complex could form around a shorter crRNA. It is also possible that the crRNA is intact, but the subcomplex forms around only part of the crRNA. At higher resolutions, we might be able to see the remaining crRNA at the top of the complex. However, the resolution of our Cmr map is too low to be able to verify this conclusion. Running denaturing gel electrophoresis or UREA-PAGE on the sample would elucidate the quality of the crRNA.

Another potential reason for the observation of a subcomplex is sampling bias. Imaging of the sample via cryo-EM was interrupted by a server error, such that only about 14% of the data was collected and processed. It is possible, although unlikely, that the particles that were imaged were unrepresentative of the entire sample. Ultimately, the best way to determine the proper structure, stoichiometry, and conformation of the Cmr complex would be to optimize the protocol. One potential method to encourage the formation of the expected Cmr complex would be to grow additional liters of _E. coli_ expressing plasmids Cmr4D26A and Cmr5. After confirming the presence of soluble protein in the clarified lysate of these cultures, lysing these additional liters with the rest of the subunits of the complex may push the complex toward its expected stoichiometry.
Chapter 5

Conclusion and Future Directions
Conclusion and Future Directions

Type III-B CRISPR-Cas systems fight invading genetic elements by surveying the cellular environment, binding, and cleaving complementary RNA (Hale et al., 2009; Zhang et al., 2012; Hale et al., 2012; Staals et al., 2013). The Cmr complex, which consists of protein subunits Cmr1-6 and a crRNA, can cleave DNA and RNA (Taylor et al., 2015; Samai et al., 2015). Previous work in our laboratory has established the basic biochemistry that enables ssDNA cleavage in the presence of a complementary RNA target, and mismatch tolerance between the crRNA and its RNA target to elicit the complex’s activity (Estrella et al., 2016; Johnson et al., 2019). The objective of this thesis was to characterize the structure of the Cmr complex and validate these biochemical findings.

The Cmr complex cleaves target RNA at four sites separated by 6-nt intervals (Ramia et al., 2014; Osawa et al., 2015; Taylor et al., 2015; Estrella et al., 2016). This indicates the presence of four catalytic subunits, Cmr4, in the Thermotoga maritima complex. However, the structure obtained using this protocol has only three Cmr4 subunits (Figure 8). Additionally, other Type III systems have three Cmr5 subunits, whereas the density map shown in Figure 8 shows only two Cmr5 subunits (Osawa et al., 2015; Jia et al., 2019b; Sofos et al., 2020). The expected stoichiometry of the complex is Cmr1;2;3;4;5;61, although some research suggests that the stoichiometry of CRISPR-Cas complexes is flexible and the number of Cmr4 and Cmr5 subunits can vary based on the length of the crRNA (Staals et al., 2013; Sofos et al., 2020; Wittig et al., 2020). Based on the stoichiometry observed here, it is possible that the crRNA was degraded in vitro due to nuclease contamination prior to complex assembly. This issue could be remediated by using an uncontaminated sample of crRNA and by minimizing the potential for contamination using nuclease-free techniques. Lysing additional liters of Cmr4- and Cmr5-
expressing cultures may help encourage the formation of the expected stoichiometry instead of the subcomplex, assuming that the cultures express and yield soluble protein.

In the future, the purification protocol should be optimized to confirm the presence of each individual protein subunit prior to complex assembly. The quality of the crRNA should be assessed as well. After complex assembly, imaging and processing a full set of particles would help to obtain a higher resolution structure of the Cmr complex. Further processing using programs such as Phenix to refine and DeepEMhancer to sharpen density maps would elucidate the finer molecular details of the complex’s activity. Future experiments would allow for the assembly of the Cmr complex with a complementary target RNA bound, as well as with a non-complementary target RNA. Higher resolution maps could show the interactions that the catalytic aspartate residues of multiple Cmr4 subunits have with target RNA, which would validate the biochemistry previously established in this lab. These maps could also elucidate the biochemical role of Cmr2 in ssDNA cleavage and target RNA recognition.

In conclusion, although the ultimate objective of this thesis was not achieved, valuable insights into the purification, assembly, and cryo-EM imaging of the non-target bound structure of the TmaCmr complex were gained. Future work will help establish a reliable protocol for the purification and assembly of the complex, and will characterize the target bound structures at higher resolutions.
References


