UNIVERSITY OF CALIFORNIA SAN DIEGO

Identifying the function of cGAS-Cap2b protein complex in Type-II Short CBASS Systems

A Thesis submitted in partial satisfaction of the requirements
for the degree Master of Science

in

Chemistry

by

David Garcia

Committee in charge:

Professor Elizabeth A. Komives, Chair
Professor Kevin D. Corbett
Professor Brian Leigh

2022

The Thesis of David Garcia is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2022

DEDICATION

I dedicate this thesis to my parents, Efrain Garcia and Reyna Garcia, my grandparents, Maria Leon, Manuel Garcia, and Margarita Garcia. You sacrificed everything to pave the first roads in a new country for your children and grandchildren to succeed. Without your love and kindness, I would have never been able to make it this far. Even though I will be the first in our family to achieve such an educational milestone, your sacrifices will assure I won't be the last. And to all my family, beyond every border, this degree is yours as much as it is mine.

Ellos pintaron la raya
para que yo la brincara
y me llaman invasor.
Es un error bien marcado,
nos quitaron ocho estados.
¿Quién es aquí el invasor?

Soy extranjero en mi tierra y no vengo a darles guerra,
soy hombre trabajador.

Y si no miente la historia,
aquí se sentó en la gloria
la poderosa nación
entre guerreros valientes,
indios de dos continentes
mezclados con español.
Y si a los siglos nos vamos:
Somos más americanos,
somos más americanos
que el hijo de anglosajón.

- Los Tigres del Norte "Somos Mas Americanos"

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

It has been a tough, but beautiful journey to get to this stage of my education. I would like to thank Professor Kevin D. Corbett for his constant patience, support, and guidance. Beyond the many research skills he taught me, I am most grateful for Dr. Corbett's leadership and care for his lab members. Through the never-ending bumps in the road I faced, Professor Corbett's kindness and understanding always gave me the reassurance I needed to make it here. His approach as a PI and professor is the best and purest definition of what it means to be a great scientist and a great leader.

I would also like to thank my wonderful advisor, Professor Elizabeth Komives, who is solely the reason why I can say I have a Master of Science degree in Biochemistry. Dr. Komives' support and mentorship allowed me to be accepted into the ACS Bridge Program and write this thesis you are now reading. I am forever indebted to her for allowing me this amazing opportunity, and I am extremely grateful that she saw something in me from the start.

A special thank you to everyone in the ACS Bridge Program. Our amazing faculty director Professor Brian Leigh, who allowed me to teach his undergraduate lab course for 3 quarters and was always there to be a voice for us when we felt overlooked. To Professor Haim Weizmann for his support and lessons on navigating through graduate school. To my cohort friends, Javier Sanlley and Stanley Buczynski, thank you for always being there when we had to navigate these new roads. Similarly, to Nichelle Camden, Georges Leconte, Alexa Perez, Alexis Wright, and Victoria Lerda thank you for being great and I wish you all the same success as the next ACS Bridge cohort!

I would like to thank everyone in the Corbett Lab for welcoming me into lab and teaching me everything that made me the scientist I am today. First, to Rebecca Lau and Kevin Corbett for the hours they spent teaching me all the steps in protein expression and purification. Also, to Chelsea Blankenchip who taught me the PCR and T4 protocols that I used endlessly as a graduate student. Amar Deep and Yajie Gu who were there to rescue me when I was lost in the maze of working the FPLC or needed help setting up an experiment. I owe the most to Qiaozhen Ye who taught me how to make buffers, optimize my protocols, and let me borrow any and all of her solutions when I needed them. Finally, Lisa Liang, Alan West, Sarah Ur, Alyssa Rodriguez, Eray Enustun, Katie Chau, Justin Nguyen, and Kevin Gong, who have all been welcoming and helped me and the lab be great in general. Honorable mention to Kaori Ego for leaving behind her lab key card so I could get into lab when I didn't yet have mine.

I owe even more thanks to my friends and teachers from Imperial Valley College. My physics and math teacher, Professor David Rosas, for inspiring me to continue my education at a community college even after I was rejected to every university after high school. I want to thank my chemistry professor, Dr. Jim Fisher, for being a great mentor and my inspiration to pursue a Biochemistry degree at UCSD. Thank you to all my Organic Chemistry friends who were there for hours on end as we studied for material none of us had any idea how to do. Special thanks to my friends, Kevin Cota and Brandon Lem, for the late nights spent at McDonalds, studying and doing homework for our Calculus and Physics classes. To an awesome friend, Andrea Bernal, thanks for listening to my rants about school and for your never-ending words of encouragement that kept me going.

Additionally, I am grateful to my best friends from San Diego and the Imperial Valley. The Seeley/Yucca guys, Jared, Alvaro, G, Nick, Eddie, Leo, Cory, Claudio, and Jacob, thanks for being my escape from the toughness of college and hearing me vent for hours on end about school. You guys gave me my sense of purpose and helped me feel worthy of being at UCSD. To Kytzia, Norma, and Gerardo thank you for being around when I needed a study buddy on those late nights at Geisel, Biomed or some random 24-hour coffee shop studying for exams. Special thanks to my best friend at UCSD, Toma, for the best memories as roommates and all the nights spent enjoying the fun parts of college. Thank you to my Rous Zoo family, Matt, Ricky, Casey, Waylon, Evan, Chris, and Rae for being the best housemates and friends, and for keeping me sane when I was stressed or depressed about school.

Finally, to my parents and siblings, thank you for your constant support of me and being the inspiration behind my goal of pursuing higher education. Mom, thank you for showing me what it means to be hardworking, determined, and kind, even in times of struggle you always made sure we had a bed to sleep in and food on our table. To my dad, thank you for your dedication and all you sacrificed for us, and for your support during the times I needed some extra love to stay standing. Lastly, to my brother, Gio, and my sister, Steph, thank you for constantly reassuring me of my worth and for inspiring me to be as great as you are.

VITA

2017 Associates of Science, Chemistry, Imperial Valley College, Imperial, CA

2020 Bachelor of Science, Biochemistry, University of California, San Diego

2022 Master of Science, Chemistry, University of California, San Diego

FIELD OF STUDY

Major Field: Biochemistry

Studies in Biochemistry and Biophysics

Professor Elizabeth A. Komives

ABSTRACT OF THE THESIS

Identifying the function of cGAS-Cap2b protein complex in Type-II Short CBASS Systems

by

David Garcia

Master of Science in Chemistry

University of California San Diego, 2022

Professor Elizabeth A. Komives, Chair

Two sides, in an infinite battle for survival, as one side gets stronger, so does the other. As described by the Red Queen hypothesis, this is the dilemma between bacteria and bacteriophages (or phages), where both sides are constantly evolving to try and eliminate their opponent. Phages are simple organisms and outnumber all other organisms on earth, and their means of survival requires the destruction of bacteria. During infection, phages inject their DNA into bacterium and as a response bacterium will destroy itself but not before the phage has successfully replicated. This allows for further bacterial infection upon the destruction of bacterium from initial phage infection and can be recurrent until all bacteria has been destroyed. As a means for survival, bacteria have evolved a myriad of defense systems that allow it to fight phage infection. In this thesis I will cover some of the most well-known

defense systems: the restriction-modification (R-M) systems that target phage sequences, the CRISPR-Cas systems where cells obtain immunity from previous phage infection, and the abortive infection systems (Abi) that upon phage detection will program cell death. Although each of these systems is unique and important for phage defense, I will focus on a type of abortive infection systems that were recently discovered in bacteria known as cyclic-oligonucleotide based antiphage signalling systems or 'CBASS'.

In this thesis, I also discuss the defense mechanisms of the different types of CBASS currently identified and discuss my study of a unique type of CBASS systems described as Type II (short) CBASS systems. Previously identified Type II (long) systems were discovered to have operons containing ancillary genes that share similarities with the non-canonical E1 and E2 signaling machinery described in the noncanonical autophagic ubiquitination pathway. Type II (short) CBASS systems have similar operons but with only one ancilliary gene related to ubiquitin E2 proteins. I was able to demonstrate how the bacterial E2-like protein, Cap2b, and the nucleotidyltransferase, cGAS, in Type II (short) systems using the artificial intelligence (AI) structural prediction tool, AlphaFold. Using the structural prediction, I focused on three strains to study and clone these proteins in order to isolate and purify them for structural determination using X-ray crystallography.

CHAPTER I: INTRODUCTION

A. Bacteria Phage Conflict

Bacteria have evolved various defense pathways for survival against their ever-lasting enemy and recurrent invaders, bacteriophages (phages). Phages are the most diverse and abundant biological entities on Earth, with their population outnumbering bacteria by almost tenfold (Doron et al. 2018). In response to bacteria evolving multiple defense mechanisms against phages, phages have in turn evolved to avert these mechanisms to survive and replicate. The continual conflict between bacteria and phages has allowed for both of these organisms to undergo continuous cycles of co-evolution. In some occasions, bacterial lineage is preserved by emerging phage-resistant hosts, while simultaneously, new bacterial strains with high phage sensitivity may not survive.

Thus, phage resistance has become a target area of research aimed at understanding bacteriophage resistance mechanisms. Some of the anti-phage defense mechanisms identified, use methods such as restriction-modification (R-M) which targets specific DNA modifications, CRISPR-Cas, where cells obtain immunity from previous phage infection, and abortive infection systems (Abi) that upon phage detection cause cell death or metabolic arrest (Wilson et al. 1991). A better understanding of these myriad systems is critical for development of phage therapeutics, in which phages are used to combat persistent drug-resistant bacterial infections in critically-ill patients.

I. Restriction-Modification Systems

In the history of bacteriophage resistance, restriction-modification (R-M) systems were some of the first identified defense mechanisms, and are the most widespread among bacteria. The discovery of R-M systems can be traced back to the early 1950s, with the discovery of bacterial strains that could "restrict" or inhibit replication of viruses propagated on other bacterial strains, due to the activity of sequence-specific endonucleases. The endonucleases produce discrete DNA fragments after cleavage and could be used for analyzing and rearranging DNA, leading to the discovery of over 200 different specificities (Wilson et al 1991), and over 100 restriction modification systems being cloned (Wilson et al 1991**)**. R-M systems typically comprise enzyme pairs, with an endodeoxyribonuclease (ENase) and a DNA methyltransferase (MTase), which perform opposing intracellular enzyme activities (Wilson et al 1991). The enzymes target specific sequences in DNA usually with 4-8 defined nucleotides, which can be varying in continuity or symmetry. Additionally, double-stranded (ds) and single-stranded (ss) DNA could be recognized by these ENases and MTases. They may also exist as "cognate" enzymes (derived from the same system) and target the same sequence. However, these enzymes are rarely found as a single, multi-subunit complex and tend to be separate in most R-M systems as restriction endonucleases and modification methyltransferases.

Restriction endonuclease cleavage of double-stranded DNA occurs on every location of a recognition sequence, where it hydrolyzes a phosphate-deoxyribose bond on each DNA strand backbone. This cleavage reaction usually requires no energy input, and occurs at a fixed position within the recognition sequence, or a few bases on either side. Other systems may have hydrolysis occur on indefinite distances from the recognition site. Cleavage by these endonucleases commonly require a divalent cation such as Mg^{2+}, and stimulus by ATP or S-

adenosylmethionine (AdoMet). DNA is usually cleaved on the 5'-side of the phosphate, creating fragments with 5'-phosphoryl and 3'-hydroxyl termini, which can be rejoined by DNA ligase as opposed to a 3'-phosphoryl/5'-hydroxyl end.

The modification methyltransferases in these systems are responsible for methylation of one nucleotide on each strand of the recognition sequence. A methyl group is added to a nucleotide from methyl donor and essential cofactor, AdoMet. Methylation of recognition sequences causes disruption of cleavage by the R-M endonucleases. If methylation is cognate, conferred by the natural partner of the endonuclease, cleavage is always restricted to protect the cell's own DNA from digestion. Non-cognate methylation, however, occurs elsewhere in the recognition sequence and will not always prevent cleavage. In cases of hemimethylation (cognate methylation on only one DNA strand), cleavage may still be impeded but methylation usually occurs on both recognition sequence strands.

Mainly, R-M systems function to contain viral DNA infection within clonal bacteria populations (Wilson et al 1991). Although it seems like a basic manner to determine the presence of foreign DNA through methylation sensors, it can be quite effective as the cell's own DNA will likely be methylated before restriction. The distinction between exogenous and endogenous DNA was proposed some time ago (Meselson et al 1972), but this form of phage resistance is considered to be more protective of bacterial cells in close vicinity of the infected cell. Infection may still occur in a given bacterial cell, leading to death, but R-M systems also function to prevent the replication of the infecting phage, preventing the formation of phage progeny which would otherwise infect neighboring cells. Therefore, R-M systems may not be enough to stop phages, but bacteria have more soldiers ready for another form of battle.

II. CRISPR-Cas Defense

Clustered regularly interspaced short palindromic repeats or CRISPRs and the CRISPR associated (*Cas*) genes were first discovered in *E. coli* in 1987 by Yoshizumi Ishino, who accidentally cloned the repetitive sequence with interspersed spacer sequences while studying genes involved in phosphate metabolism. Due to limited DNA sequence data available at the time it took several decades for the CRISPR-Cas loci function to be identified. CRISPR-Cas loci were found to have a general composition consisting of 21-40 bp direct repeats with non-repetitive spacers (20-58 bp) between repeats, and a varying number of *cas* genes on either side. It was not until 2006 that researchers identified CRISPR as a defense mechanism that provides protection against phages in prokaryotes (Makarova et al. 2006). Additionally, the CRISPR-Cas system has been used most recently in genome editing research and our understanding of the complex immunity system has grown.

CRISPR-Cas systems are divided into two classes, 1 and 2, depending on the effector proteins present in the system. Class 1 CRISPR-Cas systems are widespread in archaea and bacteria, and containing multi-subunit effector complexes composed of 4-7 Cas proteins. This class makes up ~90% of all identified CRISPR-*cas* loci, and is broken up into types I, III, and IV (Ishino et al 2018). Each type (I, III, and IV) in Class I CRISPR-Cas systems have unique Cas proteins: Type I with Cas3, Type III with Cas9, and Type IV with Cas10. Each type contains a subtype (e.g I-A to F, and II-A to C, and more) that is defined by particular additional genes or gene arrangements. Type I and III systems contain complexes known as CRISPR-associated complex for antiviral defense (Cascade), but not the type IV systems which are mostly uncharacterized and believed to use crRNAs from different CRISPR arrays (Ishino et al 2018).

Class 2 CRISPR-Cas systems are also broken down into types, including types II, V, and VI with subtypes also dependent on the presence and arrangement of particular genes. Class 2 systems have a simpler architecture than that of Class 1, mostly consisting of a multidomain effector protein, Cas9, which is a crRNA-dependent endonuclease. The two endonuclease domains are termed RuvC and HNH after the nuclease families each domain falls into, and their function is to together cleave the two strands of a DNA recognized by the crRNA bound to the enzyme. Another unique feature of the class 2 CRISPR-Cas system is in its type II system that encodes a *trans*-activating crRNA (tracrRNA). This tracrRNA is an essential tool in type II systems for recognizing targets and pre-crRNA processing. Type V systems contain a different effector protein derived from the gene *cpf1* that's adjacent to *cas1*, *cas2*, and a CRIPSR-array. Cas12a (Cpf1) is a single-RNA-guided nuclease similar to Cas9, but comprises two RuvC domains, with no HNH domain present. This effector protein has a different cleavage pattern and protospacer adjacent motif (PAM) recognition used for binding target strands.

The CRISPR-Cas defense systems are extensive, and aside from phage protection mechanisms, have become the source for advancing genome editing research. With their nucleases' ability to target DNA specifically, much of the focus of CRISPR research has shifted from phage defense mechanisms to the discovery of variants within each system that may be useful for gene editing or other biotechnologies. Some variants discovered are of the Cas9 and Cas12a proteins. Bioinformatics led to the discovery of Cas12b, Cas13a and Cas13b effector proteins in type III and type VI systems, which have been shown to specifically target RNA. The Cas13a and Cas13b effector proteins contain a pair of higher eukaryote and prokaryote nucleotide-binding (HEPN) domains. These domains exist in lieu of RuvC-like nuclease domains and mediate RNA interference.

While much more is yet to be discovered about the CRISPR systems, the complexity is still being used to understand its role in bacteriophage resistance. In addition to R-M systems, these systems also serve as another line of defense for bacterial cells when faced with adversity. Now, what happens when these systems both fail to inactivate phages? What does the bacterial cell confer next? Ultimately, it has one final option that is not one of survival, but to protect its larger bacterial population at the cost of its own life.

III. Abortive infection systems (Abi)

Amongst their myriad defense mechanisms, bacteria also have abortive infection (Abi) systems. Abi systems consist of heterologous proteins through which phage resistance is attained by detecting phage infection and promoting growth arrest, and cell suicide in the most extreme cases. These abortive infection systems are different from R-M or CRISPR-Cas systems, because cell death is not considered an optimal defense strategy but rather a necessary strategy for survival of the broader bacterial community. Abi is the only system that favors cell death as a target for preventing phage multiplication in cells by committing suicide before an infecting phage completes its replication cycle. Due to such a benevolent act, Abi systems can be regarded as the last line of defense for bacterial cells if all other mechanisms fail and cell survival is highly unfavorable.

Abortive infection systems broadly mediate two activities, first to sense phage infection and second to promote cell death or metabolic arrest once phage is detected. Abi systems can detect phage infection in several ways, for example recognizing highly conserved phage structural proteins, phage genome replication intermediates, and/or high-level transcription from phage DNA. Once any of these components are detected, Abi defense mechanisms begin the

cell-death/metabolic arrest processes rapidly in order to suppress the production of progeny phage. Just as sensing mechanisms are diverse, so too are the modes of defense, with different systems alternatively mediating degradation of the cell's inner membrane, degrading transfer RNAs (tRNAs), or degrading phage and host DNA and messenger RNAs (mRNAs). In some cases it was reported that the cell death pathway may be reversible and used as a deterrent for the other systems, R-M and CRISPR-Cas, to aid in phage deactivation (Lopatina et al 2020).

One of the most important organisms in all areas of research is the amazing *Esherichia coli*. This organism has been used as a model for too many studies to name. For bacterial defense, it was essential for studying phage infections and specifically helped discover much of what we know about abortive infection systems. The first Abi systems discovered in *E. coli* were the Rex system, PifA, Lit and PrrC. The genes *rexA* and *rexB* are essential for defense in the Rex system, which is activated by a repressed lambda prophage. RexA protein senses DNA-protein complex, possibly from phage replication intermediate, and RexA activates the RexB protein responsible for the membrane degradation process. This system has shown to be efficient against plaque formation by particular strains of T4, T7, and T5 phages, with an exception to wildtype T4 phage which encodes proteins RIIA and RIIB that can counter the Rex defense system (Lopatina et al 2020**)** .

Other known Abi systems include Late Inhibition of T4 (Lit) and PrrC, which can impede host translation to cause cell death. Lit is a protease that is activated by phage T4 head protein *gp23* expression that binds to translation elongation factor, EF-Tu, forming the Lit-activating complex. Lit will cleave EF-Tu at its conserved nucleotide-binding site, and this causes inhibition of bacterial growth and impeding cellular protein translation. (Lopatina et al 2020). Similarly, *prrC* is a unique Abi gene that is activated by the binding-inhibition of type I

restriction endonuclease *Eco*prrI. The restriction enzyme *Eco*prrI is inhibited by the T4 phage peptide, Stp, and this activates the PrrC ribonuclease domain that cleaves $tRNA^{lys}$ and halts protein synthesis (Lopatina et al 2020).

PifA in *E. coli* also functions as an Abi system by impeding late gene transcription and destroying DNA replication. Once PifA is activated, it does not stop T7 infection initially, but rather at the midway of T7 infection cycle by releasing ATP from the cell after degradation of the cell membrane. This protein does not contain transmembrane helices in its sequence but is associated with the membrane and has an ATP/GTP binding domain essential for defense. Two T7 proteins responsible for activating PifA toxicity are T7 capsid protein gp10 and gp1.2, which inhibits deoxynucleotide hydrolysis in the cell. This system is described as one of the toxin-antitoxin systems in Abi, but it's ability to sense the T7 proteins has not been elucidated.

Lastly, a large class of Abi systems are the toxin-antitoxin (TA) systems. These are made up of genes usually transcribed from the same operon. The mechanisms of toxin-antitoxin systems are explained by the name itself, as the first gene transcribed is toxic, and the second gene provides immunity against that toxicity. There are six major TA system types, with the most common being type II. In type II systems there are toxin and antitoxin proteins, where the antitoxin protein binds to the toxin and reduces toxicity. When cells undergo stress due to phage infection, the antitoxin loses binding affinity and allows for toxicity to increase and cause either growth arrest or cell death.

Among these different abortive infection systems, one has been recently discovered which uses second messenger molecules to activate effector proteins that lead to cell death. These antiphage systems are of huge interest and have brought new insights into a form of Abi systems that can detect phage separately from the protein responsible for activating cell death

mechanism. They are known as cyclic oligonucleotide based antiphage signaling systems or CBASS. One family of CBASS systems share ancestry with the cyclic GMP-AMP synthase (cGAS)-STING innate immune pathway in animals. In animals, the cGAS-STING pathway uses a cGAS protein as a viral DNA sensor that when activated produces cyclic GMP-AMP (cGAMP). This second messenger in turn binds to the STING protein, activating an innate-immune response. A similar mechanism has been detected in *Vibrio cholerae* and other bacteria, and further identified as a new form of defense.

IV. Cyclic-oligonucleotide based antiphage signaling systems (CBASS)

Across all three superkingdoms of life, nucleotides function as extra- and intra-cellular signals with varying functions. The role of secondary messengers such as cyclic AMP (cAMP) and cyclic GMP (cGMP) are well understood for signaling protein binding, regulation of transcription or ion channel conductance, and multiple developmental processes in animals and bacteria (Burroughs et al. 2015). Additionally, cyclic di-nucleotides c-di-GMP and c-di-AMP have been identified as major signaling molecules in bacteria. The cyclic di-nucleotide, cyclic GMP-AMP (cGAMP), functions in producing type I interferons in eukaryotes for antiviral response. In bacteria, a variety of cGAS-like enzymes can generate diverse second messengers, including cGAMP and many other cyclic di- or trinucleotides (Whiteley et al. 2019). All of these second messengers are now understood to activate effector proteins within their cognate CBASS operons, defining a broad family of anti-phage defense systems (Cohen et al. 2019).

From the previous defense systems discussed, we know bacterial antiphage systems exist in defense islands where neighboring genes are expected to also play a role in phage protection. With this knowledge, researchers discovered a homolog of the cGAS encoding gene in *V.*

cholerae, dinucleotide cyclase in *Vibrio* (*dncV*) around known defense genes (Davies et. al 2012)Similarly, they discovered a bacterial effector phospholipase, or cGAMP-activated phospholipase in *Vibrio* (CapV), which is encoded in the *V. cholerae* genome adjacent to the *dncV* gene encoding bacterial cGAS (Severin et. al 2018). Two additional highly conserved genes were identified next to the *capV-dncV* gene pair which implicated this operon may require all four genes for defense. The two additional genes are of proteins with E1 and E2 domains, and a JAB domain all resembling the eukaryotic ubiquitin system. Initial discovery of bacterial cGAMP signaling also showed that this four-gene system can protect against phage (Cohen et al. 2019), and led to the discovery of multiple variants of CBASS systems in bacterial and archaeal genomes.

CBASS systems can be classified into four main configuration types (I, II, III, and IV) based on their core operon composition, effector activity, and signaling molecule product of oligonucleotide cyclase (Millman 2020). Type I CBASS systems are the most common and have a compact two-gene arrangement comprising an oligonucleotide cyclase and an effector gene. Protection against phage by type I CBASS is unique with its absence of ancillary genes as all other systems have them present and are necessary for defense. The effector proteins in these type I systems work primarily by detecting the production of cyclic oligonucleotides which begins cell membrane degradation. Major families of effector proteins in Type I CBASS include patatin-like phospholipases that degrade inner cell membrane phospholipids, and transmembrane proteins that are thought to oligomerize to form pores in the cell membrane.

B. Classification of CBASS Systems

While Type I CBASS systems encode only a cGAS-like system and an effector, the majority of CBASS systems additionally encode ancillary genes thought to regulate activation of

the system. Type II CBASS systems encode two ancillary genes that have domains related to eukaryotic ubiquitination machinery in addition to the core cyclase-effector pair. The first gene is *cap2* which encodes CBASS-associated protein 2 (Cap2) containing E1-like and E2-like domains which resemble ubiquitin activating and ubiquitin conjugating domains, respectively. The second gene is *cap3* which encodes CBASS-associated protein 3 (Cap3) an isopeptidase protein homologous to the eukaryotic JAB/JAMM-family deubiquitinases. It was previously unknown what protein the ubiquitin handling domains of Cap2 and Cap3 target, as ubiquitin and ubiquitin like-proteins are absent in *E. coli* and *V. cholerae* (Millman 2020). However, recent studies discovered that Cap2 forms a stable complex with cGAS protein *in vivo*, links this protein to an unknown target in a process termed cGASylation, and activates cGAMP synthesis by cGAS in an as-yet unknown manner (Ledvina et al. 2022). A 2.1Å x-ray crystal structure revealed the Cap2-cGAS complex is formed through the binding of cGAS C-terminus to the adenylation active site of Cap2 E1's domain with a conjugated AMP molecule (Ledvina et al. 2022). Similarly, it was found that the Cap3 protein can regulate cGAS activation by cleaving cGAS-target conjugates. (Ledvina et al. 2022). These findings indicate that Type II CBASS systems require Cap2 and Cap3 activities for phage defense, by activating cGAMP synthesis by cGAS. cGAMP in turn activates the effector phospholipase and promotes inner cell membrane degradation.

A minor class of Type II CBASS systems encodes only one ancillary gene,. which is distantly related to ubiquitin E2 proteins (Millman 2020). These "Type II-short" CBASS systems are mechanistically mysterious: most E2 proteins act on products of E1-protein adenylation and cysteine thioester formation. Without an E1 or an obvious ubiquitin-like protein, these systems' functions remain completely mysterious.

Type III CBASS systems are also composed of a core cyclase-effector pair, but similar to Type II systems, have ancillary genes that encode proteins related to eukaryotic signaling proteins. Type III CBASS systems encode one or two proteins homologous to HORMA-domain proteins in eukaryotes responsible for the formation of signaling complexes that control steps in meiosis, mitosis, and DNA repair (Millman 2020, Gu et al. 2022). In addition, the *cap6* gene encodes a AAA+ ATPase related to the TRIP13/Pch2 protein that is known for regulation of HORMA-domain activity in eukaryotes. A Type III CBASS system from *E. coli* MS115-1 allowed for structural identification of these proteins and the essential role of HORMA-domain protein Cap7 in phage-lambda protection (Ye et al. 2020). During phage infection, the oligonucleotide cyclase is physically bound and activated by the HORMA-domain protein and produces cyclic triadenylate molecules (cAAA) that activate an endonuclease effector protein responsible for cell death by degrading phage and host DNA. In absence of phage, the TRIP13-like protein Cap6 works as a negative regulator of cAAA production in these systems by unfolding the HORMA domain N-terminus and dissociating the HORMA:oligonucleotide cyclase complex. When Cap7 detects phage by binding an as-yet unidentified phage protein, a conformational change to the "closed" state of the HORMA domain enables it to bind and activate the cGAS-like oligonucleotide cyclase (Ye et al. 2020, and Millman 2020?). Type III CBASS systems have an additional configuration that consists of a Cap7 HORMA-domain protein (HORMA2) and a Cap8 HORMA-domain protein (HORMA3) that is considerably divergent and does not activate the cyclase in vitro. The function of Cap8/HORMA3 is mostly unknown, but it's hypothesized to function as a HORMA2 stabilizer or scaffold of signaling complexes (Ye et al. 2020).

Finally, Type IV CBASS systems are mainly found in archaea and Firmicutes and are the rarest amongst all of these systems, representing only 2.5% of CBASS systems identified (Millman 2020). This system is also composed of a core cyclase-effector pair, but contains ancillary genes *cap9* and *cap10* that encode proteins with nucleotide-modifying domains. The gene *cap9* encodes a predicted QueC enzymatic domain that converts modified base 7-carboxy-7-deazaguanine (CDG) to 7-cyano-7deazaguanine ($preQ_0$), while *cap10* encodes the predicted enzyme TGT, known for base exchange of a guanine residue in tRNA molecules using $preQ_0$ (Millman 2020). Some type IV CBASS systems have a *cap11* gene in their operon that encodes an N-glycosolase/DNA lyase (OGG) known for removing damaged guanine bases in DNA and to nick the DNA in apyrimidinic sites. However, no studies to date have demonstrated phage defense in type IV systems, and the function of their ancillary genes also remains a mystery.

The discovery of new and diverse anti-phage systems has exploded in recent years as microbial genome sequencing in non-model organisms has been expanded. (Millman 2020). Following the model of R-M systems and CRISPR-Cas systems, these new defense systems may reveal useful mechanisms for the development of new biotechnological tools in research. Similarly, understanding the dynamics of bacteriophages and these antiviral systems could help in phage therapy design to combat multidrug resistant bacterial pathogens.

CHAPTER 2: IDENTIFYING THE FUNCTION OF TYPE II (SHORT) CBASS SYSTEMS

A. Ubiquitination Pathways

It has been over 40 years since the discovery of ubiquitin (Ub), a protein that conjugates to other proteins via activity of E1, E2, and E3 enzymes that mediate conjugation. Since then, several protein families with evolutionary relation to ubiquitin were discovered in eukaryotes as well as prokaryotes, and were termed ubiquitin-like proteins (Ubls). Ubiquitin and Ubls modify eukaryotic proteins in a three-step cascade via E1, E2, and E3 enzyme activity. The E1 activating enzyme catalyzes the adenylation of the C-terminal glycine of the Ub/Ubl, forming an intermediate that is targeted for thioester linkage by the E1's conserved cysteine residue. (Burroughs et. al 2009, Streich et al. 2014). The thioester bond between E1 and Ub/Ubl (E1~Ubl) is transferred by the E1 enzyme to a conserved cysteine residue on E2 which forms the E2~Ubl product. E2 enzymes act as ubiquitin carrying proteins (Ubcs) and are the target of E3 Ubl ligases that transfer the Ubl from the E2 to a lysine residue on target proteins (Iyer et al. 2006, Streich et al. 2014, Cappadocia et al. 2018).

In eukaryotes, ubiquitination pathways are grouped into two categories, canonical and non-canonical, with the distinctions largely determined by the domain architecture of the major players. In autophagy, non-canonical E1 and E2 enzymes function similarly to those in canonical ubiquitination machinery. The eukaryotic non-canonical E1 enzyme in the autophagy pathway, termed ATG7, activates the Ubls ATG8 (termed LC3 in humans) and ATG12 for their transfer to non-canonical E2 enzymes ATG3 and ATG10, respectively. ATG7 activates the Ubl ATG12 for transfer to the ATG10 E2-like enzyme, allowing for the eventual conjugation of ATG12 to ATG5. Similarly, ATG7 can activate the Ubl ATG8 enzyme for transfer to the ATG3 E2-like

enzyme, which eventually mediates ATG8 conjugation to a phospholipid (Noda et al. 2009). The E1 enzyme ATG7 and the E2s ATG10 and ATG3 have several key differences in structure and function to the canonical E1 and E2 proteins. Unlike the canonical E1 protein adenylation domains, ATG7 lacks a C-terminal ubiquitin fold domain (UFD), which in canonical E1 proteins holds the catalytic cysteine residue. Instead, ATG7's catalytic cysteine is located on a "crossover loop" which reaches over the adenylation active site of E1. ATG3 and ATG10 also differ in structural features to canonical E2's, lacking an intact UBC fold and showing a distinctive location of the conserved catalytic residues typical of canonical E2 enzymes.

All bacteria encode sulfer-transfer pathways with evolutionary predecessors of ubiquitin E1 enzymes, termed ThiF and MoeB. These proteins adenylate the C-terminus of their respective substrates, the ubiquitin-like proteins ThiS and MoaD (Xi 2001, Lake 2001). ThiS and MoaD are not, however conjugated to target proteins in a manner similar to eukaryotic ubiquitination pathways. In 2006, researchers used bioinformatics analysis of bacterial genomes to identify a set of novel bacterial operons that encode Ub-related proteins, E1, E2, and predicted ubiquitin protease-like proteins (Iyer 2006). To date, however, these proteins have not been demonstrated to perform ubiquitination-like protein conjugation. Additionally, while these pathways have been postulated to function in defense pathways to protect against bacteriophage, no biological functions have been experimentally identified. Thus, the biochemical activities and biological functions of these novel pathways have remained mysterious.

B. Type II CBASS Systems and their Ubiquitin-like Proteins

From the defense systems discussed in the previous chapter, we know researchers discovered a homolog of the cGAS encoding gene in *V. cholerae*, dinucleotide cyclase in *Vibrio* (*dncV*) around known defense genes, and a bacterial effector phospholipase, or cGAMP-

activated phospholipase in *Vibrio* (CapV) (Cohen et al. 2019). Two additional highly conserved genes were identified next to the *capV-dncV* gene pair, which implied this operon may require all four genes for defense. The two additional genes are Cap2 (CBASS-associated protein 2) which contains predicted E1 and E2-like domains, and Cap3, a predicted JAB family (iso)peptidase. Initial discovery of bacterial cGAMP signaling also showed that this four-gene system can protect against phage (Cohen et al. 2019), and led to the discovery of multiple variants of CBASS systems in bacterial and archaeal genomes. Cap2/Cap3-encoding CBASS systems are termed Type II CBASS systems, and will be the focus of this chapter. Specifically, our focus is on a subset of Type II CBASS systems, termed Type II (short), whose Cap2 protein contains only a predicted E2 domain (and it therefore termed Cap2b), and which do not encode a Cap3 homolog. Our goals are to determine the molecular mechanisms of Type II (short) CBASS systems and understand their potential roles in bacteriophage protection.

Recently, our lab established the molecular mechanisms of Type II CBASS systems' Cap2 protein, which shares similarities with the non-canonical E1 and E2 signaling machinery described in the noncanonical autophagic ubiquitination pathway (Ledvina et al. 2022). Type II CBASS systems fall into two subclasses, "Type II (long)" and "Type II (short)". Type II (long) CBASS systems have an operon composed of a core cyclase-effector pair, and ancillary genes that encode for proteins Cap2 and Cap3 which resemble the eukaryotic ubiquitin machinery. Cap2 encodes a distinctive fusion of ubiquitin E1-like and E2-like domains, and is capable of mediating ubiquitin-like linkage of its cognate cGAS enzyme's C-terminus to an unknown target. This process, termed "cGASylation," primes cGAS for activation of second messenger synthesis by an unknown additional signal. In these systems, Cap3 resembles a JAB/JAMM

family ubiquitin peptidase, and antagonizes Cap2 activity by cleaving cGAS-target conjugates (Ledvina et al. 2022).

While the role of Type II (long) CBASS systems and its proteins Cap2 and Cap3 are now defined, whether and how this functional paradigm applies to Type II (short) CBASS systems is unknown. Type II (short) CBASS systems encode only one ancillary protein, Cap2b, that is related to ubiquitin E2 proteins. The typical function of an E2 protein is to mediate ubiquitin transfer from an E1 protein to a target molecule through thioester transfer via its catalytic cysteine residue. While Cap2b has a highly conserved catalytic cysteine, Type II (short) CBASS systems do not encode an E1-like protein. Therefore, the role of Cap2b, specifically whether it catalyzes a chemical reaction via its catalytic cysteine and whether that reaction is required for cGAS activation, is unknown.

C. Identifying Type II (short) CBASS systems

To address the role of Cap2b in Type II (short) CBASS systems, we first performed structural modeling using the AlphaFold protein structure prediction server (Jumper et al. 2021). We chose three Type II (short) CBASS systems to study, from *Vibrio cholerae* strain EDC_800 ("Gamma"), *Bradyrhizobium japonicum* strain S 370 ("Brady"), and *Vibrio vulnificus* strain WAPHLVBRA00020 ("Vibrio") (Figure 2.1A). AlphaFold predictions of all three operons' Cap2b proteins revealed strong similarity to canonical E2 proteins (not shown). Since CBASS-associated ancillary genes often bind to or otherwise act on their cognate cGAS enzymes, we performed AlphaFold predictions for a cGAS:Cap2b complex for all three systems. In all three cases, AlphaFold confidently predicted that the disordered C-terminal region of cGAS docks into the catalytic cleft of Cap2. The AlphaFold predictions for the "Brady" and "Vibrio" cGAS:Cap2 complexes were strongly similar, and showed Cap2b docking against cGAS via the same surface

that interacts with ubiquitin in canonical E2 proteins (Figure 2.1B-D). Curiously, both predictions showed the cGAS C-terminus positioned 3-4 residues away from the catalytic cysteine, unlike canonical E2 proteins which form a thioester linkage to the C-terminal carbonyl group of ubiquitin (Figure 2.1D). The catalytic cysteine is positioned immediately beside a highly conserved glycine residue close to the Cap2b C-terminus.

Our AlphaFold predictions strongly suggested that in Type II (short) CBASS systems, Cap2b binds directly to cGAS. The predictions do not address the question of whether the highly-conserved Cap2b catalytic cysteine catalyzes a chemical reaction, and also do not address whether or how Cap2b controls second messenger synthesis by cGAS. We therefore cloned complexes of Cap2b with cGAS from all three Type II (short) CBASS systems listed above and attempted to purify these complexes to better understand how these systems function. Some of the key questions we aimed to address were whether the cGAS and Cap2b proteins form a complex, whether cGAS is more or less active when bound to Cap2b, and whether Cap2b's E2 domain mediates higher-order assembly of cGAS. We also sought to determine the role of Cap2b's catalytic cysteine and whether it can mediate a chemical reaction related to the canonical biochemistry of other E2 proteins.

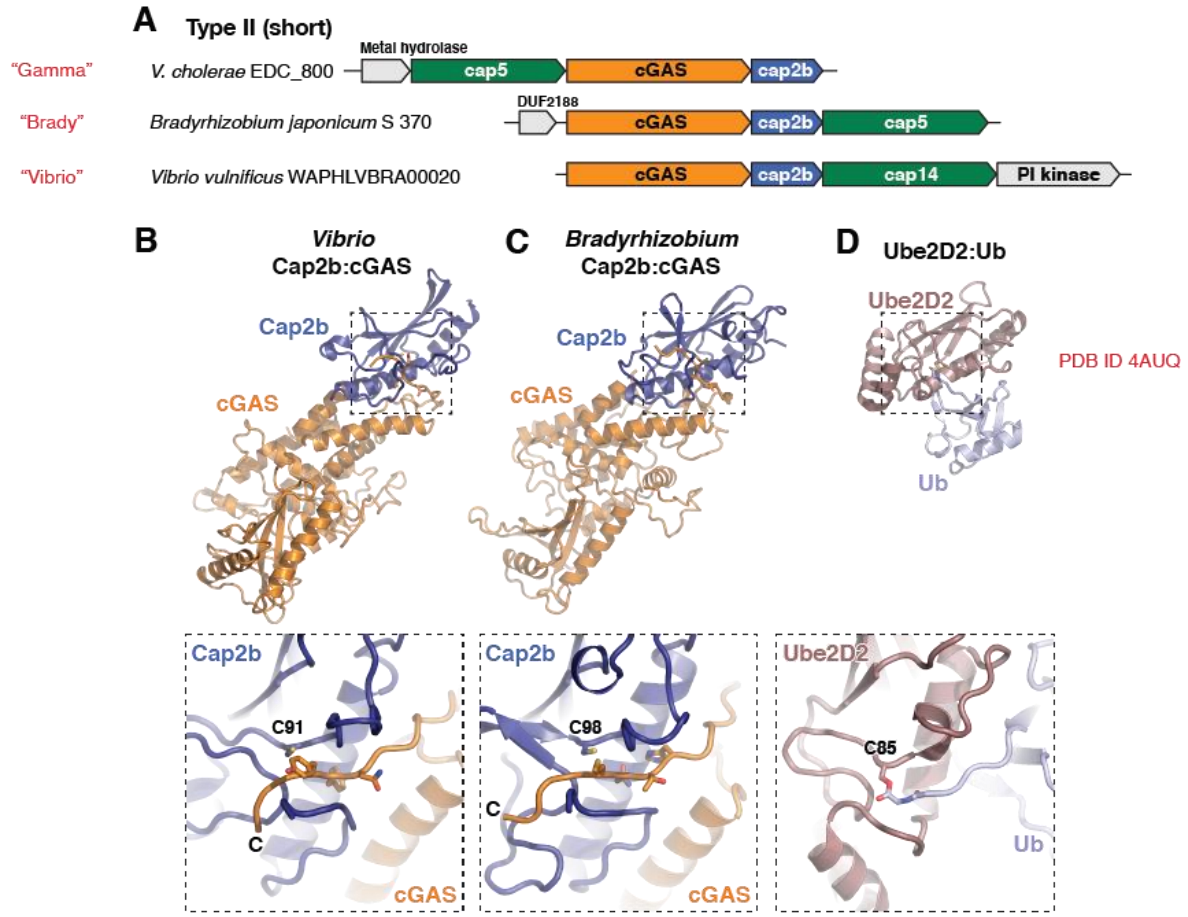

Figure 2.1: Alpha Fold Predicted Structures (a) Operon structure of three Type II (short) CBASS operons. cGAS is shown in orange, Cap2b in blue, predicted effector proteins in green, and proteins of unknown function in gray. (b) AlphaFold-predicted structure of the *Vibrio vulnificus* cGAS:Cap2b structure, with cGAS orange and Cap2b in blue. Inset (bottom) shows a closeup view of the cGAS C-terminus bound to the Cap2b catalytic cleft, with the putative catalytic cysteine (C91) shown as sticks. (c) AlphaFold-predicted structure of the *Bradyrhizobium japonicum* cGAS:Cap2b structure, with cGAS orange and Cap2b in blue. Inset (bottom) shows a closeup view of the cGAS C-terminus bound to the Cap2b catalytic cleft, with the putative catalytic cysteine (C98) shown as sticks. (d) Structure of the canonical E2 protein Ube2D2 (brown) in the same orientation as Cap2b in panels (b) and (c), bound to ubiquitin (light blue). Inset (bottom) shows a closeup view of the ubiquitin C-terminus covalently linked to the Cap2b catalytic cysteine (C85), mutated to serine in this structure to capture the transient covalent intermediate state (PDB ID 4AUQ).

I. Expression of cGAS and Cap2b Proteins

To identify the function of Type II (short) CBASS systems, we first cloned cGAS and Cap2b from three different strains for coexpression in *E. coli*: *Vibrio cholerae* strain EDC_800 ("Gamma"), *Bradyrhizobium japonicum* strain S 370 ("Brady"), and *Vibrio vulnificus* strain WAPHLVBRA00020 ("Vibrio"). The cGAS and Cap2b proteins for all three strains were cloned into *E. coli* expression vectors, with cGAS proteins fused to an N-terminal TEV protease-

cleavable His_6-tag, and Cap2b proteins untagged (see **Methods**). As a first step, the Brady-, Gamma-, and Vibrio-cGAS proteins were expressed on their own in *E. coli* for purification. Initial results from expression tests of cGAS proteins from Brady and Gamma showed good protein presence in pilot-scale purifications using Nickel-NTA affinity resin (Figure 2.2B and 2.2C), while the Vibrio-cGAS showed poor expression (Figure 2.2A).

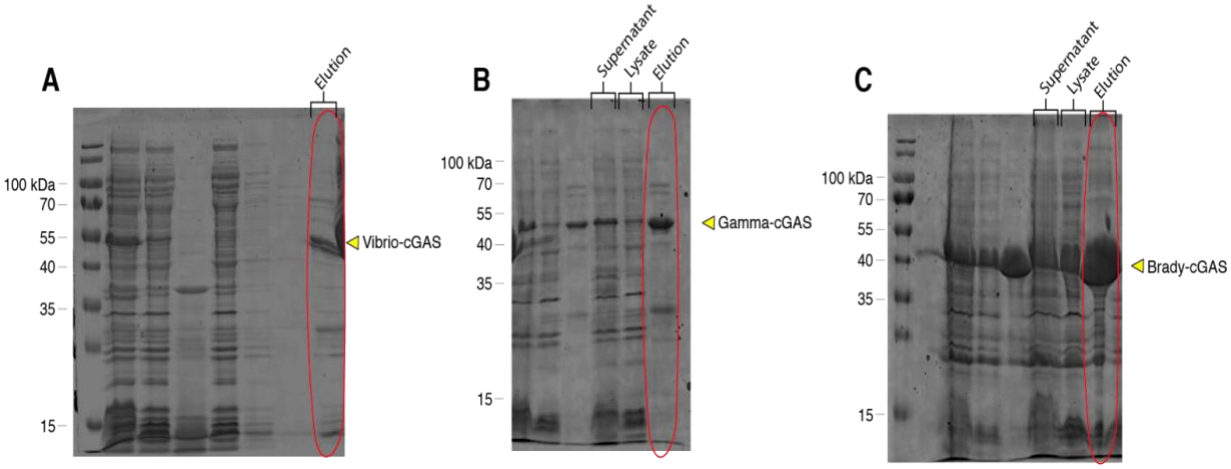

Figure 2.2: Initial SDS-PAGE analysis of cGAS protein from Vibrio, Gamma, and Brady. (A) Results from Ni-NTA column on SDS-PAGE for Vibrio-cGAS protein with gel band observed at 47kDa in the well containing the final elution (circled). (B) SDS-PAGE results from pilot pulldown of Gamma-cGAS showing bands of low protein concentration at 45kDa in the well containing the final elution from the nickel column (circled). (C) Initial SDS-PAGE results showing presence of bands with high protein concentration for Brady-cGAS protein at 46kDa in the well containing the final elution from the Ni-NTA column (circled).

Next, we co-expressed cGAS and Cap2b in *E. coli* for all three strains to determine if they form complexes *in vitro*. However, the cGAS:Cap2b complex for Gamma and Vibrio showed very little protein expression and were not pursued for further purification due to better expression from the Brady-cGAS:Cap2b complex (Figure 2.3A-C). Hence, we focused on the Brady-cGAS/Cap2b co-expression complex due to gel electrophoresis results that indicated purification had low contaminants (Figure 2.3C). After several rounds of purification of the Brady-cGAS:Cap2b complex, we were successful in isolating the non-covalent complex, but with an excess of the tagged cGAS protein (Figure 2.3C). The non-covalent cGAS:Cap2b

complex could not be concentrated sufficiently for crystallization trials due to precipitation of the proteins in low-salt buffers. As a next step, we decided to make new constructs for cGAS and Cap2b to determine whether we could improve cGAS-Cap2b complex stability for structural determination.

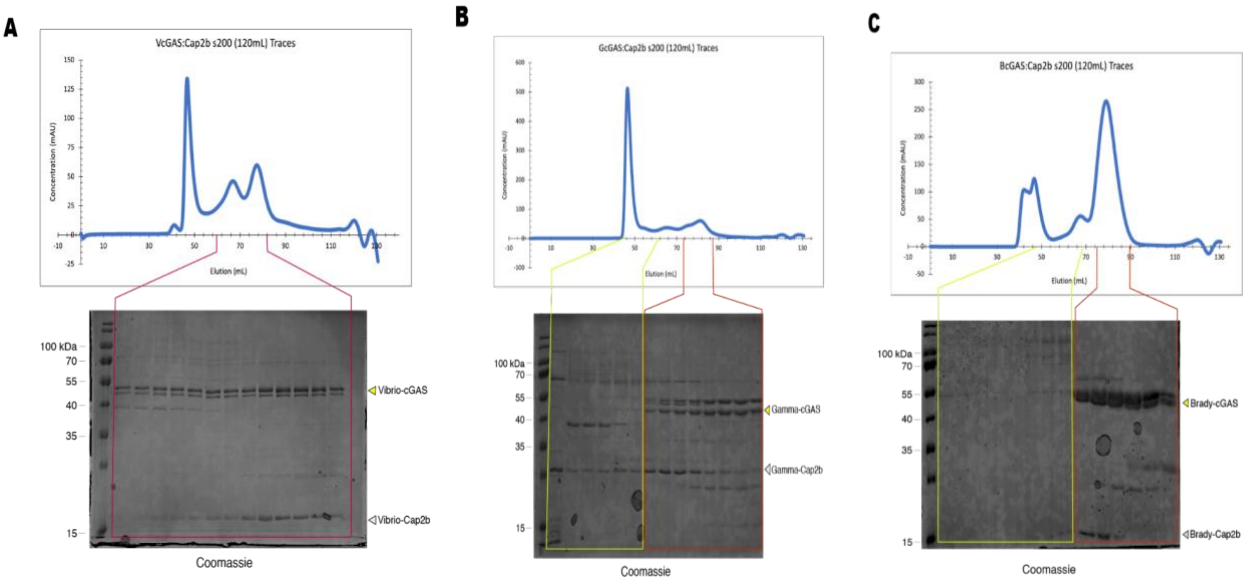

Figure 2.3: SDS-PAGE and Size Exclusion Chromatography (SEC) results for Gamma, Vibrio, and Brady-cGAS in N-terminal His6-tagged vector and -Cap2b in untagged vector. (A) SDS-PAGE analysis (bottom) showed the presence of bands for Vibrio-cGAS and -Cap2b at 47kDa and 18 kDa, indicated by the yellow triangle and gray triangle on the right, respectively. SEC results for the Vibrio-cGAS:Cap2b non-covalent complex had elution concentrations indicating protein size of each column elution (top) where the SEC peak fractions seen around 60-70mL (red box) correspond to the Vibrio-cGAS and Vibrio-Cap2b bands seen in SDS-PAGE (bottom). (B) SDS-PAGE analysis of the Gamma-cGAS and-Cap2b (bottom) showed the presence of Gamma-cGAS and -Cap2b at 45kDa and 26 kDa, indicated by the yellow triangle and gray triangle on the right, respectively. SEC results for the Gamma cGAS:Cap2b non-covalent complex had elution concentrations indicating protein size of each column elution (top) where SEC peak fractions seen around 70-85mL (red box) correspond to the Gamma-cGAS and Brady-Cap2b bands seen in SDS-PAGE (bottom). Small peaks seen around 45-60mL were also ran on SDS-PAGE and are shown in the yellow box. (C) SDS-PAGE analysis (bottom) of SEC results showed the presence of Brady-cGAS and -Cap2b at 46kDa and 16kDa, indicated by the yellow triangle and gray triangle on the right, respectively. SEC traces for the Brady cGAS:Cap2b non-covalent complex had elution concentrations indicating protein size of each column elution (top) where the SEC peak fractions seen around 75-85mL (red box) correspond to the Brady-cGAS and Brady-Cap2b bands seen in SDS-PAGE (bottom). Small peaks seen around 45-60mL were also ran on SDS-PAGE and are shown in the yellow box.

The first clones of cGAS in N-terminal His_6-tag vectors and Cap2b in untagged vectors showed instability at high concentrations. In an attempt to increase stability, we cloned cGAS and Cap2b proteins from the same three strains of Brady, Gamma, and Vibrio into different vectors. The new constructs of cGAS were cloned into untagged vectors, while Cap2b proteins were cloned with N-terminal His_6-tags. Due to previous experiments showing the Brady-

cGAS:Cap2b complexes to be the most stable, we focused on the new constructs from Brady for further tests. The Brady-cGAS in the untagged vector, and Brady-Cap2b in the N-terminal His_6-tag vector, were co-expressed in *E. coli*. Initial results showed a similar non-covalent complex of the cGAS and Cap2b, with an additional band in SDS-PAGE observed at ~65 kDa (Figure 2.4). Because 65 kDa is roughly the sum of the Brady-cGAS (46 kDa) and Brady-Cap2b (16 kDa), we theorized that this band could represent a covalently linked complex between the two proteins. Thus, we performed multiple purification methods to try and isolate this complex for crystallography and biochemical assays.

II. Purification of the cGAS-Cap2b complexes from *Brady*

From previous cloning and co-expression of Brady-cGAS in untagged vector and Brady-Cap2b in N-terminal His_6-tag vector, we observed gel electrophoresis bands of ~65 kDa, 46 kDa, and ~16 kDa that we hypothesized to be the cGAS-Cap2b complexes predicted from AlphaFold structures. In order to try and isolate these complexes we noted the theoretical pI from ExPASY ProtParam of the Brady-cGAS:Cap2b complex to be 6.72, and made buffers with a corresponding pH of 7.5 for optimal purification. Expression tests and Ni-NTA purification verified the presence of the 65 kDa, 46 kDa, and 16 kDa electrophoresis bands hypothesized to be the cGAS-Cap2b covalent and non-covalent complexes. The next step in purification was to run the Ni-NTA elution through an ion-exchange (IEX) column to remove contaminants and attempt to isolate the covalent cGAS-Cap2b complex. First we performed purification of the proteins using an Ni-NTA column, then we diluted the eluates for ion-exchange (IEX) purification. During IEX the protein complexes eluted in the initial flowthrough when loaded onto the IEX column. The new elution from IEX flowthrough contained both the cGAS-Cap2b

covalent and non-covalent complexes with fewer contaminants and was confirmed again by presence of the 65 kDa, 46 kDa, and 16 kDa gel electrophoresis bands. The flowthrough elution with cGAS-Cap2b complexes was concentrated to ~2-mL for gel filtration in order to separate the complexes and isolate the Brady-cGAS:Cap2b covalent complex for crystallization trials. After gel filtration of the cGAS-Cap2b complexes, gel electrophoresis was performed to determine the presence of the covalent and non-covalent complexes. Results showed the 65kDa gel band to be present, with all elutions also containing the non-covalent cGAS-Cap2b complex seen in the previous clones with cGAS in N-terminal His_6-tagged vector and Cap2b in untagged vector (Figure 2.4). The covalently linked Brady-cGAS:Cap2b was present in the gel filtration elutions as was hoped, however gel filtration was not able to separate the covalent and non-complexes. An additional unknown protein gel band at ~25 kDa believed to be a contaminant was also in the final gel filtration elutions. These results indicated that our process was not sufficient in isolating the covalent Brady-cGAS:Cap2b complex from the non-covalent form and due to the low covalent complex concentration, further isolation was not performed.

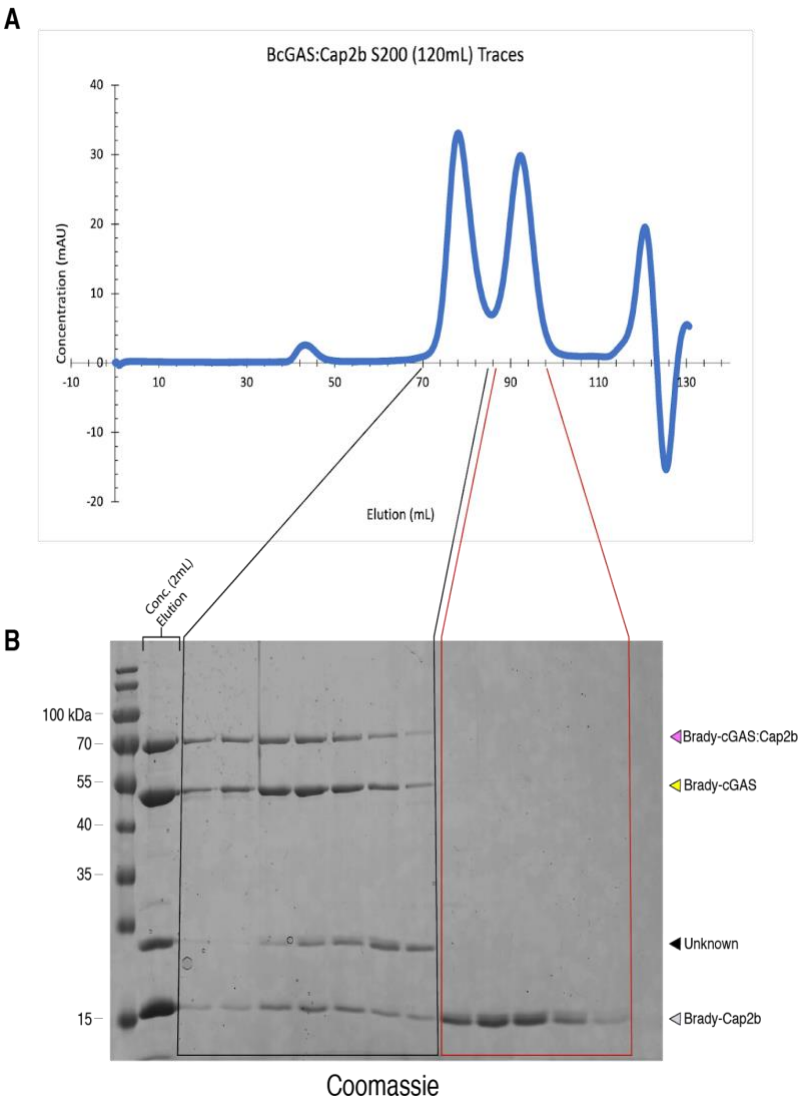

Figure 2.4: Size Exclusion Chromatography (SEC) and SDS-PAGE results for co-expression of Brady-cGAS in untagged vector and Brady-Cap2b in His6-tagged and vector. (A) SEC results for the concentrated 2-mL Ni-NTA elution of Brady cGAS:Cap2b co-expression with elution concentrations per elution volume indicating protein size of each elution. Brady-cGAS:Cap2b covalent complex, Brady-cGAS:Cap2b non-covalent complex elutions correspond to the peak seen around 70-80mL (black box), while the Brady-Cap2b elutions correspond to the peak observed at 90-95mL (red box). (B) SDS-PAGE analysis of the 2-mL Ni-NTA elution after SEC, using the peak fractions for analyzing presence of covalent and non-covalent Brady cGAS-Cap2b complex. SEC results from the coexpression of Brady-cGAS:Cap2b confirmed the presence of the 65kDa gel band hypothesized to be the covalent cGAS:Cap2b complex, indicated by the pink triangle on the right. The 46kDa band for Brady-cGAS and 16kDa band for Brady-Cap2b also confirmed the presence of the non-covalent complex indicated by the yellow triangle and gray triangle, respectively.

Another round of cloning was performed specifically for the Brady-cGAS/Cap2b proteins in an attempt to isolate the complex from all other contaminants observed in elutions. For these constructs, Brady-cGAS was cloned in untagged vector once more, but we added the sequence for eight-residue Strep-tag II (WSHPQFEK) at its C-terminus to perform purification using

Strep-Tactin affinity resin (Iba Life Sciences). The same Brady-Cap2b in N-terminal His_6-tag vector from previous constructs was used for this round of co-expression and purification. The objective with using the Brady-cGAS with Strep-tag II was to isolate the non-covalent cGAS-Cap2b complex from the covalent cGAS-Cap2b complex. Based on a recently-published study describing a similar system (see next section), we theorized that in the covalent complex, the Cap2b cysteine cleaves cGAS near its C-terminus when it covalently attaches, resulting in the observed 65 kDa band in previous purifications. With Strep-tag II tagged cGAS, the tag would be cleaved by this reaction, enabling us to separate covalent and non-covalent complexes based on the presence or absence of the tag. That is, the covalent complex would be eluted during the wash steps of the Strep-tactin affinity column while the non-covalent cGAS-Cap2b complex would bind to the Strep-tactin column (Figure 5). This method would allow for a fast isolation method of the covalent Brady-cGAS:Cap2b complex (present in the wash steps) and the non-covalent Brady-cGAS:Cap2b complex (present in the Strep-Tactin column elution). The purification was performed on the Strep-Tactin column, and the presence of the 65 kDa electrophoresis band hypothesized to be the cGAS-Cap2b covalent complex was observed. As expected, the Brady-cGAS:Cap2b non-covalent complex was present in the final elutions indicating that it was successful in binding the strep-tactin resin, but the elution concentrations were not optimal for further isolation. (Figure 2.5). Other contaminants were also noted, but due to time constraints, this purification method was not pursued further. Instead we aimed once more for optimizing the purification using Brady-cGAS in untagged vector and Brady-Cap2b in N-terminal His_6-tag vector.

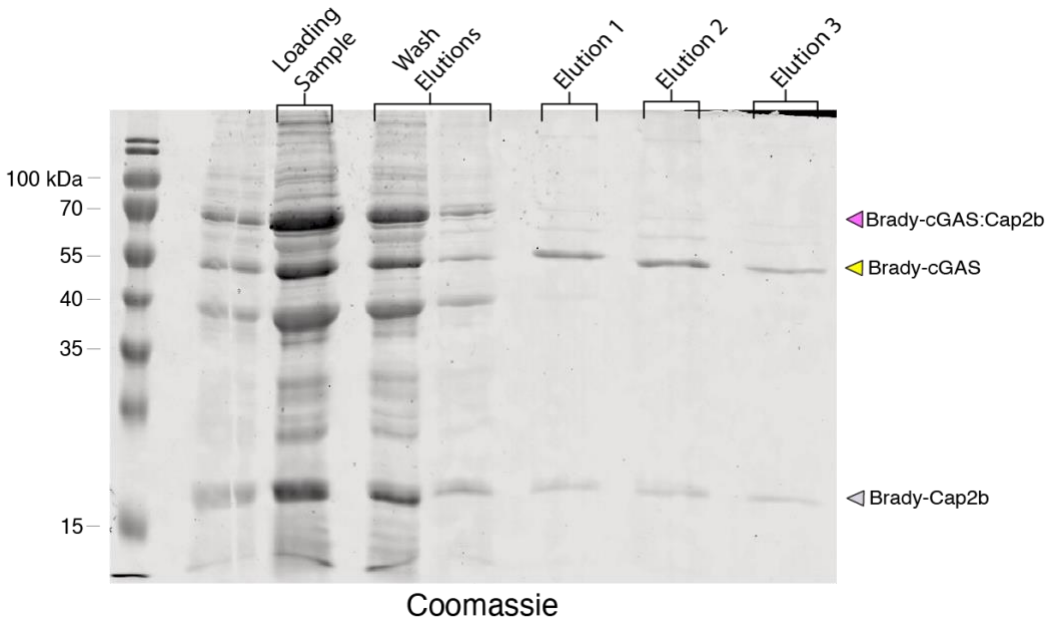

Figure 2.5: SDS-PAGE analysis of Strep-Tactin®XT 4Flow® results for Brady-cGAS with Strep-tag II and -Cap2b in untagged vector. SDS Page results of the Strep-tactin column for the Brady cGAS:Cap2b complex showing the presence of the Brady-cGAS:Cap2b covalent complex band at 65kDa and non-covalent complex at 47 kDa and 16 kDa in the loading sample, indicated by the pink, yellow, and gray triangles on the right, respectively. The covalent complex concentration decreases significantly as observed in the last three lanes corresponding to the three final elutions. SDS-PAGE analysis of the Strep-Tactin elutions also showed isolation of the non-covalent complex of Brady-cGAS and Brady-Cap2b in the final elutions which corresponds to the presence of Strep-II tag, indicated by the yellow triangle and gray triangle on the right, respectively.

A final round of purification was performed using Brady-cGAS in untagged vector and Brady-Cap2b in N-terminal His_6-tag vector to try and isolate the complexes. After several rounds, the presence of all three bands were still observed: the gel electrophoresis band at 65kDa for cGAS-Cap2b covalent complex, gel electrophoresis band at 46kDa and 16kDa for cGAS-Cap2b non-covalent complex (Figure 2.6). For this purification, we attempted to perform TEV cleavage of N-terminal His_6-tags on the Brady-Cap2b proteins after gel filtration. However, this reduced the presence of covalent cGAS-Cap2b complex, confirmed by lack of 65kDA band seen in gel electrophoresis. Interestingly, TEV cleavage results showed an increased presence of non-covalent cGAS-Cap2b complexes which may be indicative of the Brady-cGAS/Cap2b complex losing binding affinity due to cleavage of the His_6-tag. This was confirmed by SDS-PAGE of the final gel filtration elutions, which contained the 46kDa band of Brady-cGAS protein in high concentration (Figure 2.6).

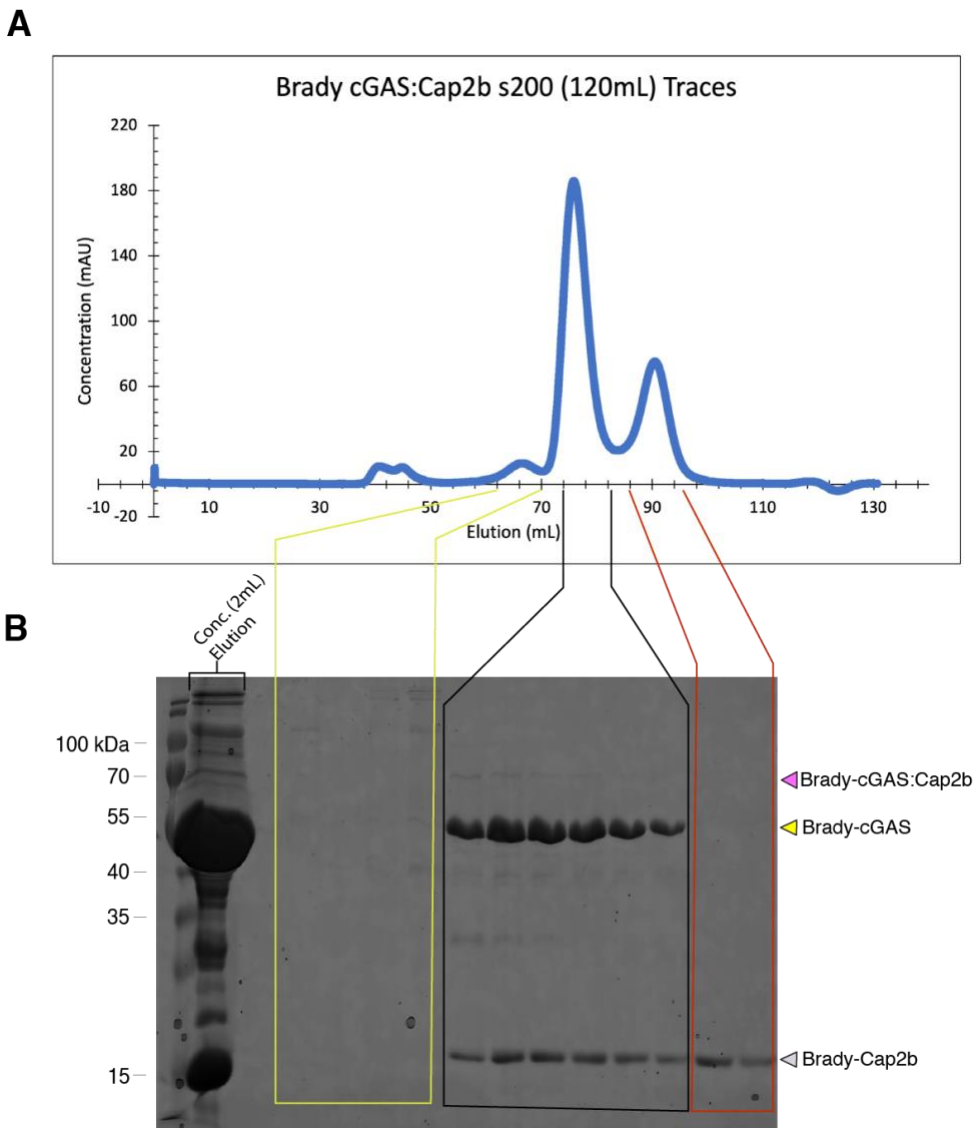

Figure 2.6: Size exclusion chromatography (SEC) results after TEV cleavage of the His6-tag on Brady-Cap2b and SDS-PAGE analysis of Brady-cGAS:Cap2b complex from SEC elutions. (A) SEC results for the concentrated 2-mL Ni-NTA elution from the Brady cGAS:Cap2b co-expression purification. Graph demonstrates the concentrations of each volume eluted, and the protein size of each elution is indicated by the peak at each volume . Brady-cGAS:Cap2b non-covalent complex elutions correspond to the peak seen around 75-80mL (black box), while the Brady-Cap2b elutions correspond to the peak observed at 90-95mL (red box). Small peak seen around 60-70mL was also ran on SDS-PAGE and is shown in the yellow box. (B) SDS-PAGE analysis of the 2-mL Ni-NTA elution after SEC, using the peak fractions for analyzing presence of covalent and non-covalent Brady cGAS-Cap2b complex. SEC results from the coexpression of Brady-cGAS:Cap2b confirmed the presence of the 46kDa and 16 kDa gel band hypothesized to be the non-covalent cGAS:Cap2b complex, indicated by the yellow triangle and gray triangle on the right, respectively. The covalent Brady-cGAS:Cap2b complex had very low concentrations confirmed by the presence of 65kDa gel band indicated by the pink triangle on the right. The 46kDa band for Brady-cGAS and 16kDa band for Brady-Cap2b also confirmed the presence of the non-covalent complex (black), with the addition of the Brady-Cap2b protein (red box) isolated in the last two SDS-PAGE lanes.

D. Summary of Findings & Future Directions

AlphaFold structure predictions showed us that Type II (short) CBASS systems likely form a complex consisting of the cGAS protein and Cap2b (ubiquitin E2-like) protein in various bacterial strains. The complex forms through the binding of the cGAS C-terminus to the catalytic cysteine (C98) on Cap2b via the same surface that interacts with ubiquitin in canonical E2 proteins (Figure 2.1B-D). Curiously, both predictions showed the cGAS C-terminus positioned 3-4 residues away from the catalytic cysteine, unlike canonical E2 proteins which form a thioester linkage to the C-terminal carbonyl group of ubiquitin (Figure 2.1D). From the three chosen strains to clone from AlphaFold structure predictions, the Brady constructs showed better co-expression of the cGAS and Cap2b than Vibrio and Gamma constructs. Initially, we aimed to determine whether the cGAS:Cap2b complex was more active in second messenger production than the cGAS alone in vitro, but due to low concentrations of the cGAS:Cap2b complex further activity assays were not performed. However, our results did indicate that in co-expression, the Brady-cGAS:Cap2b non-covalent complex was always higher in concentration than the Brady-cGAS:Cap2b covalent complex (Figure 2.2-2.6). It is possible that there may be a missing component needed for stabilizing the covalent complex formation and thus we see the covalently linked cGAS:Cap2b complex decrease in concentration. Similarly, the cGAS:Cap2b covalent complex forms best when the cGAS is cloned in untagged vector and the Cap2b is cloned in N-terminal His_6-tag vector (Figure 2.2A, Figure 2.4B). Ultimately, we found that the non-covalently linked cGAS-Cap2b complex is present in higher concentrations in vitro. Even when covalent cGAS:Cap2b is observed, the covalent complex either is lost or the complex prefers to be in the non-covalent form. In the final steps of our protein purification even when covalently linked

cGAS-Cap2b complex was previously present in solution, the non-covalent cGAS-Cap2b is the only complex remaining in solution. (Figure 2.4).

The next steps in these experiments would be to try and isolate the covalently linked cGAS:Cap2b complex separately from the non-covalently linked complex. An option for accomplishing this is to optimize the purification of cGAS-Cap2b complex where the cGAS is cloned with a C-terminal Strep-tag II for purification with Strep-tactin resin. Due to time constraints, this method of purification was not fully optimized and the lack of proper purification tools could have impacted the results we observed initially. Similarly, previous experiments only performed this purification using the Brady-cGAS and Brady-Cap2b, whereas the other strains may show more positive results. Ultimately, isolating the cGAS-Cap2b complex should be the aim for any future experiments and one could also aim to use different strains from those identified previously by Aaron Whiteley's lab (Whiteley et al. 2019) that may yield higher concentrations of the complex for structural determination.

Recent studies were published while our lab was working on these Type II (short) CBASS systems, where researchers successfully identified some functions of these CBASS systems. Similar to our studies, they used AlphaFold to predict the structure of the cGAS:Cap2b complex, and they determined the covalent link of cGAS and Cap2b, where they termed cGAS as *Sm*CdnG and Cap2b as *Sm*E2 in their paper (Yan et al. 2022). Furthermore, their studies found that in the covalent linkage between cGAS C-terminus and Cap2b cysteine, there is an energy-free thioester bond that links the two proteins via a “glycinated” cysteine at the Cap2b active site and cGAS C-terminal glycine residue. It was proposed that cGAS/*Sm*CdnG is a 3’,2’-cGAMP synthase and that it is structurally unstable unless attached to the Cap2b/*Sm*E2. They hypothesized that the Cap2b/*Sm*E2 protein is responsible for the stability and regulation of

cGAS/*Sm*CdnG, where self-cleavage occurs on the thioester-bound complex after increase in PPi and ADP from phage infection. Additionally, this self-cleavage promotes the cGAS/*Sm*CdnG protein to synthesize secondary messenger cyclic GMP-AMP(cGAMP), and this activates downstream effector proteins (Yan et al. 2022). However, they were also unable to determine the structure of the complex, leaving some uncertainty on the details of the self-cleaving mechanism proposed.

The previous study proposed similar conclusions to those our lab had hypothesized, and their work demonstrated the results we might have obtained had we isolated the complex. In all our purification studies we saw the same three bands present in SDS-PAGE: 1) the covalently linked cGAS-Cap2b complex, and separate 2) cGAS and 3)Cap2b proteins forming a non-covalent complex. If their hypothesis that stand-alone cGAS produces secondary messengers is true, then activating the effector proteins could be causing the covalent linkage of our complex to be disrupted, resulting in loss of the cGAS:Cap2b complex. The model they propose falls in line with what we thought about these Type II (short) CBASS systems, which is that they can confer defense even in absence of an ubiquitin transfer protein, E1. Usually the E2 mediates ubiquitin transfer from an E1 to a substrate, but in these Type II (short) systems it could be working to stabilize the cGAS to prevent secondary messenger production. However, it will be necessary for future studies to confirm the covalent linkage and mechanism of action between cGAS and Cap2b through structural determination, to further understand how these systems function in anti-phage defense. Furthermore, the structure can help elucidate the key components that allow for activation of secondary messengers and which protein/s is needed for their production. Isolating the complex should be followed by performance of activity assays and infection assays

in order to determine protection efficiency and to identify the span of phages that it confers protection against.

Acknowledgements

This work was possible thanks to the UCSD ACS Bridge Program funded by the National Science Foundation and Genentech Foundation. Additional funding was also provided by the UCSD Chemistry & Biochemistry Department. The author thanks the Corbett Lab for their input on this research topic, and the principal investigator, Kevin D. Corbett who was supported by the Ludwig Institute for Cancer Research and the University of California, San Diego.

MATERIALS AND METHODS

Materials

For studying Type-II (short) CBASS systems, accession numbers of different CD-NTase (cGAS) and Cap2b's (or E2's) provided by Aaron Whiteley. The chosen systems were of *Vibrio cholerae* strain EDC_800 ("Gamma", NCBI #WP_000072410.1) , *Bradyrhizobium japonicum* strain S 370 ("Brady", NCBI #WP_011082904.1), and *Vibrio vulnificus* strain WAPHLVBRA00020 ("Vibrio", NCBI #WP_017790128.1).

Structure Prediction

The predicted structures of cGAS and Cap2b proteins from the "Gamma", "Brady", and "Vibrio" strains were obtained from AlphaFold/ColabFold. For the cGAS-Cap2b complex structures, multiple sequence alignments (MSA) were performed using MMseqs2 (UniRef + Environmental) in AlphaFold.

Plasmid Construction for Cap2b proteins

The three constructs with gene encoding Cap2b were derived and codon optimized from *Vibrio cholerae* TM 11079-80 (NCBI #WP_000072410.1 (Whiteley et al., 2019)) for Gamma-Cap2b; *Bradyrhizobium diazoefficiens* USDA 110 (NCBI #WP_011082904.1 (Whiteley et al., 2019)) for Brady-Cap2b; and *Vibrio vulnificus* (NCBI #WP_017790128.1 (Whiteley et al., 2019)) for Vibrio-Cap2b. Primer DNA for each Cap2b construct was synthesized and ordered from Integrated DNA Technologies (IDT). PCR was performed on a set of primers from each construct, followed by T4 DNA ligation. Each construct was cloned into UC Berkeley MacroLab vectors 2AT(Addgene #29665) for protein expression with TEV-cleavable N-terminal His_6-tags, transformed into NovaBlue competent cells, and confirmed via Sanger sequencing. Geneblocks for each untagged Cap2b construct synthesized and ordered from Integrated DNA Technologies (IDT) using the same coding sequences from His-tagged Cap2b constructs. T4 DNA ligation was performed for cloning the genes from each strain into UC Berkeley MacroLab vector 2BT(Addgene #29665) for protein expression with no peptide fusions.

Plasmid Construction for cGAS proteins

The gene encoding cGAS was codon optimized for "Gamma" from *Vibrio cholerae* TM 11079-80 (NCBI #WP_000072410.1 (Whiteley et al., 2019)); "Brady" from *Bradyrhizobium diazoefficiens* USDA 110 (NCBI #WP_011082904.1 (Whiteley et al., 2019); and "Vibrio" from *Vibrio vulnificus* (NCBI #WP_017790128.1 (Whiteley et al., 2019)). Constructs with N-terminal His_6-tags were synthesized and cloned into UC Berkeley MacroLab vectors 2BT(Addgene #29666) for protein expression with TEV-cleavable N-terminal His_6-tags. Performed PCR on

His-tagged constructs, transformed into NovaBlue competent cells, and confirmed via Sanger sequencing (Azenta). Same procedure for cGAS in untagged vectors, but cloning done into UC Berkeley Macrolab vectors 2AT (Addgene #29665) for protein expression with no peptide fusions.

Protein Expression and Purification

The cGAS and Cap2b proteins were expressed independently in *E. coli* strain Rosetta 2 (DE3) pLysS (EMD Millipore). Cultures for expression and of proteins were grown at 37ºC until A_{600} = 0.8, and induced using 0.25 mM IPTG with temperature at 20ºC for 18h. Overnight incubation was followed by cell harvest by centrifugation then resuspension in buffer A (25mM Tris pH 7.5, 5mM imidazole, 300mM NaCl, 10% glycerol, 5mM $MgCl_2$, and 5mM β-mercaptoethanol). Purification of proteins was performed on Ni^{2+} -affinity (HisTrap HP, Cytiva) washed with buffer A plus, 15mM imidazole and eluted with buffer A plus 395mM Imidazole. Elutions were diluted with buffer B (25mM Tris pH 7.5, 10% glycerol, 5mM $MgCl_2$, and 5mM β-mercaptoethanol) then concentrated by ultrafiltration (Amicon Ultra, EMD Millipore) to 2-mL and passed over gel filtration column (Superdex 200, Cytiva) in buffer GF (25mM Tris pH 7.5, 150mM NaCl, 10% glycerol, 5mM $MgCl_2$, and 1mM DTT). Protein presence confirmed through SDS-PAGE analysis with Coomassie blue (Bio-Rad) staining.

Protein Co-Expression and Purification

The cGAS and Cap2b proteins were co-expressed in *E. coli* strain Rosetta 2 (DE3) pLysS (EMD Millipore). Cultures for co-expression of proteins were grown at 37ºC until A_{600} = 0.8, and induced using 0.25 mM IPTG with temperature at 20ºC for 18h. Overnight incubation was followed by cell harvest by centrifugation then resuspension in buffer A (25mM Tris pH 7.5,

5mM imidazole, 300mM NaCl, 10% glycerol, 5mM $MgCl_2$, and 5mM β-mercaptoethanol). Purification of proteins was performed on Ni^{2+} -affinity (HisTrap HP, Cytiva) washed with buffer A plus, 15mM imidazole and eluted with buffer A plus 395mM Imidazole. Elutions were diluted with buffer B (25mM Tris pH 7.5, 10% glycerol, 5mM $MgCl_2$, and 5mM β-mercaptoethanol) then passed through ion-exchange column (HiTrap Q HP, Cytiva), collecting only the loading flowthrough. Proteins were concentrated by ultrafiltration (Amicon Ultra, EMD Millipore) to 2-mL and passed over gel filtration column (Superdex 200, Cytiva) in buffer GF (25mM Tris pH 7.5, 150mM NaCl, 10% glycerol, 5mM $MgCl_2$, and 1mM DTT). Protein presence confirmed through SDS-PAGE analysis with Coomassie blue (Bio-Rad) staining.

Protein Co-Expression and Purification using Strep-tag® II

The cGAS in untagged vector with the Strep-tag® II sequence was constructed using the same protocol from the plasmid construction above. Cap2b with N-terminal His_6-tags used for co-expression with Strep-tag® II-cGAS in *E. coli* strain Rosetta 2 (DE3) pLysS (EMD Millipore). Cultures for co-expression of proteins were grown at 37ºC until A_{600} = 0.8 and induced using 0.25 mM IPTG with temperature at 20ºC for 18h. Overnight incubation was followed by cell harvest by centrifugation then resuspension in buffer A (25mM Tris pH 7.5, 5mM imidazole, 300mM NaCl, 10% glycerol, 5mM $MgCl_2$, and 5mM β-mercaptoethanol). Protein purification was performed using Strep-Tactin®XT 4Flow® resin in 5-mL gravity column. Washes and elutions were performed using the IBA Life Sciences protocol "Protein purification with Strep-Tactin®XT resins" included with Strep-Tactin®XT 4Flow® resin. Protein presence confirmed through SDS-PAGE analysis with Coomassie blue (Bio-Rad) staining.

REFERENCES

1. Brüssow, Harald, and Roger W. Hendrix. 2002. “Phage Genomics: Small Is Beautiful.” *Cell* 108 (1): 13–16.

2. Burroughs, A. Maxwell, Lakshminarayan M. Iyer, and L. Aravind. 2009. “Natural History of the E1-like Superfamily: Implication for Adenylation, Sulfur Transfer, and Ubiquitin Conjugation.” *Proteins* 75 (4): 895–910.

3. Burroughs, A. Maxwell, Dapeng Zhang, Daniel E. Schäffer, Lakshminarayan M. Iyer, and L. Aravind. 2015. “Comparative Genomic Analyses Reveal a Vast, Novel Network of Nucleotide-Centric Systems in Biological Conflicts, Immunity and Signaling.” *Nucleic Acids Research* 43 (22): 10633–54.

4. Cappadocia, Laurent, and Christopher D. Lima. 2018. “Ubiquitin-like Protein Conjugation: Structures, Chemistry, and Mechanism.” *Chemical Reviews* 118 (3): 889–918.

5. Cohen, Daniel, Sarah Melamed, Adi Millman, Gabriela Shulman, Yaara Oppenheimer-Shaanan, Assaf Kacen, Shany Doron, Gil Amitai, and Rotem Sorek. 2019. “Cyclic GMP-AMP Signalling Protects Bacteria against Viral Infection.” *Nature* 574 (7780): 691–95.

6. Davies, Bryan W., Ryan W. Bogard, Travis S. Young, and John J. Mekalanos. 2012. “Coordinated Regulation of Accessory Genetic Elements Produces Cyclic Di-Nucleotides for V. Cholerae Virulence.” *Cell* 149 (2): 358–70.

7. Doron, Shany, Sarah Melamed, Gal Ofir, Azita Leavitt, Anna Lopatina, Mai Keren, Gil Amitai, and Rotem Sorek. 2018. “Systematic Discovery of Antiphage Defense Systems in the Microbial Pangenome.” *Science* 359 (6379). https://doi.org/10.1126/science.aar4120.

8. Gu, Yajie, Arshad Desai, and Kevin D. Corbett. 2022. “Evolutionary Dynamics and Molecular Mechanisms of HORMA Domain Protein Signaling.” *Annual Review of Biochemistry* 91 (June): 541–69.

9. Ishino, Yoshizumi, Mart Krupovic, and Patrick Forterre. 2018. “History of CRISPR-Cas from Encounter with a Mysterious Repeated Sequence to Genome Editing Technology.” *Journal of Bacteriology* 200 (7). https://doi.org/10.1128/JB.00580-17.

10. Iyer, Lakshminarayan M., A. Maxwell Burroughs, and L. Aravind. 2006. “The Prokaryotic Antecedents of the Ubiquitin-Signaling System and the Early Evolution of Ubiquitin-like Beta-Grasp Domains.” *Genome Biology* 7 (7): R60.

11. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R.,

Adler, J., Back, T., Peterson, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P., Hassabis, D. 2021. “Highly Accurate Protein Structure Prediction with AlphaFold.” *Nature* 596 (7873): 583–89.

12. Lake, M. W., M. M. Wuebbens, K. V. Rajagopalan, and H. Schindelin. 2001. “Mechanism of Ubiquitin Activation Revealed by the Structure of a Bacterial MoeB-MoaD Complex.” *Nature* 414 (6861): 325–29.

13. Ledvina, Hannah E., Qiaozhen Ye, Yajie Gu, Yun Quan, Rebecca K. Lau, Huilin Zhou, Kevin D. Corbett, and Aaron T. Whiteley. 2022. “cGASylation by a Bacterial E1-E2 Fusion Protein Primes Antiviral Immune Signaling.” *bioRxiv*. https://doi.org/10.1101/2022.03.31.486616.

14. Lopatina, Anna, Nitzan Tal, and Rotem Sorek. 2020. “Abortive Infection: Bacterial Suicide as an Antiviral Immune Strategy.” *Annual Review of Virology* 7 (1): 371–84.

15. Makarova, Kira S., Nick V. Grishin, Svetlana A. Shabalina, Yuri I. Wolf, and Eugene V. Koonin. 2006. “A Putative RNA-Interference-Based Immune System in Prokaryotes: Computational Analysis of the Predicted Enzymatic Machinery, Functional Analogies with Eukaryotic RNAi, and Hypothetical Mechanisms of Action.” *Biology Direct* 1 (March): 7.

16. Meselson, M., R. Yuan, and J. Heywood. 1972. “Restriction and Modification of DNA.” *Annual Review of Biochemistry* 41: 447–66.

17. Millman, Adi, Sarah Melamed, Gil Amitai, and Rotem Sorek. 2020. “Diversity and Classification of Cyclic-Oligonucleotide-Based Anti-Phage Signalling Systems.” *Nature Microbiology* 5 (12): 1608–15.

18. Noda, Nobuo N., Yoshinori Ohsumi, and Fuyuhiko Inagaki. 2009. “ATG Systems from the Protein Structural Point of View.” *Chemical Reviews* 109 (4): 1587–98.

19. Severin, G. B., Ramliden, M. S., Hawver, L. A., Wang, K., Pell, M. E., Kieninger, A. K., Khataokar, A., O'Hara, B. J., Behrmann, L. V., Neiditch, M. B., Benning, C., Waters, C. M., & Ng, W. L. (2018). “Direct Activation of a Phospholipase by Cyclic GMP-AMP in El Tor *Vibrio Cholerae*.” *Proceedings of the National Academy of Sciences of the United States of America* 115 (26): E6048–55.

20. Streich, Frederick C., Jr, and Christopher D. Lima. 2014. “Structural and Functional Insights to Ubiquitin-like Protein Conjugation.” *Annual Review of Biophysics* 43: 357–79.

21. Whiteley, Aaron T., James B. Eaglesham, Carina C. de Oliveira Mann, Benjamin R. Morehouse, Brianna Lowey, Eric A. Nieminen, Olga Danilchanka, King, D. S., Lee, A. S. Y., Mekalanos, J. J., & Kranzusch, P. J. 2019. “Bacterial cGAS-like Enzymes Synthesize Diverse Nucleotide Signals.” *Nature* 567 (7747): 194–99.

22. Wilson, G. G., and N. E. Murray. 1991. “Restriction and Modification Systems.” *Annual Review of Genetics* 25: 585–627.

23. Xi, J., Y. Ge, C. Kinsland, F. W. McLafferty, and T. P. Begley. 2001. “Biosynthesis of the Thiazole Moiety of Thiamin in Escherichia Coli: Identification of an Acyldisulfide-Linked Protein--Protein Conjugate That Is Functionally Analogous to the ubiquitin/E1 Complex.” *Proceedings of the National Academy of Sciences of the United States of America* 98 (15): 8513–18.

24. Yan, Yan, Fengtao Huang, Bingbing Yu, Rui Cheng, Hui Wu, Xueling Lu, Xionglue Wang, Jun Xiao, Longfei Wang, and Bin Zhu. 2022. “A Bacterial E2 Enzyme Regulates an Antiviral cGAS through a Peptide Exchange–ligation–cleavage Mechanism.” *bioRxiv*. https://doi.org/10.1101/2022.05.25.493364.

25. Ye, Qiaozhen, Rebecca K. Lau, Ian T. Mathews, Erica A. Birkholz, Jeramie D. Watrous, Camillia S. Azimi, Joe Pogliano, Mohit Jain, and Kevin D. Corbett. 2020. “HORMA Domain Proteins and a Trip13-like ATPase Regulate Bacterial cGAS-like Enzymes to Mediate Bacteriophage Immunity.” *Molecular Cell* 77 (4): 709–22.e7.