UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Molecular Mechanisms and Regulation of Bacterial Cyclic Oligonucleotide-Based Anti-Phage Signaling Systems

Permalink

https://escholarship.org/uc/item/18f90023

Author

Lau, Rebecca Kavaler

Publication Date

2022

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA SAN DIEGO

Molecular Mechanisms and Regulation of Bacterial Cyclic Oligonucleotide-Based Anti-Phage Signaling Systems

A Dissertation submitted in partial satisfaction of the requirements
for the degree Doctor of Philosophy

in

Biomedical Sciences

by

Rebecca Kavaler Lau

Committee in charge:

Professor Kevin Corbett, Chair
Professor Partho Ghosh
Professor Mohit Jain
Professor Joe Pogliano
Professor Manuela Raffatellu

2022

Copyright

Rebecca Kavaler Lau, 2022

All rights reserved.

The Dissertation of Rebecca Kavaler Lau is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2022

DEDICATION

This thesis is dedicated to my family.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would first like to thank my thesis advisor, Kevin Corbett, for all his guidance through this project. Dr. Corbett has been an amazing mentor throughout my entire time in the Biomedical Sciences program. His enthusiasm for this project, and science in general, has been an inspiration and has encouraged me to persevere even when experiments were not working. Without his guidance and mentorship, I would not have applied for or received the many fellowships that funded my time in the lab such as the T32 Training Grant in Quantitative Biology and the F31 through NIGMS. It has been great to work in a lab with a leader who cares about the science, but also the well-being of the people in the lab. I could not have chosen a better lab to complete my thesis research work in.

I would also like to thank my thesis committee, Partho Ghosh, Mohit Jain, Joe Pogliano, and Manuela Raffatellu, for all their input on this project. In the end, many of my thesis committee members ended up as collaborators in the project, which helped improve the work immensely. Partho Ghosh was also my mentor through my first rotation when arriving at UC San Diego, and I appreciate his weekly meetings with me, which cultivated a great appreciation for the field of structural biology.

I am so thankful to everyone in the Corbett Lab for all their input on this work. I would especially like to thank Qiaozhen Ye, who was my mentor when first rotating in the lab and helped me crystallize my first protein, and who I worked very closely with on the first CBASS story. I would also like to thank Yajie Gu for his thoughts and

discussions on the various CBASS projects and Chelsea Blankenchip, who was enthusiastic to join the CBASS team. Mentoring Chelsea on her project allowed me to better think about mine.

I would like to thank my friends from the Biomedical Sciences program Gabriela Goldberg, Tiani Louis, Celeste Allaband, Margaret Burns, and Alanna Koehler. It has been amazing to be part of a supportive group of friends.

Finally, I would like to thank my family who have always supported me. My parents, Evy Kavaler and John Lau, have always valued education and inspired my love for learning. I also want to thank my husband, Marc Russell, who loved and supported me the whole way through.

Chapter 1, in full, is a reprint of the material as it appears in *Molecular Cell* 2020. Ye, Qiaozhen; Lau, Rebecca K.; Mathews, Ian T.; Birkholz, Erica A.; Watrous, Jeramie D.; Azimi, Camillia S.; Pogliano, J.; Jain, Mohit; Corbett, Kevin D. “HORMA Domain Proteins and a Trip13-like ATPase Regulate Bacterial CGAS-like Enzymes to Mediate Bacteriophage Immunity”. The dissertation/thesis author was a primary investigator and author of this paper.

Chapter 2, in full, is a reprint of the material as it appears in *Molecular Cell* 2020. Lau, Rebecca K.; Ye, Qiaozhen; Birkholz, Erica A.; Berg, Kyle R.; Patel, Lucas; Mathews, Ian T.; Watrous, Jeramie D.; Ego, Kaori; Whiteley, Aaron T.; Lowey, Brianna; Mekalanos, John J.; Kranzusch, Philip J.; Jain, Mohit; Pogliano, Joe; Corbett, Kevin D. “Structure and Mechanism of a Cyclic Trinucleotide-Activated Bacterial Endonuclease Mediating Bacteriophage Immunity”. The dissertation/thesis author was a primary

investigator and co-first author of this paper.

Chapter 3, in full, is a reprint of the material as it is published in *The EMBO Journal* 2022. Lau, Rebecca K.; Enustun, Eray; Gu, Yajie; Nguyen, Justin V.; Corbett, Kevin D. “A Conserved Signaling Pathway Activates Bacterial CBASS Immune Signaling in Response to DNA Damage”. The dissertation/thesis author was the primary investigator and author of this paper.

VITA

2013 Bachelor of Science in Biochemistry & Molecular Biology, University of California Davis

2022 Doctor of Philosophy in Biomedical Sciences, University of California San Diego

PUBLICATIONS

Lau, R.K., Ye, Q., Birkholz, E.A., Berg, K.R., Patel, L., Mathews, I.T., Watrous, J.D., Ego, K., Whiteley, A.T., Lowey, B., Mekalanos, J.J., Kranzusch, P.J., Jain, M., Pogliano, J., Corbett, K.D. (2020) Structure and Mechanism of a Cyclic Trinucleotide-Activated Bacterial Endonuclease Mediating Bacteriophage Immunity. *Mol Cell* 77, 723-733.e6.

Ye, Q., **Lau, R.K.**, Mathews, I.T., Birkholz, E.A., Watrous, J.D., Azimi, C.S., Pogliano, J., Jain, M., Corbett, K.D. (2020) HORMA Domain Proteins and a Trip13-like ATPase Regulate Bacterial cGAS-like Enzymes to Mediate Bacteriophage Immunity. *Mol Cell* 77, 709-722.e7.

Blankenchip C.L., Nguyen J.V., **Lau R.K.**, Ye Q., Gu Y. & Corbett K.D. (2022) Control of bacterial immune signaling by a WYL domain transcription factor. *Nucleic Acids Res* 50: 5239–5250

Liang, Q., Richey, S.T., Ur, S.N., Ye, Q., **Lau, R.K.**, Corbett, K.D., & Kevin Corbett, C. D. (2022) Structure and activity of a bacterial defense-associated 3′-5′ exonuclease. *Protein Science*, 31(7), e4374.

Lau R.K., Enustun E., Gu Y., Nguyen J.V. & Corbett K.D. (2022) A conserved signaling pathway activates bacterial CBASS immune signaling in response to DNA damage. *EMBO J*, e111540

Ledvina H.E., Ye Q., Gu Y., Quan Y., **Lau R.K.**, Zhou H., Corbett K.D. & Whiteley A.T. (2022) cGASylation by a bacterial E1-E2 fusion protein primes antiviral immune signaling. *bioRxiv*: 2022.03.31.486616

ABSTRACT OF THE DISSERTATION

Molecular Mechanisms and Regulation of Bacterial Cyclic Oligonucleotide-Based Anti-Phage Signaling Systems

by

Rebecca Kavaler Lau

Doctor of Philosophy in Biomedical Sciences

University of California San Diego, 2022

Professor Kevin Corbett, Chair

Bacteria sense and respond to their environment through many signaling pathways. Response to bacteriophage infection is vital to bacterial survival, therefore they have a variety of ways in which they can respond. Classical responses to

bacteriophage infection such as restriction modification systems have been well characterized. However, with the increase in sequencing capabilities, a vast number of new bacteriophage defense systems have been annotated. These systems often have only putative identification of genes responsible for their protection. In this dissertation, I characterize a novel class of bacteriophage defense systems that are now called Cyclic Oligonucleotide-Based Antiphage Signaling Systems (CBASS).

Each CBASS system encodes an enzyme called a cGAS/DncV-like nucelotidyltransferase (CD-NTase) that synthesizes a cyclic oligonucleotide second messenger molecule which allows it to amplify the original signal coming from the bacteriophage. Here, I describe the molecular mechanisms of a CBASS system found in the *E. coli* strain MS115-1. This CBASS system contains a CD-NTase in a family which has no members that have shown activity in vitro, CdnC. The system also contains a putative endonuclease, NucC, and proteins that have structural similarity to eukaryotic signaling proteins in the HORMA family and a AAA+ ATPase Trip13. I show that the HORMA protein functions in a similar mechanism to HORMA proteins in eukaryotes through the ability to change between an open and closed state. In the closed state, the HORMA binds CdnC and this binding is required for CdnC activity. CdnC synthesizes the cyclic olignucleotide second messenger cAAA. Once synthesized by CdnC, cAAA binds to a nuclease NucC, which forms a unique homotrimeric structure. The cAAA-bound form of NucC forms a homohexamer which cleaves double stranded DNA indiscriminately, leading to the destruction of both the

bacterial genome as well as the phage genome. By killing themselves, the bacteria thereby prevent propagation of phage to surrounding cells.

The CBASS system in *E. coli* MS115-1 along with 6.5% of other CBASS systems encode two genes upstream, called *capP* and *capH,* whose function have not been characterized. I show that these two genes work as transcriptional regulators of CBASS genes in response to DNA damage. I determine that CapH is a transcriptional repressor that binds to the promoter region upstream of CdnC. Upon binding single-stranded DNA, CapP, which is a metalloprotease, cleaves CapH which leads to expression of CBASS genes. CapP and CapH are additionally found upstream of other bacteriophage defense systems, suggesting that they may play a role in regulating these systems as well. However, because they are not found to be associated with all bacteriophage defense systems, they may not be required, but rather may be a mechanism for the bacteria to adjust the range of phages to which they respond.

INTRODUCTION

Bacteria, like other organisms, have evolved to survive in their environment. One stress that is constant regardless of the environment is the exposure to bacteriophages (phages), which are virus-like particles that can adhere to and infect bacteria. The infecting phage typically replicate within their bacterial host and then lyse the host cell to release phage progeny, resulting in cell death. To combat these phage infections, bacteria have developed a wide array of defense systems, with more being discovered regularly (Bernheim and Sorek, 2020; Doron et al., 2018; Hampton et al., 2020; Makarova et al., 2013; Tesson et al., 2022). In response, phages have developed systems to counter or overcome this immunity, which the bacteria will then need to respond to again (Hobbs et al., 2022). This everlasting battle between phage and bacteria has been coined the "Red Queen Hypothesis" in reference to the quote from Lewis Caroll's *Through the Looking Glass* where the Red Queen says to Alice: "here, you see, it takes all the running you can do to keep in the same place" (Carroll, 1900).

One widespread family of defense systems found in bacteria that have been well studied are restriction modification (R-M) systems (Dryden, 2013; Kovall and Matthews, 1999; Loenen et al., 2014; Murray, 2000). These systems typically encode a methyltransferase that methylates the host genome at specific sequences and a second enzyme, a restriction endonuclease, which cleaves the same DNA sequence if it is not protected by methylation. The restriction enzyme recognizes a specific DNA sequence, which is usually a palindrome, and cleaves both strands of DNA. When a bacteriophage

infects the host bacterium, the pattern of methylation on its DNA is usually different and therefore not protected from the resident restriction endonuclease. In this way, the host cell is able to discern its own genome from that of an invading phage and specifically cleave only the phage DNA, preventing its replication within the cell.

More recently, the discovery of CRISPR-Cas systems has garnered more interest in bacterial defense systems and their uses in biology and biotechnology. CRISPR arrays, which stands for Clustered Regularly Interspaced Short Palindromic Repeats, are located in the genomes of bacteria as part of their immune systems (Brouns et al., 2008; Haurwitz et al., 2010; van der Oost et al., 2009). When bacterial cells encounter a new phage, if they are able to detect that the phage DNA is different from their own they cut it into small pieces. Some of these small fragments are incorporated into the bacterial genome in the CRISPR arrays. Because the fragments are transferred to the host genome, the information can be passed on to future generations of the bacteria and the immunity is maintained. The small fragments can then be transcribed into small RNAs that are bound by Cas proteins, and are used as templates to recognize the sequence if the same phage infects the bacteria or its offspring again. In the case of Type II CRISPR-Cas systems, the Cas protein is responsible for both detecting and cleaving the phage DNA (Chylinski et al., 2014; Makarova et al., 2011a). Other CRISPR-Cas systems, such as Type I, III, and IV have one protein that binds the small guide RNA and others that form a scaffold and are ultimately responsible for cleaving the foreign DNA (Arslan et al., 2014, p. 1; Kazlauskiene et al., 2017; Niewoehner et al., 2017; Özcan et al., 2019).

The advance of sequencing technology has allowed thousands of bacterial genomes to be sequenced and comparative genomics techniques are used to search for new bacteriophage defense systems. These phage defense systems are often located close together in the genome, in clusters called “defense islands,” which allows us to search near already characterized anti-phage systems such as R-M or CRISPR-Cas systems to search for genes with no known functions and test to see if they are able to defend against phage (Makarova et al., 2011b). This has opened up a new field of discovering and characterizing phage defense systems with novel mechanisms, some of which have proteins that have carried over their functions into eukaryotes.

Phage life cycles

There are many different types of phages, but most follow one of two life cycles after adhering to and infecting a bacterial cell. These two life cycles are the lytic pathway and the lysogenic pathway (Benler and Koonin, 2020; Campbell, 2003; Casjens and Hendrix, 2015). In the lytic pathway, phages inject their nucleic acid into the cell and begin to replicate using the host machinery (Figure I.1A). In some cases, such as in T4 phage, the phage first degrades the host DNA, then begins to synthesize copies of its own (Kutter et al., 2018). As it replicates its DNA, the phage also hijacks host protein synthesis machinery to make phage structural proteins, including capsid and tail proteins, which assemble into virions that are filled with newly-replicated DNA. After a number of new phage particles are synthesized and fully assembled, the phage

enacts its final signal which causes the bacterial cells to lyse and the phage to go on to infect new bacterial cells.

In certain conditions, some phage enter the lysogenic lifestyle, in which they integrate their genome into the host genome and remain dormant, replicating along with the host (Lwoff, 1953) (Figure I.1B). The phage can replicate like this undetected, as what is called a “prophage.” The prophage remains integrated in the host genome until it is triggered to excise itself and enter the lytic cycle in a process called prophage induction. This induction can be caused by a variety of stress signals, all of which generally cause DNA damage. In the case of bacteriophage lambda, a commonly studied phage which has both a lytic and lysogenic life cycle, DNA damage is sensed through the detection of activated RecA (Galkin et al., 2009). In response to DNA damage and upon binding activated RecA, transcription factor cI, which is responsible for repressing the lytic cycle genes, self-cleaves. Self-cleavage relieves repression, and the lytic cycle genes can be expressed, thus the phage enters the lytic cycle. The newly created phage particles are capable of infecting new bacterial cells, after which they will either replicate through the lytic or lysogenic life cycle.

CBASS

One bacteriophage defense system that has recently been characterized, and which is the basis of this thesis, is the cyclic oligonucleotide-based antiphage signaling system, also known as CBASS. CBASS systems all encode a cGAS/DncV-like nucleotidyltransferase (CD-NTase) which is capable of synthesizing a cyclic

oligonucleotide second messenger (Cohen et al., 2019; Whiteley et al., 2019; Ye et al., 2020).

CBASS systems exist in single operons typically encoding two to five genes, and have been organized into distinct classes based largely on the regulatory proteins encoded in the operon (Millman et al., 2020) (Figure I.2). Type I CBASS operons are the simplest, encoding only a CD-NTase and a receptor effector protein. Little is known about how these systems are activated or regulated. Type II and Type III CBASS operons have a CD-NTase and effector, as well as ancillary regulatory proteins. The ancillary proteins in Type II CBASS operons are an E1-E2 fusion protein related to eukaryotic ubiquitin ligases which has been named Cap2 (CBASS-associated protein 2), and a JAB/JAMM family protease (Cap3) (Ledvina et al., 2022). A subcategory of Type II CBASS, Type II (short), encodes a Cap2-like protein with only an E2 domain, and lacks the JAB/JAMM family protease.

Type III CBASS encode one or two HORMA domain proteins (Cap7 and Cap8) and a AAA+ ATPase related to eukaryotic Pch2/TRIP13 (Cap6). HORMA domain proteins (HORMADs) function as signaling molecules in various eukaryotic processes such as meiosis and autophagy (Rosenberg and Corbett, 2015). In eukaryotes, HORMADs can switch between two conformations, an “open” state and “closed” state (Yang et al., 2008). The “closed” state is stabilized by binding to a peptide called a “closure motif”. The closure motif is generally a disordered sequence within a different protein, or a motif within the same protein. Once bound and in the closed state, the HORMAD is in the “on” state of signaling. HORMADs can be reverted to the open state

through the activity of a Pch2/Trip13 AAA+ ATPase, which binds the N-terminus of the HORMAD and destabilizes the interactions maintaining the closed state (Ye et al., 2017, 2015). While much is known about the functions and mechanisms of HORMADs and Pch2/Trip13 proteins in eukaryotic cells, their role in bacteria has not been characterized.

Type IV CBASS are the most rare occurrences of CBASS within prokaryotes, and are more often found in archaea than bacteria (Millman et al., 2020). The genes in Type IV CBASS have not been characterized, but are predicted to be a QueC domain protein (Cap9), a TGT-like protein (Cap10), and an OGG protein (Cap11).

CD-NTases

Second messenger signaling in bacteria is well known to modulate bacterial behavior. The small molecule second messenger cyclic di-GMP (c-di-GMP) is found throughout the bacterial world and can alter many responses including the switch from motile to biofilm, attachment to surfaces, and the ability to enter various stages of the cell cycle (Jenal et al., 2017). Bacterial cells often encode more than one c-di-GMP synthase, which are characterized by their GGDEF catalytic domain (Hengge, 2009). C-di-GMP is synthesized when two GGDEF domain-containing enzymes dimerize, each with one GTP in the active site. Bacteria also contain an array of phosphodiesterases which degrade the c-di-GMP and are involved in regulation of c-di-GMP levels (Pesavento et al., 2008).

While most dinucleotide synthases are dimers, a novel class of dinucleotide synthases was discovered in *Vibrio cholerae* that exists as a monomer (Kato et al., 2015). This enzyme, termed DncV (dinucleotide cyclase in *Vibrio*) was found to synthesize the small molecule cyclic GMP-AMP (cGAMP) (Davies et al., 2012). This second messenger molecule has two 3'-5' linkages, and while other signaling cascades involving cyclic nucleotides and dinucleotides appear to have multiple targets within the cell to effect a global rearrangement of cellular processes, cGAMP often has only one target effector protein, CapV, which is usually found in the same operon as DncV itself (Davies et al., 2012; Diner et al., 2013). DncV is a member of a larger family of proteins that are termed CD-NTases (cGAS/DncV-like nucleotidyltransferases), which have been shown to synthesize a wide array of cyclic oligonucleotides, ranging from dinucleotides to tetranucleotides, and use different combinations of the four RNA bases (guanine, uridine, adenine, cytidine) (Whiteley et al., 2019). Based on sequence alignments, Whiteley et al categorized bacterial CD-NTases into different clades which can be used to predict their second messenger product. Enzymes from CdnA family have been shown to synthesize cGAMP. CD-NTases from the CdnC and CdnD families have been shown to synthesize cyclic trinucleotides cAAA and cAAG, respectively (Whiteley et al., 2019; Ye et al., 2020).

CD-NTases are also found in mammalian cells. One such protein, called cGAS (cyclic GMP-AMP synthase), is a DNA-sensing enzyme that synthesizes cGAMP upon binding double-stranded DNA in the cytosol (Sun et al., 2013; Wu et al., 2013). cGAS shows low sequence conservation with DncV, but shares overall structural similarity,

including the bi-lobed structure and catalytic site (Kranzusch et al., 2013). Although cGAS, like DncV, synthesizes cGAMP, the cGAMP created by cGAS contains an asymmetric bond pattern, with one bond being 2'-5' and the other being 3'-5' (Diner et al., 2013). Once synthesized, cGAMP binds to STING (stimulator of interferon genes), which then activates an innate immune response (Cai et al., 2014; Sun et al., 2013; Wu et al., 2013). STING has been shown to bind a variety of cyclic oligonucleotides including the 3'-5' linked cGAMP synthesized in bacteria, and is therefore considered a signal integrator that can sense cGAS activation and other bacterial-derived second messengers (Burdette et al., 2011).

Along with cGAS, bacterial CD-NTases are also structurally related to the mammalian innate immune protein oligo-adenylate synthetase (OAS). OAS is part of an antiviral response in mammalian cells in the innate immune system (Lohöfener et al., 2015). Rather than binding dsDNA, OAS is activated by binding dsRNA of viral origin, and synthesizes linear 2'-5' oligoadenylate. Linear oligoadenylate binds and activates RNase L, which then degrades the double stranded RNA from invading viruses (Morin et al., 2010).

The mammalian homologs of DncV, cGAS and OAS, both bind and are activated by nucleic acid that is either mis-localized to the cytoplasm or has been introduced to the cytoplasm by an RNA virus. cGAS binds dsDNA and OAS binds dsRNA and the binding surface appears to be conserved between the two enzymes (Lohöfener et al., 2015; Zhou et al., 2018). However, the nucleic acid binding face is not conserved in bacterial CD-NTases, and the mechanism of activation of bacterial CD-NTases remains

unknown. DncV appears to have constitutive activity *in vitro*, however many of the bacterial CD-NTases in other families require incubation at pH 9.4 to synthesize cyclic oligonucleotide products (Davies et al., 2012; Whiteley et al., 2019). It is unlikely that bacterial CD-NTases are activated by binding directly to nucleic acids not only because they do not have the conserved binding surface, but also because bacteria contain no nucleus. Therefore, bacterial CD-NTases are constantly exposed to nucleic acid. I hypothesize that these bacterial CD-NTases are activated through a different mechanism which may be dependent on the activity of the other components of the CBASS operons.

CBASS Effectors

The effector proteins in CBASS systems are involved in cellular destruction. These proteins are typically predicted to be phospholipases, nucleases, or pore-forming molecules. The first characterized CBASS effector is the phospholipase effector in the DncV operon in *Vibrio cholerae* (Davies et al., 2012; Kato et al., 2015). The phospholipase CapV is activated specifically by cGAMP produced by DncV. Once activated by cGAMP, CapV catabolizes phospholipids to release free fatty acids (Severin et al., 2018). Cells overexpressing both DncV and CapV become very sick and are unable to grow, due to a reduction in membrane integrity caused by the activation of CapV.

Another class of CBASS effectors consists of endonucleases. Endonucleases cleave double stranded DNA internally. There are at least three characterized

endonuclease types associated with CBASS, one of which (NucC) is the subject of Chapter 2 of this thesis (Lau et al., 2020; Lowey et al., 2020). A separate family of CBASS endonucleases contains one characterized member called Cap4. Cap4 is unique from NucC as it contains two domains, a nuclease domain a SAVED domain. The nuclease domains of both NucC and Cap4 are related to those of restriction endonucleases. The SAVED domain of Cap4 was first identified by Burroughs et al. and is predicted to be an oligonucleotide binding domain and is thus named SMODS-associated and fused to various effector domains. SAVED domains can be found attached to many CBASS effectors, including other endonucleases such as HNH homing endonucleases (Burroughs and Aravind, 2020).

A third common effector type found in CBASS operons is a pore-forming protein with one or more predicted transmembrane segments. These proteins are thought to oligomerize when bound to the cyclic oligonucleotide second messengers synthesized by their cognate CD-NTases (Duncan-Lowey et al., 2021). The oligomer complexes form pores in the cell membranes, causing a disruption in the usually tightly regulated exchange of molecules between the cytosol and extracellular environment.

Motivation/Hypothesis

CBASS operons have been identified in many bacterial genomes, but prior to our work it had not yet been demonstrated that they are indeed bacterial defense systems. The first characterized DncV/CapV system was shown to suppress growth when overexpressed and CapV to degrade phospholipids, but the response to phage was not

characterized (Severin et al., 2018). Furthermore, we do not know how CBASS systems are activated and regulated. While some CD-NTases can be activated *in vitro*, non-biologically relevant conditions are required (Whiteley et al., 2019). Many CD-NTases cannot even be activated in these harsh conditions, so the extent of CD-NTase products remains to be discovered. As mentioned above, Type III CBASS systems have HORMA and Trip13, which we hypothesize are responsible for regulating the activity of the CD-NTase. In this thesis, I address the questions surrounding the purpose and mechanism of a Type III CBASS found in *E. coli* MS115-1.

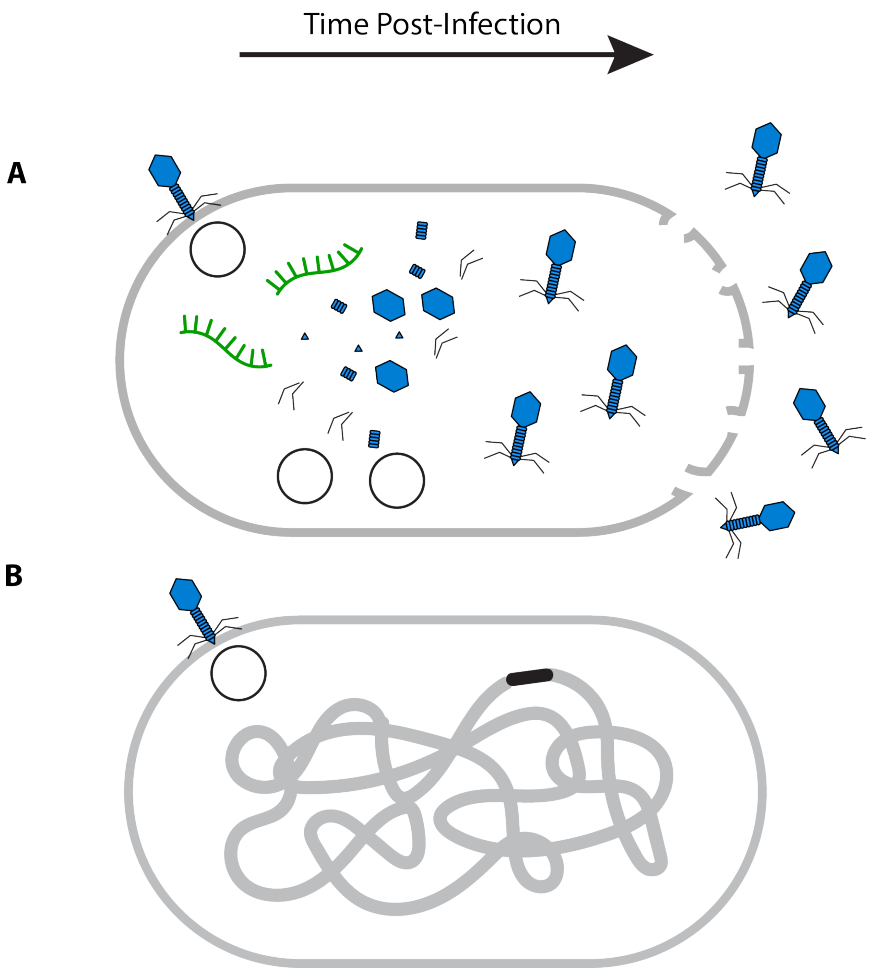


Figure I.1 Bacteriophage Lifecycles

(A) A bacteriophage undergoing the lytic cell cycle. Phage adhere to and infect bacteria by injecting their DNA (black). Phage replicate their genomic DNA and begin transcribing (green) and translating their proteins (blue). They assemble new phage particles and package the newly synthesized DNA, then lyse the cells releasing the fully formed particles.
(B) A bacteriophage undergoing the lysogenic cell cycle. After injecting their DNA, the phage genome integrates into the bacterial genome and replicates along with the cell.

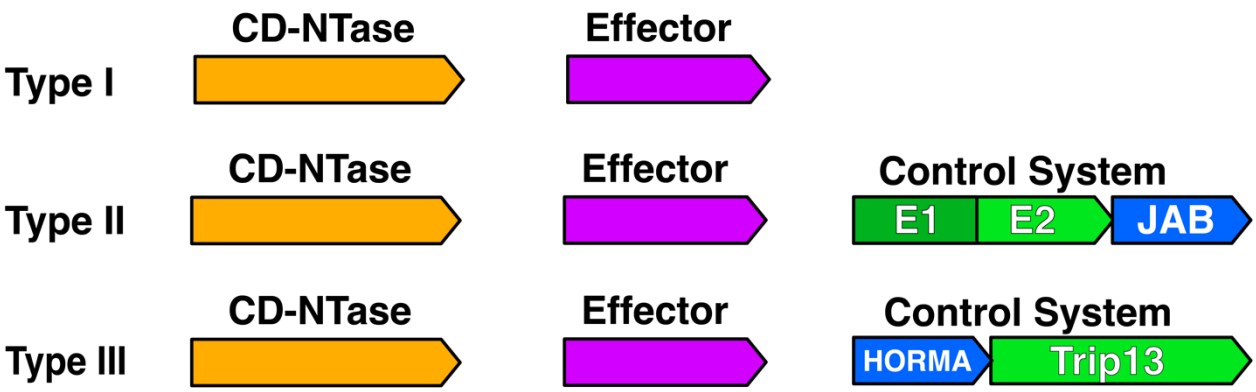


Figure I.2 Types of CBASS systems

Schematic of each type of CBASS system.

References

Arslan, Z., Hermanns, V., Wurm, R., Wagner, R., Pul, Ü., 2014. Detection and characterization of spacer integration intermediates in type I-E CRISPR–Cas system. Nucleic Acids Research 42, 7884–7893. https://doi.org/10.1093/nar/gku510

Benler, S., Koonin, E.V., 2020. Phage lysis-lysogeny switches and programmed cell death: Danse macabre. BioEssays 42, 2000114. https://doi.org/10.1002/BIES.202000114

Bernheim, A., Sorek, R., 2020. The pan-immune system of bacteria: antiviral defence as a community resource. Nature Reviews Microbiology 18, 113–119. https://doi.org/10.1038/s41579-019-0278-2

Brouns, S.J.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J.H., Snijders, A.P.L., Dickman, M.J., Makarova, K.S., Koonin, E.V., Van Der Oost, J., 2008. Small CRISPR RNAs guide antiviral defense in prokaryotes. Science 321, 960–964. https://doi.org/10.1126/science.1159689

Burdette, D.L., Monroe, K.M., Sotelo-Troha, K., Iwig, J.S., Eckert, B., Hyodo, M., Hayakawa, Y., Vance, R.E., 2011. STING is a direct innate immune sensor of cyclic di-GMP. Nature 478, 515–518. https://doi.org/10.1038/nature10429

Burroughs, A.M., Aravind, L., 2020. Identification of Uncharacterized Components of Prokaryotic Immune Systems and Their Diverse Eukaryotic Reformulations. Journal of Bacteriology 202, e00365-20. https://doi.org/10.1128/JB.00365-20

Cai, X., Chiu, Y.-H., Chen, Z.J., 2014. The cGAS-cGAMP-STING Pathway of Cytosolic DNA Sensing and Signaling. Molecular Cell 54, 289–296. https://doi.org/10.1016/j.molcel.2014.03.040

Campbell, A., 2003. The future of bacteriophage biology. Nat Rev Genet 4, 471–477. https://doi.org/10.1038/nrg1089

Carroll, L., 1832-1898, 1900. Through the looking-glass and what Alice found there. Chicago : W.B. Conkey Co., [1900] ©1900.

Casjens, S.R., Hendrix, R.W., 2015. Bacteriophage lambda: early pioneer and still relevant. Virology 0, 310–330. https://doi.org/10.1016/j.virol.2015.02.010

Chylinski, K., Makarova, K.S., Charpentier, E., Koonin, E.V., 2014. Classification and evolution of type II CRISPR-Cas systems. Nucleic Acids Res 42, 6091–6105. https://doi.org/10.1093/nar/gku241

Cohen, D., Melamed, S., Millman, A., Shulman, G., Oppenheimer-Shaanan, Y., Kacen, A., Doron, S., Amitai, G., Sorek, R., 2019. Cyclic GMP–AMP signalling protects bacteria against viral infection. Nature 574, 691–695. https://doi.org/10.1038/s41586-019-1605-5

Davies, B.W., Bogard, R.W., Young, T.S., Mekalanos, J.J., 2012. Coordinated Regulation of Accessory Genetic Elements Produces Cyclic Di-Nucleotides for V. cholerae Virulence. Cell 149, 358–370. https://doi.org/10.1016/J.CELL.2012.01.053

Diner, E.J., Burdette, D.L., Wilson, S.C., Monroe, K.M., Kellenberger, C.A., Hyodo, M., Hayakawa, Y., Hammond, M.C., Vance, R.E., 2013. The innate immune DNA sensor cGAS produces a noncanonical cyclic dinucleotide that activates human STING. Cell reports 3, 1355–61. https://doi.org/10.1016/j.celrep.2013.05.009

Doron, S., Melamed, S., Ofir, G., Leavitt, A., Lopatina, A., Keren, M., Amitai, G., Sorek, R., 2018. Systematic discovery of antiphage defense systems in the microbial pangenome. Science 359, eaar4120. https://doi.org/10.1126/science.aar4120

Dryden, D.T.F., 2013. The architecture of restriction enzymes. Structure 21, 1720–1721. https://doi.org/10.1016/j.str.2013.09.009

Duncan-Lowey, B., McNamara-Bordewick, N.K., Tal, N., Sorek, R., Kranzusch, P.J., 2021. Effector-mediated membrane disruption controls cell death in CBASS antiphage defense. Molecular Cell 81, 5039-5051.e5. https://doi.org/10.1016/j.molcel.2021.10.020

Galkin, V.E., Yu, X., Bielnicki, J., Ndjonka, D., Bell, C.E., Egelman, E.H., 2009. Cleavage of Bacteriophage λ cI Repressor Involves the RecA C-terminal Domain. J Mol Biol 385, 779–787. https://doi.org/10.1016/j.jmb.2008.10.081

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

Haurwitz, R.E., Jinek, M., Wiedenheft, B., Zhou, K., Doudna, J.A., 2010. Sequence- and structure-specific RNA processing by a CRISPR endonuclease. Science 329, 1355–1358. https://doi.org/10.1126/science.1192272

Hengge, R., 2009. Principles of c-di-GMP signalling in bacteria. Nature Reviews Microbiology 7, 263–273. https://doi.org/10.1038/nrmicro2109

Hobbs, S.J., Wein, T., Lu, A., Morehouse, B.R., Schnabel, J., Leavitt, A., Yirmiya, E., Sorek, R., Kranzusch, P.J., 2022. Phage anti-CBASS and anti-Pycsar nucleases

subvert bacterial immunity. Nature 2022 1–8. https://doi.org/10.1038/s41586-022-04716-y

Jenal, U., Reinders, A., Lori, C., 2017. Cyclic di-GMP: second messenger extraordinaire. Nat Rev Microbiol 15, 271–284. https://doi.org/10.1038/nrmicro.2016.190

Kato, K., Ishii, R., Hirano, S., Ishitani, R., Nureki, O., 2015. Structural Basis for the Catalytic Mechanism of DncV, Bacterial Homolog of Cyclic GMP-AMP Synthase. Structure 23, 843–850. https://doi.org/10.1016/j.str.2015.01.023

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

Kovall, R.A., Matthews, B.W., 1999. Type II restriction endonucleases: Structural, functional and evolutionary relationships. Current Opinion in Chemical Biology 3, 578–583. https://doi.org/10.1016/S1367-5931(99)00012-5

Kranzusch, P.J., Lee, A.S.-Y., Berger, J.M., Doudna, J.A., 2013. Structure of human cGAS reveals a conserved family of second-messenger enzymes in innate immunity. Cell reports 3, 1362–8. https://doi.org/10.1016/j.celrep.2013.05.008

Kutter, E., Bryan, D., Ray, G., Brewster, E., Blasdel, B., Guttman, B., 2018. From Host to Phage Metabolism: Hot Tales of Phage T4's Takeover of E. coli. Viruses 10, 387. https://doi.org/10.3390/v10070387

Lau, R.K., Ye, Q., Birkholz, E.A., Berg, K.R., Patel, L., Mathews, I.T., Watrous, J.D., Ego, K., Whiteley, A.T., Lowey, B., Mekalanos, J.J., Kranzusch, P.J., Jain, M., Pogliano, J., Corbett, K.D., 2020. Structure and Mechanism of a Cyclic Trinucleotide-Activated Bacterial Endonuclease Mediating Bacteriophage Immunity. Molecular Cell 77, 723-733.e6. https://doi.org/10.1016/j.molcel.2019.12.010

Ledvina, H.E., Ye, Q., Gu, Y., Quan, Y., Lau, R.K., Zhou, H., Corbett, K.D., Whiteley, A.T., 2022. cGASylation by a bacterial E1-E2 fusion protein primes antiviral immune signaling. bioRxiv 2022.03.31.486616. https://doi.org/10.1101/2022.03.31.486616

Loenen, W.A.M., Dryden, D.T.F., Raleigh, E.A., Wilson, G.G., 2014. Type I restriction enzymes and their relatives. Nucleic Acids Research 42, 20–44. https://doi.org/10.1093/nar/gkt847

Lohöfener, J., Steinke, N., Kay-Fedorov, P., Baruch, P., Nikulin, A., Tishchenko, S., Manstein, D.J., Fedorov, R., 2015. The Activation Mechanism of 2′-5′-

Oligoadenylate Synthetase Gives New Insights Into OAS/cGAS Triggers of Innate Immunity. Structure 23, 851–862.

Lowey, B., Whiteley, A.T., Keszei, A.F.A., Morehouse, B.R., Mathews, I.T., Antine, S.P., Cabrera, V.J., Kashin, D., Niemann, P., Jain, M., Schwede, F., Mekalanos, J.J., Shao, S., Lee, A.S.Y., Kranzusch, P.J., 2020. CBASS Immunity Uses CARF-Related Effectors to Sense 3′–5′- and 2′–5′-Linked Cyclic Oligonucleotide Signals and Protect Bacteria from Phage Infection. Cell 182, 38-49.e17. https://doi.org/10.1016/J.CELL.2020.05.019

Lwoff, A., 1953. Lysogeny. Bacteriological Reviews 17, 269–337. https://doi.org/10.1128/br.17.4.269-337.1953

Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J.M., Wolf, Y.I., Yakunin, A.F., van der Oost, J., Koonin, E.V., 2011a. Evolution and classification of the CRISPR–Cas systems. Nature Reviews Microbiology 9, 467–477. https://doi.org/10.1038/nrmicro2577

Makarova, K.S., Wolf, Y.I., Koonin, E.V., 2013. Comparative genomics of defense systems in archaea and bacteria. Nucleic Acids Research 41, 4360–4377. https://doi.org/10.1093/nar/gkt157

Makarova, K.S., Wolf, Y.I., Snir, S., Koonin, E.V., 2011b. Defense Islands in Bacterial and Archaeal Genomes and Prediction of Novel Defense Systems. Journal of Bacteriology 193, 6039–6056. https://doi.org/10.1128/JB.05535-11

Millman, A., Melamed, S., Amitai, G., Sorek, R., 2020. Diversity and classification of cyclic-oligonucleotide-based anti-phage signalling systems. Nature Microbiology 2020 5:12 5, 1608–1615. https://doi.org/10.1038/s41564-020-0777-y

Morin, B., Rabah, N., Boretto-Soler, J., Tolou, H., Alvarez, K., Canard, B., 2010. High yield synthesis, purification and characterisation of the RNase L activators 5'-triphosphate 2′-5′-oligoadenylates. Antiviral Research 87, 345–352. https://doi.org/10.1016/j.antiviral.2010.06.003

Murray, N.E., 2000. Type I Restriction Systems: Sophisticated Molecular Machines (a Legacy of Bertani and Weigle). Microbiology and Molecular Biology Reviews 64, 412–434. https://doi.org/10.1128/mmbr.64.2.412-434.2000

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

Özcan, A., Pausch, P., Linden, A., Wulf, A., Schühle, K., Heider, J., Urlaub, H., Heimerl, T., Bange, G., Randau, L., 2019. Type IV CRISPR RNA processing and effector complex formation in Aromatoleum aromaticum. Nat Microbiol 4, 89–96. https://doi.org/10.1038/s41564-018-0274-8

Pesavento, C., Becker, G., Sommerfeldt, N., Possling, A., Tschowri, N., Mehlis, A., Hengge, R., 2008. Inverse regulatory coordination of motility and curli-mediated adhesion in Escherichia coli. Genes & development 22, 2434–46. https://doi.org/10.1101/gad.475808

Rosenberg, S.C., Corbett, K.D., 2015. The multifaceted roles of the HORMA domain in cellular signaling. The Journal of cell biology 211, 745–55. https://doi.org/10.1083/jcb.201509076

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, E6048–E6055. https://doi.org/10.1073/pnas.1801233115

Sun, L., Wu, J., Du, F., Chen, X., Chen, Z.J., 2013. Cyclic GMP-AMP synthase is a cytosolic DNA sensor that activates the type I interferon pathway. Science (New York, N.Y.) 339, 786–91. https://doi.org/10.1126/science.1232458

Tesson, F., Hervé, A., Mordret, E., Touchon, M., d'Humières, C., Cury, J., Bernheim, A., 2022. Systematic and quantitative view of the antiviral arsenal of prokaryotes. Nat Commun 13, 2561. https://doi.org/10.1038/s41467-022-30269-9

van der Oost, J., Jore, M.M., Westra, E.R., Lundgren, M., Brouns, S.J.J., 2009. CRISPR-based adaptive and heritable immunity in prokaryotes. Trends in Biochemical Sciences 34, 401–407. https://doi.org/10.1016/j.tibs.2009.05.002

Whiteley, A.T., Eaglesham, J.B., de Oliveira Mann, C.C., Morehouse, B.R., Lowey, B., Nieminen, E.A., Danilchanka, O., King, D.S., Lee, A.S.Y., Mekalanos, J.J., Kranzusch, P.J., 2019. Bacterial cGAS-like enzymes synthesize diverse nucleotide signals. Nature 567, 194–199. https://doi.org/10.1038/s41586-019-0953-5

Wu, J., Sun, L., Chen, X., Du, F., Shi, H., Chen, C., Chen, Z.J., 2013. Cyclic GMP-AMP is an endogenous second messenger in innate immune signaling by cytosolic DNA. Science 339, 826–830. https://doi.org/10.1126/science.1229963

Yang, M., Li, B., Liu, C.-J., Tomchick, D.R., Machius, M., Rizo, J., Yu, H., Luo, X., 2008. Insights into Mad2 Regulation in the Spindle Checkpoint Revealed by the Crystal

Structure of the Symmetric Mad2 Dimer. PLOS Biology 6, e50. https://doi.org/10.1371/journal.pbio.0060050

Ye, Q., Kim, D.H., Dereli, I., Rosenberg, S.C., Hagemann, G., Herzog, F., Tóth, A., Cleveland, D.W., Corbett, K.D., 2017. The AAA + ATP ase TRIP 13 remodels HORMA domains through N-terminal engagement and unfolding. The EMBO Journal 36, 2419–2434. https://doi.org/10.15252/embj.201797291

Ye, Q., Lau, R.K., Mathews, I.T., Birkholz, E.A., Watrous, J.D., Azimi, C.S., Pogliano, J., Jain, M., Corbett, K.D., 2020. HORMA Domain Proteins and a Trip13-like ATPase Regulate Bacterial cGAS-like Enzymes to Mediate Bacteriophage Immunity. Molecular Cell 77, 709-722.e7. https://doi.org/10.1016/j.molcel.2019.12.009

Ye, Q., Rosenberg, S.C., Moeller, A., Speir, J.A., Su, T.Y., Corbett, K.D., 2015. TRIP13 is a protein-remodeling AAA+ ATPase that catalyzes MAD2 conformation switching. eLife 4, e07367. https://doi.org/10.7554/eLife.07367

Zhou, W., Whiteley, A.T., de Oliveira Mann, C.C., Morehouse, B.R., Nowak, R.P., Fischer, E.S., Gray, N.S., Mekalanos, J.J., Kranzusch, P.J., 2018. Structure of the Human cGAS–DNA Complex Reveals Enhanced Control of Immune Surveillance. Cell 174, 300-311.e11. https://doi.org/10.1016/J.CELL.2018.06.026

CHAPTER 1: HORMA Domain Proteins and a Trip13-like ATPase Regulate Bacterial cGAS-like Enzymes to Mediate Bacteriophage Immunity

Summary

Bacteria are continually challenged by foreign invaders including bacteriophages, and have evolved a variety of defenses against these invaders. Here, we describe the structural and biochemical mechanisms of a bacteriophage immunity pathway found in a broad array of bacteria, including *E. coli* and *Pseudomonas aeruginosa*. This pathway employs eukaryotic-like HORMA domain proteins that recognize specific peptides, then bind and activate a cGAS/DncV-like nucleotidyltransferase (CD-NTase) to generate a cyclic triadenylate (cAAA) second messenger; cAAA in turn activates an endonuclease effector, NucC. Signaling is attenuated by a homolog of the AAA+ ATPase Pch2/TRIP13, which binds and disassembles the active HORMA-CD-NTase complex. When expressed in non-pathogenic *E. coli*, this pathway confers immunity against bacteriophage λ through an abortive infection mechanism. Our findings reveal the molecular mechanisms of a bacterial defense pathway integrating a cGAS-like nucleotidyltransferase with HORMA domain proteins for threat sensing through protein detection, and negative regulation by a Trip13 ATPase.

Keywords

Second messenger signaling, bacteriophage immunity, abortive infection, CD-NTase, HORMA domain, AAA+ ATPase remodeler

Introduction

All cellular life depends on signaling pathways that can rapidly sense and respond to changes in both a cell's internal state and its external environment. Many such signaling pathways rely on diffusible second messenger molecules, which are synthesized by sensor proteins and activate effector proteins to drive a physiological response. In bacteria, the cyclic dinucleotide second messengers cyclic di-AMP, cyclic di-GMP, and cyclic GMP-AMP (cGAMP) play key roles in cellular homeostasis and pathogenesis (Corrigan and Gründling, 2013; Davies et al., 2012; Hengge, 2009). In mammals, cGAMP and linear oligoadenylate are important innate-immune signals, synthesized by cGAS (cyclic GMP-AMP synthase) and OAS (oligoadenylate synthase) family proteins in response to cytoplasmic DNA or double-stranded RNA (Chen et al., 2016; Hornung et al., 2014). Nucleotide-based second messengers also mediate cross-kingdom signaling, with mammalian innate immune receptors able to recognize and respond to a variety of bacterially-generated second messengers (Burdette et al., 2011; McFarland et al., 2017; Whiteley et al., 2019).

A major family of second messenger biosynthetic enzymes found across kingdoms is distantly related to DNA polymerase β and includes bacterial DncV (Dinucleotide cyclase in *Vibrio*), which synthesizes cGAMP (Davies et al., 2012; Zhu et al., 2014), and mammalian cGAS and OAS. Comparative genomics studies in bacteria have identified over 6000 distinct cGAS/DncV-like nucleotidyltransferases (CD-NTases) in environmental and human patient-derived bacterial strains, with examples found in over 10% of sequenced bacterial genomes (Burroughs et al., 2015; Cohen et al., 2019;

Whiteley et al., 2019). These enzymes synthesize a marked variety of second messengers, including cyclic dinucleotides with both purine and pyrimidine bases, cyclic trinucleotides, and other as-yet uncharacterized molecules (Whiteley et al., 2019). The majority of bacterial CD-NTases are found within “defense islands” alongside known anti-phage defense systems (Cohen et al., 2019), and are encoded in operons that also encode putative effector proteins including predicted nucleases, proteases, and phospholipases (Burroughs et al., 2015; Whiteley et al., 2019). Two CD-NTase containing operons, one from *V. cholerae* encoding a phospholipase effector, and another from *Bacillus cereus* encoding a predicted transmembrane protein effector, have been shown to confer immunity against a range of bacteriophages (Cohen et al., 2019). In both cases, immunity is mediated by abortive infection, in which infected cells die prior to completion of phage replication and thereby prevent additional rounds of infection (Cohen et al., 2019). This new class of anti-bacteriophage defense systems has been termed CBASS (Cyclic oligonucleotide-Based Anti-phage Signaling System) (Cohen et al., 2019).

Biochemical analysis of 66 purified bacterial CD-NTases from a variety of CBASS systems showed that the majority of these enzymes are inactive in vitro, suggesting that their activity is tightly regulated in cells (Whiteley et al., 2019). Intriguingly, about 10% of CBASS systems encode one or two proteins with weak homology to eukaryotic HORMA domain proteins, along with an ortholog of the AAA+ ATPase regulator of HORMA domain proteins, Pch2/TRIP13 (Burroughs et al., 2015). In eukaryotes, HORMA domains (named for the first three proteins shown to possess this

domain: Hop1, Rev7, and Mad2) bind short regions within their interaction partners called “closure motifs,” and assemble into signaling complexes that control inter-homolog recombination in meiosis (Hop1), DNA repair (Rev7), the mitotic spindle assembly checkpoint (Mad2), and autophagy (Atg13, Atg101) (Aravind and Koonin, 1998; Rosenberg and Corbett, 2015). The HORMA domain possesses two distinct folded states: an “open” state unable to bind closure motifs, and a “closed” state in which the HORMA domain C-terminus wraps entirely around a bound closure motif (Luo and Yu, 2008; Rosenberg and Corbett, 2015). The Pch2/TRIP13 ATPase disassembles HORMA:closure motif complexes by partially unfolding the N-terminus of the HORMA domain, dissociating the complex and in some cases converting the HORMA domain from the “closed” to the “open” state (Ma and Poon, 2016; Ye et al., 2017, 2015).

While HORMA domains play key signaling roles in eukaryotes, these proteins have not been previously described in bacteria. The presence of putative HORMA domain proteins in a subset of CBASS systems suggests that these proteins may mediate phage sensing, perhaps utilizing the HORMA domain’s peptide binding and conformational conversion abilities to control second messenger synthesis by their cognate CD-NTases. Here, we show that this is indeed the case: bacterial HORMA domain proteins adopt the canonical HORMA domain fold, undergo conformational conversion between open and closed states, and bind specific closure motif peptides. Binding of a “closed” HORMA domain protein to its cognate CD-NTase activates synthesis of a cyclic tri-AMP (cAAA) second messenger, which in turn activates an effector endonuclease also encoded in the operon. Bacterial Trip13-like ATPases

function as negative regulators of signaling, binding and likely disassembling the active CD-NTase:HORMA complex. Finally, we show that a HORMA+Trip13 CBASS system from a patient-derived *E. coli* strain confers immunity against bacteriophage λ infection via abortive infection. Together, these results define the molecular mechanisms of a new bacterial defense system that employs HORMA domain proteins as a regulator of second messenger signaling.

Results

Bacterial HORMA domain proteins bind and activate second messenger synthesis by their associated CD-NTases

CBASS systems encoding putative HORMA domain proteins and an ortholog of the AAA+ ATPase Pch2/TRIP13 (hereon termed Trip13) fall into two major classes, one of which encodes a single HORMA domain protein (here termed HORMA1) and the second of which encodes two such proteins (HORMA2 and HORMA3) (**Figure 1.1A** and **Figure A.A.S1A**) (Burroughs et al., 2015). The HORMA1 and HORMA2/3 operons also encode diverged Trip13 ATPases, suggesting that the HORMA and Trip13 genes have co-evolved within each family of operons (**Figure A.A.S1A-C**) (Tromer et al., 2019). Bacterial CD-NTases have been grouped into different clades based on their sequence similarity (Whiteley et al., 2019), and we find that HORMA1 operons mainly encode clade C01 CD-NTases (hereon termed CdnC), while HORMA2/3 operons encode clade D05 CD-NTases (termed CdnD). To address the function of these operons, we chose a CdnC+HORMA1+Trip13 operon from *E. coli* strain MS115-1, and a CdnD+HORMA2/3+Trip13 operon from *Pseudomonas aeruginosa* strain ATCC27853, for further study (**Figure 1.1A**). The *E. coli* operon is found in several patient-derived *E. coli* strains, and the *P. aeruginosa* operon is found in over a quarter of patient-derived *P. aeruginosa* strains. Both operons also encode related putative endonuclease effectors, termed REase by Burroughs et al. (Burroughs et al., 2015) and here termed NucC (Nuclease, CD-NTase associated).

As a first step, we performed yeast two-hybrid analysis with the *E. coli* MS115-1 (*Ec*) proteins, and found that *Ec* CdnC and *Ec* HORMA bind one another (**Figure 1.1B** and **Figure A.A.S2A**). When co-expressed in *E. coli*, *Ec* CdnC and *Ec* HORMA form two complexes with both 1:1 and 2:2 stoichiometry (**Figure A.A.S2B**). In a yeast three-hybrid experiment, a stabilized hexamer mutant of *Ec* Trip13 (Walker B motif residue Glu159 to Gln; Trip13^{EQ}) binds to an *Ec* CdnC:HORMA complex (**Figure 1.1B** and **Figure A.A.S2A**), and in vitro reconstitution reveals that this ternary complex has a stoichiometry of 6 TRIP13^{EQ}:1 CdnC:1 HORMA (**Figure A.A.S2C**). Similar yeast two-hybrid analysis with the *P. aeruginosa* ATCC27853 (*Pa*) proteins showed that *Pa* HORMA2 and HORMA3 interact with one another, and that both HORMA proteins interact with *Pa* CdnD (**Figure A.A.S2E**). In vitro reconstitutions reveal that these proteins form a CdnD:HORMA2:HORMA3 complex with 1:1:1 stoichiometry (**Figure A.A.S2F-H**). Finally, both *Ec* NucC and *Pa* NucC self-interact in yeast two-hybrid assays, suggesting that NucC forms homodimers or larger oligomers (**Figure A.A.S2A** and **Figure A.A.S2E**).

Both *Ec* CdnC and *Pa* CdnD were previously found to be inactive in vitro (Whiteley et al., 2019), leading us to test if their activity depends on their cognate HORMA and Trip13 proteins. For this purpose, we tested purified proteins in an in vitro second messenger synthesis assay based on prior work with a related CD-NTase, *E. cloacae* CdnD02 (**Figure 1.1C**) (Whiteley et al., 2019). We first tested for activity of *Ec* CdnC and the *Ec* CdnC:HORMA complex (both 1:1 and 2:2 complexes). As the related cGAS and OAS enzymes both require nucleic acid binding for activation (Hornung et al.,

2014), we also included double-stranded DNA as a potential activator. We find that while CdnC is inactive on its own (**Figure 1.1C**, sample i), the 2:2 complex of *Ec* CdnC and HORMA synthesizes two phosphatase-resistant products in the presence of DNA (**Figure 1.1C**, sample v). Curiously, the 1:1 CdnC:HORMA complex is inactive (**Figure 1.1C**, sample iv). CdnC+HORMA is activated by long double-stranded DNAs, but not by single-stranded DNA or short (40 bp) double-stranded DNA (**Figure A.A.S3A**). CdnC activity is disrupted by mutation of the putative active site residues Asp72 and Asp74, demonstrating that CdnC uses a similar biosynthetic mechanism as related CD-NTases (**Figure 1.1C**, sample vi). Finally, we found that addition of sub-stoichiometric amounts of *Ec* Trip13 (1:60 molar ratio of active Trip13 hexamers to CdnC+HORMA) strongly reduced second messenger synthesis (**Figure 1.1C**, sample vii). Together, these data support a model in which *Ec* CdnC is activated by binding its cognate HORMA protein and DNA, and inactivated by Trip13, potentially through direct disassembly of the CdnC:HORMA complex (**Figure 1.1F**).

We next used liquid chromatography coupled to tandem mass spectrometry (LC MS/MS) to analyze the major and minor products of *Ec* CdnC. We identified the major product as cyclic tri-AMP (cAAA; **Figure 1.1D-E**), and the minor product as cyclic di-AMP (cAA; not shown). *Ec* CdnC synthesizes only cAAA/cAA even in the presence of both GTP and ATP, indicating that the enzyme uses only ATP as a substrate (not shown). Thus, *Ec* CdnC synthesizes a cyclic trinucleotide second messenger upon activation by its cognate HORMA domain protein and double-stranded DNA.

We next tested the activity of *Pa* CdnD in the presence of *Pa* HORMA2, *Pa* HORMA3, and *Pa* Trip13. While *Pa* CdnD shows lower activity than *Ec* CdnC in vitro and requires Mn^{2+} instead of Mg^{2+} as a cofactor, we find that CdnD also primarily synthesizes cAAA (**Figure A.A.S3B**). *Pa* CdnD requires HORMA2 for activity, but is not activated by HORMA3 (**Figure A.A.S3B**), perhaps reflecting functional specialization of the two HORMA domain proteins in this operon. As with the *E. coli* proteins, addition of *Pa* Trip13 to a reaction with *Pa* CdnD and HORMA2 strongly inhibits second messenger synthesis (**Figure A.A.S3B**). Finally, we find that in contrast to *Ec* CdnC, DNA is not required for *Pa* CdnD activity, nor does addition of DNA further stimulate cAAA production (**Figure A.A.S3B**). Thus, while *Ec* CdnC and *Pa* CdnD are regulated differently, the two enzymes share a common dependence on HORMA domain protein binding for activation, and generate the same cyclic trinucleotide second messenger.

Structures of HORMA-associated CD-NTases

To address the structural and biochemical mechanisms by which *Ec* CdnC and *Pa* CdnD synthesize cAAA in a regulated manner, we first purified and determined high-resolution crystal structures of both enzymes (**Figure 1.2A-B**). Both *Ec* CdnC and *Pa* CdnD adopt a polymerase β-like nucleotidyltransferase fold, with two lobes sandwiching a large active site. *Ec* CdnC and *Pa* CdnD adopt similar overall structures, with a 3.95 Å Cα r.m.s.d. (over 250 residues), and are more structurally related to one another than to either DncV (CD-NTase clade A) or CdnE, both of which synthesize cyclic dinucleotide second messengers (**Figure 1.2C**).

In *Ec* CdnC, the active site is occupied by one molecule of ATP that was co-purified with the protein (**Figure 1.2A**); we were unable to remove this tightly-bound nucleotide. We could purify *Pa* CdnD without a bound nucleotide, and found by isothermal titration calorimetry that the enzyme binds strongly to ATP, but not to GTP (**Figure A.A.S4A**). These data agree with our findings that both *Ec* CdnC and *Pa* CdnD primarily synthesize cAAA. We determined structures of *Pa* CdnD in both Apo and ATP-bound states, and found that the two states differ by a 3.9° rotation of the enzyme's N- and C-lobes with respect one another, with the ATP-bound state more "closed" than the Apo state (**Figure 1.2B**).

Our biochemical data show that *Ec* CdnC requires dsDNA for activation, similar to mammalian cGAS (**Figure 1.1C**). We modelled DNA binding to CdnC by overlaying our structure with a cGAS:DNA complex (Gao et al., 2013). In the resulting model, DNA contacts a surface of *Ec* CdnC that is strongly positively charged and highly conserved (**Figure 1.2D** and **Figure A.A.S4B**). To test the idea that this surface is responsible for DNA binding, we mutated several positively-charged residues on this surface of CdnC and measured both DNA binding and DNA-stimulated cAAA synthesis. The four mutants we tested had variable effects on DNA binding (**Figure A.A.S4C**), but all four eliminated cAAA synthesis by CdnC in the presence of HORMA+DNA (**Figure 1.2E**). These data support a role for this surface in DNA binding and activation of CdnC.

In mammalian cGAS, DNA binding to α-helices α7 and α8 restructures an "activation loop" in the enzyme's active site to promote substrate binding and second messenger synthesis (**Figure 1.2F**) (Civril et al., 2013; Kato et al., 2013; Li et al., 2013;

Zhang et al., 2014). In our structures of *Ec* CdnC and *Pa* CdnD, the conformation of the activation loop is equivalent to the activated state of cGAS (**Figure 1.2G-I**). Thus, while DNA binding is required for activation of *Ec* CdnC, the structural mechanism of activation is likely distinct from that of mammalian cGAS.

Pa CdnD does not require DNA for cAAA synthesis in vitro. Nonetheless, *Pa* CdnD shows a similar, albeit less pronounced, pattern of high conservation and positive charge on the surface equivalent to the cGAS DNA-binding surface (**Figure A.A.S4D-E**). We therefore tested DNA binding directly, but did not detect binding (**Figure A.A.S4C**). These data recall earlier findings with *V. cholerae* DncV, which possesses an equivalent conserved, positively charged surface, but whose synthesis of cGAMP is not stimulated by DNA (Kranzusch et al., 2014).

Bacterial HORMA domain proteins adopt canonical open and closed states

We next sought to determine whether the putative bacterial HORMA domain proteins share the canonical HORMA domain fold and mechanism of closure motif binding (**Figure 1.3A**). We first determined a structure of *Pa* HORMA2, and found that the protein adopts a canonical "closed" HORMA domain structure, with a central three-stranded β-sheet (β-strands 4, 5, and 6) backed by three α-helices (αA, B, and C) (**Figure 1.3B**). Characteristic of closed HORMA domains, the N-terminus of *Pa* HORMA2 is disordered, and the C-terminus wraps around the protein and forms a β-hairpin (β-strands 8′ and 8′′) against strand β5. Strikingly, the C-terminal safety belt of *Pa* HORMA2 embraces a peptide from the disordered N-terminus of a crystallographic symmetry-related molecule in the canonical closure motif binding site (**Figure 1.3B**).

Thus, *Pa* HORMA2 represents a *bona fide* HORMA domain with a closure-motif binding C-terminal safety belt.

To identify preferred closure motif sequences for bacterial HORMA proteins, we performed phage display with purified *Ec* HORMA, *Pa* HORMA2, and *Pa* HORMA3, and a library of 7-mer peptides. We identified consensus binding motifs enriched in hydrophobic amino acids for both *Pa* HORMA2 and *Ec* HORMA (**Figure A.A.S5A-B**), and confirmed binding by coexpression and purification of peptide-HORMA complexes (**Figure A.A.S5C-D**). Thus, both *Ec* HORMA and *Pa* HORMA2 can likely adopt the closed conformation and bind specific closure motif peptides.

To determine whether *Pa* HORMA2 can also adopt an "open" HORMA domain state, we generated a truncated construct lacking the C-terminal safety belt (*Pa* HORMA2-ΔC). The crystal structure of *Pa* HORMA2-ΔC shows that in contrast to full-length *Pa* HORMA2, the N-terminus of *Pa* HORMA2-ΔC is well-ordered and packs as a β-strand (β1) against strand β5 (**Figure 1.3C**). This transition of the N-terminus from disordered to an ordered β1 strand is characteristic of open HORMA domains (Rosenberg and Corbett, 2015). Thus, bacterial HORMA domain proteins can adopt both the open and closed states, and bind closure motif peptides equivalently to their eukaryotic relatives.

We next determined a structure of *Pa* HORMA3, and found that this protein also adopts an open state with an ordered N-terminal β1 strand (**Figure 1.3D**). Based on the sequence of the *Pa* HORMA3 N-terminus and the predicted disorder of its poorly-conserved C-terminus (**Figure 1.3D**), we propose that this protein may be "locked" in

the open state similarly to the eukaryotic HORMA domain protein Atg101 (Qi et al., 2015; Suzuki et al., 2015). Consistent with this idea, we were unable to identify *Pa* HORMA3-binding peptide sequences by phage display (not shown). Our data therefore suggests that in HORMA2/3 containing CBASS systems, HORMA2 can adopt the open and closed states and bind closure motif peptides, while HORMA3 adopts an open state and does not bind peptides.

We next determined a structure of a *Pa* HORMA3:HORMA2:peptide complex (Peptide 1: EVMEFNP). Strikingly, the *Pa* HORMA3:HORMA2 dimer bears no resemblance to dimers of eukaryotic HORMA domains, which assemble through their αC helices (Mapelli et al., 2007; Rosenberg and Corbett, 2015). Rather, *Pa* HORMA3 and *Pa* HORMA2 interact through both proteins' αA and αB helices, plus the short α-helix in the C-terminal region of *Pa* HORMA3 (**Figure 1.3E**). The interface does not involve any elements of *Pa* HORMA2 that change conformation significantly between the open and closed states, and yeast two-hybrid analysis shows that the open *Pa* HORMA2-ΔC construct can also bind *Pa* HORMA3 (**Figure A.A.S5E-G**). Thus, while *Pa* HORMA3 apparently does not bind closure motif peptides or undergo conformational conversion, it can likely scaffold larger complexes by binding *Pa* HORMA2 and CdnD.

HORMA proteins bind their cognate CD-NTases specifically in the closed state

We next sought to understand how bacterial HORMA domain proteins regulate their cognate CD-NTases. We first purified and determined the structure of a complex containing *Pa* CdnD, HORMA2, and Peptide 1, and found that *Pa* HORMA2 binds the C-lobe of CdnD (**Figure 1.4A**). Compared to its Apo state, HORMA2-bound *Pa* CdnD

shows a 6.6° rotation of the N-lobe toward the C-lobe, consistent with the idea that binding of HORMA2 may aid activation of CdnD. HORMA2 interacts with CdnD through its β8′ strand and αC helix (**Figure 1.4B**); as the β8′ strand does not form in the open state, this binding mode is likely dependent on HORMA2 adopting a closed conformation. Indeed, we found by yeast two-hybrid analysis that *Pa* HORMA2-ΔC, which adopts an open conformation, does not interact with *Pa* CdnD (**Figure 1.4C**). Consistent with our ability to reconstitute a CdnD:HORMA2:HORMA3 complex, we could model a structure of *Pa* HORMA2 bound to both CdnD and HORMA3 with no apparent clashes (**Figure A.A.S6A**).

We next purified and determined the structure of an *Ec* CdnC:HORMA-ΔN complex, which lacks the disordered N-terminal 11 residues of *Ec* HORMA. In this complex, the structure of *Ec* HORMA closely resembles *Pa* HORMA2 in the closed state, with a well-ordered C-terminal safety belt forming a β-hairpin (β8′-8′′) against strand β5 (**Figure 1.4D**). Notably, in this structure we find that *Ec* HORMA-ΔN is bound to a short peptide from the N-terminal His_6-tag of a symmetry-related HORMA-ΔN molecule, resulting in a CdnC:HORMA-ΔN complex with 2:2 stoichiometry (**Figure A.A.S6B**). Unlike our structure of *Ec* CdnC alone, in which the protein is bound to a single ATP molecule co-purified from *E. coli*, the two copies of *Ec* CdnC in the CdnC:HORMA-ΔN structure contain either no ligand or AMP in the active site (**Figure 1.4D**). This suggests that in comparison to isolated *Ec* CdnC, the CdnC:HORMA complex may be more conformationally flexible and enzymatically active.

The interaction of *Ec* HORMA with *Ec* CdnC closely resembles that of *Pa* HORMA2 with *Pa* CdnD (**Figure 1.4D**). *Ec* HORMA binds the CdnC C-lobe, adjacent to CdnC's putative DNA-binding surface. *Ec* HORMA binds CdnC through a similar interface as the one used by *Pa* HORMA2, binding CdnC largely through its αA helix (**Figure 1.4D-E**). *Ec* HORMA's β8′ strand does not contact CdnC, yet truncation of the protein's C-terminus (*Ec* HORMA-ΔC, lacking residues 141-172) disrupts CdnC binding, suggesting that the *Ec* HORMA-CdnC interaction likely also depends on the closed HORMA domain state (**Figure 1.4F**).

In our structure of CdnC:HORMA-ΔN, *Ec* HORMA binds a symmetry-related HORMA N-terminus as a short β-strand, but with an N-to-C polarity opposite that observed in our structure of *Pa* HORMA2 bound to Peptide 1 (and indeed opposite that of all known eukaryotic HORMA:closure motif complexes). We interpret this switched polarity of closure motif binding as an artifact of a non-biological complex assembly, rather than an indication that this HORMA domain binds closure motifs in an unique manner. We modelled binding of a consensus closure motif sequence identified by phage display (HGKILLT) with the canonical N-to-C polarity, revealing that the hydrophobic residues in the *Ec* HORMA-binding consensus would all fit into hydrophobic pockets on *Ec* HORMA (**Figure A.A.S6C**). This supports the idea that our phage display experiment identified sequences that bind *Ec* HORMA as a closure motif.

The observation of a 2:2 complex in the structure of *Ec* CdnC:HORMA-ΔN suggests that the 2:2 complex we observe when purifying *Ec* CdnC:HORMA may be assembled through reciprocal binding of HORMA N-termini to opposite HORMA

protomers. To test this idea, we purified an *Ec* CdnC:HORMA-ΔN complex in which the N-terminal His_6-tag was removed, and found that this complex forms only 1:1 complexes and is only minimally active (**Figure 1.4G** and **Figure A.A.S6D**). When we co-purified the same complex with a consensus closure motif peptide (HGKILLT), we found that it too forms a 1:1 CdnC:HORMA complex, but is now active in vitro (**Figure 1.4G** and **Figure A.A.S6D**). These data explain why in our earlier assays, only 2:2 CdnC:HORMA complexes were active for cAAA synthesis: these complexes' HORMA proteins are bound to closure motif-mimicking peptides, while the HORMA proteins in 1:1 complexes are unbound. Thus, these data support the idea that activation of *Ec* CdnC requires a closure motif-bound HORMA protein.

Structure of a bacterial Trip13 engaged with its CD-NTase:HORMA substrate

The Trip13-like ATPase encoded by bacterial HORMA-containing operons negatively regulates the production of cAAA by the HORMA-CD-NTase complex (**Figure 1.1C**). To address the structural mechanism of this activity, and to compare this mechanism to that of eukaryotic Pch2/TRIP3 in control of the eukaryotic HORMA domain protein Mad2, we determined the structures of both an isolated bacterial Trip13 hexamer (**Figure A.A.S7A-D**) and of an *Ec* CdnC:HORMA:$Trip13^{EQ}$ complex (**Figure 1.5A**; the Trip13 E159Q mutation stabilizes hexamer assembly and substrate interactions by disallowing ATP hydrolysis (Ye et al., 2017, 2015)). The crystal structure of *Ec* CdnC:HORMA:$Trip13^{EQ}$ reveals a mode of substrate binding and engagement for bacterial Trip13 that is similar to eukaryotic Pch2/TRIP13, with the high resolution of 2.6

Å providing unprecedented insight into substrate engagement and unfolding by AAA+ ATPases.

A number of recent studies have reported moderate-resolution (3.2-5.0 Å) cryo-electron microscopy structures of AAA+ ATPase remodelers engaged with their substrates, including mammalian TRIP13 bound to a p31^{comet}:Mad2 complex (Alfieri et al., 2018; Deville et al., 2017; Gates et al., 2017; Han et al., 2017; Puchades et al., 2017; Ripstein et al., 2017; White et al., 2018; Yu et al., 2018). These structures have revealed that active AAA+ remodelers adopt a right-handed spiral or "lock-washer" conformation, with 4-5 nucleotide-bound subunits forming a tight helical spiral and engaging an extended substrate peptide with their "pore loops". In the case of TRIP13, the structures revealed that the "adapter" protein p31^{comet} binds the top surface of the TRIP13 hexamer, thereby positioning the substrate Mad2 such that its extended N-terminus drapes into the hexamer pore for engagement and remodeling (Alfieri et al., 2018; Ye et al., 2017). In our structure of *Ec* CdnC:HORMA:Trip13^{EQ}, the *Ec* Trip13^{EQ} hexamer adopts a right-handed spiral conformation, with four subunits (monomers A-D) bound to ATP, and two subunit (monomers E and F) unbound (**Figure A.A.S8A-B**). The four ATP-bound subunits overlay closely, while monomers E and F show rotation of the large and small AAA subdomains with respect to one another (**Figure A.A.S8C**). Monomer F is positioned at the "bottom" of the spiral, and as such shows the most significant conformational differences compared to the other five subunits. Overall, the structure of the *Ec* Trip13^{EQ} hexamer closely resembles that of other substrate-engaged AAA+ ATPases, including mammalian TRIP13 (Alfieri et al., 2018).

As noted above, eukaryotic TRIP13 recognizes its HORMA domain substrate Mad2 through the adapter protein p31^{comet} (Alfieri et al., 2018; Ye et al., 2017). In the *Ec* CdnC:HORMA:Trip13^{EQ} complex, CdnC serves as the adapter, binding the top face of the Trip13 hexamer at the interface of monomers E and F (**Figure 1.5A**). CdnC-Trip13 binding positions *Ec* HORMA over the Trip13 hexamer pore, with its extended N-terminus draping into the pore and interacting with the pore loops from Trip13 monomers A-E (**Figure 1.5B**). These pore loops form a tight spiral and engage the extended HORMA N-terminus with two-residue periodicity: Trip13 pore loop 1 residues Gly119 and Val121 in monomers A-E engage even-numbered residues from HORMA (Ser2, Tyr4, Tyr6, Val8, and Glu10) through main-chain hydrogen bonding (**Figure 1.5C-D**). At the same time, conserved histidine residues (His173) in pore loop 2 of Trip13 monomers A-D form a "ladder" with each histidine side-chain forming two hydrogen bonds with the side-chains of serine/threonine residues at odd-numbered positions in the HORMA N-terminus (Ser3, Ser5, and Thr7) (**Figure 1.5C-D**). In keeping with these highly specific interactions, we observe a striking pattern of conservation in the N-termini of HORMA1 proteins (which includes *Ec* HORMA), with conserved serine/threonine residues in every second position (**Figure 1.5E**). This conservation pattern extends to *Ec* HORMA residue 17, strongly suggesting that *Ec* Trip13 can undergo at least 3-4 cycles of ATP hydrolysis and translocation before disengaging, sufficient to partially unfold helix αA and disrupt both HORMA-CdnC binding (which primarily involves helix αA) and HORMA-closure motif binding (by destabilizing the C-terminal safety belt). HORMA2 proteins do not possess this pattern of conserved

serine/threonine residues in their N-termini, and Trip13 proteins in HORMA2/3 operons lack a conserved pore loop 2 histidine residue (**Figure A.A.S7E-F**), suggesting a different mode of HORMA-Trip13 recognition in these operons. Thus, our structure of the *Ec* CdnC:HORMA:Trip13^{EQ} complex supports a model in which Trip13 disassembles a CD-NTase:HORMA complex by remodeling the HORMA domain protein, and provides the highest-resolution view to date of a AAA+ ATPase hexamer engaged with its substrate.

Our biochemical data shows that addition of Trip13 inactivates second messenger synthesis by *Ec* CdnC+HORMA+DNA, with Trip13 effectively inhibiting second messenger synthesis at a 1:60 ratio of Trip13 hexamer to CdnC+HORMA (**Figure 1.1C**, sample vi). This data is consistent with the idea that Trip13 catalytically disassembles and inactivates the CdnC:HORMA complex, likely through engagement and unfolding of the HORMA N-terminus. In agreement with this idea, we found that Trip13 is unable to effectively inhibit second messenger synthesis when ATP binding is disrupted by a mutation to the Walker A motif (K87A; **Figure 1.5F**, sample iii). Further, Trip13 is unable to inhibit second messenger synthesis by an *Ec* CdnC:HORMA-ΔN complex, which lacks the extended N-terminus of *Ec* HORMA and thus cannot be engaged by Trip13 (**Figure 1.5F**, samples iv-v). These data support a model in which bacterial Trip13 enzymes attenuate signaling by their cognate CD-NTases by disassembling an active CD-NTase:HORMA complex.

NucC is a cAAA-activated DNA endonuclease

Bacterial CBASS systems encode one of several putative effector proteins, which include proteases, phospholipases, and at least three families of putative endo- and exonucleases (Burroughs et al., 2015). Both the *E. coli* MS115-1 and *P. aeruginosa* ATCC27853 systems encode a restriction endonuclease-related protein we term NucC (Nuclease, CD-NTase associated). We purified *E. coli* MS115-1 NucC and tested its ability to degrade purified plasmid DNA in the presence of second messengers including cyclic and linear di-AMP, and the cyclic trinucleotides cAAA and cAAG. We found that NucC degrades DNA to ~50-100 bp fragments in the presence of low-nanomolar concentrations of cAAA, and is also activated to a lesser extent by cyclic di-AMP and cAAG (**Figure 1.6A**). NucC is apparently insensitive to DNA methylation: the enzyme is equally active on unmethylated PCR product, plasmid purified from a K-12 based *E. coli* strain (NovaBlue), and plasmid purified from strain MS115-1, which harbors the CBASS system and several predicted DNA methylases not found in K-12 strains (**Figure A.A.S9A**). Thus, NucC is DNA endonuclease activated by cAAA, the second messenger product of its cognate CD-NTases.

The *E. coli* MS115-1 CBASS system confers bacteriophage immunity through an abortive infection mechanism

Recent work has demonstrated that two CBASS systems with different architectures provide immunity against a range of bacteriophages, acting through an abortive infection mechanism to kill cells prior to the completion of phage replication (Cohen et al., 2019). Our structural and biochemical data suggest that in HORMA+Trip13 containing CBASS systems, HORMA domain proteins recognize

specific peptides from a bacteriophage, then bind and activate their cognate CD-NTase to produce cAAA, which in turn activates a defensive response by the effector nuclease NucC. We tested the resistance of *E. coli* MS115-1 to bacteriophage λ*cI-* (lacking the lysogeny gene *cI*), and found that in contrast to a laboratory strain of *E. coli* (JP313; (Economou et al., 1995)), *E. coli* MS115-1 is immune to bacteriophage λ*cI-* infection as judged by plaque formation (**Figure 1.6B-C**). We next inserted the entire *E. coli* MS115-1 operon (including CdnC, HORMA, Trip13, and NucC) into *E. coli* JP313 on a plasmid vector, and found that the operon confers robust immunity to bacteriophage λ*cI-* (**Figure 1.6B-C**). Immunity was compromised when we disrupted second messenger synthesis by CdnC (CdnC D72N/D74N mutant), HORMA-CdnC binding (HORMA-ΔC or R32A/R35A mutants), or NucC's DNA cleavage activity (NucC D73N mutant; based on sequence alignments with type II restriction endonucleases), demonstrating that immunity depends on both the sensor and effector modules in this pathway (**Figure 1.6B-C**). Disruption of Trip13's ATP hydrolysis activity (Trip13^{EQ} mutant) did not compromise immunity, consistent with the idea that Trip13 is a negative regulator of this pathway (**Figure 1.6B-C**).

To determine whether the *E. coli* MS115-1 CBASS system functions by abortive infection, we tracked the growth of bacterial cultures infected at high multiplicity of infection (MOI) with bacteriophage λ*cI-*. In cells lacking the CBASS operon infected at an MOI of 2.5, we observed complete culture collapse beginning ~90 minutes post-infection due to phage-induced cell lysis (**Figure 1.6D**). In contrast, cells harboring the CBASS operon showed culture collapse beginning ~60 minutes post-infection, fully 30

minutes prior to phage-induced cell lysis. At an MOI of 0.25, we observed a sharp drop in growth rate at ~60 minutes post-infection in CBASS-harboring cells, followed by recovery of growth, while cells lacking CBASS showed culture collapse beginning at ~90 minutes post-infection (**Figure 1.6E**). Introduction of the NucC D73N active-site mutation eliminated these effects (**Figure 1.6D-E**). Together, these data support the idea that the *E. coli* MS115-1 CBASS system functions by abortive infection, causing the death of infected cells prior to the completion of phage replication. The requirement for NucC further suggests that cell death is caused by NucC-mediated host cell genome destruction.

Discussion

The recent discovery and classification of a diverse family of cGAS/DncV-like nucleotidyltransferases (CD-NTases) that are sparsely distributed among important environmental and pathogenic bacteria raised many important questions about these proteins' molecular mechanisms and the biological functions of their associated operons (Burroughs et al., 2015; Cohen et al., 2019; Whiteley et al., 2019). Two such operons have been shown to confer immunity against bacteriophage infection, acting through an abortive infection mechanism in which infected cells die prior to completion of the bacteriophage replication cycle, and are now termed CBASS (Cyclic oligonucleotide-Based Anti-phage Signaling System) (Cohen et al., 2019). Here, we address the molecular mechanisms of CBASS defense systems and in particular define the mechanisms of a CBASS family encoding eukaryotic-like HORMA domain proteins and Trip13-like ATPase regulators. We demonstrate CD-NTase activation by binding of a HORMA-closure motif complex, identify the second messenger product of these enzymes, and show that second messenger production activates the effector nuclease NucC. Together, these proteins constitute a pathway that confers robust bacteriophage immunity onto its bacterial host through an abortive infection mechanism.

Our data support a model for HORMA/Trip13 regulation of CD-NTase activity in bacteria that is remarkably reminiscent of the roles of MAD2 and Pch2/TRIP13 in the eukaryotic spindle assembly checkpoint. In this checkpoint pathway, the HORMA domain protein MAD2 is kept in its inactive open state through constant conformational conversion by TRIP13 (Kim et al., 2018; Ma and Poon, 2018, 2016). Upon entrance into

mitosis, large-scale assembly of MAD2 into a “mitotic checkpoint complex” with the closure motif-bearing protein CDC20 overwhelms the disassembly capacity of TRIP13, resulting in checkpoint activation. After successful kinetochore-microtubule attachment and cessation of mitotic checkpoint complex assembly, the checkpoint is inactivated as existing complexes are disassembled by TRIP13 and an alternative pathway (Kim et al., 2018). In a similar manner, we propose that in the absence of a foreign threat, bacterial HORMA domain proteins are maintained in the inactive “open” state by continual activity of Trip13 (**Figure 1.7**, step 1), thereby inhibiting CD-NTase activation and second messenger synthesis. Maintenance of the HORMA proteins in the open state by Trip13 primes them for binding to specific bacteriophage proteins upon infection or lysogenic induction. Following binding and conversion to the closed state, the HORMA domain protein binds and activates its cognate CD-NTase to synthesize cAAA (**Figure 1.7**, step 2). cAAA in turn activates the effector nuclease NucC, which indiscriminately destroys both the invading phage genome and the bacterial host genome, resulting in cell death and abortive infection (**Figure 1.7**, step 3). In related work, we demonstrate how NucC is allosterically activated by cAAA binding, and confirm that NucC destroys cellular DNA in infected cells to cause cell death (Lau et al., 2019).

While we demonstrate the structural mechanisms of bacterial HORMA domain protein-mediated CD-NTase activation, important questions remain regarding these proteins’ evolution and function. First, it will be important to identify the closure motif-containing proteins recognized by bacterial HORMA domain proteins. The consensus closure motif sequences we identified by phage display do not allow clear identification of likely

targets by sequence searches, so more direct tests will be required to identify these targets. The operon's utility to bacterial survival would be maximized if the HORMA proteins were to recognize a widely-conserved bacteriophage protein, rather than a protein that is only found in a few bacteriophages. It would also be beneficial to recognize phage infection as early as possible, potentially by recognizing initial invasion through tail spike detection, or by recognizing a protein that is transcribed and translated early in the phage infection cycle.

A key question is the evolutionary origin of bacterial HORMA and Trip13 proteins, and their relationship with eukaryotic HORMA domain proteins and Pch2/TRIP13 ATPases. A recent bioinformatics analysis of bacterial and eukaryotic HORMA proteins showed that the bacterial proteins fall into two major families (HORMA1 and HORMA2), and that these two families co-occur with diverged Trip13 ATPases (Tromer et al., 2019). This study also showed that eukaryotic HORMA domain proteins – including Mad2, p31^{comet}, Rev7, Atg13/101, and the meiotic HORMAD proteins – form a monophyletic group within the bacterial HORMA2 family, and that eukaryotic Pch2/TRIP13 ATPases form a monophyletic group within HORMA2-associated bacterial Trip13 proteins (**Figure A.A.S8G**) (Tromer et al., 2019). These findings indicate that HORMA domain proteins and Trip13 ATPases originated in bacteria, and their presence in CBASS systems indicates an ancestral function in protein sensing in bacteriophage defense.

Our structural data have enabled us to identify a third family of bacterial HORMA protein, which we term HORMA3, that is found in CBASS systems alongside HORMA2

proteins. We find that these proteins are likely “locked” in an open state, and form heterodimeric complexes with their cognate HORMA2 proteins. As HORMA2 binds closure motif peptides and can adopt both open and closed conformations, this protein likely functions as the peptide sensor in HORMA2/3 operons. The biological role of HORMA3 remains unknown, but it may function either to stabilize HORMA2 or scaffold larger signaling complexes.

Well-characterized CD-NTases including bacterial DncV and mammalian cGAS catalyze cyclic dinucleotide synthesis by binding two nucleotides, and catalyzing nucleophilic attack of each nucleotide’s 2′ or 3′ ribose oxygen by the other nucleotide’s α-phosphate. OAS proteins, on the other hand, synthesize linear 2′-5′ linked oligoadenylates through a similar mechanism that likely involves a single nucleophilic attack, followed by translocation and binding/attack by another ATP (Lohöfener et al., 2015). In this context, a key unanswered question is how bacterial CD-NTases in clades C and D – including *Ec* CdnC, *Pa* CdnD, and *E. cloacae* CdnD02 (Whiteley et al., 2019) – catalyze specific synthesis of cyclic trinucleotide second messengers. The enzymes’ active sites likely cannot accommodate three nucleotides, necessitating translocation of a linear intermediate during synthesis prior to the final cyclization reaction. Both *Ec* CdnC and *Pa* CdnD synthesize cAA as a minor product, suggesting that the enzymes can cyclize both 2-base or 3-base linear intermediates. Additional studies will be required to understand the mechanism and specificity of this reaction.

Here, we have defined the molecular mechanisms and biological role of a class of bacterial CBASS systems with associated HORMA and Trip13-like regulators and the

nuclease effector NucC. Bacterial CBASS systems are extraordinarily diverse, however, with multiple regulatory/sensor systems including the HORMA/Trip13 systems described here and a putative ubiquitin-conjugation system, and a variety of putative effectors including nucleases, proteases, and phospholipases (Burroughs et al., 2015). While we have demonstrated that HORMA+TRIP13-containing CBASS systems likely sense phage proteins through their HORMA domain proteins, how other CBASS systems sense a bacteriophage infection, and whether all CBASS systems act via abortive infection remain key open questions. Finally, more work will be required to determine whether all CBASS systems function in bacteriophage defense or whether some have evolved other functions, protecting bacteria against other threats, contributing to homeostasis of bacterial communities, or even contributing to pathogenicity in unanticipated ways.

Acknowledgements

The authors thank the staffs of the Stanford Synchrotron Radiation Lightsource and the Advanced Photon Source NE-CAT beamlines for assistance with crystallographic data collection; A. Bobkov (Sanford Burnham Prebys Medical Discovery Institute, Protein Analysis Core) for assistance with isothermal titration calorimetry; M. Raffatellu for valuable advice on bacterial genetics and experimental design; E. Tromer for input on HORMA/Trip13 sequence analysis; and A. Desai, M. Daugherty, and P. Kranzusch for critical reading and helpful suggestions. RKL was supported by the UC San Diego Quantitative and Integrative Biology training grant (NIH T32 GM127235).

ITM was supported by NIH T32 HL007444, T32 GM007752 and F31 CA236405. JDW was supported by NIH K01 DK116917 and P30 DK063491. MJ was supported by NIH S10 OD020025 and R01 ES027595. JP was supported by NIH R01 GM129245. KDC was supported by the Ludwig Institute for Cancer Research and the University of California, San Diego.

Author contributions

Conceptualization, QY and KDC; Methodology, QY, ITM, JDW, MJ, KDC; Investigation, QY, RKL, ITM, EB, JDW, CSA, KDC; Writing – Original Draft, QY, RKL, KDC; Writing – Review & Editing, QY, RKL, MJ, KDC; Visualization, QY, ITM, JDW, KDC; Supervision, MJ and KDC; Project Administration, KDC; Funding Acquisition, JP, MJ and KDC.

Declaration of Interests

The authors declare no competing interests.

Figure 1.3 HORMA+Trip13-associated CD-NTases synthesize cAAA

(A) Schematic of CD-NTase+HORMA+Trip13 operons from *E. coli* MS115-1 (top) and *P. aeruginosa* ATCC27853 (bottom). See **Figure A.A.S1** for phylogenetic analysis of Trip13 and HORMA proteins. See **Table A.A.S4** for protein sequences. **(B)** Summary of yeast two-hybrid and three-hybrid assays with the *E. coli* MS115-1 CdnC, HORMA, and Trip13. See **Figure A.A.S2A** for yeast two-hybrid and three-hybrid results, **Figure A.A.S2B** for purification and stoichiometry of a CdnC:HORMA complex, and **Figure A.A.S2C** for purification of a CdnC:HORMA:Trip13^{EQ} complex. See **Figure A.A.S2D-G** for equivalent assays with the *P. aeruginosa* ATCC27853 operon. **(C)** *Top:* Schematic of second messenger synthesis assays. *Ec* CdnC was incubated with potential regulators (HORMA, plasmid DNA, and Trip13) plus ATP, then the products were separated by anion-exchange chromatography and analyzed by liquid chromatography with tandem mass spectrometry (LC-MS/MS). *Bottom:* Anion exchange elution profiles from second messenger synthesis assays using *Ec* CdnC, HORMA, DNA, and Trip13. Blue lines show absorbance at 254 nm, and red lines show conductivity for the 0.2-2.0 M ammonium acetate gradient elution. CdnC D72N/D74N contains aspartate-to-asparagine mutations in the putative active-site residues 72 and 74. See **Figure A.A.S3A** for assays with *Ec* CdnC and different DNAs, and **Figure A.A.S3B** for equivalent assays with *Pa* CdnD. **(D)** Liquid chromatography elution profile of the major product of *Ec* CdnC (peak 2 from panel C, sample v), with measured MS1 m/z and theoretical mass of cAAA. **(E)** MS2 fragmentation spectrum of the major product of *Ec* CdnC, annotated according to expected fragments of cAAA. The m/z of the extracted ion (corresponding to the [M+H] adduct of cAAA) was 988.1648, consistent with cAAA (monoisotopic mass = 987.1576 amu; molecular weight = 987.6263 amu). The minor product of *Ec* CdnC (peak 1 from panel C, sample v) was also analyzed by LC MS/MS and confirmed to be cAA (not shown). **(F)** Model for CdnC activation by HORMA and DNA binding, with negative regulation by Trip13.

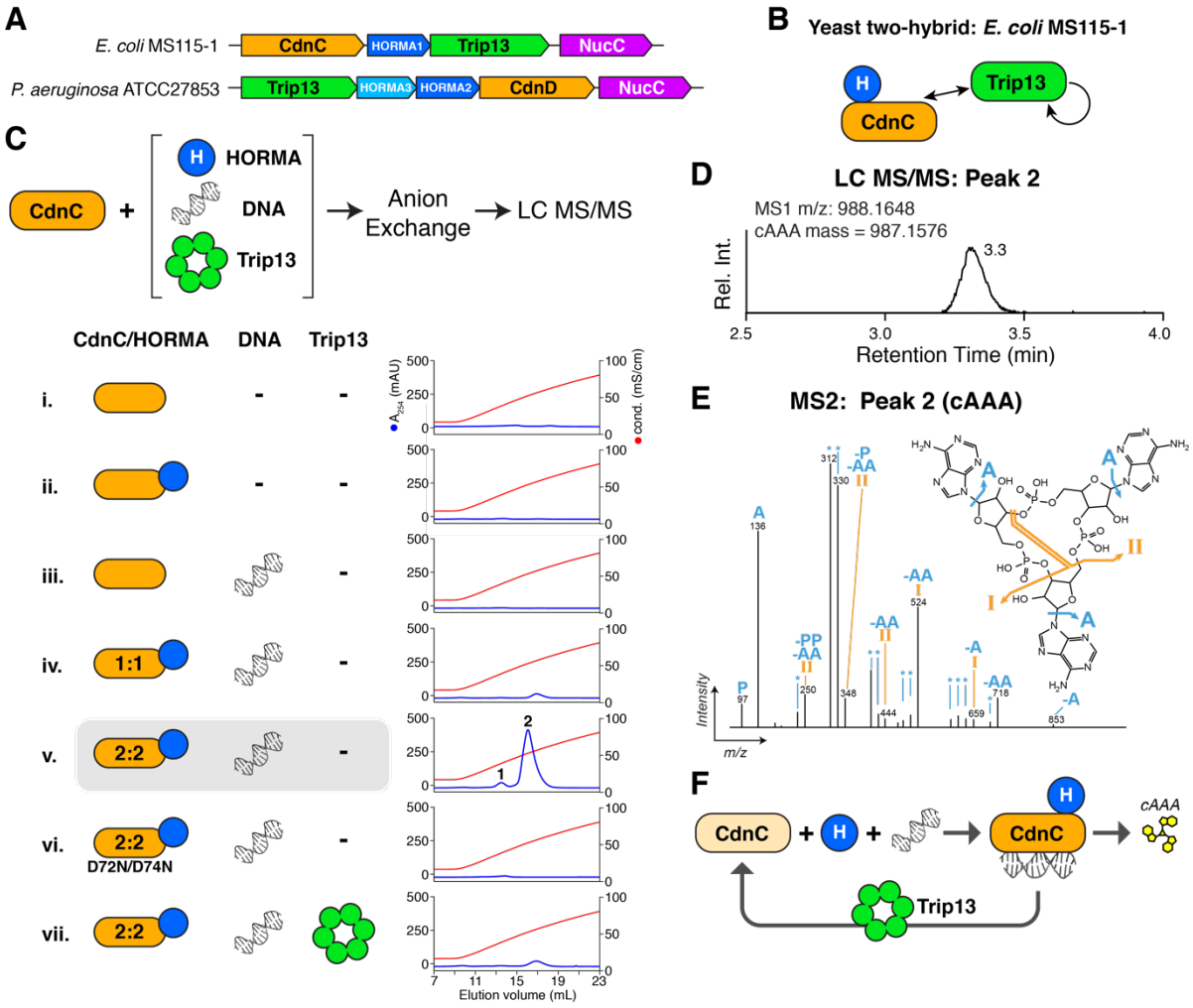

A
E. coli MS115-1
CdnC
HORMA1
Trip13
NucC
P. aeruginosa ATCC27853
Trip13
HORMA3
HORMA2
CdnD
NucC
B
Yeast two-hybrid: E. coli MS115-1
H
CdnC
Trip13
C
CdnC
+
H
HORMA
DNA
Trip13
Anion Exchange
LC MS/MS
CdnC/HORMA
DNA
Trip13
i.
ii.
iii.
iv.
1:1
v.
2:2
vi.
2:2
D72N/D74N
vii.
2:2
A254 (mAU)
cond. (mS/cm)
Elution volume (mL)
D
LC MS/MS: Peak 2
MS1 m/z: 988.1648
cAAA mass = 987.1576
Rel. Int.
Retention Time (min)
E
MS2: Peak 2 (cAAA)
Intensity
m/z
F
CdnC
H
CdnC
cAAA
Trip13

Figure 1.4 Structures of HORMA+Trip13-associated CD-NTases

(A) Structure of *Ec* CdnC, with bound ATP·Mg^{2+} shown as sticks. **(B)** Overlay of *Pa* CdnD in the Apo state (gray) and bound to ATP (orange), with bound ATP·Mg^{2+} shown as sticks. *Right:* Closeup view of ATP·Mg^{2+} binding to *Pa* CdnD. See **Figure A.A.S4A** for ITC assays measuring nucleotide binding to *Pa* CdnD. **(C)** Schematic of structural similarity in Pol-β type nucleotidyltransferases. CD-NTases are shaded in pink, and bacterial CD-NTases are shaded in orange. **(D)** Reverse view of *Ec* CdnC, showing surface charge and DNA (gray) modelled from a superposition with *M. musculus* cGAS bound to DNA (PDB 4K9B) (Gao et al., 2013). For surface conservation of *Ec* CdnC, see **Figure A.A.S4B**. For surface conservation and charge of *Pa* CdnD, see **Figure A.A.S4C-D**. Closeup view (right) shows positively-charged residues, with helices α7 and α8 (equivalent to mammalian cGAS DNA-binding surface) labeled. **(E)** Anion exchange elution profiles from second messenger synthesis assays with wild-type *Ec* CdnC (sample i) and mutants to the putative DNA-binding surface (samples ii-v). **(F)** Overlay of inactive *H. sapiens* cGAS (gray; PDB ID 4O69 (Zhang et al., 2014)) with active, DNA-bound *M. musculus* cGAS (pink; PDB ID 4O6A (Zhang et al., 2014)) showing motion of the cGAS activation loop upon DNA binding. **(G)** Overlay of ATP-bound *Ec* CdnC (orange) with active cGAS (pink). **(H)** Overlay of Apo *Pa* CdnD (gray) with active cGAS (pink). **(I)** Overlay of ATP-bound *Pa* CdnD (orange) with active cGAS (pink).

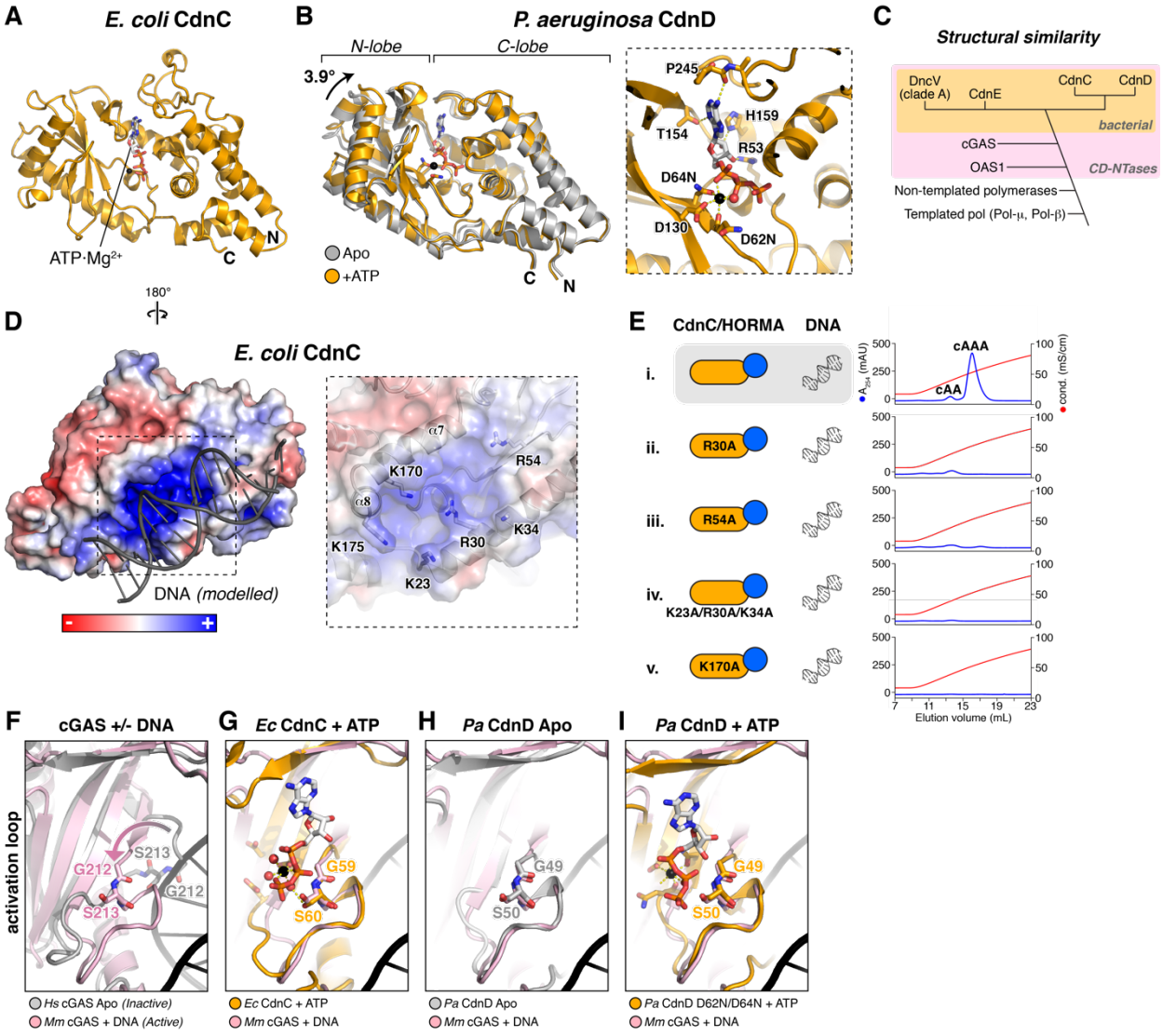

A
E. coli CdnC
ATP·Mg2+
N
C
B
P. aeruginosa CdnD
N-lobe
C-lobe
3.9°
Apo
+ATP
P245
H159
T154
R53
D64N
D130
D62N
C
N
C
Structural similarity
DncV (clade A)
CdnE
CdnC
CdnD
bacterial
cGAS
OAS1
CD-NTases
Non-templated polymerases
Templated pol (Pol-μ, Pol-β)
180°
D
E. coli CdnC
DNA (modelled)
-
+
α7
K170
R54
α8
K175
R30
K34
K23
E
CdnC/HORMA
DNA
i.
ii.
R30A
iii.
R54A
iv.
K23A/R30A/K34A
v.
K170A
cAAA
cAA
A260 (mAU)
cond. (mS/cm)
Elution volume (mL)
F
cGAS +/- DNA
activation loop
S213
G212
G212
S213
Hs cGAS Apo (Inactive)
Mm cGAS + DNA (Active)
G
Ec CdnC + ATP
G59
S60
Ec CdnC + ATP
Mm cGAS + DNA
H
Pa CdnD Apo
G49
S50
Pa CdnD Apo
Mm cGAS + DNA
I
Pa CdnD + ATP
G49
S50
Pa CdnD D62N/D64N + ATP
Mm cGAS + DNA

Figure 1.5 Structure and closure motif binding of bacterial HORMA proteins

(A) *Top:* Schematic of HORMA domain primary structure, with the HORMA domain core gray, N-terminus blue, and C-terminal safety belt red. *Bottom:* Schematic of the open and closed states of the canonical HORMA domain protein MAD2. The transition from open to closed involves movement of both the N-terminus (blue) and C-terminal safety belt (red), and binding of a closure motif peptide (yellow) by the safety belt. **(B)** *Left:* Structure of *Pa* HORMA2 in the closed state, colored as in panel (A). Secondary-structure elements are labeled as in MAD2, except for strand β5′, which is not observed in other HORMA domain proteins. *Right:* Closeup view of *Pa* HORMA2 binding the extended N-terminus of a symmetry-related *Pa* HORMA2 molecule, which mimics a closure motif (yellow). **(C)** Structure of *Pa* HORMA2-ΔC (lacking residues 134-166) in the open state. **(D)** *Top:* Schematic of *Pa* HORMA3, with plot showing the Jalview alignment conservation score (3-point smoothed; gray) (Livingstone and Barton, 1993) and DISOPRED3 disorder propensity (red) for aligned bacterial HORMA3 proteins (Jones and Cozzetto, 2015). *Bottom:* Structure of *Pa* HORMA3 in the open state. **(E)** Structure of the *Pa* HORMA3:HORMA2:Peptide 1 complex. See **Figure A.A.S5** for identification of closure motif sequences for *Pa* HORMA2 and *Ec* HORMA by phage display.

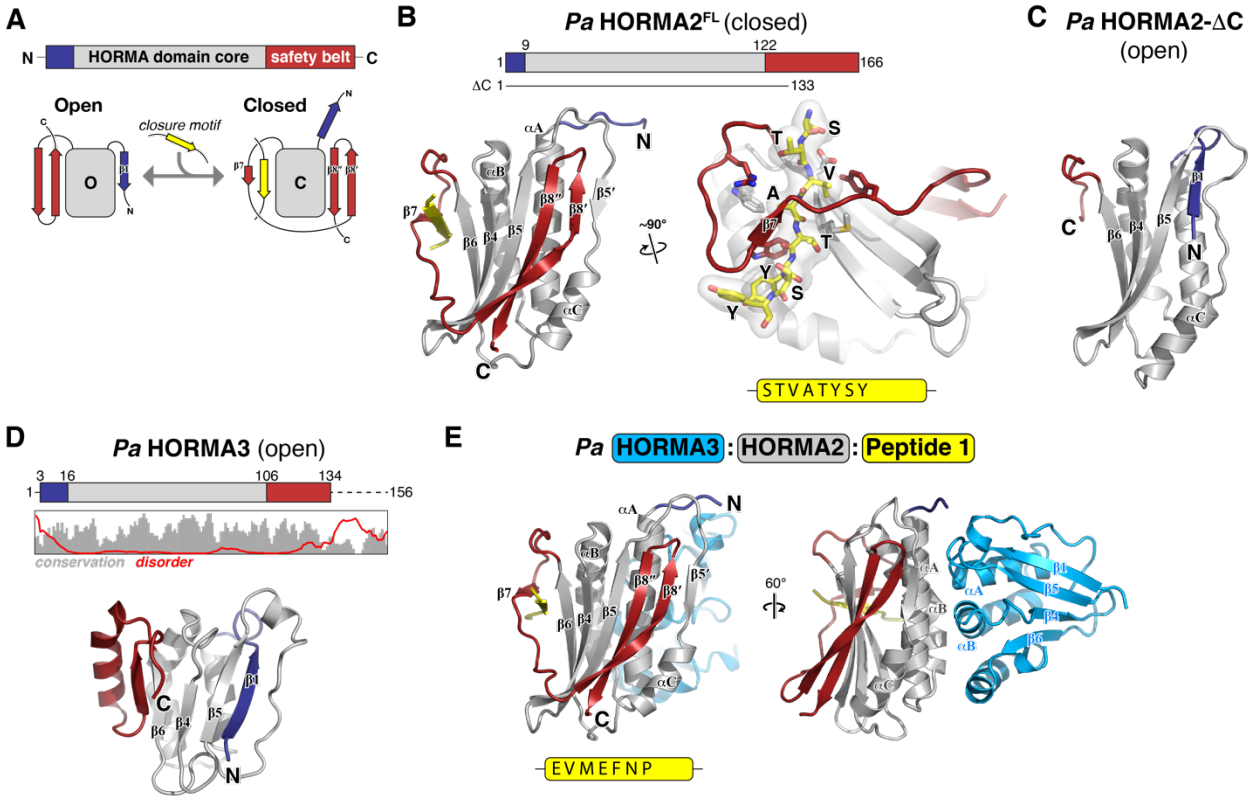

A
N
HORMA domain core
safety belt
C
Open
closure motif
Closed
O
C
B
Pa HORMA2FL (closed)
1
9
122
166
ΔC 1
133
αA
N
αB
β7
β8″
β8′
β5′
β6
β4
β5
αC
C
~90°
T
S
V
A
T
Y
S
Y
STVATYSY
C
Pa HORMA2-ΔC
(open)
β1
β6
β4
β5
N
αC
D
Pa HORMA3 (open)
3
16
106
134
1
156
conservation
disorder
C
β1
β5
β4
β6
N
E
Pa HORMA3 : HORMA2 : Peptide 1
αA
N
αB
β7
β8″
β8′
β5′
β6
β4
β5
αC
C
EVMEFNP
60°
αA
αB
αC
β1
β5
β4
β6
αB

Figure 1.6 Structures of CD-NTase:HORMA complexes

(A) Overall structure of the *Pa* CdnD:HORMA2:Peptide 1 complex, with CdnD orange, HORMA2 gray with safety-belt red and N-terminus blue, and Peptide 1 yellow. CdnD is overlaid with Apo CdnD (gray), showing the 6.6° N-lobe rotation in the HORMA2-bound state. **(B)** Closeup view of the *Pa* CdnD:HORMA2 interface. **(C)** Yeast two-hybrid analysis of the *Pa* CdnD-HORMA2 interaction. BD: fusion to the Gal4 DNA-binding domain; AD: fusion to the Gal4 activation domain. -LT: media lacking leucine and tryptophan (non-selective); -LTHA: media lacking leucine, tryptophan, histidine, and adenine (stringent selection). **(D)** Overall structure of the *Ec* CdnC:HORMAΔN complex, with CdnC orange, HORMA gray with safety-belt red and N-terminus blue, and bound peptide yellow (see **Figure A.A.S6B**). **(E)** Closeup view of the Ec CdnC:HORMA interface. **(F)** Yeast two-hybrid analysis of the *Ec* CdnC:HORMA interaction. **(G)** Anion exchange elution profiles from second messenger synthesis assays with separately-purified 1:1 (sample i) and 2:2 (sample ii) samples of the full-length *Ec* CdnC:HORMA complex (with intact His_6-tag on HORMA, same as **Figure 1.1C**, samples iv and v.) and *Ec* CdnC:HORMA-ΔN with His6-tag removed, without addition of peptide (sample iii) or with addition of consensus HORMA-binding closure motif peptide HGKILLT (sample iv). See **Figure A.A.S6D** for size exclusion elution profiles of *Ec* CdnC:HORMA-ΔN. Residual activity in sample iii is likely due to incomplete dissociation of the cleaved His_6-tag from HORMA-ΔN.

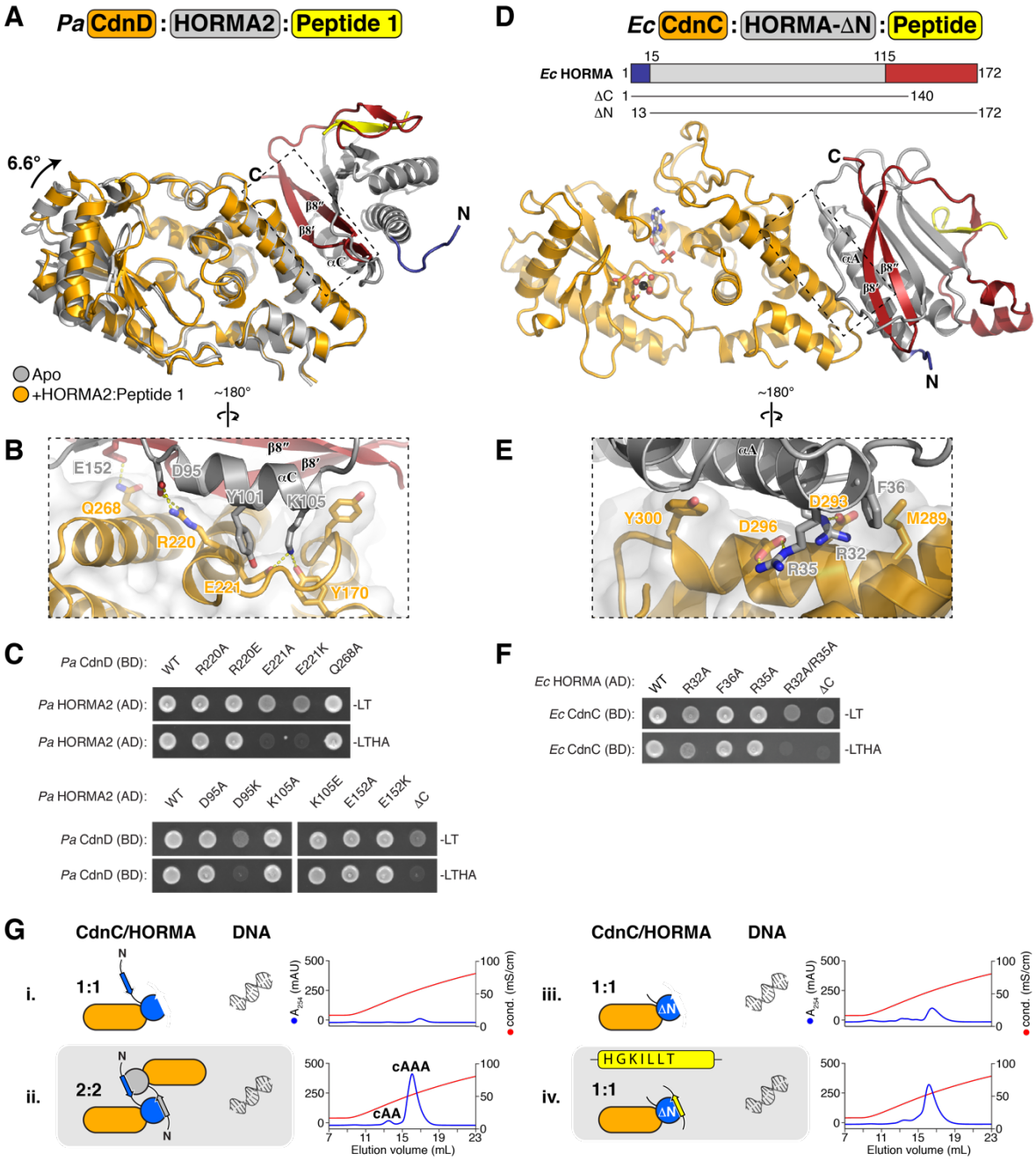
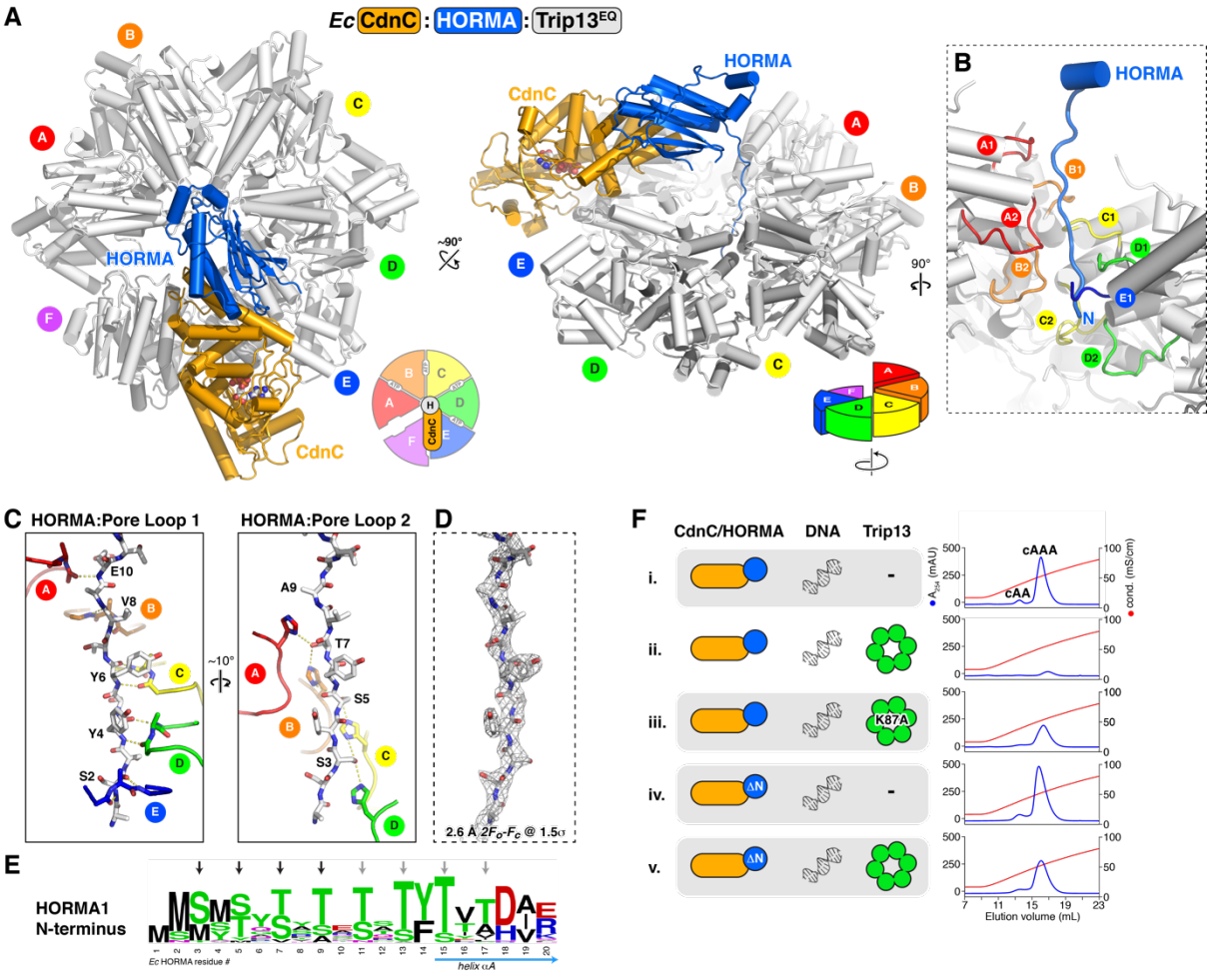

A
Pa CdnD : HORMA2 : Peptide 1
6.6°
C
N
β8″
β8′
αC
Apo
+HORMA2:Peptide 1
~180°
D
Ec CdnC : HORMA-ΔN : Peptide
Ec HORMA 1
15
115
172
ΔC 1
140
ΔN 13
172
C
αA
β8″
β8′
N
~180°
B
E152
D95
β8″
β8′
αC
Y101
K105
Q268
R220
E221
Y170
E
αA
F36
D293
Y300
D296
M289
R32
R35
C
Pa CdnD (BD):
WT
R220A
R220E
E221A
E221K
Q268A
Pa HORMA2 (AD):
-LT
-LTHA
Pa HORMA2 (AD):
WT
D95A
D95K
K105A
K105E
E152A
E152K
ΔC
Pa CdnD (BD):
-LT
-LTHA
F
Ec HORMA (AD):
WT
R32A
F36A
R35A
R32A/R35A
ΔC
Ec CdnC (BD):
-LT
-LTHA
G
CdnC/HORMA
DNA
i.
1:1
N
ii.
2:2
N
N
cAAA
cAA
A260 (mAU)
cond. (mS/cm)
Elution volume (mL)
iii.
1:1
ΔN
iv.
HGKILLT
1:1
ΔN

Figure 1.7 Structure of the *Ec* CdnC:HORMA:Trip13^{EQ} complex

(A) Top and side views of the *Ec* CdnC:HORMA:Trip13^{EQ} complex, with CdnC orange, HORMA blue, and Trip13^{EQ} white. See **Figure A.A.S8** for structural analysis of Trip13^{EQ}. Trip13 forms a spiral conformation, with its A subunit at top and F subunit at bottom (see schematics). CdnC binds the top surface of Trip13 subunits E and F (see **Figure A.A.S8B** for buried surface area). See **Figure A.A.S6E-G** for yeast two-hybrid analysis of the CdnC-Trip13 and HORMA-TRIP13 interfaces. **(B)** Closeup view of the *Ec* HORMA N-terminus (blue) engaged by Trip13 pore loops 1 (upper) and 2 (lower) from chains A-E. **(C)** Detail views of interactions between the *Ec* HORMA N-terminus (white sticks) with Trip13 pore loop 1 (left) and pore loop 2 (right). HORMA residues Ser4, Tyr4, Tyr6, Val8, and Glu10 form an extended backbone hydrogen-bonding network with Trip13 pore loop 1 residues Gly119 (main-chain carbonyl) and Val121 (main-chain amine) from monomers A-E (Trip13 Arg120 side-chains are not shown for clarity). HORMA residues Ser3, Ser5, Thr7, and Ala9 form an extended hydrogen-bond network with Trip13 pore loop 2 residue His173 from monomers A-D. **(D)** View equivalent to panel (C) showing 2*Fo-Fc* electron density for the HORMA N-terminus at 1.5 σ (2.6 Å resolution). **(E)** Sequence logo showing conservation of the N-termini of HORMA1 proteins. See **Figure A.A.S7E-F** for analysis of Trip13 pore loop 1 and 2 conservation, and equivalent logos of Trip13 and HORMA from two-HORMA operons. **(F)** Anion exchange elution profiles from second messenger synthesis assays with wild-type proteins (samples i-ii), Trip13 Walker A mutant K87A (sample iii), and HORMA-ΔN (missing N-terminal residues 1-12; samples iv-v). See **Figure A.A.S3B** for equivalent assays with *Pa* CdnD and Trip13 mutants.


A
Ec CdnC:HORMA:Trip13EQ
HORMA
CdnC
~90°
90°
B
HORMA
A1
B1
A2
C1
D1
B2
E1
C2
N
D2
C HORMA:Pore Loop 1
HORMA:Pore Loop 2
E10
V8
Y6
Y4
S2
~10°
A9
T7
S5
S3
D
2.6 Å 2Fo-Fc @ 1.5σ
E
HORMA1
N-terminus
Ec HORMA residue #
helix αA
F
CdnC/HORMA
DNA
Trip13
i.
ii.
iii.
iv.
v.
K87A
ΔN
cAAA
cAA
A280 (mAU)
cond. (mS/cm)
Elution volume (mL)

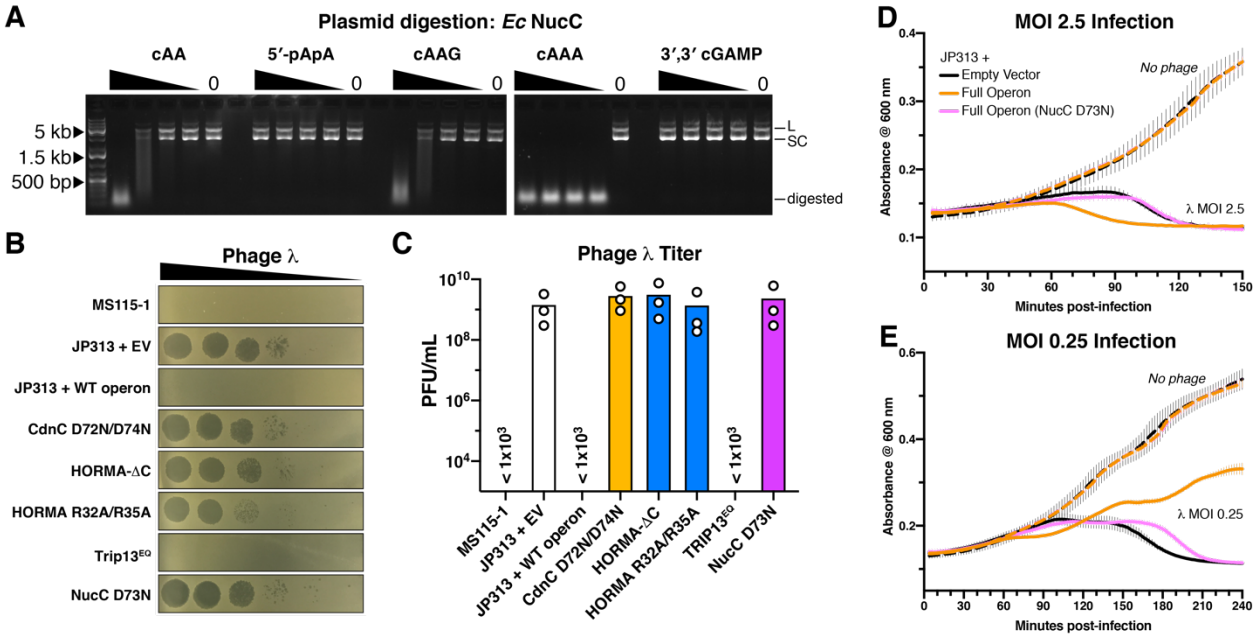


Figure 1.8 The *E. coli* MS115-1 CBASS system confers bacteriophage immunity through activation of a DNA endonuclease

(A) Plasmid digestion assay with *Ec* NucC (10 nM) and the indicated second messenger molecules at 400/100/25/6.25/0 nM. L: linear plasmid, SC: supercoiled plasmid. **(B)** Dilution of phage λ on *E. coli* MS115-1 (top) and JP313 (wild-type laboratory strain) with the plasmid-encoded *E. coli* MS115-1 CBASS system with wild-type proteins (WT) or the indicated mutations. EV: empty vector. Six 10-fold phage dilutions are shown. **(C)** Quantitation of phage λ infectivity. Shown is the average +/- standard deviation of three trials at a single bacteriophage dilution. The three strains marked "< $1x10^3$" showed no plaques with the highest-tested bacteriophage concentration. Bars represent mean of three individual measurements (white circles). **(D)** Growth curves for *E. coli* JP313 transformed with the indicated plasmids and infected at 0 minutes with bacteriophage λ at an MOI of 2.45 (solid lines) or without phage addition (dashed lines). Lines represent the mean of four replicates, and error bars indicate standard deviation. **(E)** Growth curves of cell cultures infected at an MOI of 0.245.

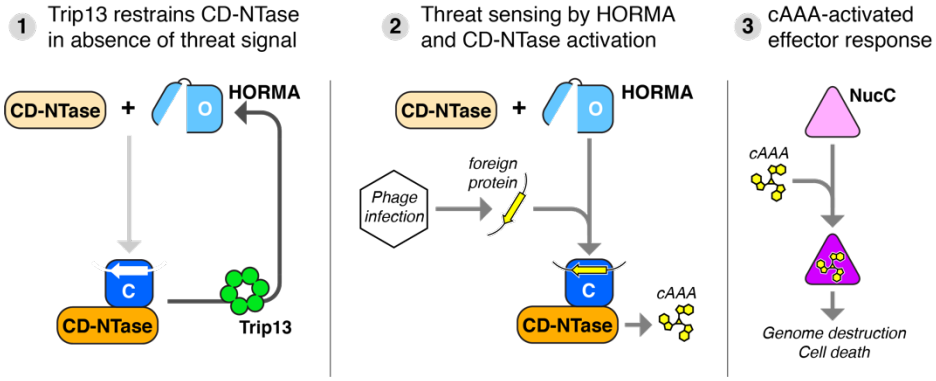


Figure 1.9 Model for bacteriophage sensing and immunity by CD-NTase+HORMA+Trip13 operons

(1) In the absence of a bacteriophage threat, Trip13 maintains HORMA proteins in the open state (light blue) by disassembling complexes of closed-HORMA (dark blue) + CdnC, thereby restraining CD-NTase activation. (2) Upon infection, open HORMA proteins recognize closure motif sequences in foreign proteins (yellow), convert to the closed state, then bind their cognate CD-NTase (orange). The CD-NTase:HORMA complex (which further requires bound DNA in some cases; not shown) synthesizes a nucleotide-based second messenger (cAAA). (3) Second messengers bind and activate effector proteins including NucC (Lau et al., 2019), causing cell death and abortive infection.

STAR Methods

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Kevin D. Corbett (kcorbett@ucsd.edu). All unique/stable reagents generated in this study are available from the Lead Contact with a completed Materials Transfer Agreement.

EXPERIMENTAL MODELS AND SUBJECT DETAILS

All proteins were produced in *E. coli* strain Rosetta2 pLysS (EMD Millipore). Cells were grown in standard media (2XYT broth) with appropriate antibiotics, and standard temperatures (37°C for growth, 20°C for protein expression induction). For cellular assays with *E. coli* MS115-1 (BEI Resources #HM-344) and JP313 (Economou et al., 1995), cells were grown in LB broth at 30°C with appropriate antibiotics.

METHOD DETAILS

Yeast Two-hybrid

Coding sequences for genes from operon 300414846 from *E. coli* MS 115-1 (CD-NTase018 from (Whiteley et al., 2019)) were synthesized (GeneArt) and inserted into vectors pBridge and pGADT7 (Clontech) by isothermal assembly. Coding sequence for genes from *P. aeruginosa* strain ATCC27853 (CD-NTase023 from (Whiteley et al.,

2019)) were amplified from genomic DNA (American Type Culture Collection) and inserted into vectors pBridge and pGADT7 by isothermal assembly. See **Table A.A.S4** for all protein sequences. pBridge vectors were transformed into *S. cerevisiae* strain AH109 and selected on SC media lacking tryptophan (-TRP). pGADT7 AD vectors were transformed into *S. cerevisiae* strain Y187 and selected on SC media lacking leucine (-LEU). Haploid strains were mated and diploids selected on SC -TRP/-LEU. Diploid cells were diluted in water and replated onto SC -TRP/-LEU (control), -TRP/-LEU/-HIS (histidine) (low stringency), and -TRP/-LEU/-HIS/-ADE (adenine) (high stringency), incubated for 2-3 days at 30°C, then examined for growth.

For yeast three-hybrid assays, pBridge vectors containing one protein in MCS I were further modified by NotI cleavage at the MCS II site followed by isothermal assembly-mediated insertion of a second gene, resulting in a single vector encoding two genes, with the Gal4-BD tag fused to the N-terminus of the gene in MCS I. These vectors were transformed into AH109 and mated with pGADT7 AD vectors encoding other proteins.

Protein Expression, Purification, and Characterization

All proteins were cloned into UC Berkeley Macrolab vectors 2AT (for untagged expression) or 2BT (encoding an N-terminal TEV protease-cleavable His_6-tag). Co-expression cassettes were assembled by amplifying genes from these vectors by PCR, and re-inserting into vector 2BT so that one protein is tagged. For coexpression of HORMA proteins with closure motif peptides, the peptide (Peptide 1: EVMEFNP; Peptide 2: HGKILLT) was cloned into UC Berkeley Macrolab vector 2CT (encoding an

N-terminal TEV protease-cleavable His_6-maltose binding protein tag), then assembled into co-expression vectors with *Ec* HORMA or *Pa* HORMA2 by PCR. All point-mutants were generated by PCR-mediated mutagenesis.

Proteins were expressed in *E. coli* strain Rosetta 2 (DE3) pLysS (EMD Millipore, Billerica MA). Cultures were grown at 37°C to A600=0.5, then induced with 0.25 mM IPTG and shifted to 20°C for 15 hours. Cells were harvested by centrifugation and resuspended in buffer A (20 mM Tris pH 7.5 (Tris pH 8.5 for CdnC/CdnD), 10% glycerol) plus 400 mM NaCl, 10 mM imidazole, and 5 mM β-mercaptoethanol. Proteins were purified by Ni^{2+}-affinity (Ni-NTA agarose, Qiagen) then passed over an anion-exchange column (Hitrap Q HP, GE Life Sciences, Piscataway NJ) in Buffer A plus 5 mM β-mercaptoethanol and 0.1-1 M NaCl, collecting flow-through or peak fractions. Tags were cleaved with TEV protease (Tropea et al., 2009), and cleaved protein was passed over another Ni^{2+} column (collecting flow-through fractions) to remove uncleaved protein, cleaved tags, and tagged TEV protease. The protein was passed over a size exclusion column (Superdex 200, GE Life Sciences) in buffer GF (buffer A plus 300 mM NaCl and 1 mM dithiothreitol (DTT)), then concentrated by ultrafiltration (Amicon Ultra, EMD Millipore) to 10-20 mg/ml and stored at 4°C. For selenomethionine derivatization, protein expression was carried out in M9 minimal media supplemented with amino acids plus selenomethionine prior to IPTG induction (Van Duyne et al., 1993), and proteins were exchanged into buffer containing 1 mM tris(2-carboxyethyl)phosphine (TCEP) after purification to maintain the selenomethionine residues in the reduced state.

For characterization of oligomeric state by size exclusion chromatography coupled to multi-angle light scattering (SEC-MALS), 100 μL of purified protein/complex at 2-5 mg/mL was injected onto a Superdex 200 Increase 10/300 GL or Superose 6 Increase 10/300 GL column (GE Life Sciences) in a buffer containing 20 mM HEPES pH 7.5, 300 mM NaCl, 5% glycerol, and 1 mM DTT. Light scattering and refractive index profiles were collected by miniDAWN TREOS and Optilab T-rEX detectors (Wyatt Technology), respectively, and molecular weight was calculated using ASTRA v. 6 software (Wyatt Technology).

Crystallization and structure determination

E. coli CdnC

We obtained crystals of *Ec* CdnC using two different methods. First, small-scale dialysis of *Ec* CdnC at 10 mg/mL from buffer GF to a buffer containing 0% glycerol and 200 mM NaCl resulted in the formation of large prism crystals. Second, mixing CdnC (10 mg/mL in buffer GF) 1:1 with a well solution containing 20 mM HEPES pH 8.1, 10-12% PEG 4000, 100 mM NaCl, 10 mM $MgCl_2$ (plus 1 mM TCEP for selenomethionine-derivatized protein) in hanging-drop format resulted in smaller prism crystals. All crystals were cryoprotected by the addition of 2-Methyl-2,4-pentanediol (MPD; 30% for dialysis crystals, 15% for hanging-drop), and flash-frozen in liquid nitrogen. We collected diffraction data at the Advanced Photon Source NE-CAT beamline 24ID-E (support statement below), and all datasets were processed with the RAPD data-processing pipeline, which uses XDS (Kabsch, 2010) for data indexing and reduction, AIMLESS (Evans and Murshudov, 2013) for scaling, and TRUNCATE (Winn et al., 2011) for

conversion to structure factors. We determined the structure by single-wavelength anomalous diffraction methods using a 1.92 Å dataset from selenomethionine-derivatized protein, in the PHENIX Autosol wizard (Terwilliger et al., 2009). We manually rebuilt the initial model in COOT (Emsley et al., 2010), and refined against a 1.44 Å native dataset in phenix.refine (Afonine et al., 2012) using positional and individual anisotropic B-factor refinement (**Table A.A.S3**). Since ATP was not added to *Ec* CdnC purification or crystallization buffers, we conclude that ATP observed in the *Ec* CdnC active site was co-purified with the enzyme.

P. aeruginosa CdnD

Crystals of wild-type *Pa* CdnD (Apo state) were obtained by mixing protein (20 mg/mL) in crystallization buffer (25 mM Tris-HCl pH 7.5, 200 mM NaCl, 5 mM $MgCl_2$, 1 mM TCEP) 1:1 with well solution containing 0.1 M HEPES pH 7.5 (or imidazole pH 8.0) and 1.8-2.0 M Ammonium sulfate in hanging-drop format. Crystals were cryoprotected by the addition of 30% glycerol, and flash-frozen in liquid nitrogen. Diffraction data were collected at the Advanced Photon Source beamline 24ID-E and processed with the RAPD data-processing pipeline. Crystals of *Pa* CdnD D62N/D64N (ATP-bound) were obtained by mixing protein (20 mg/mL) in crystallization buffer plus 10 mM $MgCl_2$ and 5 mM ATP 1:1 with well solution containing 0.1 M Tris pH 8.5, 0.2 M sodium formate, and 20-22% PEG 6000 in hanging drop format. Crystals were cryoprotected by the addition of 10% glycerol and flash-frozen in liquid nitrogen. Diffraction data were collected at the Advanced Photon Source beamline 24ID-E and processed with the RAPD data-processing pipeline.

We determined the structure of Apo *Pa* CdnD using a 2.05 Å-resolution single-wavelength dataset from a crystal of selenomethionine-derivatized protein, using the PHENIX Autosol wizard. We manually rebuilt the initial model in COOT, followed by refinement in phenix.refine using positional and individual B-factor refinement. The structure of ATP-bound *Pa* CdnD D62N/D64N was determined by molecular replacement in PHASER (McCoy et al., 2007).

P. aeruginosa HORMA3

Crystals of *Pa* HORMA3 were obtained by mixing protein (10 mg/mL) in crystallization buffer 1:1 with well solution containing 100 mM imidazole pH 8.0, 200 mM $CaCl_2$, and 32% PEG 3350 in hanging-drop format. Crystals were flash-frozen in liquid nitrogen directly from the crystallization drop. We collected diffraction data at the Advanced Photon Source NE-CAT beamline 24ID-C, and all datasets were processed with the RAPD pipeline. In order to obtain phases by selenomethionine derivatization, we designed several mutants with conserved hydrophobic residues mutated to methionine. One mutant, V102M/L146M, crystallized more robustly than wild-type protein and was used for all later crystallographic analysis. We determined the structure of *Pa* HORMA3 V102M/L146M from a 1.86 Å resolution dataset from crystals derivatized with NaBr (1 minute soak in cryoprotectant solution containing 1 M NaBr). Data was processed with the RAPD pipeline, Br sites were identified with hkl2map (Pape et al., 2004; Sheldrick, 2010), and the structure was determined using the phenix AUTOSOL wizard. The initial models was manually rebuilt with COOT and refined against a 1.64 Å resolution native dataset using phenix.refine.

P. aeruginosa HORMA2

We obtained crystals of *Pa* HORMA2 by mixing protein (10 mg/mL) in crystallization buffer 1:1 in hanging drop format with well solution containing 100 mM Imidazole pH 8.0, 0.8 M NaH_2PO_4, and 0.55-0.8 M KH_2PO_4. Crystals were cryoprotected with an additional 30% glycerol and flash-frozen in liquid nitrogen. Native data (to 2.1 Å resolution) were collected at APS beamline 24ID-C, and a single-wavelength anomalous diffraction dataset (to 2.5 Å resolution) on a crystal grown from selenomethionine-derivatized protein was collected at SSRL beamline 9-2. Data were processed by the RAPD pipeline (native) or the autoxds pipeline (selenomethionine). The structure was determined by SAD phasing using the PHENIX AutoSol wizard, models were manually rebuilt with COOT, and refined against the native dataset using phenix.refine.

P. aeruginosa HORMA2-ΔC

We purified *Pa* HORMA2-ΔC (residues 2-133) as above (replacing Tris-HCl with HEPES pH 8.0 in all buffers), then dimethylated surface lysine residues by mixing protein at 1 mg/mL with 50 mM freshly-prepared dimethylamine borane complex and 100 mM formaldehyde, incubating at 4°C for one hour, then quenching by addition of 25 mM glycine. Dimethylated *Pa* HORMA2-ΔC was then re-concentrated to 17 mg/mL and buffer-exchanged into GF buffer to remove residual reaction components.

We obtained crystals of dimethylated *Pa* HORMA2-ΔC by mixing mixed protein (17 mg/mL) in crystallization buffer 1:1 in hanging drop format with well solution

containing 100 mM Tris-HCl pH 8.5, 10 mM $NiCl_2$, and 20% PEG 2000 MME. Crystals were cryoprotected with an additional 10% glycerol or 10% PEG 400, then looped and frozen in liquid nitrogen. Diffraction data were collected at APS beamline 24ID-C and processed by the RAPD pipeline. We determined the structure by molecular replacement using the structure of *Pa* HORMA2 as a search model (the successful model lacked the C-terminal safety belt region and the N-terminus, which both undergo conformational changes between the open and closed states. Initial models were manually rebuilt with COOT and refined using phenix.refine.

P. aeruginosa HORMA3:HORMA2:Peptide 1

A complex of *Pa* HORMA3 (residues 1-133), HORMA2, and Peptide 1 (EVMEFNP) was obtained by co-expressing and purifying His_6-MBP-tagged Peptide 1 with untagged HORMA3 and HORMA2, with the tag cleaved prior to crystallization. Crystals were obtained by mixing protein (10 mg/mL) in crystallization buffer 1:1 in hanging drop format with well solution containing 100 mM HEPES pH 7.5, 0.5 M sodium acetate, 31% PEG 3350, and 1 mM ATP. Crystals were flash-frozen in liquid nitrogen directly from the crystallization drop. We collected a 2.0 Å diffraction dataset on SSRL beamline 9-2, processed the data using the autoxds pipeline, and determined the structure by molecular replacement with PHASER, using the structures of *Pa* HORMA3 and *Pa* HORMA2. Initial models were manually rebuilt with COOT and refined using phenix.refine.

Rhizobiales sp. Trip13

Rh Trip13 was purified as above, except all purification buffers contained 5 mM $MgCl_2$. We obtained crystals of selenomethionine-derivatized *Rh* Trip13 by mixing protein (24 mg/mL) in crystallization buffer 1:1 in hanging drop format with well solution containing 100 mM sodium citrate pH 6.0, 1.6 M NH_2SO_4, and 0.2 M sodium/potassium tartrate. Crystals were cryoprotected with 30% glycerol, and flash-frozen in liquid nitrogen. We collected a 2.05 Å-resolution single-wavelength anomalous diffraction dataset on APS beamline 24ID-E, processed the data with the RAPD pipeline, and determined the structure using the PHENIX Autosol wizard. Initial models were manually rebuilt with COOT and refined using phenix.refine. The structure of *Rh* Trip13 is shown in **Figure A.A.S7**.

P. aeruginosa HORMA2:Peptide 1:CdnD

We obtained crystals of *Pa* HORMA2:Peptide 1:CdnD by mixing purified complex (15 mg/mL) in crystallization buffer 1:1 in hanging drop format with well solution containing 100 mM Bis-Tris pH 5.5, 0.2 M ammonium acetate, and 30% PEG 3350. Crystals were flash-frozen in liquid nitrogen directly from the crystallization drop. Diffraction data were collected at APS beamline 24ID-C, and processed by the RAPD pipeline. The structure was determined by molecular replacement in PHASER using the structures of *Pa* HORMA2 and CdnD as search models. Initial models were manually rebuilt with COOT, and refined using phenix.refine.

E. coli HORMA:CdnC:Trip13^{EQ}

To assemble a HORMA:CdnC:Trip13^{EQ} complex, we separately purified *Ec* HORMA:CdnC (coexpressed, with an N-terminal His_6-tag on HORMA) and *Ec* Trip13^{EQ} (E159Q), then mixed at a ratio of one 1:1 HORMA:CdnC complex per Trip13^{EQ} hexamer in buffer GF plus 1 mM ATP and 5 mM $MgCl_2$ at 4°C. After a one-hour incubation, the mixture was passed over a Superdex 200 size exclusion column (GE Life Sciences), and assembled complexes were collected and concentrated by ultrafiltration. Crystals were obtained by mixing protein (10 mg/mL) 1:1 with well solution containing 0.1 M MOPS pH 7.0, 0.2 M $MgCl_2$, and 14% PEG 3350 in hanging-drop format. We collected a 2.64 Å resolution dataset at APS beamline 24ID-E, and processed the data using the RAPD pipeline. Crystals adopt space group $P2_1$, with one 1 HORMA, 1 CdnC, and 6 Trip13^{EQ} monomers per asymmetric unit. We determined the structure by molecular replacement in PHASER, using the structure of *Ec* CdnC and a Trip13 hexamer model built by separately overlaying the large and small AAA subdomains of six *Rh* Trip13 monomers onto a recent cryo-EM structure of human TRIP13 bound to p31^{comet}:MAD2 (Alfieri et al., 2018). A model of *Pa* HORMA2 was used as a guide to manually build *Ec* HORMA into difference density obtained from initial refinement. Initial models were manually rebuilt with COOT and refined using phenix.refine.

E. coli CdnC:HORMA-ΔN

We obtained crystals of *Ec* CdnC:HORMA-ΔN by mixing protein (12 mg/mL) in crystallization buffer 1:1 in hanging drop format with well solution containing 0.1 M CHES pH 9.5, 0.1 M magnesium acetate, and 17% PEG 3350. Crystals were cryoprotected by the addition of 20% glycerol and flash-frozen in liquid nitrogen.

Diffraction data were collected at APS beamline 24ID-C, and processed by the RAPD pipeline. The structure was determined by molecular replacement in PHASER using the structures of *Ec* HORMA and *Ec* CdnC as search models. Initial models were manually rebuilt with COOT, and refined using phenix.refine.

APS NE-CAT Support Statement

This work is based upon research conducted at the Northeastern Collaborative Access Team beamlines, which are funded by the National Institute of General Medical Sciences from the National Institutes of Health (P41 GM103403). The Eiger 16M detector on 24-ID-E beam line is funded by a NIH-ORIP HEI grant (S10OD021527). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357.

SSRL Support Statement

Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research, and by the National Institutes of Health, National Institute of General Medical Sciences (including P41GM103393). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of NIGMS or NIH.

Isothermal Titration Calorimetry

Isothermal titration calorimetry was performed at the Sanford Burnham Prebys Medical Discovery Institute Protein Analysis Core. Measurements were performed on a MicroCal iTC200 in buffer GF, by injecting ATP or GTP (1 mM) into a solution containing 0.1 mM *Pa* CdnD (wild-type or D62N/D64N mutant). Two injections were performed per protein/nucleotide combination.

DNA Binding

For DNA binding assays, a 40 bp double-stranded DNA was produced by annealing 5′-FAM Oligo #1 and Oligo #2 (see **Key Resources Table**). Binding reactions (30 μL) contained 50 nM dsDNA probe and 0.39-50 μM protein in binding buffer (25 mM Tris-HCl pH 8.5, 50 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, and 5% glycerol), and were incubated 30 minutes at room temperature for binding. Pre-treated nitrocellulose (top; GE Healthcare) and nylon (GE Healthcare Hybond-N; bottom) membranes were assembled into a Bio-Rad bio-dot 96-well vacuum filtration apparatus. Samples were passed through the filtration apparatus, washed three times with 200 μL binding buffer, then imaged using a Bio-Rad Chemi-Doc gel imager. For K_d value calculation, images were quantified with ImageJ (Schneider et al., 2012), then binding curves were fit in Prism v. 8 (GraphPad Software).

Phage Display

For phage display, we used a commercial assay kit (Ph.D.-7 Phage Display Peptide Library Kit, New England Biolabs) and followed the recommended protocol for

"solution phase panning with affinity bead capture" with the following modifications. We used His_6-tagged *Ec* HORMA, *Pa* HORMA2, and *Pa* HORMA3 as bait proteins. For affinity purification, we diluted 2 μL of bait protein at 0.2 mM and 10 μL phage library into 200 μL volume in TBST (50mM Tris pH7.5, 150mM NaCl, plus 0.1% Tween (first round) or 0.5% Tween (subsequent rounds)). After a 20 minute incubation at room temperature, 50 μL of magnetic Ni-NTA beads (Qiagen; pre-blocked with blocking buffer (0.1M $NaHCO_3$, pH 8.6, 5mg/mL BSA, 0.02% NaN_3) and washed 3x with TBST) were added, and the mixture was incubated a further 15 minutes at room temperature. Beads were washed 10x with wash buffer (50 mM Tris-HCl pH 7.5, 300 mM NaCl, plus 0.1% Tween (first round) or 0.5% Tween (subsequent rounds)). For elution, 1 mL elution buffer (0.2M Glycine-HCl, pH 2.2, 1mg/mL BSA) was added, incubated for 10 minutes at room temperature, then neutralized the eluate by adding 150 μL Tris-HCl pH 9.1. Eluted phage were titered and amplified according to standard protocols, and the selection was repeated, Following four rounds of selection, phages were isolated and the variable peptide region sequenced for at least 20 individual clones.

Second messenger synthesis

Cyclic trinucleotide synthesis was performed essentially as described (Whiteley et al., 2019). Briefly, 2 mL synthesis reactions contained 1 μM (for *Ec* CdnC) or 3 μM (for *Pa* CdnD) CD-NTase or CD-NTase:HORMA complex and 0.25 mM ATP (plus 0.25 mM GTP for *Pa* CdnD) in reaction buffer with 12.5 mM NaCl, 20 mM $MgCl_2$, 1 mM DTT, and 10 mM Tris-HCl pH 8.5 (or HEPES-NaOH pH 7.5 for reactions with *Pa* Trip13). For reactions with DNA, plasmid DNA (pUC18) was used at 0.15 μg/mL, 4 kb linear DNA

was used at 0.1 μg/mL, single-stranded DNA (sheared salmon sperm DNA; Invitrogen #AM9680) was used at 1 μg/mL, and 40 bp dsDNA (annealed Oligos #1 and #2; see **Key Resources Table**) was used at 13 μg/mL. For reactions with Trip13, *Ec* Trip13 was used at 100 nM monomer/16.67 nM hexamer and *Pa* Trip13 was used at 300 nM monomer/50 nM hexamer; this represents a 60-fold lower molar concentration of active Trip13 hexamers compared to CdnC/HORMA in the reactions. Reactions were incubated at 37°C for 16 hours, then 2.5 units/mL reaction volume (5 units for 2 mL reaction) calf intestinal phosphatase was added and the reaction incubated a further 2 hours at 37°C. The reaction was heated to 65°C for 30 minutes, and centrifuged 10 minutes at 15,000 RPM to remove precipitated protein. Reaction products were separated by ion-exchange chromatography (1 mL Hitrap Q HP, GE Life Sciences) using a gradient from 0.2 to 2 M ammonium acetate. Product peaks were pooled, evaporated using a speedvac, and analyzed by LC-MS/MS.

Mass spectrometry

To characterize the products of CD-NTase reactions, we performed liquid chromatography-tandem mass spectrometry (LC-MS/MS). LC-MS/MS analysis was performed using a Thermo Scientific Vanquish UHPLC coupled to a Thermo Scientific Q Exactive™ HF Hybrid Quadrupole-Orbitrap™ Mass Spectrometer, utilizing a ZIC-pHILIC polymeric column (100 mm x 2.1 mm, 5 μm) (EMD Millipore) maintained at 45°C and flowing at 0.4 mL/min. Separation of cyclic trinucleotide isolates was achieved by injecting 2 μL of prepared sample onto the column and eluting using the following linear gradient: (A) 20 mM ammonium bicarbonate in water, pH 9.6, and (B) acetonitrile; 90%

B for 0.25 minutes, followed by a linear gradient to 55% B at 4 minutes, sustained until 6 minutes. The column was re-equilibrated for 2.50 minutes at 90% B.

Detection was performed in positive ionization mode using a heated electrospray ionization source (HESI) under the following parameters: spray voltage of 3.5 kV; sheath and auxiliary gas flow rate of 40 and 20 arbitrary units, respectively; sweep gas flow rate of 2 arbitrary units; capillary temperature of 275°C; auxiliary gas heater temperature of 350 °C. Profile MS1 spectra were acquired with the following settings; mass resolution of 35,000, AGC volume of $1x10^6$, maximum IT of 75 ms, with a scan range of 475 to 1050 m/z to include z=1 and z=2 ions of cyclic trinucleotides. Data dependent MS/MS spectra acquisition was performed using collision-induced dissociation (CID) with the following settings: mass resolution of 17,500; AGC volume of $1x10^5$; maximum IT of 50 ms; a loop count of 5; isolation window of 1.0 m/z; normalized collision energy of 25 eV; dynamic exclusion was not used. Data reported is for the z=1 acquisition for each indicated cyclic trinucleotide.

Nuclease Assays

For all nuclease assays, UC Berkeley Macrolab plasmid 2AT (Addgene #29665; 4,731 bp) was used. *Ec* NucC (10 nM unless otherwise indicated) and second messenger molecules were mixed with 1 μg plasmid DNA or PCR product in a buffer containing 10 mM Tris-HCl (pH 7.5), 25 mM NaCl, 10 mM $MgCl_2$, and 1 mM DTT (50 μL reaction volume), incubated 10 min at 37°C, then separated on a 1.2% agarose gel. Gels were stained with ethidium bromide and imaged by UV illumination. Second messenger molecules were either purchased from Invivogen or synthesized

enzymatically by *Enterobacter cloacae* CdnD02 and purified as previously described (Lau et al., 2019; Whiteley et al., 2019).

Plaque Assays

To measure bacteriophage immunity, the four-protein CBASS operon from *E. coli* MS115-1 (See **Methods A.A.S1** for sequence) was PCR-amplified and inserted into vector pLAC22 for IPTG-inducible expression, then modified by PCR-based mutagenesis to generate mutant/tagged protein variants. Vectors were transformed into *E. coli* strain JP313 (Economou et al., 1995) for plaque assays.

For bacteriophage infection plaque assays, a single bacterial colony was picked from a fresh LB agar plate and grown in LB broth containing proper selection antibiotic at 37°C to an OD_{600} of 0.1-0.2. CdnC operon expression was induced through the addition of 0.2 mM IPTG, followed by further growth for one hour to an OD_{600} of 0.6-0.7. For phage dilution assays, 500 μL of cells were mixed with 4.5 mL of of 0.35% LB top agar was added and entire sample was poured on LB plates containing appropriate selection and/or IPTG (0.1 mM). Plates were spotted with 5 μL of bacteriophage λ*cI*- (cI gene deleted to inhibit lysogeny and promote clear plaque formation) diluted in phage buffer (10 mM Tris, pH 7.5, 10 mM $MgCl_2$, 68 mM NaCl) plus 1 mM $CaCl_2$ at six dilutions: ~$2.4x10^6$ PFU/mL and five 10-fold dilutions thereof. Plates were incubated at 37°C for 18 hours, then imaged. For plaque numbers, 500 μL of cells at OD_{600} 0.6-0.7 were mixed with 10 μL of bacteriophage λ*cI* (stock concentration ~$2.4x10^7$ PFU/mL as measured by infection of JP313, diluted 1000-fold prior to mixing) diluted in phage buffer (10 mM Tris, pH 7.5, 10 mM $MgCl_2$, 68 mM NaCl) plus 1 mM $CaCl_2$, and

incubated at room temperature for 30 min. 4.5 mL of 0.35% LB top agar was added and entire sample was poured on LB plates containing appropriate selection and IPTG (0.1 mM) and incubated at 37°C. Plaques were counted after 18 hours of growth.

Bacterial Growth Assays

pLAC22 vectors were constructed encoding either the full CBASS operon from *E. coli* strain MS115-1 (see **Supplemental Note**) or a variant encoding the NucC D73N mutant under the control of an IPTG-inducible promoter. *E. coli* strain JP313 (a derivative of *E. coli* MC4100) (Economou et al., 1995) was transformed with these vectors and incubated in 5 mL LB broth with 0.1 mg/ml ampicillin in culture tubes at 30°C with rolling. OD at 600nm was measured after 3 hours and cells were diluted to an OD_{600} of 0.1 in a reservoir containing LB with 100ug/ml ampicillin and 0.2mM IPTG. 100 μl of cells were loaded into a 96-well plate (four replicates per strain) and 5 μl of bacteriophage λ*cI-* diluted in phage buffer with 1mM $CaCl_2$ was added (final multiplicity of infection either 0.25 or 2.5). The 96-well plate was shaken at 30°C between measurements of OD_{600}, as an average of 25 flashes every 2 minutes, by a Tecan Infinite M200 plate reader.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis of primary diffraction datasets was performed with standard measures including R_{sym}, R_{meas}, and $CC_{1/2}$ to determine data quality and resolution cutoffs. Statistical analysis of refined crystal structures used standard measures

including *R* and R_{free} factors, and geometry analysis by the MolProbity server (Chen et al., 2010).

DATA AND CODE AVAILABILITY

All primary diffraction datasets are available at the SBGrid Data Bank (https://data.sbgrid.org) under the accession numbers noted in **Table A.A.S3** and the **Key Resources Table**. All reduced diffraction datasets and refined molecular coordinates are available at the RCSB Protein Data Bank (https://www.rcsb.org) under the accession numbers noted in **Table A.A.S3** and the **Key Resources Table**. The original imaging data for **Figure 1.1** and **Figure A.A.S1** is available at Mendeley Data (https://data.mendeley.com) under DOI 10.17632/dr8mjtwh2d.1.

References

Afonine, P. V, Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H., Adams, P.D., IUCr, 2012. Towards automated crystallographic structure refinement with phenix.refine. Acta crystallographica Section D, Biological crystallography 68, 352–367. https://doi.org/10.1107/S0907444912001308

Alfieri, C., Chang, L., Barford, D., 2018. Mechanism for remodelling of the cell cycle checkpoint protein MAD2 by the ATPase TRIP13. Nature 559, 274–278. https://doi.org/10.1038/s41586-018-0281-1

Aravind, L., Koonin, E. V, 1998. The HORMA domain: a common structural denominator in mitotic checkpoints, chromosome synapsis and DNA repair. Trends in biochemical sciences 23, 284–286.

Burdette, D.L., Monroe, K.M., Sotelo-Troha, K., Iwig, J.S., Eckert, B., Hyodo, M., Hayakawa, Y., Vance, R.E., 2011. STING is a direct innate immune sensor of cyclic di-GMP. Nature 478, 515–518. https://doi.org/10.1038/nature10429

Burroughs, A.M., Zhang, D., Schäffer, D.E., Iyer, L.M., Aravind, L., 2015. Comparative genomic analyses reveal a vast, novel network of nucleotide-centric systems in biological conflicts, immunity and signaling. Nucleic acids research 43, gkv1267--10654. https://doi.org/10.1093/nar/gkv1267

Chen, Q., Sun, L., Chen, Z.J., 2016. Regulation and function of the cGAS–STING pathway of cytosolic DNA sensing. Nature Immunology 17, 1142–1149. https://doi.org/10.1038/ni.3558

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

Civril, F., Deimling, T., de Oliveira Mann, C.C., Ablasser, A., Moldt, M., Witte, G., Hornung, V., Hopfner, K.-P., 2013. Structural mechanism of cytosolic DNA sensing by cGAS. Nature 498, 332–337. https://doi.org/10.1038/nature12305

Cohen, D., Melamed, S., Millman, A., Shulman, G., Oppenheimer-Shaanan, Y., Kacen, A., Doron, S., Amitai, G., Sorek, R., 2019. Cyclic GMP-AMP signaling protects

bacterial against viral infection. Nature 574, 691–695. https://doi.org/10.1038/s41586-019-1605-5

Corrigan, R.M., Gründling, A., 2013. Cyclic di-AMP: another second messenger enters the fray. Nature reviews Microbiology 11, 513–524. https://doi.org/10.1038/nrmicro3069

Davies, B.W., Bogard, R.W., Young, T.S., Mekalanos, J.J., 2012. Coordinated regulation of accessory genetic elements produces cyclic di-nucleotides for V. cholerae virulence. Cell 149, 358–370. https://doi.org/10.1016/j.cell.2012.01.053

Deville, C., Carroni, M., Franke, K.B., Topf, M., Bukau, B., Mogk, A., Saibil, H.R., 2017. Structural pathway of regulated substrate transfer and threading through an Hsp100 disaggregase. Science advances 3, e1701726. https://doi.org/10.1126/sciadv.1701726

Economou, A., Pogliano, J.A., Beckwith, J., Oliver, D.B., Wickner, W., 1995. SecA membrane cycling at SecYEG is driven by distinct ATP binding and hydrolysis events and is regulated by SecD and SecF. Cell 83, 1171–81.

Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K., 2010. Features and development of Coot. Acta crystallographica Section D, Biological crystallography 66, 486–501. https://doi.org/10.1107/S0907444910007493

Evans, P.R., Murshudov, G.N., 2013. How good are my data and what is the resolution? Acta crystallographica Section D, Biological crystallography 69, 1204–1214. https://doi.org/10.1107/S0907444913000061

Gao, P., Ascano, M., Wu, Y., Barchet, W., Gaffney, B.L., Zillinger, T., Serganov, A.A., Liu, Y., Jones, R.A., Hartmann, G., Tuschl, T., Patel, D.J., 2013. Cyclic [G(2',5')pA(3',5')p] is the metazoan second messenger produced by DNA-activated cyclic GMP-AMP synthase. Cell 153, 1094–1107. https://doi.org/10.1016/j.cell.2013.04.046

Gates, S.N., Yokom, A.L., Lin, J., Jackrel, M.E., Rizo, A.N., Kendsersky, N.M., Buell, C.E., Sweeny, E.A., Mack, K.L., Chuang, E., Torrente, M.P., Su, M., Shorter, J., Southworth, D.R., 2017. Ratchet-like polypeptide translocation mechanism of the AAA+ disaggregase Hsp104. Science 357, 273–279. https://doi.org/10.1126/science.aan1052

Han, H., Monroe, N., Sundquist, W.I., Shen, P.S., Hill, C.P., 2017. The AAA ATPase Vps4 binds ESCRT-III substrates through a repeating array of dipeptide-binding pockets. eLife 6, 50. https://doi.org/10.7554/eLife.31324

Hengge, R., 2009. Principles of c-di-GMP signalling in bacteria. Nature reviews Microbiology 7, 263–273. https://doi.org/10.1038/nrmicro2109

Hornung, V., Hartmann, R., Ablasser, A., Hopfner, K.-P., 2014. OAS proteins and cGAS: unifying concepts in sensing and responding to cytosolic nucleic acids. Nature reviews. Immunology 14, 521–528. https://doi.org/10.1038/nri3719

Jones, D.T., Cozzetto, D., 2015. DISOPRED3: precise disordered region predictions with annotated protein-binding activity. Bioinformatics (Oxford, England) 31, 857–863. https://doi.org/10.1093/bioinformatics/btu744

Kabsch, W., 2010. XDS. Acta crystallographica Section D, Biological crystallography 66, 125–132. https://doi.org/10.1107/S0907444909047337

Kato, K., Ishii, R., Goto, E., Ishitani, R., Tokunaga, F., Nureki, O., 2013. Structural and functional analyses of DNA-sensing and immune activation by human cGAS. PLoS ONE 8, e76983. https://doi.org/10.1371/journal.pone.0076983

Kim, D.H., Han, J.S., Ly, P., Ye, Q., McMahon, M.A., Myung, K., Corbett, K.D., Cleveland, D.W., 2018. TRIP13 and APC15 drive mitotic exit by turnover of interphase- and unattached kinetochore-produced MCC. Nature Communications 9, 4354. https://doi.org/10.1038/s41467-018-06774-1

Kranzusch, P.J., Lee, A.S.Y., Wilson, S.C., Solovykh, M.S., Vance, R.E., Berger, J.M., Doudna, J.A., 2014. Structure-guided reprogramming of human cGAS dinucleotide linkage specificity. Cell 158, 1011–1021. https://doi.org/10.1016/j.cell.2014.07.028

Lau, R.K., Ye, Q., Berg, K.R., Mathews, I.T., Watrous, J.D., Ego, K.M., Whiteley, A.T., Lowey, B., Mekalanos, J.J., Kranzusch, P.J., Jain, M., Corbett, K.D., 2019. Structure and mechanism of a cyclic trinucleotide-activated bacterial endonuclease mediating bacteriophage immunity. Submitted.

Li, X., Shu, C., Yi, G., Chaton, C.T., Shelton, C.L., Diao, J., Zuo, X., Kao, C.C., Herr, A.B., Li, P., 2013. Cyclic GMP-AMP synthase is activated by double-stranded DNA-induced oligomerization. Immunity 39, 1019–1031. https://doi.org/10.1016/j.immuni.2013.10.019

Livingstone, C.D., Barton, G.J., 1993. Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation. Computer applications in the biosciences : CABIOS 9, 745–56.

Lohöfener, J., Steinke, N., Kay-Fedorov, P., Baruch, P., Nikulin, A., Tishchenko, S., Manstein, D.J., Fedorov, R., 2015. The Activation Mechanism of 2'-5'-Oligoadenylate Synthetase Gives New Insights Into OAS/cGAS Triggers of Innate Immunity. Structure (London, England : 1993) 23, 851–862. https://doi.org/10.1016/j.str.2015.03.012

Luo, X., Yu, H., 2008. Protein metamorphosis: the two-state behavior of Mad2. Structure (London, England : 1993) 16, 1616–1625. https://doi.org/10.1016/j.str.2008.10.002

Ma, H.T., Poon, R.Y.C., 2018. TRIP13 Functions in the Establishment of the Spindle Assembly Checkpoint by Replenishing O-MAD2. Cell Reports 22, 1439–1450. https://doi.org/10.1016/j.celrep.2018.01.027

Ma, H.T., Poon, R.Y.C., 2016. TRIP13 Regulates Both the Activation and Inactivation of the Spindle-Assembly Checkpoint. Cell Reports 14, 1086–1099. https://doi.org/10.1016/j.celrep.2016.01.001

Mapelli, M., Massimiliano, L., Santaguida, S., Musacchio, A., 2007. The Mad2 Conformational Dimer: Structure and Implications for the Spindle Assembly Checkpoint. Cell 131, 730–743. https://doi.org/10.1016/j.cell.2007.08.049

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., Read, R.J., 2007. Phaser crystallographic software. Journal of applied crystallography 40, 658–674. https://doi.org/10.1107/S0021889807021206

McFarland, A.P., Luo, S., Ahmed-Qadri, F., Zuck, M., Thayer, E.F., Goo, Y.A., Hybiske, K., Tong, L., Woodward, J.J., 2017. Sensing of Bacterial Cyclic Dinucleotides by the Oxidoreductase RECON Promotes NF-κB Activation and Shapes a Proinflammatory Antibacterial State. Immunity 46, 433–445. https://doi.org/10.1016/J.IMMUNI.2017.02.014

Pape, T., Schneider, T.R., IUCr, 2004. HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs. Journal of applied crystallography 37, 843–844. https://doi.org/10.1107/S0021889804018047

Puchades, C., Rampello, A.J., Shin, M., Giuliano, C.J., Wiseman, R.L., Glynn, S.E., Lander, G.C., 2017. Structure of the mitochondrial inner membrane AAA+

protease YME1 gives insight into substrate processing. Science 358, eaao0464. https://doi.org/10.1126/science.aao0464

Qi, S., Kim, D.J., Stjepanovic, G., Hurley, J.H., 2015. Structure of the Human Atg13-Atg101 HORMA Heterodimer: an Interaction Hub within the ULK1 Complex. Structure (London, England : 1993) 23, 1848–1857. https://doi.org/10.1016/j.str.2015.07.011

Ripstein, Z.A., Huang, R., Augustyniak, R., Kay, L.E., Rubinstein, J.L., 2017. Structure of a AAA+ unfoldase in the process of unfolding substrate. eLife 6, 43. https://doi.org/10.7554/eLife.25754

Rosenberg, S.C., Corbett, K.D., 2015. The multifaceted roles of the HORMA domain in cellular signaling. Journal of Cell Biology. https://doi.org/10.1083/jcb.201509076

Schneider, C.A., Rasband, W.S., Eliceiri, K.W., 2012. NIH Image to ImageJ: 25 years of image analysis. Nature Methods 9, 671–675. https://doi.org/10.1038/nmeth.2089

Sheldrick, G.M., 2010. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. Acta crystallographica Section D, Biological crystallography 66, 479–485. https://doi.org/10.1107/S0907444909038360

Suzuki, H., Kaizuka, T., Mizushima, N., Noda, N.N., 2015. Structure of the Atg101–Atg13 complex reveals essential roles of Atg101 in autophagy initiation. Nature Structural & Molecular Biology 22, 572–580. https://doi.org/10.1038/nsmb.3036

Terwilliger, T.C., Adams, P.D., Read, R.J., McCoy, A.J., Moriarty, N.W., Grosse-Kunstleve, R.W., Afonine, P. V, Zwart, P.H., Hung, L.-W., 2009. Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. Acta crystallographica Section D, Biological crystallography 65, 582–601. https://doi.org/10.1107/S0907444909012098

Tromer, E.C., van Hooff, J.J.E., Kops, G.J.P.L., Snel, B., 2019. Mosaic origin of the eukaryotic kinetochore. Proceedings of the National Academy of Sciences of the United States of America 116, 12873–12882. https://doi.org/10.1073/pnas.1821945116

Tropea, J.E., Cherry, S., Waugh, D.S., 2009. Expression and purification of soluble His(6)-tagged TEV protease. Methods in molecular biology (Clifton, NJ) 498, 297–307. https://doi.org/10.1007/978-1-59745-196-3_19

Van Duyne, G.D., Standaert, R.F., Karplus, P.A., Schreiber, S.L., Clardy, J., 1993. Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. Journal of molecular biology 229, 105–124. https://doi.org/10.1006/jmbi.1993.1012

White, K.I., Zhao, M., Choi, U.B., Pfuetzner, R.A., Brunger, A.T., 2018. Structural principles of SNARE complex recognition by the AAA+ protein NSF. eLife 7, 213. https://doi.org/10.7554/eLife.38888

Whiteley, A.T., Eaglesham, J.B., de Oliveira Mann, C.C., Morehouse, B.R., Lowey, B., Nieminen, E.A., Danilchanka, O., King, D.S., Lee, A.S.Y., Mekalanos, J.J., Kranzusch, P.J., 2019. Bacterial cGAS-like enzymes synthesize diverse nucleotide signals. Nature 567, 194–199. https://doi.org/10.1038/s41586-019-0953-5

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A., McNicholas, S.J., Murshudov, G.N., Pannu, N.S., Potterton, E.A., Powell, H.R., Read, R.J., Vagin, A., Wilson, K.S., 2011. Overview of the CCP4 suite and current developments. Acta crystallographica Section D, Biological crystallography 67, 235–242. https://doi.org/10.1107/S0907444910045749

Ye, Q., Kim, D.H., Dereli, I., Rosenberg, S.C., Hagemann, G., Herzog, F., Tóth, A., Cleveland, D.W., Corbett, K.D., 2017. The AAA+ ATPase TRIP13 remodels HORMA domains through N-terminal engagement and unfolding. EMBO Journal 36, 2419–2434. https://doi.org/10.15252/embj.201797291

Ye, Q., Rosenberg, S.C., Moeller, A., Speir, J.A., Su, T.Y., Corbett, K.D., 2015. TRIP13 is a protein-remodeling AAA+ ATPase that catalyzes MAD2 conformation switching. eLife 2015, 1–44. https://doi.org/10.7554/eLife.07367

Yu, H., Lupoli, T.J., Kovach, A., Meng, X., Zhao, G., Nathan, C.F., Li, H., 2018. ATP hydrolysis-coupled peptide translocation mechanism of Mycobacterium tuberculosis ClpB. Proceedings of the National Academy of Sciences 115, E9560--E9569. https://doi.org/10.1073/pnas.1810648115

Zhang, Xu, Wu, J., Du, F., Xu, H., Sun, L., Chen, Z., Brautigam, C.A., Zhang, Xuewu, Chen, Z.J., 2014. The cytosolic DNA sensor cGAS forms an oligomeric complex with DNA and undergoes switch-like conformational changes in the activation loop. Cell reports 6, 421–30. https://doi.org/10.1016/j.celrep.2014.01.003

Zhu, D., Wang, Lijun, Shang, G., Liu, X., Zhu, J., Lu, D., Wang, Lei, Kan, B., Zhang, J.-R., Xiang, Y., 2014. Structural biochemistry of a Vibrio cholerae dinucleotide cyclase reveals cyclase activity regulation by folates. Molecular cell 55, 931–937. https://doi.org/10.1016/j.molcel.2014.08.001

Chapter 1, in full, is a reprint of the material as it appears in Molecular Cell 2020. Ye, Qiaozhen; Lau, Rebecca K.; Mathews, Ian T.; Birkholz, Erica A.; Watrous, Jeramie D.; Azimi, Camillia S.; Pogliano, J.; Jain, Mohit; Corbett, Kevin D. “HORMA Domain Proteins and a Trip13-like ATPase Regulate Bacterial CGAS-like Enzymes to Mediate Bacteriophage Immunity”, Molecular Cell, 77 (4), 2020. The dissertation/thesis author was a primary investigator and author of this paper.

CHAPTER 2: Structure and Mechanism of a Cyclic Trinucleotide-Activated Bacterial Endonuclease Mediating Bacteriophage Immunity

Summary

Bacteria possess an array of defenses against foreign invaders, including a broadly-distributed bacteriophage defense system termed CBASS (Cyclic oligonucleotide-Based Anti-phage Signaling System). In CBASS systems, a cGAS/DncV-like nucleotidyltransferase synthesizes cyclic di- or tri-nucleotide second messengers in response to infection, and these molecules activate diverse effectors to mediate bacteriophage immunity via abortive infection. Here, we show that the CBASS effector NucC is related to restriction enzymes but uniquely assembles into a homotrimer. Binding of NucC trimers to a cyclic tri-adenylate second messenger promotes assembly of a NucC homohexamer competent for non-specific double-strand DNA cleavage. In infected cells, NucC activation leads to complete destruction of the bacterial chromosome, causing cell death prior to completion of phage replication. In addition to CBASS systems, we identify NucC homologs in over 30 Type III CRISPR/Cas systems, where they likely function as accessory nucleases activated by cyclic oligoadenylate second messengers synthesized by these systems' effector complexes.

Keywords

Second messenger signaling, bacteriophage immunity, abortive infection, CD-NTase, Endonuclease

Introduction

Bacteria are locked in a continual struggle to protect themselves and their genomes from foreign mobile genetic elements and bacteriophages, and as such have evolved sophisticated defenses against these invaders. Two well-characterized bacterial defense systems are restriction-modification systems and CRISPR/Cas systems, both of which rely on targeted nuclease activity to destroy foreign nucleic acids. In a typical restriction-modification system, a bacterium encodes a sequence-specific DNA endonuclease and a corresponding DNA methylase that modifies the host genome to protect it against endonuclease cleavage (Arber and Linn, 1969; Wilson and Murray, 1991). While restriction-modification systems mount a standard response to all improperly-modified DNA and as such can be likened to an “innate” immune system, the more recently-characterized CRISPR/Cas systems are more akin to an “adaptive” immune system. In these systems, effector nucleases are loaded with guide sequences that specifically target foreign nucleic acids (Koonin et al., 2017; Sorek et al., 2013; Wright et al., 2016). In Type III CRISPR/Cas systems, the multi-protein effector complex not only recognizes and cleaves foreign RNA and DNA (Koonin et al., 2017; Pyenson and Marraffini, 2017), but also synthesizes 3′-5′ linked cyclic oligoadenylate second messengers upon target binding (Kazlauskiene et al., 2017; Niewoehner et al., 2017). These second messengers in turn activate nonspecific accessory nucleases encoded in the same operons, including the RNases Csm6 and Csx1, and the DNase Can1 (Han et al., 2018; Kazlauskiene et al., 2017; Mcmahon et al., 2019; Niewoehner et al., 2017).

The newly-described CBASS (Cyclic oligonucleotide-Based Anti-phage Signaling System) bacteriophage defense systems encode a cGAS/DncV-like nucleotidyltransferase (CD-NTase), one of several diverse effectors, and optionally a set of regulatory proteins putatively responsible for sensing an infection (Burroughs et al., 2015; Cohen et al., 2019; Whiteley et al., 2019). In related work, we determined the molecular mechanisms of a CBASS system encoding regulatory proteins related to eukaryotic HORMA domain proteins and the AAA+ ATPase Pch2/TRIP13 (Ye et al., 2019). In this system, detection of specific peptides by the CBASS-encoded HORMA domain protein stimulates synthesis of the cyclic trinucleotide second messenger cyclic AMP-AMP-AMP (cAAA), which activates the DNA endonuclease NucC (Nuclease, CD-NTase associated). Both cAAA production and NucC catalytic activity are required for bacteriophage immunity, which is mediated by an abortive infection mechanism in which infected cells die prior to completion of phage replication (Ye et al., 2019). Still unanswered, however, has been how NucC activation leads to cell death, and how this restriction endonuclease-related enzyme is controlled by the second messenger cAAA.

Here, we show that NucC is a homotrimeric DNA endonuclease that binds the second messenger cAAA in a three-fold symmetric allosteric pocket. We show that nuclease activation involves the assembly of two NucC trimers into a homohexamer, which juxtaposes pairs of endonuclease active sites and promotes coordinated double-strand DNA cleavage. NucC is largely sequence non-specific and is not inhibited by DNA methylation, and its activation in infected cells leads to complete destruction of cellular DNA. Finally, we find that NucC homologs have been incorporated into Type III

CRISPR/Cas systems, defining a new family of Type III CRISPR accessory nucleases activated by cyclic oligoadenylate second messengers.

Results

NucC activation leads to genome destruction in infected cells

We have previously shown that the CBASS system from a patient-derived strain of *E. coli* (MS115-1; **Figure 2.1A**) provides immunity against bacteriophage λ infection by an abortive infection mechanism (Ye et al., 2019). To better understand the mechanism of CBASS-mediated cell death, and particularly the role of the effector nuclease NucC, we directly imaged infected cells carrying either the intact CBASS operon or a variant carrying a mutant of NucC (D73N) that lacks DNA cleavage activity in vitro (**Figure 2.1B**). We infected cells carrying these operons with bacteriophage λ at a high multiplicity of infection and visualized cell membranes and DNA in live cells. Cultures lacking CBASS show moderate cell lysis by 60 minutes post-infection, and large-scale lysis by 80 minutes post-infection (**Figure 2.1C-D, Figure A.B.S1**). In cultures carrying the intact CBASS system, however, by 40 minutes post-infection we observe a large fraction of cells in which cellular DNA appears to have been either partially or completely degraded (**Figure 2.1C-D**). These cultures show large-scale lysis by 80 minutes post-infection, but do not show the high levels of extracellular DNA characteristic of lysed cultures lacking CBASS (**Figure 2.1C**). This degradation depends on NucC's endonuclease activity, as cultures carrying CBASS with the NucC D73N mutant show only cell lysis similar to controls (**Figure 2.1 C-D**). These data strongly indicate that NucC mediates cell death by destroying DNA within the infected cell upon activation.

NucC is related to restriction endonucleases but adopts a homotrimeric architecture

To understand the structure of NucC and its mechanism of activation, we overexpressed and purified *E. coli* MS115-1 (*Ec*) NucC, and a second NucC from a related operon found in *Pseudomonas aeruginosa* strain ATCC27853 (*Pa*). *Ec* NucC and *Pa* NucC share 77% overall sequence identity, and both proteins self-associate in yeast two-hybrid assays (Ye et al., 2019). Using size exclusion chromatography coupled to multi-angle light scattering (SEC-MALS), we found that both proteins form homotrimers in solution (**Figure A.B.S2A-B**). Several families of DNA exonucleases are known to form homotrimers, including bacteriophage λ nuclease (Kovall and Matthews, 1997; Zhang et al., 2011) and the archaeal nuclease PhoExoI (Miyazono et al., 2015). In contrast, DNA endonucleases are more commonly homodimeric, with two active sites that cleave opposite DNA strands to effect double-strand cleavage (Pingoud et al., 2005).

We crystallized and determined the structure of *Ec* NucC to a resolution of 1.75 Å (**Table A.B.S1**). In agreement with prior sequence analysis (Burroughs et al., 2015), the NucC monomer adopts a restriction endonuclease-like fold (**Figure 2.2A**). A database search using the DALI structure-comparison server (Holm and Rosenström, 2010) revealed structural similarity to type II restriction endonucleases including NgoMIV (Deibert et al., 2000), Kpn2I (PDB ID 6EKR; unpublished), and SgrAI (Dunten et al., 2008). Indeed, comparing NucC with a structure of SgrAI bound to DNA (Dunten et al., 2008) reveals an overall r.m.s.d. of 1.97 Å over 132 aligned Cα atoms (**Figure 2.2A and**

Figure A.B.S2C-D). NucC also possesses a conserved active site motif of $IDx_{30}EAK$ (residues 72-106 in *Ec* NucC), and the conserved acidic residues Asp73 and Glu104 coordinate a Mg^{2+} ion in our structure (**Figure 2.2B**). Consistent with our SEC-MALS data, the structure of NucC reveals a tightly-packed homotrimer with an overall triangular architecture (**Figure 2.2B**). The three active sites are arrayed on the outer edge of the trimer. The trimer interface of NucC is composed largely of elements that are not part of the core nuclease fold, but are rather embellishments not shared with dimeric restriction endonucleases (**Figure A.B.S2E**).

NucC binds cAAA in an allosteric pocket

In vitro, *Ec* NucC is strongly activated by cAAA, and is more weakly activated by cAAG and the cyclic dinucleotide 3′,3′ cyclic di-AMP (Ye et al., 2019). We used isothermal titration calorimetry (ITC) to directly test binding of *Ec* NucC to a range of dinucleotide and trinucleotide second messengers. We found that *Ec* NucC binds strongly to 3′,3′,3′ cAAA (*Kd* = 0.7 μM), and binds more weakly to both 3′,3′ cyclic di-AMP (*Kd* = 2.6 μM) and 3′-5′ linked linear di-AMP (5′-pApA; 4.4 μM). We did not detect binding to a mixed-linkage cyclic di-AMP (2′,3′ cyclic di-AMP) or AMP (**Figure A.B.S3** and **Figure A.B.S4**). No second messenger bound with a stoichiometry higher than ~0.3 molecules per NucC monomer in our ITC assays, suggesting that the NucC trimer likely possesses a single second messenger binding site.

We next co-crystallized both *Ec* NucC and *Pa* NucC with 3′,3′,3′ cAAA and determined the structures of these complexes to a resolution of 1.66 and 1.45 Å, respectively (**Table A.B.S1**, **Figure 2.3A-C**, **Figure A.B.S5A-C**). We also determined a

1.66 Å-resolution structure of *Ec* NucC bound to 5′-pApA, which binds equivalently to cAAA in the crystals despite failing to stimulate nuclease activity in vitro (**Table A.B.S1 and Figure A.B.S5D-F**). The structures reveal that a single second messenger molecule binds in a conserved, three-fold symmetric allosteric pocket on the "bottom" of the NucC trimer, and that second messenger binding causes a series of conformational changes both within each NucC monomer and in the trimer as a whole (**Figure 2.3A-B**). Each adenine base is recognized through a hydrogen bond between its C6 amine group and the main-chain carbonyl of Val227, and through π-stacking with Tyr81 (**Figure 2.3C-D**). In addition, NucC residues Arg53 and Thr226 form a hydrogen-bond network with each phosphate group of the cyclic trinucleotide (**Figure 2.3C**). Binding of cAAA induces a dramatic conformational change in an extended hairpin loop in NucC, which we term the "gate loop", resulting in the closure of all three loops over the conserved pocket to completely enclose the bound cAAA molecule (**Figure 2.3E-F**). Two aromatic residues, His136 and Tyr141, interact through π-stacking to stabilize the gate loop's structure in both open and closed states. In the cAAA-bound structure, these two residues serve to "latch" the gate loops through tight association at the three-fold axis of the NucC trimer. Mutation of gate loop residues His136 or Tyr141, or cAAA-binding residues Arg53, Tyr81, or Thr226 completely disrupts *Ec* NucC's cAAA-activated endonuclease activity (**Figure 2.3G**).

cAAA-mediated NucC hexamerization activates double-strand DNA cleavage

cAAA binding causes several conformational changes within each NucC protomer, including closing of the gate loops, a shift in the position of α-helices 0 and 1,

and rotation of a small β-sheet encompassing NucC residues 85-95 (**Figure 2.3B, F**). Within the NucC trimer, each subunit undergoes a rotation of ~3° "upward" away from the cAAA binding site compared to the unbound structure (**Figure 2.3B**). In the structures of both *Ec* NucC and *Pa* NucC bound to cAAA, we also observe that pairs of NucC trimers bind face-to-face to form symmetric homohexamers (**Figure 2.4A-C and Figure A.B.S5A**). The trimer-trimer interface involves a symmetric interaction of the α0 helices of facing NucC protomers, plus a hydrophobic cluster assembled from three different NucC protomers. In the α0 interface, Leu19 from two NucC protomers forms a symmetric hydrophobic interaction across the interface, and Glu15 of each NucC protomer forms a hydrogen bond with Lys26 of the opposite protomer (**Figure 2.4B**). In the hydrophobic cluster, Phe28 from one NucC protomer docks into a hydrophobic pocket on the opposite trimer, which involves Trp4 from one NucC protomer and Phe60, Phe68, and Phe86 of another (**Figure 2.4C**). The docking of Phe28 into this hydrophobic pocket causes residues 23-34 linking α0 and α1, which are disordered in our NucC Apo structure, to become well-ordered in the cAAA-bound structure. Our structure of *Pa* NucC bound to cAAA shows equivalent association of two NucC trimers into a hexamer (**Figure A.B.S5A**).

We tested for NucC hexamer assembly in solution by SEC-MALS, and found that both *Ec* NucC and *Pa* NucC form hexamers when pre-incubated with cAAA (**Figure 2.4D** and **Figure A.B.S5G**). To test whether hexamer assembly and nuclease activation are functionally linked, we designed specific mutations in *Ec* NucC to disrupt hexamer assembly. We mutated Leu19 in the α0 helix to alanine or aspartate, and mutated Trp4

and Phe28 in the hydrophobic cluster to alanine. We also mutated Ala27, which borders the hydrophobic cluster and is tightly packed in the trimer-trimer interface, to either glutamate or lysine in order to disrupt this interface. We purified six mutant proteins and tested both their ability to undergo cAAA-mediated hexamerization (**Figure 2.4E** and **Figure A.B.S6A-F**) and their cAAA-stimulated DNA cleavage activity (**Figure 2.4F and Figure A.B.S6G**). Four of the six mutants (L19D, A27E, A27K, and F28A) strongly disrupt both hexamer formation and DNA cleavage activity (**Figure 2.4H**). The W4A mutant partially disrupts hexamer assembly but retains robust cAAA-activated DNA cleavage activity, while the L19A mutant does not affect either hexamer assembly or DNA cleavage (**Figure 2.4G**). Finally, we found by plaque assay that the NucC F28A mutation, which eliminates hexamer formation and DNA cleavage activity, also completely eliminates the protective effect of the CBASS system in cells (**Figure 2.4G**). Overall, these data strongly indicate that the active form of NucC is a hexamer whose assembly is promoted by cAAA binding.

To better understand why hexamer formation activates NucC, we modelled DNA binding to the NucC hexamer by overlaying each NucC monomer with a structure of DNA-bound SgrAI (Dunten et al., 2008). Hexamer formation juxtaposes pairs of NucC active sites from opposite trimers ~25 Å apart (**Figure 2.5A**), and we could model a continuous bent DNA duplex that contacts a pair of juxtaposed NucC active sites (**Figure 2.5B**). In this model, the putative scissile phosphate groups are positioned two base pairs apart on opposite DNA strands (**Figure 2.5B**), such that coordinated cleavage at these two sites would yield a double-strand break with a two-base 3′

overhang. To determine whether NucC generates single- or double-strand breaks, we used limiting Mg^{2+} conditions to enrich for singly-cut plasmid in a plasmid digestion assay (**Figure 2.5C**). While NucC did generate a nicked species in this assay, we also observed a significant fraction of linear DNA that could only arise through cleavage of both DNA strands at a single site. These data support the idea that hexamerization of NucC activates double-strand DNA cleavage through juxtaposition of pairs of active sites.

To gain further information about NucC sequence-specificity and cleavage products, we next completely digested a plasmid with NucC and deep-sequenced the resulting short fragments. To avoid losing information about DNA ends by "blunting" during sequencing library preparation, we denatured the fragments and sequenced individual DNA strands (see **STAR Methods**). We thereby mapped ~1 million cleavage events, enabling us to assemble a picture of any preferred sequence motifs at NucC cleavage sites. We observed a striking two-fold symmetry in the preferred cleavage sites, with a strong preference for "T" at the 5′ end (+1 base) of the cleaved fragment, and an equally strong preference for "A" at the –3 position (**Figure 2.5D**). Similarly, we observed a weak preference for "G" at the +5 and +6 bases, and for "C" at the –7 and –8 bases. The two-fold symmetry of preferred NucC cleavage sites strongly suggests that two NucC active sites cooperate to cleave both strands of DNA. Moreover, the location of the breaks on each strand (as judged by the first sequenced base of the fragments) suggests that double-stranded cleavage by NucC results in two-base 3′ overhangs, consistent with our structural modeling (**Figure 2.5D** and **Figure 2.5F**).

We used the same sequencing dataset to precisely measure the length distribution of NucC products. NucC product sizes followed a roughly Gaussian distribution with a mean of ~44 bases, in agreement with our electrophoretic results (**Figure 2.5E**). Curiously, the size distribution showed a systematic variation from a Gaussian distribution, with enrichment of product lengths around 22, 44, and 66 base pairs in length. These data suggest that in addition to coordinated DNA cleavage by pairs of NucC active sites, cleavage by the three pairs of active sites in a NucC hexamer may be weakly coordinated, perhaps through wrapping of a single DNA around the complex to contact multiple sets of active sites.

Type III CRISPR systems encode NucC homologs

While exploring the distribution of NucC homologs in bacteria, we identified a group of NucC homologs with ~50% sequence identity to *Ec* NucC and *Pa* NucC, that are encoded not within CBASS systems but rather within Type III CRISPR/Cas systems (**Figure 2.6A-C**). Prompted by this finding, we comprehensively searched the gene neighborhoods of bacterial NucC homologs and identified a total of 31 distinct NucC proteins encoded within Type III CRISPR/Cas systems (**Figure 2.6A, Table A.B.S4**). These proteins are found predominantly in Gammaproteobacteria and Firmicutes, but are also found in Alpha-, Beta-, and Deltaproteobacteria (**Table A.B.S4**). Their distribution within a phylogenetic tree of more than 500 predominantly CBASS-associated NucC proteins suggests that NucC was independently incorporated into Type III CRISPR/Cas systems at least ten times (**Figure 2.6A, Table A.B.S5**). Supporting this finding, a recent bioinformatics analysis of protein families associated

with Type III CRISPR/Cas systems identified a group of putative nucleases (denoted PD-DExK after the conserved active-site motif) that includes members of NucC and a second, likely unrelated nuclease family (**Table A.B.S4**) (Shah et al., 2019).

Type III CRISPR/Cas systems encode a multi-subunit effector complex, that when activated by target binding synthesizes a range of cyclic oligoadenylate second messengers with 2-6 bases (cA_2-cA_6), often with a high proportion of cAAA/cA_3 (Grüschow et al., 2019; Kazlauskiene et al., 2017; Niewoehner et al., 2017). These cyclic oligoadenylate second messengers have been shown to activate the accessory RNases Csm6 and Csx1 in the CRISPR operons (Han et al., 2018; Kazlauskiene et al., 2017; Niewoehner et al., 2017). More recently, a cA_4-activated DNA endonuclease was identified in Type III CRISPR/Cas systems, which generates single-stranded nicks thought to slow viral replication by causing replication fork collapse (Mcmahon et al., 2019). The presence of NucC homologs in Type III CRISPR/Cas systems suggests that these enzymes may function similarly to their relatives in CBASS systems, with activation dependent on cAAA synthesized by the CRISPR effector complexes. We cloned several Type III CRISPR-associated NucC proteins, and successfully purified one such enzyme from *Vibrio metoecus*, a close relative of *V. cholerae* (Kirchberger et al., 2014). We found that *V. metoecus* (*Vm*) NucC is a DNA endonuclease that is strongly activated by cAAA, and weakly activated by both cAA and cAAG, mirroring our findings with *Ec* NucC (**Figure 2.7A-B**) (Ye et al., 2019). By SEC-MALS, we found that *Vm* NucC forms a mix of trimers and hexamers on its own, and forms solely hexamers after pre-incubation with cAAA (**Figure 2.7C**).

We next determined crystal structures of *Vm* NucC in both the trimer and hexamer states. In the absence of cAAA, *Vm* NucC crystallizes as a homotrimer with an overall structure equivalent to the *Ec* NucC apo-state trimer (**Figure 2.7D**). *Vm* NucC shares a common active-site structure with *Ec* NucC (**Figure A.B.S7A-B**), and the *Vm* NucC trimer overlays with the *Ec* NucC trimer with an overall Cα r.m.s.d. of 1.48 Å (over 607 equivalent atoms). In the presence of cAAA, *Vm* NucC crystallizes as a homohexamer, with an overall structure equivalent to the hexamer structures of both *Ec* NucC and *Pa* NucC (**Figure 2.7E**). *Vm* NucC shares both the symmetric α0-α0 interface and hydrophobic cluster interactions that mediate trimer-trimer association in *Ec* NucC. Curiously, while *Vm* NucC required cAAA to crystallize in the hexamer form, we observed no electron density for cAAA in the allosteric pocket despite 100% conservation of cAAA-binding residues with *Ec* NucC and *Pa* NucC. Inspection of crystal packing interactions in these crystals revealed a symmetric interaction between the gate loops of neighboring hexamers that pulls the gate loops into an open conformation (**Figure A.B.S7C-D**). We propose that this open gate-loop conformation allows bound cAAA to diffuse out of the allosteric pocket after crystallization, resulting in an empty binding site. Nonetheless, we could model cAAA binding to *Vm* NucC without clashes, indicating that the *Vm* NucC trimer can accommodate cAAA.

Together, our biochemical and structural data on *Vm* NucC indicate that CRISPR-associated NucC homologs share a common activation and DNA cleavage mechanism with their relatives in CBASS systems. Thus, NucC homologs have been incorporated into Type III CRISPR/Cas systems, where they are likely activated upon

target recognition and cyclic oligoadenylate synthesis by these systems' effector complexes.

Discussion

Bacteria possess an extraordinary variety of pathways to defend themselves against foreign threats including bacteriophages, invasive DNA elements, and other bacteria. A large class of defense pathways, including restriction-modification systems and CRISPR-Cas systems, specifically defend the bacterial genome through the targeted action of DNA and RNA nucleases. Here, we describe the structure and activation mechanism of a restriction endonuclease-like protein, NucC, found primarily in the CBASS bacteriophage defense system. The intact NucC-containing CBASS system from the patient-derived *E. coli* strain MS115-1 confers immunity to bacteriophage λ infection, and this immunity requires both cAAA synthesis by this operon's CD-NTase and the DNA cleavage activity of NucC (Ye et al., 2019). Moreover, the *E. coli* MS115-1 CBASS system functions via abortive infection, in which infected cells die prior to completion of the phage replication cycle (Ye et al., 2019). Here, we show that death of infected cells is caused by NucC-mediated destruction of the bacterial chromosome. Thus, NucC is part of a widespread bacterial defense pathway that mediates cell death in response to bacteriophage infection.

While evolutionarily related to restriction endonucleases, NucC adopts a unique homotrimeric structure with an allosteric binding pocket for its cyclic trinucleotide activator. This pocket is assembled from embellishments to the core nuclease fold, which include a "gate loop" that closes over the bound second messenger. cAAA binding mediates a subtle conformational change within the NucC trimer, enabling the symmetric association of two NucC trimers into a hexamer and juxtaposing pairs of

active sites to activate double-strand DNA cleavage. This exquisitely specific activation mechanism likely ensures that NucC is kept inactive in cells prior to an infection. Once activated by cAAA binding, the nuclease possesses only very weak sequence specificity and in vitro, cleaves DNA to an average length of ~50 base pairs.

Bacterial CBASS systems are widespread among environmental and pathogenic bacteria, but are sparsely distributed and often associated with transposase or integrase genes and/or conjugation systems, suggesting that they are mobile DNA elements that confer some selective advantage on their bacterial host (Burroughs et al., 2015; Whiteley et al., 2019). These operons encode diverse putative effector proteins including proteases, exonucleases, and two families of endonucleases (Burroughs et al., 2015). Given the structure of these operons, it is likely that each effector is specifically activated by the second messenger synthesized by its associated CD-NTase (Whiteley et al., 2019). Bacterial CD-NTases have been classified into eight major clades, and while members of several clades have been shown to synthesize specific cyclic di- and trinucleotides (Whiteley et al., 2019), the full spectrum of products synthesized by these enzymes remains an important open question. Given our data on NucC structure and activation, we propose that all bacterial CD-NTases associated with NucC effectors – which includes members of clades C01 (including *E. coli* MS115-1 CdnC), C03, and D05 (including *P. aeruginosa* ATCC27853 CdnD) (Whiteley et al., 2019) – likely synthesize cAAA. While CD-NTases in clade C03 are primarily associated with NucC effectors, clade C01 and D05 CD-NTases are associated with a range of effectors (Whiteley et al., 2019). One key question is whether all of these effectors are

activated by cAAA, or whether their associated CD-NTases synthesize distinct second messengers. More broadly, a full accounting of bacterial CD-NTase products, and of the effector proteins they activate, remains an important area for future study.

Type III CRISPR/Cas systems uniquely encode nonspecific accessory RNases (Csm6 and Csx1) and DNases (Can1) that are activated by cyclic tetra- and hexa-adenylates synthesized by their effector complexes (Han et al., 2018; Kazlauskiene et al., 2017; Mcmahon et al., 2019; Niewoehner et al., 2017). The accessory RNases aid immunity by non-specifically degrading both host and invader transcripts, causing growth arrest in an infected cell , while Can1 is thought to slow viral DNA replication (Mcmahon et al., 2019). We identified a set of Type III CRISPR/Cas systems that encode clear homologs of NucC, with their distribution suggesting that NucC has been independently incorporated into these systems at least ten times. We show that one such NucC, from *Vibrio metoecus*, is a cAAA-stimulated double-stranded DNA endonuclease, demonstrating a common mechanism of action with CBASS-associated NucC. Notably, NucC is the first reported Type III CRISPR effector nuclease to be activated by cAAA, a major second messenger product of many Type III CRISPR effector complexes. The biological role of NucC in Type III CRISPR systems remains unclear. It is possible that incorporation of NucC into a CRISPR system converts this system from a specific defensive pathway into an abortive infection system; indeed, recent reports have shown that certain type IF and type VI CRISPR systems can act via abortive infection (Meeske et al., 2019; Watson et al., 2019). Alternatively, NucC's cleavage activity may be spatially or temporally restrained in these systems to limit its

potential to destroy the host genome. Discriminating between these models will require direct tests with NucC-encoding Type III CRISPR/Cas systems.

Acknowledgements

The authors thank the staffs of the Stanford Synchrotron Radiation Lightsource, the Advanced Photon Source NE-CAT beamlines, and the Advanced Light Source beamline 8.3.1 for assistance with crystallographic data collection; A. Bobkov (Sanford Burnham Prebys Medical Discovery Institute, Protein Analysis Core) for assistance with isothermal titration calorimetry; and A. Desai and M. Daugherty for critical reading and helpful suggestions. RKL was supported by the UC San Diego Quantitative and Integrative Biology training grant (NIH T32 GM127235). ITM was supported by NIH T32 HL007444, T32 GM007752 and F31 CA236405. JDW was supported by NIH K01 DK116917 and P30 DK063491. MJ was supported by NIH S10 OD020025 and R01 ES027595. ATW was supported as a fellow of The Jane Coffin Childs Memorial Fund for Medical Research. BL was supported as a Herchel Smith Graduate Research Fellow. JJM was supported by NIH/NIAID R01AI018045 and R01AI026289. PJK was supported from the Richard and Susan Smith Foundation and the Parker Institute for Cancer Immunotherapy. JP was supported by NIH R01 GM129245. KDC was supported by the Ludwig Institute for Cancer Research and the University of California, San Diego.

Author contributions

Conceptualization, RKL, QY, KDC; Methodology, RKL, QY, EAB, LP, ITM, JDW, KE, ATW, BL, MJ, PJK, KDC; Investigation, RKL, QY, EAB, LP, KRB, ITM, JDW, KDC; Writing – Original Draft, RKL, QY, KDC; Writing – Review & Editing, RKL, QY, MJ, PJK,

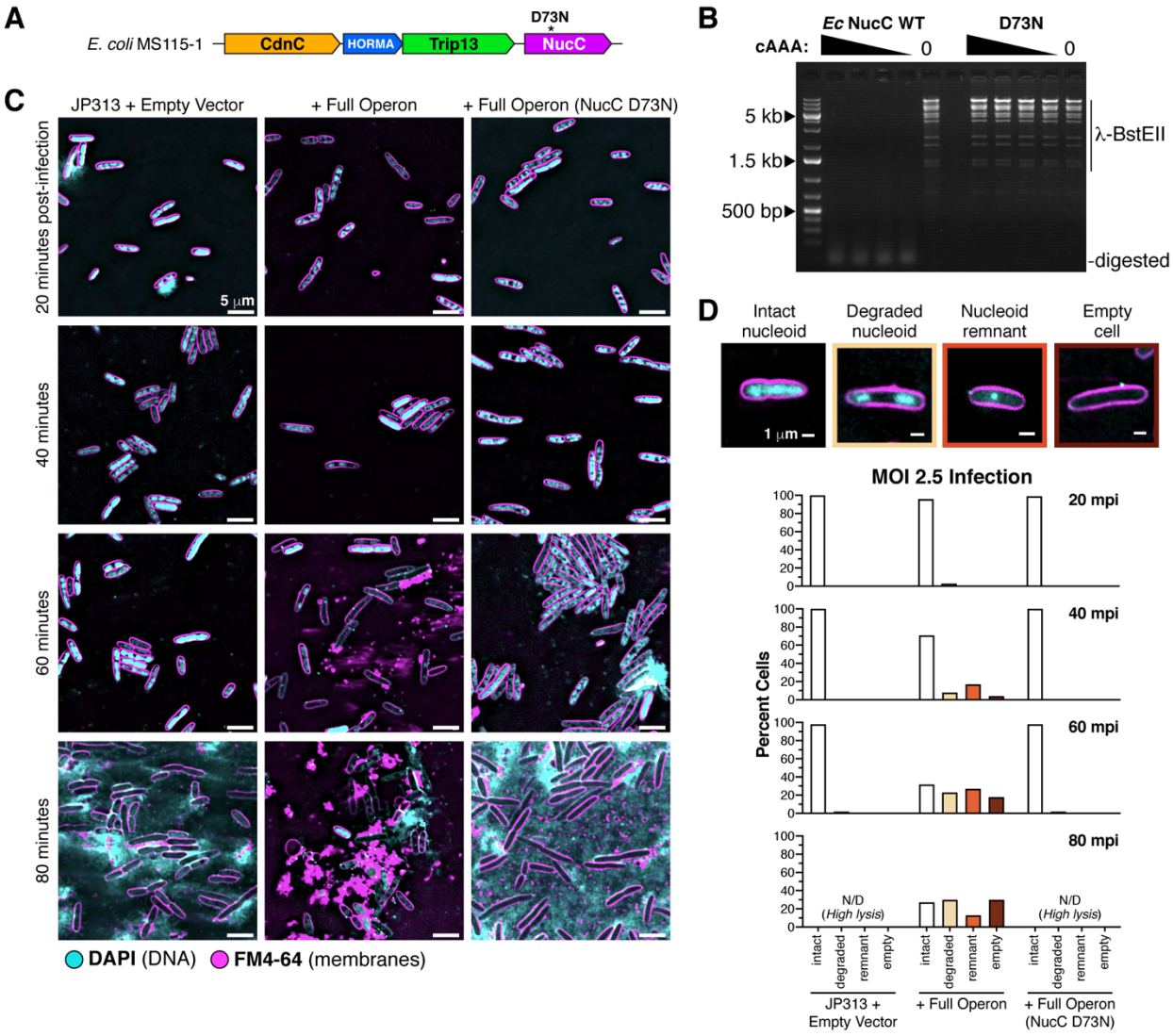
JP, KDC; Visualization, RKL, QY, ITM, JDW, KDC; Supervision, MJ, JJM, PJK, JP, KDC; Project Administration, KDC; Funding Acquisition, MJ, JJM, PJK, JP, KDC.

Declaration of Interests

The authors declare no competing interests.

Figure 2.10 NucC activation leads to chromosome destruction in infected cells

(A) Schematic of the CBASS system from *E. coli* MS115-1 harboring NucC. Asterisk denotes the position of the D73N nuclease-dead mutant (Ye et al., 2019). **(B)** Digestion of a bacteriophage λ BstEII ladder by wild-type *Ec* NucC (10 nM) and nuclease-dead D73N mutant, in the presence of cAAA (0, 6.25, 25, 100, and 400 nM). **(C)** Fluorescence microscopy of *E. coli* JP313 cells transformed with empty pLAC22 vector or vectors encoding intact *E. coli* MS115-1 CBASS or a variant with the NucC D73N mutant. DAPI staining is shown in cyan, and FM4-64 staining (membranes) is shown in magenta. See **Figure A.B.S1A** for separate channels. Strong DAPI-staining dots represent bacteriophage particles. **(D)** *Top:* Representative images of cells in different classes: Intact nucleoid, Degraded nucleoid, Nucleoid remnant, or Empty Cell (no nucleoid). *Bottom:* Quantitation of the results shown in panel (C), showing that nucleoid destruction occurs in cells harboring the full CBASS operon (n=243-1385 cells per condition in ≥10 fields over two separate experiments). N/D: not determined (high levels of cell lysis precluded scoring at 80 minutes post-infection in Empty Vector and NucC D73N conditions).


A
E. coli MS115-1
CdnC
HORMA
Trip13
NucC
D73N
*
B
Ec NucC WT
D73N
cAAA:
0
0
5 kb
1.5 kb
500 bp
λ-BstEII
-digested
C
JP313 + Empty Vector
+ Full Operon
+ Full Operon (NucC D73N)
20 minutes post-infection
40 minutes
60 minutes
80 minutes
5 μm
DAPI (DNA)
FM4-64 (membranes)
D
Intact nucleoid
Degraded nucleoid
Nucleoid remnant
Empty cell
1 μm
MOI 2.5 Infection
20 mpi
40 mpi
60 mpi
80 mpi
Percent Cells
N/D (High lysis)
intact
degraded
remnant
empty
JP313 + Empty Vector
+ Full Operon
+ Full Operon (NucC D73N)

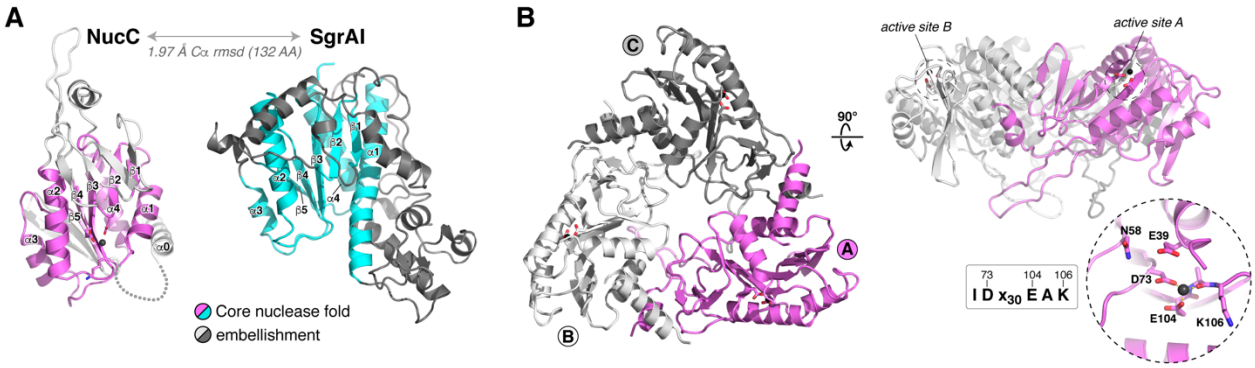


Figure 2.11 NucC is a homotrimeric relative of restriction endonucleases

(A) Structure of an *Ec* NucC monomer and the related restriction endonuclease SgrAI (PDB ID 3DW9; (Dunten et al., 2008)). Shared secondary structural elements in the core nuclease fold are colored violet (NucC) or cyan (SgrAI), and family-specific embellishments are shown in gray. **(B)** Structure of an *Ec* NucC trimer with monomers colored violet (monomer A), white (B), and gray (C). The side view (*right*) shows the active sites of the A and B monomers. *Inset:* closeup view of the NucC active site, with ordered Mg^{2+} ion shown as a black sphere. NucC possesses a conserved IDx$_{30}$EAK active-site consensus motif (See **Figure A.B.S2C-D** for comparison with the active site of SgrAI, and **Figure A.B.S2E** for NucC sequence alignment).

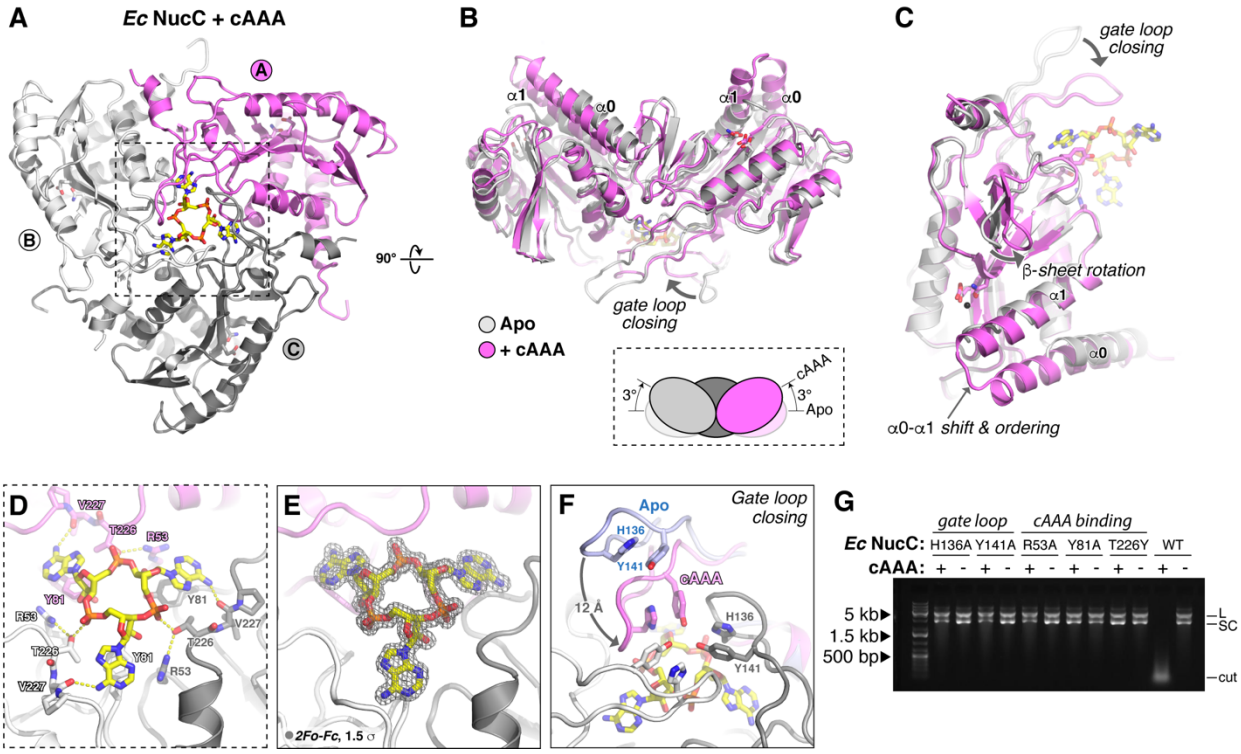
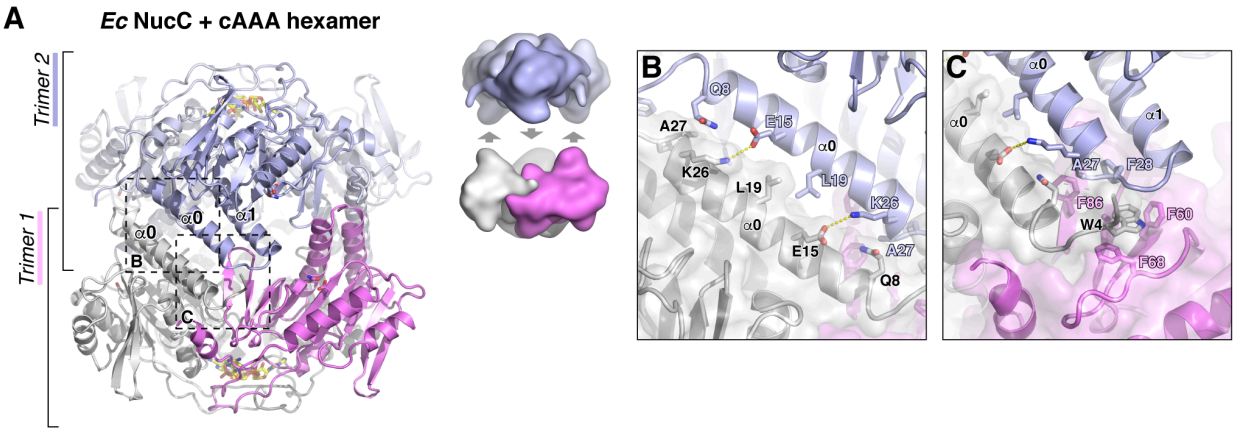


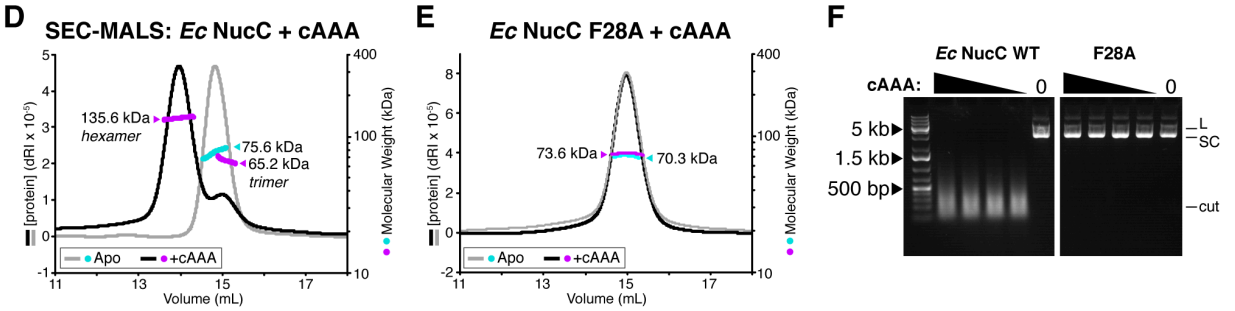
Figure 2.12 NucC binds cAAA in an allosteric pocket

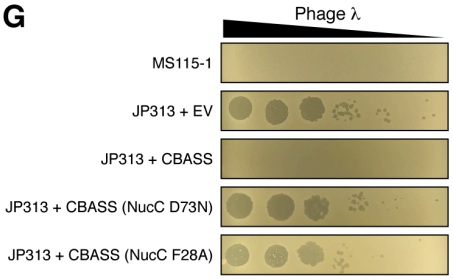
(A) Bottom view of an *Ec* NucC trimer bound to cAAA, with subunits colored violet/gray/white as in **Figure 2.2B**. See **Figure A.B.S3** for synthesis, purification, and mass spectrometry analysis of cAAA, and **Figure A.B.S4** for isothermal titration calorimetry measurements of *Ec* NucC binding second messengers. **(B)** Overlay of the NucC trimer in the Apo state (gray) and bound to cAAA (violet), showing the 3° upward rotation of each subunit induced by cAAA binding. **(C)** Overlay of the NucC monomer in the Apo state (gray) and bound to cAAA (violet), showing the major conformational differences within each subunit: gate loop closing, shifting and ordering of the α0-α1 region, and rotation of the small β-sheet. **(D)** Closeup view of cAAA binding NucC. **(E)** *2Fo-Fc* electron density for cAAA at 1.66 Å resolution, contoured at 1.5 σ. **(F)** Closeup view of gate loop closure upon cAAA binding. For one monomer, the gate loop is shown in both open (Apo, blue) and closed (cAAA-bound, violet) conformations, showing the ~12 Å motion between the two states. Pi stacking of "latch" residues H136 and Y141 (labeled in gray monomer) stabilizes gate loop conformation in both states. See **Figure A.B.S5** for structures of *Pa* NucC bound to cAAA, and of *Ec* NucC bound to 5′-pApA. **(G)** Plasmid digestion assay with wild-type and structure-based mutants of *Ec* NucC (10 nM) in the presence and absence of cAAA (100 nM).

Figure 2.13 cAAA-stimulated NucC hexamerization is required for DNA cleavage activity and bacteriophage λ immunity

(A) Cartoon view of two cAAA-bound NucC trimers (colored violet/gray/white and light blue, respectively) assembled into a homohexamer. Bound cAAA molecules are shown in yellow sticks. **(B)** Closeup view of the NucC α0-α0 interface mediating hexamer formation. **(C)** Closeup view of the NucC hydrophobic cluster interface mediating hexamer formation. **(D)** Size exclusion chromatography coupled to multi-angle light scattering (SEC-MALS) for *Ec* NucC in the absence of cAAA (gray line with molecular weight measurement in cyan) and in the presence of cAAA (black line with molecular weight measurement in violet). The molar mass of an *Ec* NucC homotrimer is 80.1 kDa, and a homohexamer is 160.3 kDa. See **Figure A.B.S5G** for SEC-MALS of *Pa* NucC in the presence and absence of cAAA. **(E)** SEC-MALS analysis of *Ec* NucC L19D mutant, which remains trimeric in the presence of cAAA. See **Figure A.B.S6A-F** for SEC-MALS analysis of other *Ec* NucC mutants. **(F)** Plasmid digestion assay with wild-type and F28A mutant *Ec* NucC (10 nM), in the presence of cAAA (for each set: 400, 100, 25, 6.25, and 0 nM cAAA). “L” denotes linear plasmid, “SC” denotes closed-circular supercoiled plasmid, and “cut” denotes fully-digested DNA. See **Figure A.B.S6G** for plasmid digestion data on other *Ec* NucC mutants. **(G)** Dilutions of bacteriophage λ on lawns of *E. coli* MS115-1 (top) and JP313 (wild-type laboratory strain) with the plasmid-encoded *E. coli* MS115-1 CBASS system encoding wild-type proteins or the indicated mutations. EV: empty vector. Six 10-fold bacteriophage dilutions are shown. **(H)** Summary of SEC-MALS, in vitro DNA cleavage activity, and bacteriophage λ immunity data for *Ec* NucC hexamer interface mutants. Shaded in gray are mutants that strongly disrupt both hexamer formation and DNA cleavage; NucC F28A further disrupts bacteriophage λ immunity. N/D: not determined.





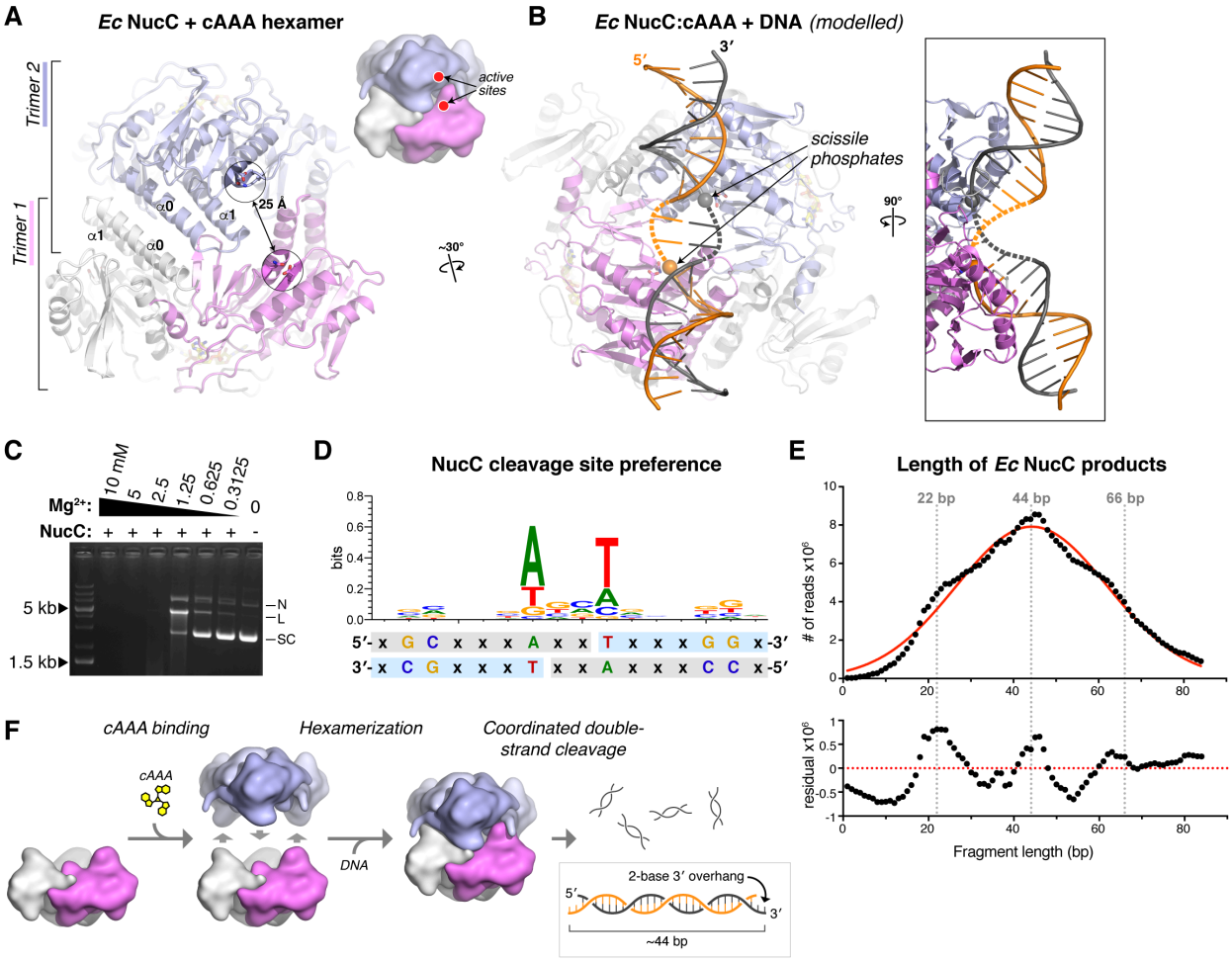


H

Ec NucC	Hexamer assembly	DNA cleavage	λ immunity
WT	+	+	+
W4A	+ / -	+	N/D
L19A	+	+	N/D
L19D	-	-	N/D
A27E	-	-	N/D
A27K	-	-	N/D
F28A	-	-	-

Figure 2.14 NucC generates double-strand cuts with two-base 3' overhangs

(A) View of the cAAA-bound *Ec* NucC hexamer, showing the juxtaposition of two active sites ~25 Å apart. **(B)** Model of a NucC hexamer bound to DNA, assembled from separately overlaying the NucC monomers colored violet and blue with a structure of SgrAI bound to DNA (PDB ID 3DW9; (Dunten et al., 2008)). The two modelled DNAs can be joined by a bent-two base-pair linker (dotted lines) to generate a contiguous dsDNA (two strands colored gray and orange). Cleavage of the scissile phosphates (spheres) at each active site would generate a double-strand break with a two-base 3′ overhang. **(C)** Plasmid cleavage by *Ec* NucC in limiting Mg^{2+}. Supercoiled DNA (SC) is converted to both nicked (N) and linear (L) forms, demonstrating significant double-strand cleavage. **(D)** Analysis of NucC DNA cleavage site preference, from mapping ~600,000 cleavage sites from fully-digested plasmid DNA. Light-blue indicates the sequenced strand; the two-fold symmetry of the cleavage site supports a model of double-strand cleavage resulting in two-base 3′ overhangs. **(E)** Length distribution of *Ec* NucC products after complete digestion, using high-throughput sequencing read lengths from a 343 million-read dataset. Red line shows a Gaussian fit, and bottom panel shows residual values from the Gaussian fit. **(F)** Model for hexamerization and coordinated double-strand cleavage by NucC. Inactive NucC trimers associate upon binding cAAA to form active hexamers. Coordinated cleavage at pairs of NucC active sites juxtaposed by hexamerization results in double-strand breaks with two-base 3′ overhangs.


A
Ec NucC + cAAA hexamer
Trimer 2
Trimer 1
active sites
α0
α1
25 Å
~30°
B
Ec NucC:cAAA + DNA (modelled)
5′
3′
scissile phosphates
90°
C
Mg2+:
10 mM
5
2.5
1.25
0.625
0.3125
0
NucC:
5 kb
1.5 kb
N
L
SC
D
NucC cleavage site preference
bits
5′- x G C x x x A x x T x x x G G x -3′
3′- x C G x x x T x x A x x x C C x -5′
E
Length of Ec NucC products
22 bp
44 bp
66 bp
# of reads x10^6
residual x10^6
Fragment length (bp)
F
cAAA binding
Hexamerization
Coordinated double-strand cleavage
cAAA
DNA
2-base 3′ overhang
5′
3′
~44 bp

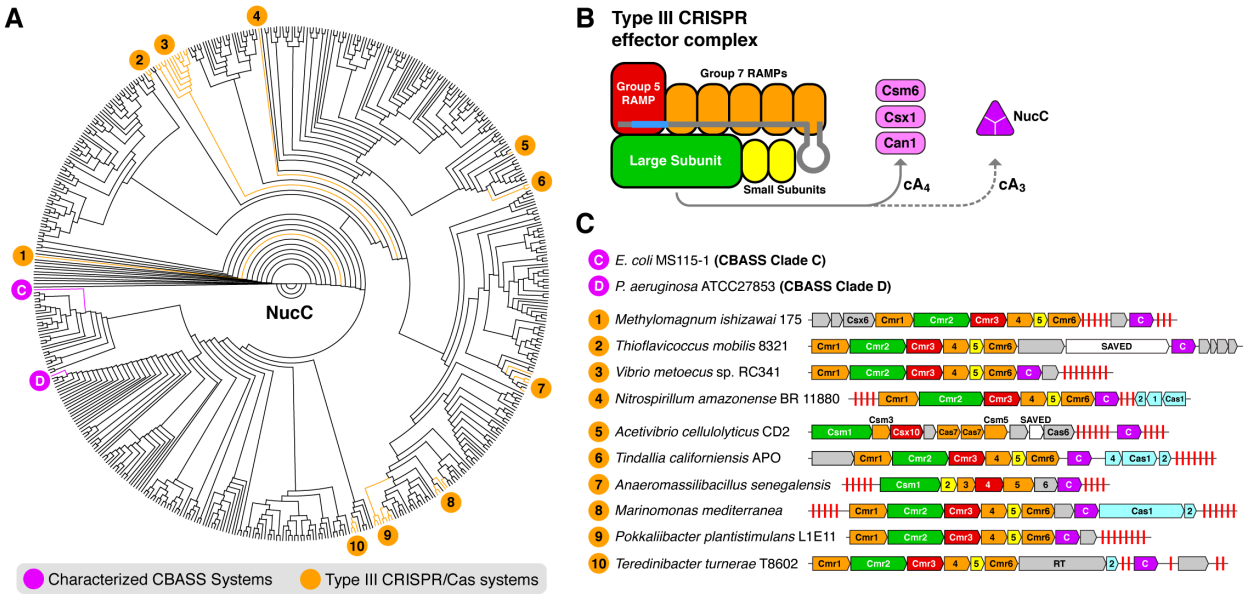


Figure 2.15 NucC is an accessory nuclease in Type III CRISPR systems

(A) Phylogenetic tree of 531 bacterial NucC proteins, with 31 NucC homologs encoded within Type III CRISPR/Cas systems highlighted in orange. Numbers denote clusters of Type III CRISPR-associated NucC homologs, whose distribution within the larger NucC tree suggests that each cluster represents an independent integration of NucC into a Type III CRISPR system. See **Table A.B.S4** for a complete list of Type III CRISPR-associated NucC homologs, and **Table A.B.S5** for the complete list of proteins used in tree construction. **(B)** Schematic of Type III CRISPR effector complexes (adapted from (Makarova et al., 2017)). The large subunit (Cmr2/Cas10; green) produces cyclic oligoadenylates that in some type III systems activate accessory RNases (Csm6 or Csx1) or DNases (Can1). **(C)** Schematics of representative Type III CRISPR operons from each cluster shown in panel (A). Core Type III genes are colored as in panel (B), with NucC homologs magenta. CRISPR arrays are denoted with red lines, predicted adaptation module genes are shown in cyan, non-core or unassigned genes are shown in gray, and two predicted SAVED domain-containing proteins (Burroughs et al., 2015) are shown in white. RT: predicted reverse transcriptase. The operon for *Acetivibrio cellulolyticus* CD2 was poorly annotated, and the gene names shown are based on the closest match from BLAST searches.

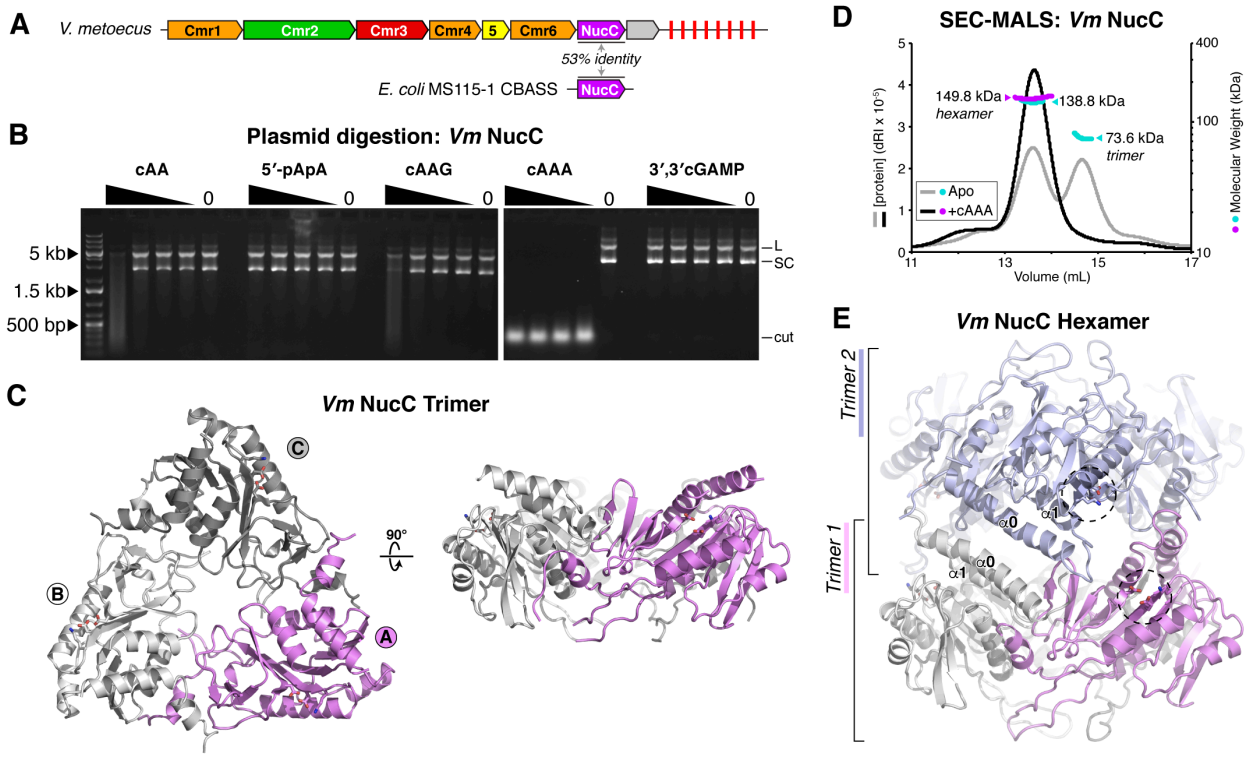


Figure 2.16 Type III CRISPR-associated NucC from V. metoecus is a cAAA-activated endonuclease

(A) Schematic of the Type III CRISPR operon from V. metoecus sp. RC341. This operon's NucC shares 53% overall sequence identity with the CBASS-associated NucC from *E. coli* MS115-1 **(B)** Plasmid digestion assay with *V. metoecus* NucC in the presence of nucleotide-based second messengers. **(C)** Two views of of the *V. metoecus* NucC apo-state trimer. The *Vm* NucC trimer shows an overall r.m.s.d. of 1.48 Å (over 607 equivalent Cα atoms) with the *Ec* NucC apo-state trimer. See **Figure A.B.S7A** for a structural comparison with *Ec* NucC, and **Figure A.B.S7B** for a close-up view of the *Vm* NucC active site. **(D)** SEC-MALS of *V. metoecus* NucC in either the Apo state (gray/cyan) or after pre-incubation with cAAA (black/magenta). Calculated MW of the NucC homotrimer = 84.1 kDa; homohexamer = 168.1 kDa. **(E)** Structure of the *Vm* NucC hexamer with trimer 1 colored as in panel (A), and trimer 2 in blue. One pair of active sites that become juxtaposed after hexamer formation are highlighted with dotted circles.

STAR Methods

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Kevin D. Corbett (kcorbett@ucsd.edu). All unique/stable reagents generated in this study are available from the Lead Contact with a completed Materials Transfer Agreement.

EXPERIMENTAL MODELS AND SUBJECT DETAILS

All proteins were produced in *E. coli* strain Rosetta2 pLysS (EMD Millipore). Cells were grown in standard media (2XYT broth) with appropriate antibiotics, and standard temperatures (37°C for growth, 20°C for protein expression induction). For cellular assays with *E. coli* MS115-1 (BEI Resources #HM-344) and JP313 (Economou et al., 1995), cells were grown in LB broth at 30°C with appropriate antibiotics.

METHOD DETAILS

Bacterial Growth and Microscopy

pLAC22 vectors were constructed encoding either the full CBASS operon from *E. coli* strain MS115-1 or a variant encoding the NucC D73N mutant under the control of an IPTG-inducible promoter, as described (Ye et al., 2019). *E. coli* strain JP313 (a derivative of *E. coli* MC4100) (Economou et al., 1995) was transformed with these vectors and incubated in 5 mL LB broth with 0.1 mg/ml ampicillin in culture tubes at

30°C with rolling. OD at 600nm was measured after 3 hours and cells were diluted to an OD_{600} of 0.1 in a reservoir containing LB with 100ug/ml ampicillin and 0.2mM IPTG. Bacteriophage λ*cI*- diluted in phage buffer with 1mM $CaCl_2$ was added for a final multiplicity of infection of 2.5, then incubated at 30°C with rolling.

For fluorescence microscopy, samples of each culture were taken at 0, 60, 90, and 120 minutes post-infection for imaging by fluorescence microscopy. Harvested samples were stained with 1 μg/mL FM4-64 (Pogliano et al., 1999) and 2 μg/mL DAPI. Cultures were centrifuged at 3,300 × g for 30 s in a microcentrifuge and resuspended in 1/10 volume of the original cultures. Three microliters of concentrated cells were transferred onto an agarose pad containing 1.2% agarose and 20% LB medium for microscopy. Microscopy was performed using a Deltavision Elite System (GE Healthcare) and images were deconvolved with SoftWoRx software. For each condition, at least five fields from each of two independent experiments were manually scored (n=243-1385 cells for each condition)

Protein Expression, Purification, and Characterization

Full length NucC genes from *E. coli* MS115-1 (Genbank ID EFJ98159; synthesized by Invitrogen/Geneart), *P. aeruginosa* ATCC 27853 (Genbank ID WP_003050273; amplified from genomic DNA), *Vibrio metoecus* sp. RC341 (IMG ID 647193063; synthesized by Invitrogen/Geneart), and *Gynuella sunshinyii* YC6258 (IMG ID 2632356247; synthesized by Invitrogen/Geneart) were cloned into UC Berkeley Macrolab vector 1B (Addgene #29653) to generate N-terminal fusions to a TEV protease-cleavable His_6-tag. Mutants of *Ec* NucC were generated by PCR mutagenesis.

Proteins were expressed in *E. coli* strain Rosetta2 pLysS (EMD Millipore) by induction with 0.25 mM IPTG at 20°C for 16 hours. For selenomethionine derivatization of *Ec* NucC, cells were grown in M9 minimal medium, then supplemented with amino acids (Leu, Ile and Val (50 mg/L), Phe, Lys, Thr (100 mg/L) and Selenomethionine (60 mg/L)) upon induction with IPTG.

For protein purification, cells were harvested by centrifugation, suspended in resuspension buffer (20 mM Tris-HCl pH 8.0, 300 mM NaCl, 10 mM imidazole, 1 mM dithiothreitol (DTT) and 10% glycerol) and lysed by sonication. Lysates were clarified by centrifugation (16,000 rpm 30 min), then supernatant was loaded onto a Ni^{2+} affinity column (HisTrap HP, GE Life Sciences) pre-equilibrated with resuspension buffer. The column was washed with buffer containing 20 mM imidazole and 100 mM NaCl, and eluted with a buffer containing 250 mM imidazole and 100 mM NaCl. The elution was loaded onto an anion-exchange column (Hitrap Q HP, GE Life Sciences) and eluted using a 100-600 mM NaCl gradient. Fractions containing the protein were pooled and mixed with TEV protease (1:20 protease:NucC by weight) plus an additional 100 mM NaCl, then incubated 16 hours at 4°C for tag cleavage. Cleavage reactions were passed over a Ni^{2+} affinity column, and the flow-through containing cleaved protein was collected and concentrated to 2 mL by ultrafiltration (Amicon Ultra-15, EMD Millipore), then passed over a size exclusion column (HiLoad Superdex 200 PG, GE Life Sciences) in a buffer containing 20 mM Tris-HCl pH 8.0, 300 mM NaCl, and 1 mM DTT. Purified proteins were concentrated by ultrafiltration and stored at 4°C for crystallization,

or aliquoted and frozen at -80°C for biochemical assays. All mutant proteins were purified as wild-type.

For characterization of oligomeric state by size exclusion chromatography coupled to multi-angle light scattering (SEC-MALS), 100 μL of purified protein at 2-5 mg/mL was injected onto a Superdex 200 Increase 10/300 GL column (GE Life Sciences) in a buffer containing 20 mM HEPES-NaOH pH 7.5, 300 mM NaCl, 5% glycerol, and 1 mM DTT. For samples with second messengers, protein was pre-incubated with 0.1 mM second messenger for 16 hours at 4°C prior to analysis. Light scattering and refractive index profiles were collected by miniDAWN TREOS and Optilab T-rEX detectors (Wyatt Technology), respectively, and molecular weight was calculated using ASTRA v. 6 software (Wyatt Technology). SEC-MALS analysis of all *Ec* NucC mutants (**Figure 3** and **Figure S5**) was performed with protein also containing the D73N active-site mutation.

Second messenger molecules

Cyclic and linear dinucleotides, including 5′-pApA, 3′,3′-cGAMP, and 3′,3′-cyclic di-AMP, were purchased from Invivogen. Cyclic trinucleotides were generated enzymatically by *Ec*-CdnD02 from *Enterobacter cloacae*, which was cloned and purified as described (Whiteley et al., 2019). Briefly, a 40 mL synthesis reaction contained 500 nM *Ec*-CdnD02 and 0.25 mM each ATP and GTP (ATP alone for large-scale synthesis of cAAA) in reaction buffer with 12.5 mM NaCl, 20 mM $MgCl_2$, 1 mM DTT, and 10 mM Tris-HCl pH 9.0. The reaction was incubated at 37°C for 16 hours, then 2.5 units/mL reaction volume (100 units for 40 mL reaction) calf intestinal phosphatase was added

and the reaction incubated a further 4 hours at 37°C. The reaction was heated to 65°C for 30 minutes, and centrifuged 20 minutes at 4,000 RPM to remove precipitated protein. Reaction products were separated by ion-exchange chromatography (Mono-Q, GE Life Sciences) using a gradient from 0 to 2 M ammonium acetate. Three major product peaks (**Figure A.B.S2A**) were pooled, evaporated using a speedvac, then resuspended in water and analyzed by LC MS/MS. cAAA synthesized by the *E. coli* MS115-1 CdnC:HORMA complex was similarly purified and verified to activate NucC equivalently to that synthesized by *Ec*-CdnD02 (Ye et al., 2019).

Mass spectrometry

Whiteley et al. previously identified the major product of *Ec*-CdnD02 as cAAG by NMR (Whiteley et al., 2019). To further characterize the products of *Ec*-CdnD02, we performed liquid chromatography-tandem mass spectrometry (LC-MS/MS) (**Figure A.B.S2B-D**). LC-MS/MS analysis was performed using a Thermo Scientific Vanquish UHPLC coupled to a Thermo Scientific Q Exactive™ HF Hybrid Quadrupole-Orbitrap™ Mass Spectrometer, utilizing a ZIC-pHILIC polymeric column (100 mm x 2.1 mm, 5 μm) (EMD Millipore) maintained at 45 °C and flowing at 0.4 mL/min. Separation of cyclic trinucleotide isolates was achieved by injecting 2 μL of prepared sample onto the column and eluting using the following linear gradient: (A) 20 mM ammonium bicarbonate in water, pH 9.6, and (B) acetonitrile; 90% B for 0.25 minutes, followed by a linear gradient to 55% B at 4 minutes, sustained until 6 minutes. The column was re-equilibrated for 2.50 minutes at 90% B.

Detection was performed in positive ionization mode using a heated electrospray ionization source (HESI) under the following parameters: spray voltage of 3.5 kV; sheath and auxiliary gas flow rate of 40 and 20 arbitrary units, respectively; sweep gas flow rate of 2 arbitrary units; capillary temperature of 275 °C; auxiliary gas heater temperature of 350°C. Profile MS1 spectra were acquired with the following settings; mass resolution of 35,000, AGC volume of 1x106, maximum IT of 75 ms, with a scan range of 475 to 1050 m/z to include z=1 and z=2 ions of cyclic trinucleotides. Data dependent MS/MS spectra acquisition was performed using collision-induced dissociation (CID) with the following settings: mass resolution of 17,500; AGC volume of 1x105; maximum IT of 50 ms; a loop count of 5; isolation window of 1.0 m/z; normalized collision energy of 25 eV; dynamic exclusion was not used. Data reported is for the z=1 acquisition for each indicated cyclic trinucleotide.

Isothermal Titration Calorimetry

Isothermal titration calorimetry was performed the Sanford Burnham Prebys Medical Discovery Institute Protein Analysis Core. Measurements were performed at 20°C on a Microcal ITC 200 (Malvern Panalytical) in a buffer containing 20 mM Tris-HCl (pH 8.5), 200 mM NaCl, 5 mM $MgCl_2$, 1 mM EDTA, and 1 mM tris(2-carboxyethyl)phosphine (TCEP). Nucleotides at 1 mM were injected into an analysis cell containing 100 μM *Ec* NucC (**Figure A.B.S3A-G, Table A.B.S2**). A second round of ITC was performed with 5′pApA and 3′,3′ c-di-AMP in a buffer containing 10 mM Tris-HCl (pH 7.5), 25 mM NaCl, 10 mM $MgCl_2$, and 1 mM TCEP (**Figure A.B.S3H-K, Table A.B.S3**).

Crystallization and structure determination

E. coli NucC

We obtained crystals of *E. coli* NucC in the Apo state in hanging drop format by mixing protein (8-10 mg/mL) in crystallization buffer (25 mM Tris-HCl pH 7.5, 200 mM NaCl, 5 mM $MgCl_2$, 1 mM TCEP) 1:1 with well solution containing 27% PEG 400, 400 mM $MgCl_2$, and 100 mM HEPES pH 7.5 (plus 1 mM TCEP for selenomethionine-derivatized protein). For cryoprotection, we supplemented PEG 400 to 30%, then flash-froze crystals in liquid nitrogen and collected diffraction data on beamline 14-1 at the Stanford Synchrotron Radiation Lightsource (support statement below). We processed all datasets with the SSRL autoxds script, which uses XDS (Kabsch, 2010) for data indexing and reduction, AIMLESS (Evans and Murshudov, 2013) for scaling, and TRUNCATE (Evans, 2006) for conversion to structure factors. We determined the structure by single-wavelength anomalous diffraction methods with a 1.82 Å resolution dataset from selenomethione-derivatized protein. We identified heavy-atom sites using hkl2map (Pape et al., 2004) (implementing SHELXC and SHELXD (Sheldrick, 2010)), then provided those sites to the PHENIX Autosol wizard (Terwilliger et al., 2009), which uses PHASER (McCoy et al., 2007) for phase calculation and RESOLVE for density modification (Thomas C Terwilliger, 2003) and automated model building (T C Terwilliger, 2003). We manually rebuilt the model in COOT (Emsley et al., 2010), followed by refinement in phenix.refine (Afonine et al., 2012) using positional, individual B-factor, and TLS refinement (statistics in **Table A.B.S1**).

We obtained crystals of *E. coli* NucC bound to 5′-pApA or cAAA in hanging drop

format by mixing protein (8-10 mg/mL) in crystallization buffer plus 0.1 mM 5′-pApA (Invivogen) or cAAA 1:1 with well solution containing 17-24% PEG 3350, 0.1 M Na/K tartrate, and 25 mM Tris-HCl pH 8.0. For cryoprotection, we added 10% glycerol, then flash-froze crystals in liquid nitrogen and collected diffraction data on beamline 24ID-E at the Advanced Photon Source at Argonne National Lab (support statement below). We processed datasets with the RAPD data-processing pipeline, which uses XDS, AIMLESS, and TRUNCATE as above. We determined the structures by molecular replacement using the structure of Apo NucC, manually rebuilt the model in COOT, and refined in phenix.refine using positional, individual B-factor, and TLS refinement (statistics in **Table A.B.S1**). For both 5′-pApA and cAAA, ligands were fully built and assigned an occupancy of 0.33 to account for their location on a three-fold crystallographic rotation axis. The ligands were assigned a "custom-nonbonded-symmetry-exclusion" in phenix.refine to allow symmetry overlaps during refinement.

For analysis of *Ec* NucC trimer and hexamer interfaces (**Figure A.B.S2E**), we used the PISA structure-analysis server (Krissinel and Henrick, 2007).

P. aeruginosa NucC

We obtained crystals of *P. aeruginosa* ATCC27853 NucC D73N mutant bound to cAAA by mixing protein (12 mg/mL) in crystallization buffer plus 0.1 mM cAAA 1:1 with well solution containing 0.1 M HEPES pH 7.5, 0.1 M L-proline, and 10% PEG 3350. Crystals were cryoprotected with 20% glycerol, flash-frozen in liquid nitrogen, and a 1.45 Å diffraction dataset was collected at beamline 8.3.1 at the Advanced Light Source at Lawrence Berkeley National Lab. We used XDS (Kabsch, 2010) for data indexing

and reduction, scaling, and conversion to structure factors, and determined the structure by molecular replacement in PHASER using the structure of cAAA-bound *Ec* NucC as a search model. We manually rebuilt the model in COOT, and refined in phenix.refine using positional, individual B-factor, and TLS refinement (statistics in **Table A.B.S1**).

V. metoecus NucC

For *Vm* NucC hexamer crystals, *Vm* NucC (20 mg/mL) in crystallization buffer plus 0.1 mM cAAA was mixed with well solution containing 0.1 M HEPES pH 7.5, 0.1 M sodium formate, and 14% PEG 3350 in hanging-drop format. Crystals were cryoprotected with the addition of 20% glycerol and flash-frozen in liquid nitrogen. Diffraction data were collected to 1.65 Å resolution at the Advanced Light Source (Berkeley, CA) beamline 8.3.1, and processed in space group $P6_322$ by XDS. Molecular replacement was performed with PHASER (McCoy et al., 2007) using the structure of *E. coli* MS115-1 NucC. The model was manually rebuilt in COOT (Emsley et al., 2010), followed by refinement in phenix.refine (Afonine et al., 2012) using positional, individual B-factor, and TLS refinement (statistics in **Table A.B.S1**).

For *Vm* NucC Apo trimer crystals, *Vm* NucC (20 mg/mL) in crystallization buffer (25 mM Tris-HCl, 0.2 M NaCl, 5 mM $MgCl_2$, 1 mM Tris-carboxyethylphosphine (TCEP)) was mixed with well solution containing 0.1 M sodium acetate pH 4.5 and 1.8 M ammonium sulfate in hanging-drop format. Crystals were cryoprotected with the addition of 30% glycerol and flash-frozen in liquid nitrogen. Diffraction data were collected to 2.15 Å resolution at the Advanced Light Source (Berkeley, CA) beamline 8.3.1, and processed in space group $P4_12_12$ by XDS. Molecular replacement was performed with

PHASER using a single chain from the *Vm* NucC Apo hexamer structure, then the model was rebuilt in COOT and refined in phenix.refine.

Synchrotron Support Statements

SSRL Support Statement Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research, and by the National Institutes of Health, National Institute of General Medical Sciences (including P41GM103393). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of NIGMS or NIH.

APS NE-CAT Support Statement This work is based upon research conducted at the Northeastern Collaborative Access Team beamlines, which are funded by the National Institute of General Medical Sciences from the National Institutes of Health (P41 GM103403). The Eiger 16M detector on 24-ID-E beam line is funded by a NIH-ORIP HEI grant (S10OD021527). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357.

ALS Beamline 8.3.1 Support Statement Beamline 8.3.1 at the Advanced Light Source is operated by the University of California Office of the President, Multicampus Research Programs and Initiatives grant MR-15-328599 the National Institutes of Health

(R01 GM124149 and P30 GM124169), Plexxikon Inc. and the Integrated Diffraction Analysis Technologies program of the US Department of Energy Office of Biological and Environmental Research. The Advanced Light Source (Berkeley, CA) is a national user facility operated by Lawrence Berkeley National Laboratory on behalf of the US Department of Energy under contract number DE-AC02-05CH11231, Office of Basic Energy Sciences.

Nuclease assays

For all nuclease assays, a λ-BstEII ladder (New England Biolabs) or UC Berkeley Macrolab plasmid 2AT (Addgene #29665; 4,731 bp; sequence in Supplemental Information) was used. *Ec* NucC (10 nM unless otherwise indicated) and second messenger molecules (10 nM unless otherwise indicated) were mixed with 1 μg DNA in a buffer containing 10 mM Tris-HCl (pH 7.5), 25 mM NaCl, 10 mM $MgCl_2$, and 1 mM DTT (50 μL reaction volume), incubated 10 min at 37°C, then separated on a 1.2% agarose gel. Gels were stained with ethidium bromide and imaged by UV illumination.

High-throughput sequencing and data analysis

For high-throughput analysis of NucC products, 1 μg of vector 2AT (see **Methods S1** for sequence) was digested to completion by *Ec* NucC, separated by agarose gel, then the prominent ~50 bp band was removed from the gel and gel-purified (Qiagen Minelute). 80 ng of purified DNA was prepared for sequencing using the Swift Biosciences Accel-NGS 1S Plus DNA Library kit, then sequenced on an Illumina HiSeq 4000 (paired-end, 100 base reads), yielding ~402 million read pairs. The Accel-NGS 1S

kit: (1) separates DNA fragments into single strands; (2) ligates the Adapter 1 oligo to the 3′ end of the single-stranded fragment using an “adaptase” step that adds ~8 bp of low-complexity sequence to the fragment before adapter ligation; (3) ligates the Adapter 2 oligo to the 5′ end of the original fragment. After sequencing, read 1 (R1) represents the 5′ end of the original fragment, and read 2 (r2) represents the 3′ end (preceded by ~ 8 bp of low-complexity sequence resulting from the adaptase step).

For NucC fragment length analysis, we took advantage of the fact that most NucC products are less than 100 base pairs in length. We identified the (reverse complement of the) 3′ end of Adapter 1 (full sequence: 5′ AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT 3′) in R1, searching for the sequence “AGATCGGA”. We calculated the length of each NucC fragment as: (length of R1 prior to AGATCGGA) minus 8 (average length of adaptase-added low-complexity sequence). The graph in **Figure 2.5E** represents the lengths of 343 million NucC fragments.

For NucC cleavage site identification (**Figure 2.5D**), we truncated a 1-million read subset of R1 to 15 bases, then mapped to the 2AT plasmid sequence (with ends lengthened by 100 bases to account for circularity of the plasmid) using bowtie (Langmead et al., 2009) with the command: “bowtie -a -n 0 -m 3 -best -t”. Of the successfully-mapped reads (594,183 of 1 million, 59.4%), 288,609 reads mapped to the forward strand and 305,574 mapped to the reverse strand. We extracted 20 bp of sequence surrounding each mapped 5′ end (first base of R1), and input the resulting

alignment of 594,183 sequences into WebLogo 3 (http://weblogo.threeplusone.com) (Crooks et al., 2004).

Identification of CRISPR-associated NucC homologs

To identify NucC homologs associated with Type III CRISPR/Cas systems, we first used the Integrated Microbial Genomes (IMG) database at the DOE Joint Genome Institute (https://img.jgi.doe.gov) to identify NucC homologs throughout bacteria. We used four diverged NucC proteins as search queries to assemble a group of 744 likely NucC proteins, each of which contained the conserved active-site motif IDx$_{30}$EAK and the characteristic gate loop. We downloaded the genomic sequence +/- 10 kb for each NucC homolog, then used NCBI Genome Workbench (https://www.ncbi.nlm.nih.gov/tools/gbench/) to perform a custom TBLASTN searches for proteins related to the Type III CRISPR protein Cmr2. We verified each hit by examining its gene neighborhood for a complete Type III CRISPR/Cas system. Verified hits were included in a set of 531 bacterial NucC proteins, aligned using MAFFT, then an average-distance tree (using the BLOSUM62 scoring matrix) was generated in Jalview (Waterhouse et al., 2009) and visualized in Dendroscope (Huson and Scornavacca, 2012). See **Table A.B.S4** for the complete list of Type III CRISPR associated NucC proteins, and **Table A.B.S5** for the complete list of bacterial NucC proteins used for the phylogenetic tree in **Figure 2.6A**.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis of primary diffraction datasets was performed with standard measures including R_{sym}, R_{meas}, and $CC_{1/2}$ to determine data quality and resolution cutoffs. Statistical analysis of refined crystal structures used standard measures including R and R_{free} factors, and geometry analysis by the MolProbity server (Chen et al., 2010).

DATA AND CODE AVAILABILITY

All primary diffraction datasets are available at the SBGrid Data Bank (https://data.sbgrid.org) under the accession numbers noted in **Table A.B.S1** and the **Key Resources Table**. All reduced diffraction datasets and refined molecular coordinates are available at the RCSB Protein Data Bank (https://www.rcsb.org) under the accession numbers noted in **Table A.B.S1** and the **Key Resources Table**. The original imaging data for Figure 1 and Figure S1 is available at Mendeley Data (https://data.mendeley.com) under DOI 10.17632/dr8mjtwh2d.1. All high-throughput sequencing data used for **Figure 2.5** is available from the Lead Contact without restriction.

References

Afonine, P. V, Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H., Adams, P.D., IUCr, 2012. Towards automated crystallographic structure refinement with phenix.refine. Acta crystallographica Section D, Biological crystallography 68, 352–367. https://doi.org/10.1107/S0907444912001308

Arber, W., Linn, S., 1969. DNA Modification and Restriction. Annual Review of Biochemistry 38, 467–500. https://doi.org/10.1146/annurev.bi.38.070169.002343

Burroughs, A.M., Zhang, D., Schäffer, D.E., Iyer, L.M., Aravind, L., 2015. Comparative genomic analyses reveal a vast, novel network of nucleotide-centric systems in biological conflicts, immunity and signaling. Nucleic acids research 43, gkv1267--10654. https://doi.org/10.1093/nar/gkv1267

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

Cohen, D., Melamed, S., Millman, A., Shulman, G., Oppenheimer-Shaanan, Y., Kacen, A., Doron, S., Amitai, G., Sorek, R., 2019. Cyclic GMP-AMP signaling protects bacterial against viral infection. Nature 574, 691–695. https://doi.org/10.1038/s41586-019-1605-5

Crooks, G.E., Hon, G., Chandonia, J.-M., Brenner, S.E., 2004. WebLogo: a sequence logo generator. Genome research 14, 1188–90. https://doi.org/10.1101/gr.849004

Deibert, M., Grazulis, S., Sasnauskas, G., Siksnys, V., Huber, R., 2000. Structure of the tetrameric restriction endonuclease NgoMIV in complex with cleaved DNA. Nature structural biology 7, 792–9. https://doi.org/10.1038/79032

Dunten, P.W., Little, E.J., Gregory, M.T., Manohar, V.M., Dalton, M., Hough, D., Bitinaite, J., Horton, N.C., 2008. The structure of SgrAI bound to DNA; recognition of an 8 base pair target. Nucleic Acids Research 36, 5405–5416. https://doi.org/10.1093/nar/gkn510

Economou, A., Pogliano, J.A., Beckwith, J., Oliver, D.B., Wickner, W., 1995. SecA membrane cycling at SecYEG is driven by distinct ATP binding and hydrolysis events and is regulated by SecD and SecF. Cell 83, 1171–81.

Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K., 2010. Features and development of Coot. Acta crystallographica Section D, Biological crystallography 66, 486–501. https://doi.org/10.1107/S0907444910007493

Evans, P., 2006. Scaling and assessment of data quality. Acta crystallographica Section D, Biological crystallography 62, 72–82. https://doi.org/10.1107/S0907444905036693

Evans, P.R., Murshudov, G.N., 2013. How good are my data and what is the resolution? Acta crystallographica Section D, Biological crystallography 69, 1204–1214. https://doi.org/10.1107/S0907444913000061

Grüschow, S., Athukoralage, J.S., Graham, S., Hoogeboom, T., White, M.F., 2019. Cyclic oligoadenylate signalling mediates Mycobacterium tuberculosis CRISPR defence. Nucleic Acids Research. https://doi.org/10.1093/nar/gkz676

Han, W., Stella, S., Zhang, Y., Guo, T., Sulek, K., Peng-Lundgren, L., Montoya, G., She, Q., 2018. A Type III-B Cmr effector complex catalyzes the synthesis of cyclic oligoadenylate second messengers by cooperative substrate binding. Nucleic Acids Research 46, 10319–10330. https://doi.org/10.1093/nar/gky844

Holm, L., Rosenström, P., 2010. Dali server: conservation mapping in 3D. Nucleic acids research 38, W545--9. https://doi.org/10.1093/nar/gkq366

Huson, D.H., Scornavacca, C., 2012. Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. Systematic biology 61, 1061–1067. https://doi.org/10.1093/sysbio/sys062

Kabsch, W., 2010. XDS. Acta crystallographica Section D, Biological crystallography 66, 125–132. https://doi.org/10.1107/S0907444909047337

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

Kirchberger, P.C., Turnsek, M., Hunt, D.E., Haley, B.J., Colwell, R.R., Polz, M.F., Tarr, C.L., Boucher, Y., 2014. Vibrio metoecus sp. nov., a close relative of Vibrio cholerae isolated from coastal brackish ponds and clinical specimens.

INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY 64, 3208–3214. https://doi.org/10.1099/ijs.0.060145-0

Koonin, E. V, Makarova, K.S., Zhang, F., 2017. Diversity, classification and evolution of CRISPR-Cas systems. Current opinion in microbiology 37, 67–78. https://doi.org/10.1016/j.mib.2017.05.008

Kovall, R., Matthews, B.W., 1997. Toroidal structure of lambda-exonuclease. Science (New York, N.Y.) 277, 1824–7.

Krissinel, E., Henrick, K., 2007. Inference of Macromolecular Assemblies from Crystalline State. Journal of Molecular Biology 372, 774–797. https://doi.org/10.1016/j.jmb.2007.05.022

Langmead, B., Trapnell, C., Pop, M., Salzberg, S.L., 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology 10, R25. https://doi.org/10.1186/gb-2009-10-3-r25

Makarova, K.S., Zhang, F., Koonin, E. V, 2017. SnapShot: Class 1 CRISPR-Cas Systems. Cell 168, 946-946.e1. https://doi.org/10.1016/j.cell.2017.02.018

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., Read, R.J., 2007. Phaser crystallographic software. Journal of applied crystallography 40, 658–674. https://doi.org/10.1107/S0021889807021206

Mcmahon, S.A., Zhu, W., Graham, S., Rambo, R., White, M.F., Gloster, T.M., 2019. Structure and mechanism of a Type III CRISPR defence DNA nuclease activated by cyclic oligoadenylate. bioRxiv 14–17. https://doi.org/10.1101/784280

Meeske, A.J., Nakandakari-Higa, S., Marraffini, L.A., 2019. Cas13-induced cellular dormancy prevents the rise of CRISPR-resistant bacteriophage. Nature. https://doi.org/10.1038/s41586-019-1257-5

Miyazono, K., Ishino, S., Tsutsumi, K., Ito, T., Ishino, Y., Tanokura, M., 2015. Structural basis for substrate recognition and processive cleavage mechanisms of the trimeric exonuclease PhoExo I. Nucleic Acids Research 43, 7122–7136. https://doi.org/10.1093/nar/gkv654

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

Pape, T., Schneider, T.R., IUCr, 2004. HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs. Journal of applied crystallography 37, 843–844. https://doi.org/10.1107/S0021889804018047

Pingoud, A., Fuxreiter, M., Pingoud, V., Wende, W., 2005. Type II restriction endonucleases: structure and mechanism. Cellular and Molecular Life Sciences 62, 685–707. https://doi.org/10.1007/s00018-004-4513-1

Pogliano, J., Osborne, N., Sharp, M.D., Abanes-De Mello, A., Perez, A., Sun, Y.-L., Pogliano, K., 1999. A vital stain for studying membrane dynamics in bacteria: a novel mechanism controlling septation during Bacillus subtilis sporulation. Molecular Microbiology 31, 1149–1159. https://doi.org/10.1046/j.1365-2958.1999.01255.x

Pyenson, N.C., Marraffini, L.A., 2017. Type III CRISPR-Cas systems: when DNA cleavage just isn't enough. Current Opinion in Microbiology 37, 150–154. https://doi.org/10.1016/J.MIB.2017.08.003

Rostøl, J.T., Marraffini, L.A., 2019. Non-specific degradation of transcripts promotes plasmid clearance during type III-A CRISPR–Cas immunity. Nature Microbiology 4, 656–662. https://doi.org/10.1038/s41564-018-0353-x

Shah, S.A., Alkhnbashi, O.S., Behler, J., Han, W., She, Q., Hess, W.R., Garrett, R.A., Backofen, R., 2019. Comprehensive search for accessory proteins encoded with archaeal and bacterial type III CRISPR- *cas* gene cassettes reveals 39 new *cas* gene families. RNA Biology 16, 530–542. https://doi.org/10.1080/15476286.2018.1483685

Sheldrick, G.M., 2010. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. Acta crystallographica Section D, Biological crystallography 66, 479–485. https://doi.org/10.1107/S0907444909038360

Sorek, R., Lawrence, C.M., Wiedenheft, B., 2013. CRISPR-Mediated Adaptive Immune Systems in Bacteria and Archaea. Annual Review of Biochemistry 82, 237–266. https://doi.org/10.1146/annurev-biochem-072911-172315

Terwilliger, Thomas C, 2003. SOLVE and RESOLVE: automated structure solution and density modification. Methods in enzymology 374, 22–37. https://doi.org/10.1016/S0076-6879(03)74002-6

Terwilliger, T C, 2003. Automated main-chain model building by template matching and iterative fragment extension. Acta crystallographica Section D, Biological crystallography 59, 38–44. https://doi.org/10.1107/S0907444902018036

Terwilliger, T.C., Adams, P.D., Read, R.J., McCoy, A.J., Moriarty, N.W., Grosse-Kunstleve, R.W., Afonine, P. V, Zwart, P.H., Hung, L.-W., 2009. Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. Acta crystallographica Section D, Biological crystallography 65, 582–601. https://doi.org/10.1107/S0907444909012098

Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., Barton, G.J., 2009. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics (Oxford, England) 25, 1189–1191. https://doi.org/10.1093/bioinformatics/btp033

Watson, B.N.J., Vercoe, R.B., Salmond, G.P.C., Westra, E.R., Staals, R.H.J., Fineran, P.C., 2019. Type I-F CRISPR-Cas resistance against virulent phage infection triggers abortive infection and provides population-level immunity. bioRxiv 679308. https://doi.org/10.1101/679308

Whiteley, A.T., Eaglesham, J.B., de Oliveira Mann, C.C., Morehouse, B.R., Lowey, B., Nieminen, E.A., Danilchanka, O., King, D.S., Lee, A.S.Y., Mekalanos, J.J., Kranzusch, P.J., 2019. Bacterial cGAS-like enzymes synthesize diverse nucleotide signals. Nature 567, 194–199. https://doi.org/10.1038/s41586-019-0953-5

Wilson, G.G., Murray, N.E., 1991. Restriction and Modification Systems. Annual Review of Genetics 25, 585–627. https://doi.org/10.1146/annurev.ge.25.120191.003101

Wright, A.V., Nuñez, J.K., Doudna, J.A., 2016. Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. Cell 164, 29–44. https://doi.org/10.1016/J.CELL.2015.12.035

Ye, Q., Lau, R.K., Mathews, I.T., Watrous, J.D., Azimi, C.S., Jain, M., Corbett, K.D., 2019. HORMA domain proteins and a Trip13-like ATPase regulate bacterial cGAS-like enzymes to mediate bacteriophage immunity. Submitted.

Zhang, J., McCabe, K.A., Bell, C.E., 2011. Crystal structures of lambda exonuclease in complex with DNA suggest an electrostatic ratchet mechanism for processivity. Proceedings of the National Academy of Sciences of the United States of America 108, 11872–7. https://doi.org/10.1073/pnas.1103467108

Chapter 2, in full, is a reprint of the material as it appears in Molecular Cell 2020. Lau, Rebecca K.; Ye, Qiaozhen; Birkholz, Erica A.; Berg, Kyle R.; Patel, Lucas; Mathews, Ian T.; Watrous, Jeramie D.; Ego, Kaori; Whiteley, Aaron T.; Lowey, Brianna; Mekalanos, John J.; Kranzusch, Philip J.; Jain, Mohit; Pogliano, Joe; Corbett, Kevin D. "Structure and Mechanism of a Cyclic Trinucleotide-Activated Bacterial Endonuclease Mediating Bacteriophage Immunity". The dissertation/thesis author was a primary investigator and co-first author of this paper.

CHAPTER 3: A conserved signaling pathway activates bacterial CBASS immune signaling in response to DNA damage

Abstract

To protect themselves from the constant threat of bacteriophage (phage) infection, bacteria have evolved diverse immune systems including restriction-modification, CRISPR-Cas, and many others. Here we describe the discovery of a two-protein transcriptional regulator module associated with hundreds of CBASS immune systems, and demonstrate that this module drives expression of its associated CBASS system in response to DNA damage. We show that the helix-turn-helix transcriptional repressor CapH binds the promoter region of its associated CBASS system to repress transcription until it is cleaved by the metallopeptidase CapP. CapP is activated in vitro by single-stranded DNA, and in cells by DNA-damaging drugs. Together, CapH and CapP drive increased expression of their associated CBASS system in response to DNA damage. We identify CapH and CapP-related proteins associated with diverse known and putative bacterial immune systems including DISARM and Pycsar anti-phage operons. Overall, our data highlight a mechanism by which bacterial immune systems can sense and respond to a universal signal of cell stress, potentially enabling multiple immune systems to mount a coordinated defensive response against an invading pathogen.

Introduction

In all organisms, survival depends on the ability of cells to sense and respond to both internal and external threats. In addition to environmental stress, bacteria are continually challenged by bacteriophages (phages), and have evolved a wide array of immune systems to protect themselves from phage infection and propagation. Many anti-phage immune systems, including restriction-modification and CRISPR-Cas systems, specifically recognize and destroy foreign DNA to prevent phage replication (Makarova et al., 2013; Mohanraju et al., 2016). Other immune systems, termed abortive infection systems, sense phage infection and respond by killing the host cell, thereby preventing phage propagation and further infection in the bacterial community (Doron et al., 2018; Dy et al., 2014; Hampton et al., 2020; Makarova et al., 2011a). In many bacteria, multiple immune systems co-exist in so-called “defense islands” (Doron et al., 2018; Makarova et al., 2011b) and may cooperate, with non-lethal systems acting as a first line of defense and abortive infection systems becoming activated only as a last resort (Bernheim and Sorek, 2020; Picton et al., 2022) .

The widespread and functionally-diverse CBASS anti-phage immune systems use an abortive infection mechanism in which a cGAS/DncV-like nucleotidyltransferase (CD-NTase) is activated upon phage infection and synthesizes a cyclic oligonucleotide secondary messenger (Cohen et al., 2019; Lau et al., 2020; Whiteley et al., 2019; Ye et al., 2020). This molecule in turn activates one of a variety of effector proteins, including phospholipases, nucleases, and pore-forming proteins, to kill the host cell (Cohen et al., 2019; Lau et al., 2020; Lowey et al., 2020; Severin et al., 2018). While so-called Type I

CBASS systems encode only a CD-NTase and a cell-killing effector protein, the majority of CBASS systems encode ancillary proteins putatively involved in infection sensing and/or CD-NTase activation (Burroughs et al., 2015; Millman et al., 2020). Type II CBASS systems encode two proteins, Cap2 and Cap3, that are related to eukaryotic ubiquitination machinery and are required for protection against phage (Cohen et al., 2019; Ledvina et al., 2022) . Type III CBASS systems, meanwhile, encode peptide-binding HORMA domain proteins (Cap7 and Cap8) that are proposed to bind specific peptides to sense infection and then activate their associated CD-NTase (Ye et al., 2020).

All CBASS systems are thought to directly sense phage infection and respond by triggering cell death. Here, we identify a pair of transcriptional regulators, termed CapH and CapP, that are associated with hundreds of CBASS systems and upregulate CBASS expression in response to DNA damage. DNA damage is a universal stress signal in bacterial cells (Benler and Koonin, 2020) and has been shown to activate various bacterial stress responses, including the SOS response (Little, 1991; Sassanfar and Roberts, 1990). We show that CapH and CapP are structurally and functionally similar to regulators that mobilize prophages and integrative and conjugative elements (ICE elements) in response to DNA damage, and to activators of the DNA damage response in radiation-resistant *Deinococcus* species. We also identify CapH and CapP-like regulators associated with a variety of known or putative bacterial immune systems, revealing that these proteins represent a conserved signaling module that regulates immune-system expression in response to DNA damage across bacteria.

Results

Identification of *capH* and *capP* genes associated with CBASS systems

We previously showed that a Type III CBASS system from *E. coli* strain MS115-1 provides robust protection against bacteriophage λ through an abortive infection mechanism (Lau et al., 2020; Ye et al., 2020). Examination of this system's genomic neighborhood revealed a pair of genes directly upstream of the core CBASS genes and encoded on the opposite strand (i.e. sharing a promoter region with the core CBASS genes), that encode a predicted helix-turn-helix (HTH) DNA binding protein and a predicted Zn^{2+} metallopeptidase (**Figure 3.1A**). We term these two genes *capH* (CBASS associated protein, Helix-turn-helix) and *capP* (CBASS associated protein, Peptidase). In their position and orientation relative to the core CBASS genes, *capH* and *capP* are similar to *capW*, a transcriptional regulator associated with a distinct subset of CBASS systems (Blankenchip et al., 2022) . BLAST searches revealed that CapP shares strong similarity to IrrE, a metallopeptidase that regulates the DNA damage response in *Deinococcus* by cleaving an HTH-family transcription factor, DdrO (Ludanyi et al., 2014; Vujičić-Žagar et al., 2009). DdrO normally binds the promoters of DNA damage response genes and suppresses their expression, but upon DNA damage IrrE becomes activated and cleaves DdrO, releasing it from DNA and activating expression of the DNA damage response genes (Blanchard et al., 2017; de Groot et al., 2019; Ludanyi et al., 2014). The similarity of CapP to IrrE, and its association with the HTH protein CapH, suggested that CapH and CapP may functionally cooperate to control expression of their associated CBASS system.

We systematically searched the genomic neighborhoods of 6,233 bacterial CBASS CD-NTases (Cohen et al., 2019) for genes related to *capP* and identified 408 CBASS systems with a predicted Zn^{2+} metallopeptidase within 10 kb of the system's CD-NTase gene. In these systems, CapP is most often annotated as a "domain of unknown function" (DUF) 955 or PFAM06114 protein. We manually inspected each system and identified a gene encoding a CapH homolog alongside *capP* in 393 of the 408 systems (in 70 cases, this gene is not annotated; **Figure 3.1A, Table A.C.S1**). The remaining 15 systems encode an apparent fusion of CapH and CapP (**Table A.C.S1**). In all cases, the *capH* and *capP* genes are encoded upstream of the core CBASS genes and on the opposite strand (**Figure 3.1A**). We identified *capH* and *capP* genes associated with Type I, Type II, and Type III CBASS systems that encode a variety of predicted effectors including phospholipases, transmembrane proteins, and endonucleases (**Figure 3.1B-C, Table A.C.S1**). In 24 systems, *capH* and *capP* are encoded alongside a predicted σ70-family σ factor (**Table A.C.S1**).

To determine whether *capH* and *capP* control expression of their associated CBASS operon, we generated a reporter construct with *capH*, *capP*, and the promoter region of the CBASS system from *E. coli* MS115-1, plus a gene encoding GFP (green fluorescent protein) in place of the core CBASS genes. When both *capH* and *capP* were present, expression of GFP in uninfected log-phase cells was too low for detection by anti-GFP immunoblotting (**Figure 3.1D**). Expression was also nearly undetectable in a strain lacking *capP*, but we observed high GFP expression in a strain lacking *capH*

(**Figure 3.1D, Figure A.C.S1A**). These data suggest that CapH is a transcriptional repressor for its associated CBASS system.

Next, we generated a separate reporter construct encoding the full six-gene CBASS system of *E. coli* MS115-1, with a N-terminal FLAG tag fused to the effector nuclease NucC (**Figure 3.1E, Figure A.C.S1B**). In agreement with our GFP reporter, expression of FLAG-NucC was undetectable by Western blot in uninfected cells (**Figure 3.1E**). FLAG-NucC expression was also undetectable after deleting *capP* in this construct (**Figure 3.1E**). We were unable to delete *capH* in this construct, with all isolated clones lacking *capH* also lacking large regions of the core CBASS genes (not shown). Given that deletion of *capH* in our GFP reporter construct results in high expression of GFP, our inability to isolate a *capH*-deleted version of the full CBASS system suggests that high expression of the system is toxic to host cells.

CapP controls CBASS expression but is not required for phage protection

Our reporter assays indicated that CapH likely acts as a transcriptional repressor for its associated CBASS operon. To determine the role of CapP in CBASS expression, we first used our FLAG-NucC reporter system to test for CBASS expression changes upon infection with an obligately lytic variant of bacteriophage λ lacking the cI gene (λ cI^-) (Rajagopala et al., 2011) . With the wild-type CBASS system encoding *capH* and *capP*, we observed a strong increase in FLAG-NucC expression starting ~60 minutes after infection and peaking around 90 minutes after infection (**Figure 3.1E**). In the absence of *capP*, we observed no such increase in FLAG-NucC expression (**Figure 3.1E**). We also observed increased FLAG-NucC expression in a system with

catalytically-dead CD-NTase (CdnC D72N/D74N), indicating that the observed expression changes do not depend on CBASS signaling (**Figure 3.1E**). These data suggest that CapP responds to phage infection by antagonizing CapH, resulting in a loss of repression and an increase in CBASS expression.

To test the role of CapH and CapP in phage protection, we compared the ability of wild-type *E. coli* MS115-1 CBASS and a mutant lacking *capP* to protect against λ cI$^-$ infection. We previously reported that when cloned into an IPTG-inducible expression vector, the four core genes from *E. coli* MS115-1 CBASS (*cdnC* (CD-NTase), *cap7* (HORMA), *cap6* (TRIP13), and *nucC*) provide strong protection against λ cI$^-$ (Lau et al., 2020; Ye et al., 2020) . We found that the native six-gene operon encoding *capH* and *capP* also provides protection against λ cI$^-$, as measured by both a reduction in viral plaque numbers (**Figure 3.1F**) and a reduction in plaque size compared to bacteria lacking CBASS (**Figure 3.1G**). While our prior study showed that the core CBASS genes under IPTG expression control reduced infection by λ cI$^-$ by over six logs (more than 10^6-fold reduction in viral plaques) (Ye et al., 2020) the full system with *capH* and *capP* provides much more modest protection, with less than a two-log reduction in viral plaques compared to a control strain (**Figure 3.1F**). Control infections with systems encoding catalytically-dead CdnC (D72N/D74N) or NucC (D73N) (**Fig A.C.S1C-D**), light microscopy analysis of infected cells (**Figure A.C.S2A-B**), and bacterial growth curves (**Figure A.C.S2C**) all confirmed that this protection was attributable to CBASS function.

Unexpectedly, we found that a mutant CBASS system lacking *capP* and therefore unable to boost CBASS expression after infection (**Figure 3.1E**) provided

protection against phage λ equivalent to the wild-type system (**Figure 3.1F-G**). This finding suggests that the boost in CBASS expression mediated by *capP* comes too late to affect the course of an infection, and that the modest protection we observe is mediated by the low basal levels of CBASS proteins already present in these cells. To confirm that CBASS genes are expressed in uninfected cells, we used qRT-PCR to show that *cdnC* mRNA is present at ~2% of that of the abundant RNA Polymerase gene *rpoA* (ΔCt = -4.8) at basal expression levels (**Figure A.C.S1E**). As we observe for protein levels by Western blot, *cdnC* mRNA levels increase dramatically upon infection with λ cI- (**Figure A.C.S1E**). Overall, these data show that while the native *E. coli* MS115-1 CBASS system does provide modest anti-phage protection, transcriptional regulation by CapH and CapP does not directly contribute to this protection. Thus, *capH* and *capP* may enable their associated CBASS system to play a defensive role other than front-line protection against viral infection.

CapH binds the promoter region of its associated CBASS system

Our reporter assays suggested that CapH acts as a transcriptional repressor for its associated CBASS system, potentially by directly binding the CBASS promoter region. In *E. coli* MS115-1 CBASS, the 182 bp region between the *capH* and *cdnC* genes contains strongly-predicted promoter sequences for both forward (toward *cdnC*) and reverse (toward *capH*) transcription (**Figure 3.2A, Figure A.C.S3A**). We used fluorescence polarization to test CapP binding to a panel of 40-bp DNAs covering this region, and identified two binding sites (Site 1 and Site 2) overlapping the predicted forward and reverse promoter sequences (**Figure 3.2A-B, Figure A.C.S3B-F**). Both

sites are conserved in the promoter regions of related CBASS operons (**Figure A.C.S3A**). Site 2 possesses a perfect 8-bp palindrome separated by 6 bp (**Figure A.C.S3A**), but Site 1 does not possess any recognizable palindromes or tandem repeats. Nonetheless, CapH binding to both sites is best fit by a cooperative binding model with a Hill coefficient of ~2, suggesting cooperative binding of multiple CapH monomers to each site (**Figure 3.2B, Figure A.C.S3B**).

Sequence analysis and 3D structure predictions indicate that CapH possesses an N-terminal helix-turn-helix (HTH) DNA binding domain followed by a short flexible linker and two conserved α-helices that may constitute an oligomerization domain (**Figure 3.2C**). While we were unable to crystallize full-length CapH, likely due to flexibility of the inter-domain linker, we crystallized and determined a 1.02 Å-resolution crystal structure of the protein's isolated HTH domain (residues 2-67) (**Figure 3.2C, Table A.C.S2**). The CapH HTH domain forms a canonical HTH fold, and modeling a DNA-bound complex based on known HTH-DNA complexes revealed several conserved residues on the DNA-binding face that that may be involved in DNA binding, including Ser32, Arg40, and Arg44 (**Figure 3.2C, Figure A.C.S3G**). We found that mutation of Ser32 completely eliminated CapH binding to the Site 2 DNA, and that mutation of either Arg40 or Arg44 reduced, but did not eliminate, binding (**Figure 3.2D**). We tested the same mutations in our GFP reporter system, and found that mutation of Ser32 resulted in high expression of GFP (**Figure 3.2E**). These data are consistent with CapH acting as a transcriptional repressor for its associated CBASS operon.

CapH oligomerization is required for DNA binding

In the *Deinococcus* DdrO-IrrE system, DdrO forms a homodimer through its C-terminal domain and this dimerization is required for DNA binding and transcriptional repression by the protein (de Groot et al., 2019; Ludanyi et al., 2014). To test whether CapH forms an oligomer, we used size exclusion chromatography coupled to multi-angle light scattering (SEC-MALS). We found that full-length CapH forms an oligomer with an overall size consistent with either a homotrimer or a mixture of dimers and tetramers (**Figure 3.3A-B, Figure A.C.S6A**). The isolated C-terminal region of CapH (residues 67-107; CapH^{CTD}) forms a similar oligomer, while the isolated N-terminal HTH domain is monomeric (**Figure 3.3B**). These data show that CapH oligomerizes through its C-terminal domain. To determine the mechanism of oligomerization, we crystallized and determined a 1.75 Å-resolution crystal structure of CapH^{CTD} (**Figure 3.3C, Table A.C.S2**). In the structure, four CapH^{CTD} protomers form a homotetramer with a dimer-of-dimers architecture. Each CapH^{CTD} protomer forms two α-helices that fold into a V shape, with the homodimer assembled by two protomers arranged antiparallel to one another with the V shapes interlocked. The CapH^{CTD} homodimer is stabilized by a hydrophobic core comprising Phe81, Tyr85, Leu96, and Leu100 of each protomer (**Figure 3.3C**). The CapH^{CTD} homodimer resembles the C-terminal dimerization domains of other dimeric bacterial transcription factors, including *Mycobacterium tuberculosis* EspR, *Bacillus subtilis* SinR, and *Citrobacter* C.Csp231I (Gangwar et al., 2014; Lewis et al., 1998; Shevtsov et al., 2015). The CapH^{CTD} homotetramer is assembled through a separate hydrophobic interface between the C-terminal α-helices of four protomers, involving residues Ile99 and Phe103 (**Figure 3.3C**).

During the course of structure determination for CapH^{CTD}, we generated a construct with a mutation of Ile99 to methionine (CapH^{CTD}(I99M)). We determined a 1.26 Å-resolution structure of this mutant, revealing a CapH homodimer equivalent to our structure of wild-type CapH^{CTD}, but lacking the tetrameric assembly (**Figure A.C.S6B**). Consistent with this finding, SEC-MALS showed that CapH(I99M) forms a stable homodimer in solution, rather than the larger oligomer observed with wild-type CapH (**Figure A.C.S6B**). Thus, the I99M mutant disrupts CapH tetramerization, but not dimerization.

We next tested the role of CapH oligomerization in DNA binding. We used fluorescence polarization to compare the DNA binding affinity of full length wild-type CapH to that of the CapH HTH domain (residues 2-67; CapH^{NTD}), which forms a monomer; and to the CapH(I99M) mutant, which forms a homodimer. We observed no binding of CapH^{NTD} to DNA, demonstrating that CapH oligomerization is required for DNA binding (**Figure 3.3D**). With both Site 1 and Site 2 DNAs, CapH(I99M) showed only a slight reduction in DNA binding affinity and cooperativity compared to wild-type CapH (**Figure 3.3E**). An electrophoretic mobility shift assay (EMSA) with Site 2 DNA showed that while wild-type CapH shows two shifted bands, CapH(I99M) shows only one (**Figure A.C.S6C**). Despite the reduced DNA binding affinity and inability to form tetramers, CapH(I99M) effectively suppressed expression of our GFP reporter system (**Figure A.C.S1A**). Together, these data suggesting that while CapH can form tetramers and that tetramer formation does play a minor role in DNA binding, tetramer formation is

not necessary for high-affinity DNA binding and suppression of CBASS gene expression.

The structure of CapP reveals an internal cysteine switch

In the *Deinococcus* DNA damage response pathway, cleavage of DdrO by the metallopeptidase IrrE results in loss of DNA binding by DdrO, enabling increased expression of DNA damage response genes (de Groot et al., 2019; Ludanyi et al., 2014). Our data on DNA binding and CBASS repression by CapH, and in particular the importance of CapH oligomerization for DNA binding, suggests a functional parallel between DdrO-IrrE and CapH-CapP. To better understand this relationship, we purified and determined a 1.35 Å-resolution crystal structure of CapP from a CBASS system in *Thauera* sp. K11 (56% identical to *E. coli* MS115-1 CapP; **Figure A.C.S4A**). The overall structure of CapP resembles that of IrrE, with the protein folding into three domains: an N-terminal Zn^{2+} metallopeptidase domain, a central linker domain with topology resembling a helix-turn-helix domain, and a C-terminal GAF domain (**Figure 3.4A-B**). The N-terminal domain closely resembles other HExxH Zn^{2+} metallopeptidases including IrrE, with five α-helices and a three-stranded β-sheet. A Zn^{2+} ion is coordinated in the conserved active site by residues His96, His100, and Glu129 (**Figure 3.4C-E**). The predicted active-site glutamate residue, Glu97, is positioned close by but not directly coordinating the bound Zn^{2+} ion. Instead, a conserved cysteine residue, Cys113, completes the coordination of the bound Zn^{2+} ion. Cys113 is located on an insertion in the metallopeptidase domain, on a fourth β-strand that drapes over the active site in the same position that substrate peptides bind in related Zn^{2+}

metallopeptidases (Cerdà-Costa and Xavier Gomis-Rüth, 2014). We term this region the cysteine switch loop, after the cysteine switch motif found in matrix metalloproteases. These enzymes are synthesized as inactive precursors with an N-terminal domain bearing a conserved cysteine residue (the cysteine switch) that coordinates the active-site Zn^{2+} ion and inhibits activity (**Figure A.C.S4B-C**). The protease is only activated upon proteolytic cleavage and dissociation of the cysteine switch domain (Cerdà-Costa and Xavier Gomis-Rüth, 2014; Springman et al., 1990; Van Wart and Birkedal-Hansen, 1990). Of the 408 CapP proteins associated with CBASS systems, 134 (33%) possess the cysteine switch loop, including *E. coli* MS115-1 CapP (**Figure 3.4F, Figure A.C.S4D**).

The presence of the cysteine switch loop in CapP suggests that the protein's peptidase activity is tightly controlled, perhaps by a conformational change that induces dissociation of the cysteine switch from the CapP active site and allows substrate binding. In *Deinococcus* IrrE, the protein's peptidase activity is thought to be activated upon DNA damage by the binding of an unknown ligand to the protein's C-terminal GAF domain, although IrrE does not contain a cysteine switch loop (Vujičić-Žagar et al., 2009). In other proteins, GAF domains are known to bind nucleotide-based second messengers including cyclic GMP, which binds the GAF domain of phosphodiesterase 6C and allosterically regulates its enzymatic activity (Gross-Langenhoff et al., 2006; Ho et al., 2000; Levdikov et al., 2009; Martinez et al., 2008, 2002b, 2002a). When we compared the structure of CapP to that of cyclic GMP-bound phosphodiesterase 6C, we observed that CapP possesses a large number of surface-exposed aromatic residues

near the putative ligand-binding site (**Figure A.C.S7**). If CapP is allosterically regulated through the GAF domain, these residues may be involved in the binding of nucleotide-based ligand(s).

CapP cleaves CapH when stimulated by single-stranded DNA

We next sought to directly test whether CapP cleaves CapH. Our initial tests using purified proteins in vitro showed no CapP-mediated cleavage of CapH (not shown), so we instead developed an assay to detect CapP activity in *E. coli* cells. We co-expressed CapP with a fusion protein comprising CapH with an N-terminal maltose binding protein (MBP) tag and a C-terminal GFP tag, and used an anti-GFP western blot to detect CapP activity. We detected a minor band representing a CapP cleavage product in the presence of wild-type CapP, but not when the CapP active site was mutated (Glu98 to glutamine; E98Q) (**Figure 3.5A**). At ~32 kDa, this band likely represents a product of CapP cleavage near the C-terminus of CapH.

Since we observed CapP-mediated CapH cleavage in cells, but not with purified proteins, we theorized that a soluble ligand present in cells is responsible for activating CapP. To test this idea, we performed in vitro cleavage assays with purified CapP and MBP-CapH-GFP in the presence of *E. coli* cell lysate. We observed robust activation of CapP in the presence of *E. coli* lysate, and stronger activation after the lysate was boiled and centrifuged to remove all proteins (**Figure A.C.S5A**). By fractionating boiled cell lysate using anion-exchange and size-exclusion chromatography, we found that CapP was most likely activated by a large, negatively-charged macromolecule (**Figure A.C.S5B**). Based on this finding, we tested whether DNA and/or RNA could directly

stimulate CapP. We found that single-stranded DNA strongly activates CapP (**Figure 3.5B**), while double-stranded DNA weakly activates CapP (**Figure 3.5B**), and single-stranded RNA does not activate CapP (**Figure A.C.S5C**). We found that single-stranded DNA as short as 5 bases long stimulated CapP, and that pyrimidines – particularly thymine – have the strongest stimulatory effect (**Figure A.C.S5C**). Finally, we found that CapP directly binds a single-stranded DNA oligonucleotide in vitro, but not an equivalent-length double-stranded DNA (**Figure A.C.S5D**). Further, CapP strongly binds poly-T DNA, weakly binds poly-C, and does not bind poly-A (**Figure A.C.S5E**). These data support a model in which CapP's peptidase activity is stimulated by the binding of single-stranded DNA, particularly T-rich DNA.

We isolated the C-terminal product from CapP-mediated cleavage of MBP-CapH-GFP, and subjected it to Edman degradation to map the CapP cleavage site (**Figure 3.5C, Figure A.C.S8**). We identified the cleavage site as between Phe82 and Arg83, within the first α-helix of the CapH C-terminal dimerization domain (**Figure 3.5C-D**). Confirming this finding, we found that a CapH Arg83 to alanine (R83A) mutant is not cleaved by CapP in vitro (**Figure 3.5E**). Finally, we found that in cells, truncation of CapH at residue 82 – mimicking CapP-cleaved CapH – resulted in strong expression as measured by both our GFP reporter system (**Figure 3.5F**) and our FLAG-NucC reporter (**Figure 3.5G**).

CapP is activated by DNA damage

Our data showing that CapP is directly stimulated by single-stranded DNA in vitro suggests that in cells, it is activated by DNA damage. To test this idea, we coexpressed

CapP and MBP-CapH-GFP in *E. coli* in the presence or absence of DNA damaging drugs. We found that in the presence of zeocin, a drug that induces DNA double-strand breaks (Chankova et al., 2007), CapH cleavage was stimulated (**Figure 3.6A**). Using our FLAG-NucC reporter system, we found that both zeocin and mitomycin C, which causes DNA double-strand breaks through a separate mechanism (Ayora et al., 2011; Kidane et al., 2004), strongly stimulate CBASS expression within 60-120 minutes of adding the drugs (**Figure 3.6B-C**). qRT-PCR for *cdnC* mRNA confirms this finding, and shows that addition of zeocin boosts CBASS expression more strongly than phage λ infection (**Figure A.C.S1E**). This boost in CBASS expression was not observed in cells lacking *capP* (**Figure 3.6B-C**). Thus, CapH and CapP mediate increased CBASS expression upon DNA damage, even in the absence of phage infection. Curiously, this increased CBASS expression does not result in increased cell death, as measured by minimum inhibitory concentration (MIC) analysis of zeocin for cells with and without the *E. coli* MS115-1 CBASS system (**Figure A.C.S1F**). This finding suggests that despite being highly expressed upon DNA damage, the CBASS system may still require a phage trigger to activate second messenger production and in turn activate NucC.

CapH and CapP are members of a broadly conserved family of DNA damage response proteins

CapH and CapP show strong structural and functional similarity to the *Deinococcus* proteins DdrO and IrrE, with both systems inducing expression of target genes upon DNA damage through metallopeptidase cleavage of a transcriptional repressor (**Figure 3.6D-E**). DNA damage is a universal signal of cell stress, and as such

is a major signal to induce lysogenic phage (prophage) to switch to the lytic life cycle, and to induce mobility of integrative and conjugative elements (ICE elements) (Auchtung et al., 2005; Baek et al., 2003). Strikingly, the use of an HTH family transcriptional repressor coupled with a DNA damage-stimulated metallopeptidase is shared in some prophages and ICE elements. For example, *Bacillus subtilis* mobile element ICE*Bs*1 and prophage ϕ105 each encode an HTH-family transcriptional repressor (ImmR and cϕ105, respectively) that strongly represses the expression of genes responsible for excision of these elements from the genome, and a Zn^{2+} metallopeptidase (ImmA) that cleaves the HTH protein upon DNA damage to relieve repression and induce excision (**Figure 3.6E**) (Bose et al., 2008; Bose and Grossman, 2011).

The structural and functional parallels between CapH/CapP, DdrO/IrrE, and ImmR/ImmA suggest that these proteins represent a broadly conserved family of DNA damage responsive transcriptional regulators. We used the FlaGs (Flanking Genes) server to search for other instances of CapH/CapP-like proteins and identify their associated operons (Saha et al., 2021). We identified a broad range of operons associated with *capH* and *capP*-like genes, with all of them sharing the location of *capH* and *capP* upstream of, and oriented on the opposite strand as, the associated operon (**Figure 3.6F**). Most of these systems represent known or putative defense systems, notably including a group of DISARM anti-phage systems associated with a *capH-capP* fusion gene, and a group of Pycsar systems associated with separate *capH* and *capP* genes. We identified three sets of operons encoding proteins of the DUF2188 or

DUF3892 families, which are uncharacterized but have been previously linked to bacterial defense (Burroughs and Aravind, 2020). Both DUF2188 and DUF3892 proteins have also been identified in operons containing HTH and metallopeptidase genes, paralleling our findings (Burroughs and Aravind, 2020). We also identified operons encoding proteins related to eukaryotic ubiquitin signaling machinery, including so-called Ub-6a systems that encode a predicted metallo-β-lactamase and a large protein with E2-like, E1-like, and JAB protease-like domains (Burroughs et al., 2009; Iyer et al., 2006). Notably, this protein shares strong homology to the Cap2 protein in Type II CBASS systems (also classified as Ub-6b systems), which conjugates the C-terminus of its cognate CD-NTase to an unknown target to regulate anti-phage signaling (Ledvina et al., 2022) . We also identified CapH and CapP-like proteins associated with Ub-6e.2 systems, which encode a protein predicted to possess multiple ubiquitin-like β-grasp domains and an E1-like protein, plus an uncharacterized protein (DUF6527; **Figure 3.6F**). Thus, CapH and CapP-like regulators are associated with a broad range of bacterial signaling pathways with known or predicted roles in anti-phage or stress responses.

Discussion

Here, we identify a pair of proteins – CapH and CapP – that are associated with hundreds of CBASS anti-phage systems, and show that they function together to regulate CBASS expression. In the basal state, the helix-turn-helix protein CapH forms oligomers and binds the promoter region of its associated CBASS system to repress transcription. With its distinctive cysteine switch motif, CapP is maintained in an inactive state in this mode. In *E. coli* MS115-1 CBASS, the resulting low-level basal expression of the CBASS core genes is apparently sufficient to provide modest protection against bacteriophage λ infection. Despite the extremely low levels of the core CBASS proteins in this repressed state, phage infection is sensed by the system's HORMA domain proteins, activating CdnC to produce a cyclic tri-AMP second messenger. Cyclic tri-AMP in turn activates NucC, which destroys the host genome to kill the cell and abort the infection (**Figure 3.7A**) (Lau et al., 2020; Ye et al., 2020).

We find that DNA damage strongly activates CapP-mediated CapH cleavage through the production of T-rich single-stranded DNA, and that this activation results in a dramatic increase in CBASS expression. Since deletion of *capP* did not affect the system's ability to directly protect against bacteriophage λ infection, this expression boost likely plays a role other than primary response to phage infection. We propose that this pathway represents a second path for CBASS activation that directly responds to DNA damage instead of, or in addition to, phage infection. DNA damage is a universal signal of bacterial cell stress and is sensed by diverse stress response pathways including the SOS response, which is broadly similar but mechanistically

distinct from the CapH/CapP system we describe. A mechanism to boost expression of immune systems upon DNA damage may enable these systems to sense and respond to a broader set of stimuli than they could without this regulation. For example, DNA damage could arise from the activation of a DNA-targeting restriction-modification or CRISPR-Cas system after phage infection. If this damage is sufficiently severe or sustained, it could activate a CapH/CapP-linked secondary immune system to aid the defensive response (**Figure 3.7B**). Alternatively, a CapH/CapP-linked immune system could piggyback on the same stress signal that activates a latent prophage to become lytic, and suppress the production of virions by pre-emptively killing the host cell or otherwise restricting the phage.

Supporting the idea that DNA damage signaling could represent a broad mechanism for activation of bacterial immune systems, we identified a diverse set of known or putative defense systems associated with CapH and CapP-like regulators. These include groups of DISARM and Pycsar systems, which function similarly to restriction-modification systems (Ofir et al., 2018) and CBASS systems (Tal et al., 2021), respectively. CapP and CapH also appear in operons encoding DUF2188 and DUF3892 proteins, which have both been predicted to participate in anti-phage defense (Burroughs and Aravind, 2020). Finally, the association of CapH and CapP with uncharacterized operons encoding proteins related to ubiquitin signaling machinery suggests that these operons, too, may play a role in defense against phage infection or other stresses.

Our identification of a mechanism enabling a single defense system to respond to multiple stimuli parallels the recent discovery and characterization of BrxR/CapW transcriptional regulators, which are associated with a variety of immune systems including BREX and CBASS anti-phage systems (Blankenchip et al., 2022; Luyten et al., 2022; Picton et al., 2022). CapW drives increased expression of its associated CBASS systems upon phage infection, but as with CapH/CapP-associated systems, this increased expression is not specifically required for protection against lytic phage (Blankenchip et al., 2022). Similarly, BrxR controls expression of its associated BREX systems, but is not required for anti-phage immunity (Luyten et al., 2022). While the activating signal of BrxR/CapW is not known, these data suggest that the protein controls activation of CBASS and BREX systems in response to signals other than phage infection. As with CapH and CapP, BrxR/CapW may enable its associated defense system to act as a second line of defense in coordination with restriction-modification or CRISPR-Cas systems (Luyten et al., 2022; Picton et al., 2022). More broadly, there may exist a range of signaling mechanisms that enable crosstalk between distinct defense systems in bacteria, mediating these systems' cooperation and integration into a comprehensive, multi-faceted immune system.

Materials and Methods

Bioinformatics

To comprehensively search CBASS systems, we exported the genomic DNA sequences within 10 kb of 6233 previously-reported CD-NTases (Cohen *et al.*, 2019) using the Integrated Microbial Genomes (IMG) database at the DOE Joint Genome Institute (https://img.jgi.doe.gov). We used NCBI Genome Workbench (https://www.ncbi.nlm.nih.gov/tools/gbench/) to perform custom TBLASTN searches for proteins related to *E. coli* MS115-1 CapP (NCBI sequence ID WP_001290439.1; **Table A.C.S1**). CBASS system type and effector assignments for each hit were taken from Cohen et al. (Cohen et al., 2019) and manually updated. Each hit was manually inspected for the presence of CapH and CapP.

For identification of other bacterial defense systems with associated CapH and CapP-like regulators, we used the FlaGs (Flanking Genes) server (Saha et al., 2021) to search for groups of genes associated with CapP (using *E. coli* MS115-1 CapP as a search sequence). We manually inspected each result to search for known or putative defense-related genes. For **Figure 3.6F**, the noted DISARM system from *Nocardia wallacei* FMUON74 encodes a *vapC*-like gene (NCBI accession # WP_187684394.1), a CapH-CapP fusion-like protein (WP_197986914.1), *drmD* (WP_187684395.1), drmMI (WP_187684396.1), drmA (WP_187684397.1), drmB ((WP_187684398.1), drmC (WP_187684399.1), and a AAA+ ATPase (WP_232110603.1). The noted Pycsar system from *Tsuneonella flava* MS1-4 encodes a CapP-like protein (WP_102155989.1), a CapH-like protein (WP_102155988.1), a DUF2188 protein (WP_007013875.1), a

predicted guanylate cyclase (WP_102155986.1), a hypothetical protein (WP_123639961.1), a predicted cyclic nucleotide monophosphate binding domain (cNMPBD) plus TIR domain protein (WP_102155984.1), and a predicted adenylate/guanylate cyclase (WP_102155983.1). The noted DUF2188 + hypothetical protein system from *Extensimonas perlucida* HX2-24 encodes a CapP-like protein (WP_144728453.1), a CapH-like protein (WP_144728455.1), two hypothetical proteins (WP_144728457.1 and WP_144728459.1), and a DUF2188 protein (WP_144729358.1). The noted DUF3892 system from *Sulfitobacter* sp. CW3 encodes a CapP-like protein (WP_037275352.1), a CapH-like protein (WP_037275354.1), and two DUF3892 proteins (WP_037275364.1 and WP_037275356.1). The noted metallo-β-lactamase (Ub-6a) system from *Methylobacterium oxalidis* NBRC 107715 encodes a CapP-like protein (WP_147028642.1), a CapH-like protein (WP_147028643.1), a hypothetical protein (WP_147028644.1), a predicted metallo-β-lactamase (WP_147028653.1), a protein with predicted E2, E1, and JAB domains (WP_147028645.1), and a second hypothetical protein (WP_147028646.1). The noted DUF6527 (Ub-6e.2) system from *Mixta intestinalis* SRCM103226 encodes a CapP-like protein (WP_160622475.1), a CapH-like protein (WP_048227226.1), a predicted multi-ubiquitin domain protein (WP_053069300.1), a predicted E1-like protein (WP_160622476.1), and a DUF6527 protein (WP_18149987.1).

Cloning, expression, and protein purification

Proteins were cloned into UC Berkeley Macrolab vector 2-BT (Addgene #29666; encoding an N-terminal TEV protease-cleavable His_6-tag), 2-ST (Addgene #29711,

encoding an N-terminal TEV protease-cleavable His_6-SUMO tag), or 2-CT (Addgene #29706, encoding an N-terminal TEV protease-cleavable His_6-MBP tag). Proteins used were: *E. coli* MS115-1 CapH (NCBI sequence ID WP_001515173.1), *E. coli* MS115-1 CapP (NCBI sequence ID WP_001290439.1), and *Thauera* sp. K11 CapP (NCBI sequence ID WP_096453114.1). Proteins were expressed in *E. coli* strain Rosetta 2 (DE3) pLysS (EMD Millipore, Billerica MA). Cultures were grown at 37°C to A_{600}=0.5, then induced with 0.25 mM IPTG and shifted to 20°C for 15 hours. Cells were harvested by centrifugation and resuspended in buffer A (20 mM Tris pH 7.5, 10% glycerol) plus 300 mM NaCl, 10 mM imidazole, and 5 mM β-mercaptoethanol (10 μM $ZnCl_2$ was added to buffers for CapP). Proteins were purified by Ni^{2+}-affinity (Ni-NTA agarose, Qiagen) then passed over an anion-exchange column (Hitrap Q HP, Cytiva) in Buffer A plus 5 mM β-mercaptoethanol and 0.1-1 M NaCl, collecting flow-through or peak fractions. Tags were cleaved with TEV protease (Tropea et al., 2009), and cleaved protein was passed over another Ni^{2+} column (collecting flow-through fractions) to remove uncleaved protein, cleaved tags, and tagged TEV protease. The protein was passed over a size exclusion column (Superdex 200, Cytiva) in buffer GF (buffer A plus 300 mM NaCl and 1 mM dithiothreitol (DTT)), then concentrated by ultrafiltration (Amicon Ultra, EMD Millipore) to 10-20 mg/ml and stored at 4°C. For selenomethionine derivatization, protein expression was carried out in M9 minimal media supplemented with amino acids plus selenomethionine prior to IPTG induction (Van Duyne et al., 1993), and proteins were exchanged into buffer containing 1 mM tris(2-

carboxyethyl)phosphine (TCEP) after purification to maintain the selenomethionine residues in the reduced state.

Crystallization and structure determination

For crystallization of *E. coli* MS115-1 CapH^{NTD} (residues 2-67), protein was concentrated to 18 mg/mL in crystallization buffer (20 mM Tris-HCl pH 8.5, 150 mM NaCl, 1 mM DTT) then mixed 1:1 with well solution containing 0.1 M Ammonium acetate pH 4.5 and 2 M Ammonium sulfate in hanging-drop format. Crystals were cryoprotected with an additional 25% sucrose and flash-frozen in liquid nitrogen. Diffraction data to 1.02 Å resolution were collected at Advanced Light Source beamline 5.0.2 (see support statement below) and processed with the DIALS data-processing pipeline (https://dials.github.io) (Beilsten-Edmands et al., 2020). We determined the structure by molecular replacement in PHASER (McCoy et al., 2007), using an ideal α-helix as a search model. We manually rebuilt the initial model in COOT (Emsley et al., 2010) , and refined in phenix.refine (Afonine et al., 2012) using individual positional and anisotropic B-factor refinement for non-hydrogen atoms, and riding hydrogens (**Table A.C.S2**).

For crystallization of *E. coli* MS115-1 CapH^{CTD}(I99M) (residues 67-107 with Ile99 to Met mutation), protein was concentrated to 8 mg/mL in crystallization buffer (20 mM Tris-HCl pH 8.5, 150 mM NaCl, 1 mM DTT) then mixed 1:1 with well solution containing 0.1 M HEPES pH 7.5, 25 mM $MgCl_2$, and 30% PEG 550 monomethyl ether in hanging-drop format. Crystals were cryoprotected with an additional 13% PEG 550 monomethyl ether and 10% glycerol, and flash-frozen in liquid nitrogen. Diffraction data to 1.26 Å resolution were collected at Advanced Photon Source beamline 24ID-C (see support

statement below) and processed with the RAPD data-processing pipeline, which uses XDS (Kabsch, 2010) for data indexing and reduction, AIMLESS (Evans and Murshudov, 2013) for scaling, and TRUNCATE (French and Wilson, 1978) for conversion to structure factors. We determined the structure by molecular replacement in PHASER, using an ideal α-helix as a search model. We manually rebuilt the initial model in COOT, and refined in phenix.refine using individual positional and anisotropic B-factor refinement for non-hydrogen atoms, and riding hydrogens (**Table A.C.S2**).

For crystallization of *E. coli* MS115-1 CapH^{CTD} (residues 67-107), protein was concentrated to 21 mg/mL in crystallization buffer (20 mM Tris-HCl pH 8.5, 150 mM NaCl, 1 mM DTT) then mixed 1:1 with well solution containing 0.1 M sodium citrate pH 3.0 and 1.6 M lithium sulfate in hanging-drop format. Crystals were cryoprotected with an additional 20% ethylene glycol and flash-frozen in liquid nitrogen. Diffraction data to 1.75 Å resolution were collected at Advanced Light Source beamline 5.0.2 (see support statement below) and processed with the DIALS data-processing pipeline. We determined the structure by molecular replacement in PHASER, using the structure of CapH^{CTD}(I99M) as a search model. We manually rebuilt the initial model in COOT, and refined in phenix.refine using individual positional and B-factor refinement, and riding hydrogens (**Table A.C.S2**).

For crystallization of *Thauera* sp. K11 CapP, protein was concentrated to 9 mg/mL in crystallization buffer (20 mM Tris-HCl pH 7.5, 100 mM NaCl, 1 mM DTT) then mixed 1:1 with well solution containing 0.1 M CHES pH 9.5, 0.3 M NaCl, and 1.8 M lithium sulfate in hanging-drop format. Crystals were cryoprotected with an additional

30% glycerol and flash-frozen in liquid nitrogen. Diffraction data for both native and selenomethionine-derivatized crystals were collected at the Stanford Synchrotron Radiation Lightsource beamline 9-2 (see support statement below) and processed with the autoxds data-processing pipeline, which uses XDS for data indexing and reduction, AIMLESS for scaling, and TRUNCATE for conversion to structure factors. We determined the structure by single-wavelength anomalous diffraction (SAD) methods in PHASER using a 1.6 Å resolution dataset from selenomethionine-derivatized protein. We manually rebuilt the initial model in COOT, and refined against a 1.35 Å resolution native dataset in phenix.refine using individual positional and anisotropic B-factor refinement, and riding hydrogens (**Table A.C.S2**).

Beamline Support Statements

ALS beamline 5.0.2: The Berkeley Center for Structural Biology is supported in part by the Howard Hughes Medical Institute. The Advanced Light Source is a Department of Energy Office of Science User Facility under Contract No. DE-AC02-05CH11231. The Pilatus detector on 5.0.1. was funded under NIH grant S10OD021832. The ALS-ENABLE beamlines are supported in part by the National Institutes of Health, National Institute of General Medical Sciences, grant P30 GM124169.

APS beamline 24ID-C: This work is based upon research conducted at the Northeastern Collaborative Access Team beamlines, which are funded by the National Institute of General Medical Sciences from the National Institutes of Health (P30 GM124165). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE

Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357.

SSRL beamline 9-2: Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research, and by the National Institutes of Health, National Institute of General Medical Sciences (P30GM133894). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of NIGMS or NIH.

SEC-MALS

For characterization of oligomeric state by size exclusion chromatography coupled to multi-angle light scattering (SEC-MALS), 100 μL of purified protein/complex at 2-5 mg/mL was injected onto a Superdex 200 Increase 10/300 GL column (Cytiva) in a buffer containing 20 mM HEPES pH 7.5, 300 mM NaCl, 5% glycerol, and 1 mM DTT. Light scattering and refractive index profiles were collected by miniDAWN TREOS and Optilab T-rEX detectors (Wyatt Technology), respectively, and molecular weight was calculated using ASTRA v. 8 software (Wyatt Technology).

DNA-binding assays

For characterization of DNA binding by fluorescence polarization assays, a 22 bp double-stranded DNA was produced by annealing complementary oligos, one of which

was 5'-6-FAM labeled (the same oligonucleotide was used without annealing for ssDNA binding studies). Binding reactions (10 μL) contained 25 mM Tris pH 7.5, 50 mM NaCl, 5% glycerol, 5 mM $MgCl_2$, 1 mM DTT, 0.01% Nonidet p40 substitute, 50 nM DNA, and the indicated amounts of His_6-MBP-tagged protein. After a 10-minute incubation at room temperature, fluorescence polarization was read using a Tecan Infinite M1000 PRO fluorescence plate reader, and binding data were analyzed with Graphpad Prism v.9.2.0 using a single-site binding model.

CapP cleavage assays

For detection of CapP activity in cells, *E. coli* MS115-1 CapP was co-expressed with a construct of *E. coli* MS115-1 CapH fused to an N-terminal His_6-maltose binding protein (MBP) tag and a C-terminal green fluorescent protein (GFP) tag (MBP-CapH-GFP) in *E. coli* Rosetta 2 (DE3) pLysS cells at 37°C. Protein expression was induced with 0.25 mM IPTG for 2-4 hours, then samples were removed for analysis by western blot (see below). For CapP activity in cells in response to zeocin, *E. coli* Rosetta 2 (DE3) pLysS cells were grown at 37°C until reaching an OD_{600} of 0.3. Protein expression was induced with 0.25 mM IPTG for 1 hour, then zeocin was added at a concentration of 100 μg/mL. 500 μL of sample was taken at 1 hour and 2 hours post-antibiotic and centrifuged 10,000xg for 1 minute. Pelleted cells were resuspended in 50 μL of 2xSDS sample buffer and analyzed by western blot (see below).

For detection of CapP activity using purified proteins, 20 μL reactions containing 10 μM purified CapP and 10 μM purified MBP-CapH-GFP in a buffer containing 50 mM Tris pH 7.5, 300 mM NaCl and 5 μM $ZnCl_2$ were incubated at 37ºC for 2 hours, then

added to 20 μL 2xSDS sample buffer. 10 μL of each sample was loaded and separated by SDS-PAGE and Coomassie-stained for visualization. For reactions containing *E. coli* cell lysate, log-phage *E. coli* Rosetta 2 (DE3) pLysS cells were lysed by sonication in a minimum volume of buffer A, then centrifuged to remove cell debris. Optionally, the lysate was incubated in a boiling water bath for 10 minutes, then centrifuged again to remove denatured proteins. For the experiment shown in **Figure A.C.S5B**, lysate was fractionated on a 5 mL HiTrap Q HP column (Cytiva). 10 μL of cell lysate was added to reactions with purified CapP and MBP-CapH-GFP. For reactions containing DNA, nucleic acid was added at a concentration of 350 μg/mL and subsequent serial 5-fold dilutions, then incubated in the buffer as above for 2 hours. For reactions containing DNA or RNA oligos, nucleic acid was added at a concentration of 10 μM, then incubated in the buffer as above for 2 hours.

Edman degradation

For Edman degradation, MBP-CapH-GFP was treated with CapP plus single-stranded DNA, separated by SDS-PAGE, then transferred to a PVDF membrane and visualized with Coomassie staining. The band representing the C-terminal cleavage product of MBP-CapH-GFP was cut out and analyzed by Edman degradation at the UC Davis Molecular Structure Facility (http://msf.ucdavis.edu).

Western Blots

For CapH repressor assays, cells containing plasmids with the MS115-1 system were grown in 5 mL of LB plus the selection marker at 37°C until they reached an OD_{600} of 0.3 – 0.5. Cultures were adjusted to an OD_{600} of 0.3, then an aliquot of 500 μL was

removed and centrifuged for 1 minute at 10,000 x g to pellet the cells. The supernatant was removed, and the cells were resuspended in 50 μL of 2xSDS sample buffer (125 mM Tris pH 6.8, 20% Glycerol, 4% SDS, 200 mM DTT, 180 μM bromophenol blue). 10μL of each sample was loaded onto a 12% SDS-PAGE gel for separation and transferred to a PVDF membrane using the Bio-Rad Trans-Blot Turbo Transfer System. Membranes were blocked for 1 hour in 5% milk in TBST (100 mM Tris pH 7.5, 150 mM NaCl, 0.1% Tween-20) at room temperature with shaking, then incubated with mouse anti-FLAG M2 antibody (Sigma-Aldrich #F3165) or mouse anti-GFP antibody (Roche #11814460001) at 1:3,000 diluted in 5% milk in TBST for 1 hour at room temperature with shaking. Membranes were washed 3 times with 10 mL of TBST then incubated with goat anti-mouse IgG antibody conjugated to horseradish peroxidase (Millipore Sigma #12-349) at 1:30,000 diluted in 5% milk in TBST for 1 hour at room temperature with shaking. Membranes were again washed 3 times with 10 mL TBST then incubated with Pierce ECL Plus substrate for 2 minutes and then imaged on a Bio-Rad ChemiDoc imager. Membranes were stripped with a solution of 200 mM glycine pH 2.2, 0.1% SDS, and 1% Tween 20 for 20 minutes at room temperature, then washed with TBST and re-blocked with 5% milk overnight at 4°C. Membranes were re-blotted using the same procedure as initial blot, but replacing primary antibody with anti-RNA polymerase B mouse antibody (clone NT63; BioLegend #10019-878).

CBASS plasmid construction

The full CBASS operon was PCR-amplified from *E. coli* MS115-1 genomic DNA (BEI Resources #HM-344), using NCBI nucleotide accession number CP073624.1 (*E.*

coli AW1.7) as a reference because of ambiguous sequence in the nucleotide record for *E. coli* MS115-1 (NCBI accession number GG771785.1). An insert spanning bases 2861237-2866476 of NCBI accession number CP073624.1 was cloned between the EcoRI and BamHI sites of plasmid pBR322 using isothermal assembly. PCR-based mutagenesis was used for all point mutants, gene deletions, and to insert an N-terminal FLAG tag and short linker (DYKDDDDK-ASAS) at the N-terminus of the *nucC* gene. For the GFP reporter strain, the region encoding the core CBASS genes was replaced by a gene encoding msfGFP.

DNA damage assays

For antibiotic treatment time course western blots, *E. coli* JP313 (MC4100 ara*Δ714*) (Economou *et al*, 1995) containing reporter plasmids were grown at 37°C until reaching an OD_{600} of 0.1. Cultures were moved to 30°C for 10 minutes, then antibiotics were added at a concentration of 100 μg/mL for zeocin and 1 μg/mL for mitomycin C. At each timepoint, 1 mL of culture was removed and resuspended in 2xSDS sample buffer with the volume adjusted to an equal concentration of cells per sample.

Bacteriophage infectivity

Phage λ cI- was amplified in *E. coli* JP313 by inoculating cultures grown in LB with 1 mM $MgCl_2$ and 1 mM $CaCl_2$ to OD_{600}=0.3 with 100 μL of high-titer phage. Cells were incubated with shaking at 37°C for 5 hours or until clearing was observed. The lysed cultures were spun at 4000xg for 20 minutes, the supernatant was collected and filtered with a 0.2 μm filter, and filtered supernatant was stored at 4°C. Phage was titered in JP313 cells. Cells were grown to OD_{600}=0.3-0.5 at 37°C, then 500 μL was

aliquoted into 5 mL Eppendorf tubes. 10 μL of phage was added to each 500 μL tube, at 10-fold dilutions. Lysate was diluted in phage buffer (150 mM NaCl, 40 mM Tris pH 7.5, 10 mM $MgSO_4$ plus 1 mM $CaCl_2$). Cultures were incubated with phage for 20 minutes after which 4.5 mL of 0.35% top agar with LB was added to each culture and mixed, then poured onto LB plates. Plates were incubated at 37°C for 16 hours, then plaques were counted. For plaque quantification assays, strains were grown from single colonies in 5 mL of LB with ampicillin (100 ug/mL) until reaching log phase (OD_{600}=0.3-0.6). 500 uL of culture was transferred to a 5 mL conical tube to which 10 uL of of λ cI- at a concentration of 1.6 x 10^{10} PFU/mL (and 10-fold dilutions thereof) in phage buffer (150 mM NaCl, 40 mM Tris pH 7.5, 10 mM $MgSO_4$ plus 1 mM $CaCl_2$) was added. Tubes were incubated at room temperature for 20 minutes after which 4.5 mL of 0.35% top agar was added mixed, then poured onto LB plates containing carbenicillin. Plates were incubated at 37°C for 16 hours, then plaques were counted.

For infection time course western blots, *E. coli* JP313 cells containing reporter plasmids were grown at 37°C until reaching an OD_{600} of 0.3. Cultures were moved to 30°C for 10 minutes, then λ cI- was added at an MOI of 10. At each timepoint, 500 μL of sample was removed and centrifuged 10,000xg for 1 minute. Cell pellets were resuspended in 50 μL 2xSDS sample buffer and analyzed by western blot.

Microscopy

For fluorescence microscopy imaging, each sample was grown as liquid culture at 30°C and induced with 0.2mM IPTG 30 minutes before of imaging. Cells were infected with λ CI- at MOI 2.5 as 0 MPI. Cells were harvested at required timepoints and

were briefly centrifuged (3300 x g for 30 seconds). After resuspending the cells with 20 μl volume, 5 μl of the samples were transferred onto an agarose pad containing 1% agarose and 20% LB medium for microscopy and stained with 1 μg/mL FM4-64 and 2 mg/mL DAPI. Microscopy was performed using by a Deltavision Elite System (GE Healthcare). Cells were quantified through manual examination of at least three replicates per condition and 30-100 cells per condition.

qRT-PCR

E. coli JP313 cells containing the WT plasmid in pBR322 were grown in 25 mL LB + Amp in 125 mL flasks until log phase (OD_{600} 0.3-0.5). Cells were diluted to an OD_{600} of 0.2 in a final volume of 25 mL, then zeocin (100 ug/mL), λ CI- (MOI 10), or LB at equivalent volumes were added to each flask. Initial time points were taken at this time. For each time point, 1 mL of cells was transferred to a 1.5 mL Eppendorf tube and spun down 1 min at 10,000 x g. Supernatant was removed and pellets were immediately flash frozen in liquid N_2. Cells were grown at 30°C and time points were taken every 20 minutes in triplicate.

RNA extractions were performed using the Invitrogen PureLink RNA Mini Kit with on-column DNase treatment (Qiagen). 1 μg of RNA was reverse transcribed to make cDNA using the Applied Biosystems High-Capacity cDNA Reverse Transcription Kit. qPCR was performed in a 384-well format with technical triplicates for each sample. Each reaction was composed of IDT PrimeTime Gene Expression Master Mix, 250 nM of each primer, 100 μM of each probe, and cDNA diluted 1:250 to a final volume of 10 μL. The following primers were synthesized by Integrated DNA Technologies:

RpoA_F: CGTGGCTTTGGCCATACTCT

RpoA_R: ACGCCTTCTTTGGTGCTGTA

RpoA probe: /56-FAM/AGCGAATGA/ZEN/TTCCATCAGGTAGTCTGGC/3IABkFQ/

CdnC_F: GGAACAGGCCAAGCGATTAC

CdnC_R: AACGAGCGAAGAGCAGTTCC

CdnC probe: /5Cy5/AGCAGAATA/TAO/CGGCGCAGTGCGT/3IAbRQSp/

56-FAM: 5' 6-FAM (Fluorescein); ZEN: ZEN internal quencher (Integrated DNA Technologies); 3IABkFQ: 3' Iowa Black FQ quencher; 5Cy5: 5' Cy5; TAO: TAO internal quencher (Integrated DNA Technologies); 3IAbRQSp: 3' Iowa Black RQ quencher.

qPCR was performed using a Bio-Rad CFX384 Touch Real-Time PCR Detection System.

Electrophoretic mobility assay (EMSA)

Purified MBP-tagged CapH was mixed with 100 nM 5'-FAM-labeled DNA at concentration stated in figures and two-fold dilutions thereof, then incubated at room temperature for 20 minutes. Samples were loaded on to pre-run 8% polyacrylamide gels (pre-run 250V for 60 mins) made with 1x TBE pH 8.5 and run in 0.5 TBE pH 8.5 at 120V for 80 minutes at 4°C. Gels were imaged using a Bio-Rad ChemiDoc imager.

MIC Assays

MIC assays were performed in a 96-well format. 100 μL of LB + Ampicillin containing different concentrations of zeocin was added to each well in the top row of the plate. Concentrations of zeocin started from 100 μg/mL and were serially diluted at

1:3 across the plate in fresh LB + Amp. Cells containing either an empty vector, a plasmid with the 6-gene *E. coli* MS115-1 system, or a mutant system with catalytic-dead NucC (D73N) were grown to OD_{600}=0.1 in LB + Amp at 37°C with shaking. Cells were diluted and 100 μL of culture was added to each well such that each well contained ~2 x 10^5 CFU/mL. Each condition was run in triplicate. The plate was incubated at 37°C and OD_{600} readings were taken after 18 hours of growth.

Data Availability

The structural data produced in this study is available in the following databases:

Primary X-ray diffraction datasets: SBGrid Data Bank (https://data.sbgrid.org): *E. coli* MS115-1 CapH NTD: #866*E. coli* MS115-1 CapH CTD: #867; *E. coli* MS115-1 CapH CTD (I99M): #868; *Thauera* sp. K11 CapP: #864, 865.

Reduced X-ray diffraction datasets and refined structures: Protein Data Bank (http://www.wwpdb.org): *E. coli* MS115-1 CapH NTD: 7T5U; *E. coli* MS115-1 CapH CTD: 7T5W; *E. coli* MS115-1 CapH CTD (I99M): 7T5V; *Thauera* sp. K11 CapP: 7T5T.

Acknowledgements

The authors thank John Schulze at the UC Davis Molecular Structure Facility for assistance with Edman Degradation, Michael Baughn for advice on qPCR experiment and primer design, and members of the Corbett lab and Aaron Whiteley for helpful feedback. K.D.C. acknowledges support from the National Institutes of Health (R35 GM144121). R.K.L. was supported by the UCSD Quantitative and Integrative Physiology Training Grant (NIH T32 GM127235) and an individual National Institutes of Health Predoctoral Fellowship (F31 GM137600).

Conflict of Interest

The authors declare no competing interests.

Figure 3.17 Identification of CBASS-associated genes capH and capP and role of CapH and CapP in the CBASS antiviral response

(A) Operon schematics of four representative CBASS systems with associated *capH* (yellow) and *capP* (pink) genes. See **Table A.C.S1** for all identified systems. For each system, core CBASS genes are colored as in the key: CD-NTases orange, putative regulator(s) blue, and effector(s) green. Dotted outlines indicate unknown theoretical genes.

(B) Distribution of *capH*+*capP*-associated CBASS systems, sorted by system type as defined by Millman et al. (Millman *et al.*, 2020). Type III (1H) and Type III (2H) refer to Type III systems with one or two HORMA domain proteins, respectively.

(C) Distribution of *capH*+*capP* -associated CBASS systems, sorted by effector type as defined by Millman et al. (Millman *et al.*, 2020). 1TM, 2TM, and 4TM refer to effectors with one, two, or four predicted transmembrane segments, respectively.

(D) *Top:* Schematic of GFP expression reporter system, with the CBASS promoter, *capH*, and *capP* genes from *E. coli* MS115-1 and the CBASS core genes replaced with GFP. *Bottom:* Western blot showing GFP expression in cells with the wild-type GFP reporter or constructs lacking either *capP* or *capH* genes. RNAP: RNA polymerase loading control. See full blot in **Figure A.C.S1A**.

(E) Western blots of the CBASS expression reporter system with FLAG-NucC (see **Figure A.C.S1B**), showing FLAG-NucC expression after infection with phage λ cI- (multiplicity of infection: 10). MPI: minutes post infection. α-RNAP: anti-RNA polymerase loading control. Low RNAP expression at later time points is due to cell death.

(G) Quantitative plaque assay showing infectivity of λ cI- against cells containing no CBASS system (EV: empty vector), the wild-type *E. coli* MS115-1 CBASS system (WT), or a mutant system lacking *capP* (Δ*capP*). Bars represent average of plaque forming units per mL of purified phage (PFU/mL), from duplicate experiments.

(H) Size of phage plaques for λ cI- infecting cells containing no CBASS system (EV: empty vector), the wild-type *E. coli* MS115-1 CBASS system (WT), or a mutant system lacking *capP* (Δ*capP*). Data are shown as average and standard deviation of all plaques counted in panel G.

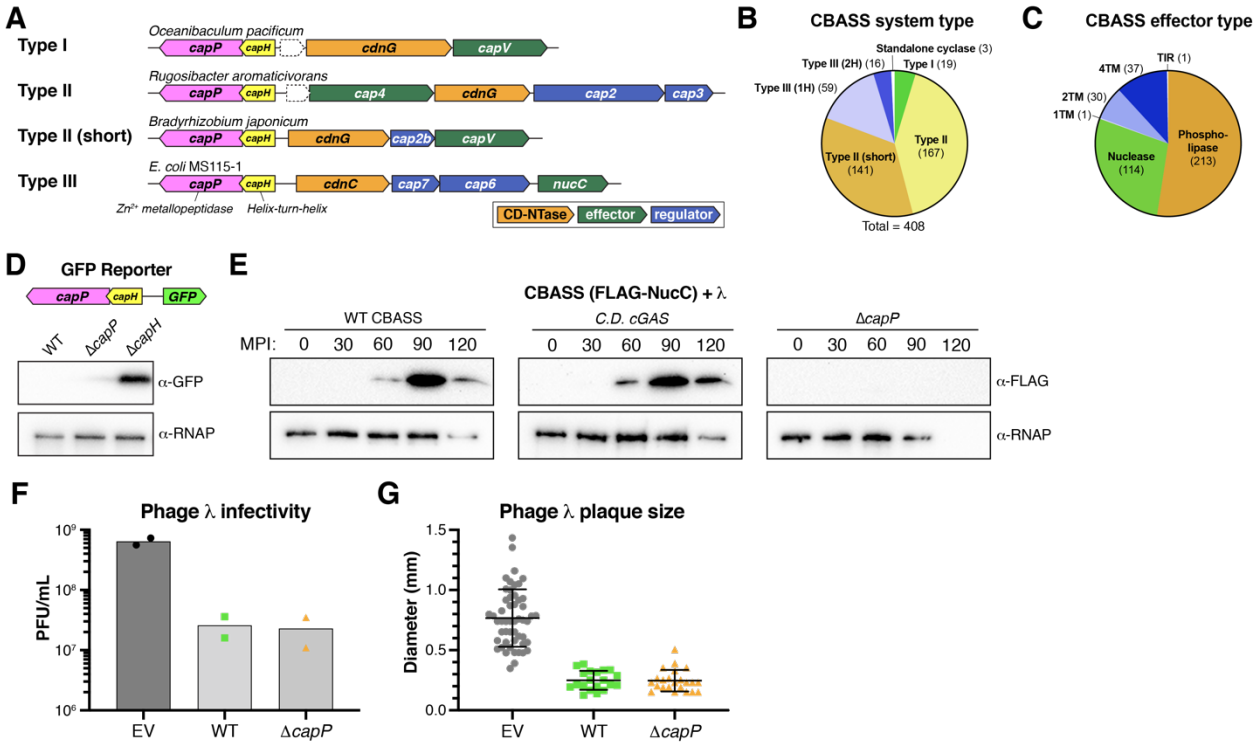

A
Type I
Oceanibaculum pacificum
capP
capH
cdnG
capV
Type II
Rugosibacter aromaticivorans
capP
capH
cap4
cdnG
cap2
cap3
Type II (short)
Bradyrhizobium japonicum
capP
capH
cdnG
cap2b
capV
Type III
E. coli MS115-1
capP
capH
cdnC
cap7
cap6
nucC
Zn2+ metallopeptidase
Helix-turn-helix
CD-NTase
effector
regulator
B
CBASS system type
Standalone cyclase (3)
Type III (2H) (16)
Type I (19)
Type III (1H) (59)
Type II (167)
Type II (short) (141)
Total = 408
C
CBASS effector type
TIR (1)
4TM (37)
2TM (30)
1TM (1)
Phospho-lipase (213)
Nuclease (114)
D
GFP Reporter
capP
capH
GFP
WT
ΔcapP
ΔcapH
α-GFP
α-RNAP
E
CBASS (FLAG-NucC) + λ
WT CBASS
C.D. cGAS
ΔcapP
MPI:
0 30 60 90 120
α-FLAG
α-RNAP
F
Phage λ infectivity
PFU/mL
EV
WT
ΔcapP
G
Phage λ plaque size
Diameter (mm)
EV
WT
ΔcapP

Figure 3.18 CapH binds the CBASS promoter region

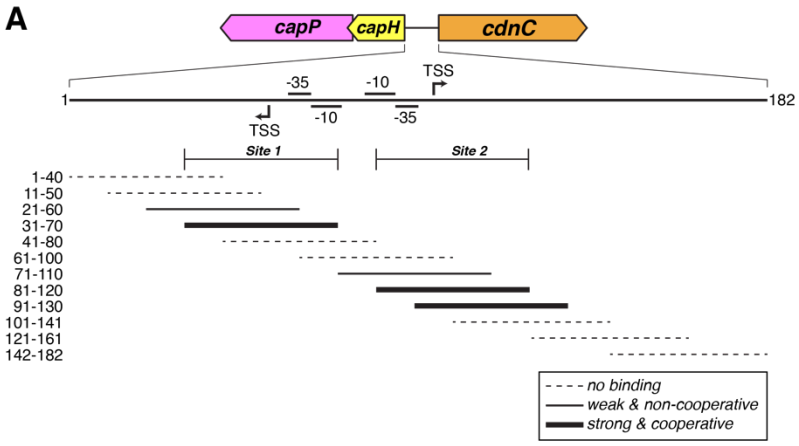
(A) *Top:* Schematic of the region between *capH* and *cdnC* genes in *E. coli* MS115-1 CBASS with predicted -35, -10, and TSS (translation start site) sites for the top (forward) and bottom (reverse) strands, predicted using BPROM (Salamov and Solovyevand, 2011). *Bottom:* Overlapping 40-41 bp DNAs tested for CapH binding by fluorescence polarization, denoted as either no binding (dotted line), weak/non-cooperative binding (thin solid line), or strong/cooperative binding (thick solid line). The boundaries of the two CapH binding sites (Site 1 and Site 2) are noted. See **Figure A.C.S3A-F** for binding curves.

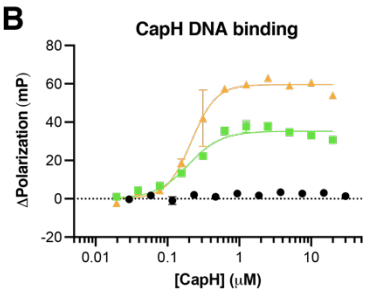
(B) Fluorescence polarization assay showing binding of *E. coli* MS115-1 CapH (His_6-MBP tagged) to three 40-bp DNAs: 1-40 (black circles), 31-70 (Site 1; green squares), and 81-120 (Site 2; orange triangles). Fit K_d and Hill coefficient for each DNA is shown.

(C) *Top:* Domain schematic of *E. coli* MS115-1 CapH, and truncation used for crystallization of the N-terminal HTH domain comprised of residues 2-67 (NTD). *Bottom:* Crystal structure of the CapH NTD (blue), with bound DNA modeled from a structural overlay with a known HTH-DNA complex structure (PDB ID 3CLC; (McGeehan et al., 2008). Shown in sticks are three conserved residues (S32, R40, and R44) putatively involved in DNA binding. See **Figure A.C.S3G** for sequence conservation of the CapH NTD.

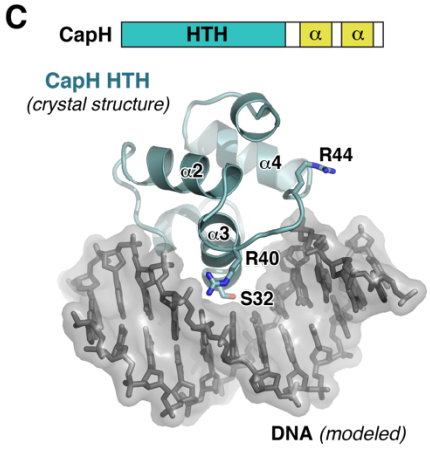
(D) Fluorescence polarization assay showing binding of *E. coli* MS115-1 CapH (His_6-MBP tagged; wild-type or indicated point mutants) to the Site 2 DNA (bases 81-120 in panel A). Fit K_d and Hill coefficient for each DNA is shown.

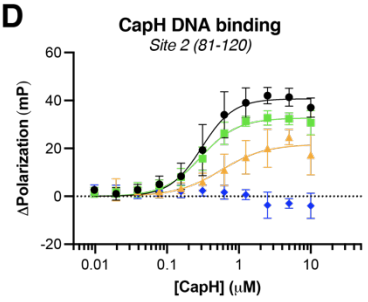
(G) GFP expression reporter assay showing loss of suppression upon mutation of CapH. RNAP: RNA polymerase loading control. See full blot in **Figure A.C.S1A**.





DNA	K_d	Hill coeff.
● 1-40	N/A	N/A
■ 31-70	0.19 +/- 0.03	1.9 +/- 0.5
▲ 81-120	0.22 +/- 0.02	2.5 +/- 0.6





CapH	K_d	Hill coeff.
● WT	0.30 +/- 0.08	2.0 +/- 0.7
■ R40A	0.30 +/- 0.07	1.7 +/- 0.5
▲ R44A	0.56 +/- 1.00	1.4 +/- 0.8
◆ S32A	N/A	N/A

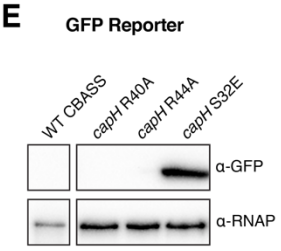


Figure 3.19 CapH oligomerization is required for DNA binding

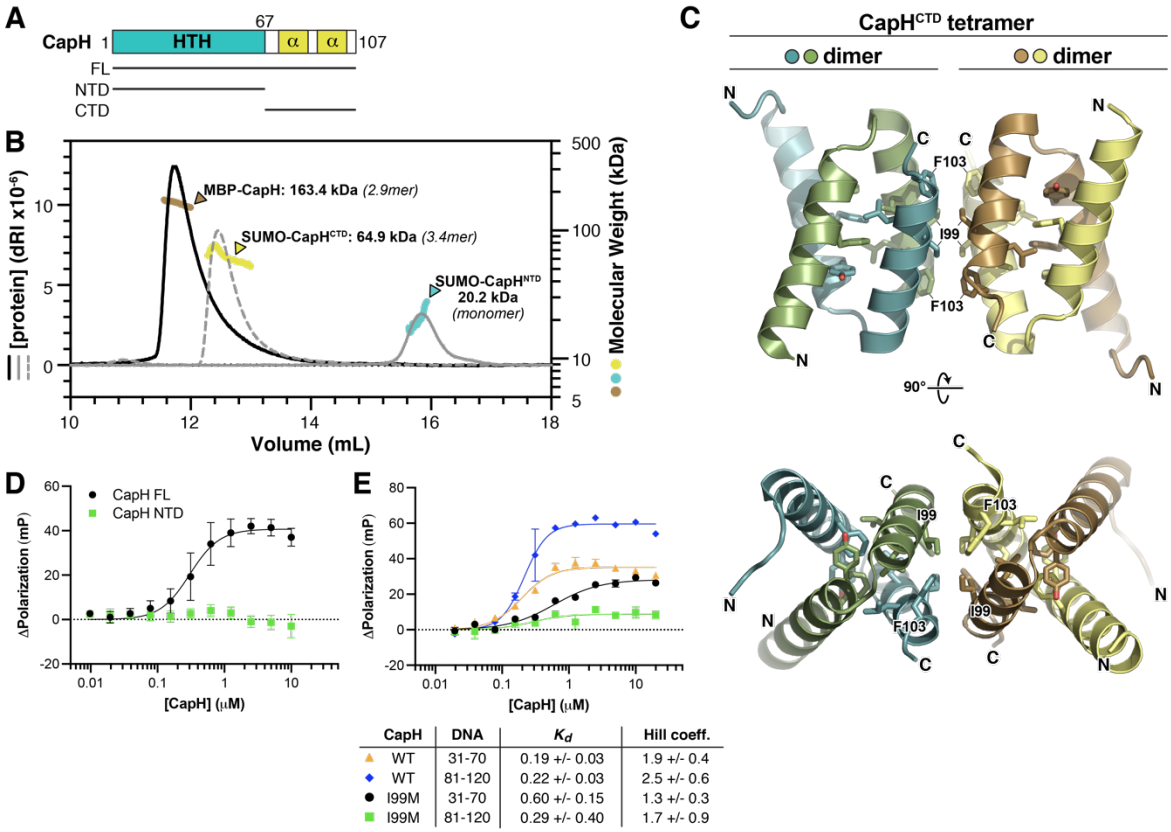
(A) Domain schematic of *E. coli* MS115-1 CapH, showing the truncations used for oligomeric state determination.

(B) Size exclusion chromatography coupled to multi-angle light scattering (SEC-MALS) determination of CapH oligomeric state. For each construct, measured molecular weight and inferred oligomer are indicated. MBP-fused full-length CapH (monomer MW=57.0 kDa) is shown in brown; SUMO-fused CapH^{CTD} (monomer MW=19.1 kDa) is yellow, and SUMO-fused CapH^{NTD} (monomer MW=22.0 kDa) is cyan. See **Figure A.C.S6A** for SEC-MALS analysis of untagged CapH.

(C) Crystal structure of the CapH^{CTD} homotetramer. Residues comprising the hydrophobic core of each dimer (Phe81, Tyr85, Leu96, and Leu100) are shown as sticks, and residues comprising the hydrophobic tetramerization interface (Ile99 and Phe103) are shown as sticks and labeled. See **Figure A.C.S6B** for structure of the CapH^{CTD}(I99M) homodimer.

(D) Fluorescence polarization assay showing binding of *E. coli* MS115-1 CapH (His_6-MBP tagged; full-length (black circles) or NTD(green squares)) to the Site 2 DNA. WT K_d = 0.30 +/- 0.08 μM, Hill coefficient = 2.0 +/- 0.7; no binding detected for CapH^{NTD}).

(E) Fluorescence polarization assay showing binding of *E. coli* MS115-1 CapH (His_6-MBP tagged) to the Site 1 (bases 31-70) and Site 2 (bases 81-120) DNAs. Wild-type CapH binding Site 1 is shown in orange triangles, and binding Site 2 is shown in blue diamonds. CapH(I99M) binding Site 1 is shown in black circles, and binding Site 2 is shown in green squares. Fit K_d and Hill coefficient for each combination is shown. See **Figure A.C.S6C** for EMSA analysis of CapH WT and I99M binding to the Site 2 DNA.



CapH	DNA	K_d	Hill coeff.
▲ WT	31-70	0.19 +/- 0.03	1.9 +/- 0.4
◆ WT	81-120	0.22 +/- 0.03	2.5 +/- 0.6
● I99M	31-70	0.60 +/- 0.15	1.3 +/- 0.3
■ I99M	81-120	0.29 +/- 0.40	1.7 +/- 0.9

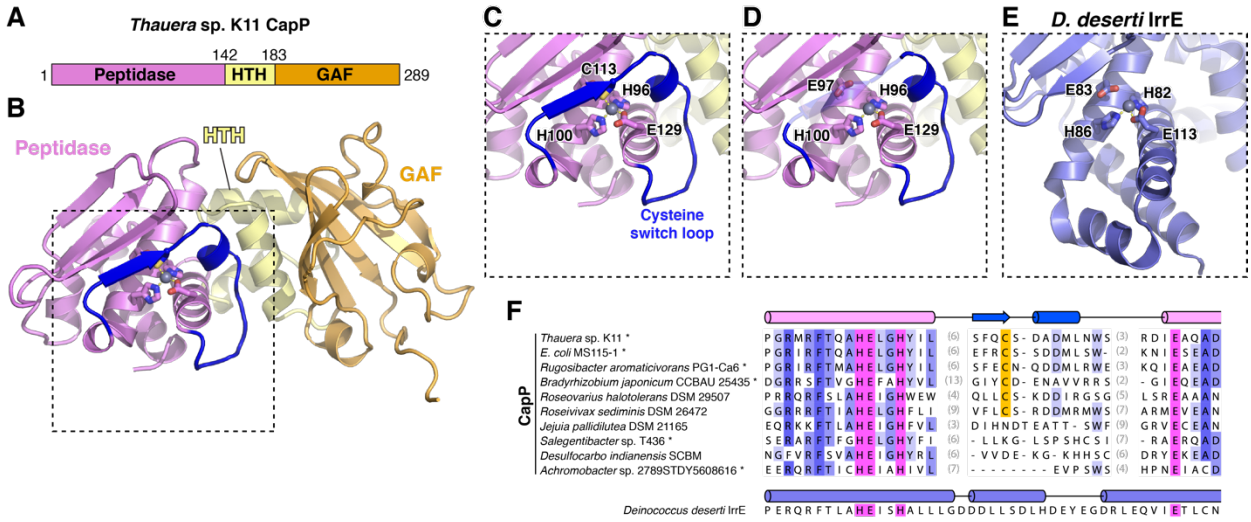


Figure 3.20 Structure of CapP reveals an internal cysteine switch

(A) Domain schematic of *Thauera* sp. K11 CapP, with N-terminal Zn^{2+} peptidase domain pink, central HTH domain yellow, and C-terminal GAF domain orange. See **Figure A.C.S4A** for comparison of the *Thauera* sp. K11 and *E. coli* MS115-1 CBASS operons.

(B) Structure of *Thauera* sp. K11 CapP with domains colored as in panel (A). Shown in sticks are active-site residues H96, E97, H100, and E129, and a bound Zn^{2+} ion is shown as a gray sphere. Shown in blue is the internal cysteine switch loop, with C113 (shown as sticks) coordinating the bound Zn^{2+} ion.

(C) Closeup view of the *Thauera* sp. K11 CapP active site. See **Figure A.C.S4B-C** for comparison with a cysteine switch-containing matrix metalloprotease.

(D) Closeup view of the *Thauera* sp. K11 CapP active site as in panel (C), with transparent cysteine switch loop to show the position of the catalytic glutamate residue E97.

(E) Equivalent closeup view of the *Deinococcus deserti* IrrE active site (PDB ID 3DTI; (Vujičić-Žagar et al., 2009), showing the active site residues and bound Zn^{2+} ion.

(F) Sequence alignment of representative CBASS-associated CapP proteins, showing the cysteine switch loop present in a subset of these proteins. See **Table A.C.S1** for full list, and **Figure A.C.S4D** for evolutionary tree of CapP with presence/absence of the cysteine switch annotated.

Figure 3.21 CapP cleaves CapH and is stimulated by single-stranded DNA

(A) Anti-GFP western blot showing coexpression in *E. coli* of an MBP-CapH-GFP fusion construct with wild-type or catalytic-dead (E98Q) CapP. Full-length MBP-CapH-GFP is indicated with a yellow arrowhead, and the C-terminal product of CapP cleavage is indicated with a white arrowhead. RNAP: RNA polymerase loading control.

(B) In vitro cleavage of purified MBP-CapH-GFP (yellow arrowhead) into N-terminal and C-terminal products (white arrowheads) by CapP is stimulated by DNA. For both ssDNA and dsDNA, the highest concentration is 350 ng/μL, followed by three 5-fold dilutions. Red asterisk indicates the band that was analyzed by Edman degradation. See **Figure A.C.5** for analysis of DNA binding and sequence-specificity for cleavage activation.

(C) Results of Edman degradation of CapH C-terminal cleavage product (red asterisk from panel (B), showing cleavage at residue R83. See **Figure A.C.S8** for full data.

(D) Cartoon view of the $CapH^{67\text{-}107}$ dimer, with R83 colored white and shown as sticks.

(E) In vitro cleavage of MBP-CapH-GFP (wild-type or R83A mutant) by CapP, in the presence of 10 μM ssDNA.

(F) GFP reporter assay showing effect of a CapP E98Q catalytic-dead mutant, CapH R83A mutant, or removal of CapH residues 83-107 (*capH* 1-82) on GFP expression. RNAP: RNA polymerase loading control. See full blot in **Figure A.C.S1A**.

(G) CBASS expression reporter system with FLAG-NucC, showing effect of a CapH R83A mutant, or removal of CapH residues 83-107 (*capH* 1-82) on FLAG-NucC expression. α-RNAP: anti-RNA polymerase loading control. See full blot in **Figure A.C.S1B**.

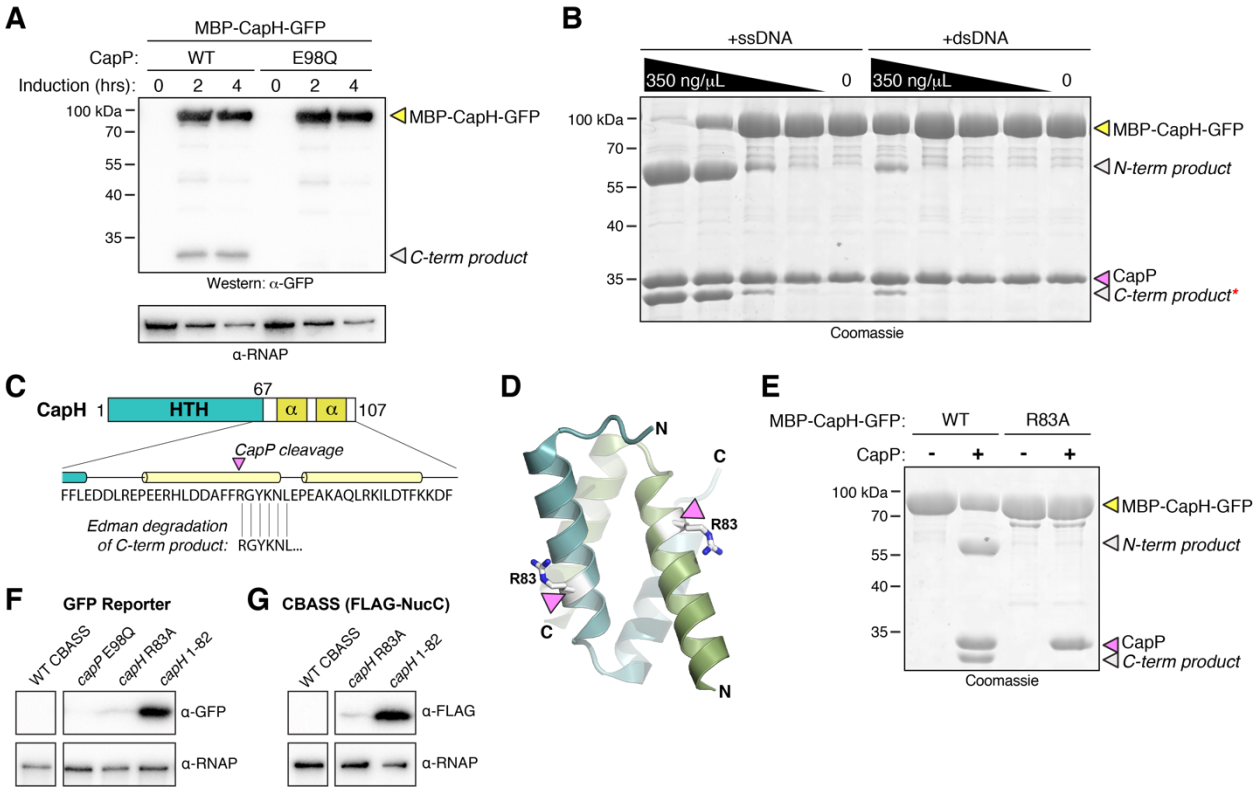

A
MBP-CapH-GFP
CapP:
WT
E98Q
Induction (hrs):
0 2 4 0 2 4
100 kDa
70
55
40
35
MBP-CapH-GFP
C-term product
Western: α-GFP
α-RNAP
B
+ssDNA
+dsDNA
350 ng/μL
0
350 ng/μL
0
100 kDa
70
55
40
35
MBP-CapH-GFP
N-term product
CapP
C-term product*
Coomassie
C
CapH
1
HTH
α
α
67
107
CapP cleavage
FFLEDDLREPEERHLDDAFFRGYKNLEPEAKAQLRKILDTFKKDF
Edman degradation of C-term product:
RGYKNL...
D
N
C
R83
R83
C
N
E
MBP-CapH-GFP:
WT
R83A
CapP:
-
+
-
+
100 kDa
70
55
40
35
MBP-CapH-GFP
N-term product
CapP
C-term product
Coomassie
F
GFP Reporter
WT CBASS
capP E98Q
capH R83A
capH 1-82
α-GFP
α-RNAP
G
CBASS (FLAG-NucC)
WT CBASS
capH R83A
capH 1-82
α-FLAG
α-RNAP

Figure 3.22 CapH and CapP induce CBASS expression in response to DNA damage

(A) Anti-GFP western blot showing coexpression in *E. coli* of an MBP-CapH-GFP fusion construct with wild-type CapP after exposure to zeocin (100 μg/mL). Full-length MBP-CapH-GFP is indicated with a yellow arrowhead, and the C-terminal product of CapP cleavage is indicated with a white arrowhead. α-RNAP: anti-RNA polymerase loading control.

(B) Western blot of the CBASS expression reporter system with FLAG-NucC, showing FLAG-NucC expression after exposure to zeocin (100 μg/mL). α-RNAP: anti-RNA polymerase loading control.

(C) Western blot of the CBASS expression reporter system with FLAG-NucC, showing FLAG-NucC expression after exposure to mitomycin C (1 μg/mL). α-RNAP: anti-RNA polymerase loading control.

(D) Proposed signaling pathway for DNA damage-responsive transcriptional control systems in bacteria.

(E) Diverse systems in bacteria that include an HTH transcriptional repressor (yellow) and a DNA damage-activated Zn^{2+} metallopeptidase (pink) that targets the transcriptional repressor for cleavage.

(F) Known or likely bacterial defense systems associated with CapH (yellow) and CapP (pink)-like genes. See **Methods** for accession numbers for each gene.


A
+zeocin
-zeocin
Hours: 0 1 2 0 1 2
100 kDa
70
55
40
35
25
MBP-CapH-GFP
C-term product
α-GFP
α-RNAP
B
CBASS (FLAG-NucC) + zeocin
WT CBASS
ΔcapP
Minutes: 0 15 30 60 120 180 240 0 15 30 60 120 180 240
α-FLAG
α-RNAP
C
CBASS (FLAG-NucC) + mitomycin C
WT CBASS
ΔcapP
Minutes: 0 15 30 60 120 180 240 0 15 30 60 120 180 240
α-FLAG
α-RNAP
D
DNA damage
Peptidase
HTH
Target(s)
E
CBASS Immunity:
E. coli MS115-1
capP
capH
cdnC
DNA Damage Response:
Deinococcus
irrE
ddrO
DDR gene
Prophage:
B. subtilis phage ϕ105
immA(ϕ105)
c ϕ105
orf4
Mobile Elements:
B. subtilis ICEBs1
int
immA
immR
xis
F
DISARM
vapC
drmD
drmMI
drmA
drmB
drmC
AAA+ ATPase
Pycsar
guanylate cyclase
hyp. protein
cNMPBD+TIR
A/G cyclase
DUF2188
DUF2188 + hypothetical protein
hyp.
hyp.
DUF2188
DUF3892
DUF3892
Metallo-β-lactamase (Ub-6a)
hyp
metallo-β-lactamase
E2 - linker - E1 - JAB
hyp.
DUF6527 (Ub-6e.2)
Ub
Ub
Ub
E1-like
DUF6527
multi-Ub domain

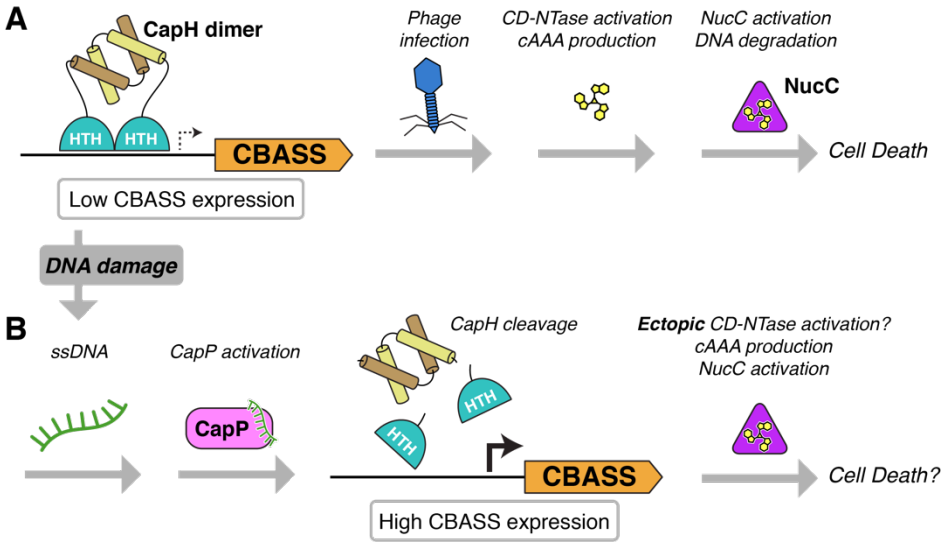


Figure 3.23 Model for CapH/CapP function in CBASS-mediated immunity

(A) In the basal state, CapH dimers (depicted) or monomers (not shown) bind their cognate CBASS promoter and repress transcription. Low-level CBASS expression is sufficient for detection of a lytic phage infection, CD-NTase activation and second messenger production (in *E. coli* MS115-1 CBASS, cyclic tri-AMP/cAAA), followed by effector activation (in *E. coli* MS115-1, NucC) and cell death.

(B) Upon DNA damage, activation of other anti-phage immune systems, or environmental stimuli, CapP is activated and cleaves CapH. CapH cleavage and release from the CBASS promoter mediates a dramatic increase in CBASS activation, resulting in ectopic CD-NTase activation, second messenger production, effector activation, and possibly cell death.

References

Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H., Adams, P.D., 2012. Towards automated crystallographic structure refinement with phenix.refine. Acta Crystallographica Section D: Biological Crystallography 68, 352. https://doi.org/10.1107/S0907444912001308

Auchtung, J.M., Lee, C.A., Monson, R.E., Lehman, A.P., Grossman, A.D., 2005. Regulation of a Bacillus subtilis mobile genetic element by intercellular signaling and the global DNA damage response. Proceedings of the National Academy of Sciences of the United States of America 102, 12554–12559. https://doi.org/10.1073/PNAS.0505835102/SUPPL_FILE/05835FIG6.PDF

Ayora, S., Carrasco, B., Cárdenas, P.P., César, C.E., Cañas, C., Yadav, T., Marchisone, C., Alonso, J.C., 2011. Double-strand break repair in bacteria: a view from Bacillus subtilis. FEMS Microbiology Reviews 35, 1055–1081. https://doi.org/10.1111/J.1574-6976.2011.00272.X

Baek, K., Svenningsen, S., Eisen, H., Sneppen, K., Brown, S., 2003. Single-cell analysis of lambda immunity regulation. J Mol Biol 334, 363–372. https://doi.org/10.1016/j.jmb.2003.09.037

Beilsten-Edmands, J., Winter, G., Gildea, R., Parkhurst, J., Waterman, D., Evans, G., 2020. Scaling diffraction data in the DIALS software package: algorithms and new approaches for multi-crystal scaling. Acta Crystallographica Section D: Structural Biology 76, 385–399. https://doi.org/10.1107/S2059798320003198

Benler, S., Koonin, E.V., 2020. Phage lysis-lysogeny switches and programmed cell death: Danse macabre. BioEssays 42, 2000114. https://doi.org/10.1002/BIES.202000114

Bernheim, A., Sorek, R., 2020. The pan-immune system of bacteria: antiviral defence as a community resource. Nature Reviews Microbiology 18, 113–119. https://doi.org/10.1038/s41579-019-0278-2

Blanchard, L., Guérin, P., Roche, D., Cruveiller, S., Pignol, D., Vallenet, D., Armengaud, J., de Groot, A., 2017. Conservation and diversity of the IrrE/DdrO-controlled radiation response in radiation-resistant Deinococcus bacteria. MicrobiologyOpen 6, e00477. https://doi.org/10.1002/mbo3.477

Blankenchip, C.L., Nguyen, J.V., Lau, R.K., Ye, Q., Gu, Y., Corbett, K.D., 2022. Control of bacterial immune signaling by a WYL domain transcription factor. Nucleic Acids Research 50, 5239–5250. https://doi.org/10.1093/nar/gkac343

Bose, B., Auchtung, J.M., Lee, C.A., Grossman, A.D., 2008. A conserved anti-repressor controls horizontal gene transfer by proteolysis. Molecular Microbiology 70, 570–582.

Bose, B., Grossman, A.D., 2011. Regulation of horizontal gene transfer in Bacillus subtilis by activation of a conserved site-specific protease. Journal of Bacteriology 193, 22–29. https://doi.org/10.1128/JB.01143-10

Burroughs, A.M., Aravind, L., 2020. Identification of Uncharacterized Components of Prokaryotic Immune Systems and Their Diverse Eukaryotic Reformulations. Journal of Bacteriology 202, e00365-20. https://doi.org/10.1128/JB.00365-20

Burroughs, A.M., Iyer, L.M., Aravind, L., 2009. Natural history of the E1-like superfamily: Implication for adenylation, sulfur transfer, and ubiquitin conjugation. Proteins: Structure, Function, and Bioinformatics 75, 895–910. https://doi.org/10.1002/PROT.22298

Burroughs, A.M., Zhang, D., Schäffer, D.E., Iyer, L.M., Aravind, L., 2015. Comparative genomic analyses reveal a vast, novel network of nucleotide-centric systems in biological conflicts, immunity and signaling. Nucleic Acids Research 43, 10633–10654. https://doi.org/10.1093/nar/gkv1267

Cerdà-Costa, N., Xavier Gomis-Rüth, F., 2014. Architecture and function of metallopeptidase catalytic domains. Protein Science 23, 123–144. https://doi.org/10.1002/pro.2400

Chankova, S.G., Dimova, E., Dimitrova, M., Bryant, P.E., 2007. Induction of DNA double-strand breaks by zeocin in Chlamydomonas reinhardtii and the role of increased DNA double-strand breaks rejoining in the formation of an adaptive response. Radiation and Environmental Biophysics 46, 409–416. https://doi.org/10.1007/S00411-007-0123-2/FIGURES/7

Cohen, D., Melamed, S., Millman, A., Shulman, G., Oppenheimer-Shaanan, Y., Kacen, A., Doron, S., Amitai, G., Sorek, R., 2019. Cyclic GMP–AMP signalling protects bacteria against viral infection. Nature 574, 691–695. https://doi.org/10.1038/s41586-019-1605-5

de Groot, A., Siponen, M.I., Magerand, R., Eugénie, N., Martin-Arevalillo, R., Doloy, J., Lemaire, D., Brandelet, G., Parcy, F., Dumas, R., Roche, P., Servant, P., Confalonieri, F., Arnoux, P., Pignol, D., Blanchard, L., 2019. Crystal structure of the transcriptional repressor DdrO: insight into the metalloprotease/repressor-controlled radiation response in Deinococcus. Nucleic Acids Research 47, 11403–11417. https://doi.org/10.1093/NAR/GKZ883

Doron, S., Melamed, S., Ofir, G., Leavitt, A., Lopatina, A., Keren, M., Amitai, G., Sorek, R., 2018. Systematic discovery of antiphage defense systems in the microbial pangenome. Science 359, eaar4120. https://doi.org/10.1126/science.aar4120

Dy, R.L., Richter, C., Salmond, G.P.C., Fineran, P.C., 2014. Remarkable mechanisms in microbes to resist phage infections. Annual Review of Virology 1, 307–331. https://doi.org/10.1146/ANNUREV-VIROLOGY-031413-085500

Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K., 2010. Features and development of Coot. Acta Crystallographica Section D: Biological Crystallography 66, 486. https://doi.org/10.1107/S0907444910007493

Evans, P.R., Murshudov, G.N., 2013. How good are my data and what is the resolution? Acta Crystallographica Section D: Biological Crystallography 69, 1204–1214. https://doi.org/10.1107/S0907444913000061

French, S., Wilson, K., 1978. On the treatment of negative intensity observations. Acta Crystallographica Section A 34, 517–525. https://doi.org/10.1107/S0567739478001114

Gangwar, S.P., Meena, S.R., Saxena, A.K., 2014. Comparison of four different crystal forms of the Mycobacterium tuberculosis ESX-1 secreted protein regulator EspR. Acta Crystallographica Section F:Structural Biology Communications 70, 433–437. https://doi.org/10.1107/S2053230X14004166/FW5444SUP1.PDF

Gross-Langenhoff, M., Hofbauer, K., Weber, J., Schultz, A., Schultz, J.E., 2006. cAMP is a ligand for the tandem GAF domain of human phosphodiesterase 10 and cGMP for the tandem GAF domain of phosphodiesterase 11. Journal of Biological Chemistry 281, 2841–2846. https://doi.org/10.1074/jbc.M511468200

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

Ho, Y.-S.J., Burden, L.M., Hurley, J.H., 2000. Structure of the GAF domain, a ubiquitous signaling motif and a new class of cyclic GMP receptor. The EMBO journal 19, 5288–99. https://doi.org/10.1093/emboj/19.20.5288

Iyer, L.M., Burroughs, A.M., Aravind, L., 2006. The prokaryotic antecedents of the ubiquitin-signaling system and the early evolution of ubiquitin-like β-grasp domains. Genome Biology 7, 1–23. https://doi.org/10.1186/GB-2006-7-7-R60/FIGURES/7

Kabsch, W., 2010. XDS. Acta Crystallographica Section D Biological Crystallography 66, 125–132. https://doi.org/10.1107/S0907444909047337

Kidane, D., Sanchez, H., Alonso, J.C., Graumann, P.L., 2004. Visualization of DNA double-strand break repair in live bacteria reveals dynamic recruitment of Bacillus subtilis RecF, RecO and RecN proteins to distinct sites on the nucleoids. Molecular Microbiology 52, 1627–1639. https://doi.org/10.1111/J.1365-2958.2004.04102.X

Lau, R.K., Ye, Q., Birkholz, E.A., Berg, K.R., Patel, L., Mathews, I.T., Watrous, J.D., Ego, K., Whiteley, A.T., Lowey, B., Mekalanos, J.J., Kranzusch, P.J., Jain, M., Pogliano, J., Corbett, K.D., 2020. Structure and Mechanism of a Cyclic Trinucleotide-Activated Bacterial Endonuclease Mediating Bacteriophage Immunity. Molecular Cell 77, 723-733.e6. https://doi.org/10.1016/j.molcel.2019.12.010

Ledvina, H.E., Ye, Q., Gu, Y., Quan, Y., Lau, R.K., Zhou, H., Corbett, K.D., Whiteley, A.T., 2022. cGASylation by a bacterial E1-E2 fusion protein primes antiviral immune signaling. bioRxiv 2022.03.31.486616. https://doi.org/10.1101/2022.03.31.486616

Levdikov, V.M., Blagova, E., Colledge, V.L., Lebedev, A.A., Williamson, D.C., Sonenshein, A.L., Wilkinson, A.J., 2009. Structural Rearrangement Accompanying Ligand Binding in the GAF Domain of CodY from Bacillus subtilis. Journal of Molecular Biology 390, 1007–1018. https://doi.org/10.1016/j.jmb.2009.05.077

Lewis, R.J., Brannigan, J.A., Offen, W.A., Smith, I., Wilkinson, A.J., 1998. An evolutionary link between sporulation and prophage induction in the structure of a repressor:anti-repressor complex. Journal of Molecular Biology 283, 907–912. https://doi.org/10.1006/JMBI.1998.2163

Little, J.W., 1991. Mechanism of specific LexA cleavage: Autodigestion and the role of RecA coprotease. Biochimie 73, 411–421. https://doi.org/10.1016/0300-9084(91)90108-D

Lowey, B., Whiteley, A.T., Keszei, A.F.A., Morehouse, B.R., Mathews, I.T., Antine, S.P., Cabrera, V.J., Kashin, D., Niemann, P., Jain, M., Schwede, F., Mekalanos, J.J., Shao, S., Lee, A.S.Y., Kranzusch, P.J., 2020. CBASS Immunity Uses CARF-Related Effectors to Sense 3′–5′- and 2′–5′-Linked Cyclic Oligonucleotide Signals and Protect Bacteria from Phage Infection. Cell 182, 38-49.e17. https://doi.org/10.1016/J.CELL.2020.05.019

Ludanyi, M., Blanchard, L., Dulermo, R., Brandelet, G., Bellanger, L., Pignol, D., Lemaire, D., de Groot, A., 2014. Radiation response in Deinococcus deserti: IrrE is a metalloprotease that cleaves repressor protein DdrO. Molecular Microbiology 94, 434–449. https://doi.org/10.1111/mmi.12774

Luyten, Y.A., Hausman, D.E., Young, J.C., Doyle, L.A., Higashi, K.M., Ubilla-Rodriquez, N.C., Lambert, A.R., Arroyo, C.S., Forsberg, K.J., Morgan, R.D., Stoddard, B.L., Kaiser, B.K., 2022. Identification and characterization of the WYL BrxR protein and its gene as separable regulatory elements of a BREX phage restriction system. bioRxiv 2021.12.19.473356. https://doi.org/10.1101/2021.12.19.473356

Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J.M., Wolf, Y.I., Yakunin, A.F., van der Oost, J., Koonin, E.V., 2011a. Evolution and classification of the CRISPR–Cas systems. Nature Reviews Microbiology 9, 467–477. https://doi.org/10.1038/nrmicro2577

Makarova, K.S., Wolf, Y.I., Koonin, E.V., 2013. Comparative genomics of defense systems in archaea and bacteria. Nucleic Acids Research 41, 4360–4377. https://doi.org/10.1093/nar/gkt157

Makarova, K.S., Wolf, Y.I., Snir, S., Koonin, E.V., 2011b. Defense Islands in Bacterial and Archaeal Genomes and Prediction of Novel Defense Systems. Journal of Bacteriology 193, 6039–6056. https://doi.org/10.1128/JB.05535-11

Martinez, S.E., Beavo, J.A., Hol, W.G.J., 2002a. GAF domains: two-billion-year-old molecular switches that bind cyclic nucleotides. Mol Interv 2, 317–323. https://doi.org/10.1124/mi.2.5.317

Martinez, S.E., Heikaus, C.C., Klevit, R.E., Beavo, J.A., 2008. The structure of the GAF a domain from phosphodiesterase 6C reveals determinants of cGMP binding, a conserved binding surface, and a large cGMP-dependent conformational change. Journal of Biological Chemistry 283, 25913–25919. https://doi.org/10.1074/jbc.M802891200

Martinez, S.E., Wu, A.Y., Glavas, N.A., Tang, X.B., Turley, S., Hol, W.G.J., Beavo, J.A., 2002b. The two GAF domains in phosphodiesterase 2A have distinct roles in dimerization and in cGMP binding. Proceedings of the National Academy of Sciences of the United States of America 99, 13260–13265. https://doi.org/10.1073/pnas.192374899

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., Read, R.J., 2007. Phaser crystallographic software. Journal of Applied Crystallography 40, 658. https://doi.org/10.1107/S0021889807021206

McGeehan, J.E., Streeter, S.D., Thresh, S.J., Ball, N., Ravelli, R.B.G., Kneale, G.G., 2008. Structural analysis of the genetic switch that regulates the expression of restriction-modification genes. Nucleic Acids Research 36, 4778–4787. https://doi.org/10.1093/nar/gkn448

Millman, A., Melamed, S., Amitai, G., Sorek, R., 2020. Diversity and classification of cyclic-oligonucleotide-based anti-phage signalling systems. Nature Microbiology 2020 5:12 5, 1608–1615. https://doi.org/10.1038/s41564-020-0777-y

Mohanraju, P., Makarova, K.S., Zetsche, B., Zhang, F., Koonin, E.V., van der Oost, J., 2016. Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. Science 353, aad5147. https://doi.org/10.1126/science.aad5147

Ofir, G., Melamed, S., Sberro, H., Mukamel, Z., Silverman, S., Yaakov, G., Doron, S., Sorek, R., 2018. DISARM is a widespread bacterial defence system with broad anti-phage activities. Nature Microbiology 3, 90–98. https://doi.org/10.1038/s41564-017-0051-0

Picton, D.M., Harling-Lee, J.D., Duffner, S.J., Went, S.C., Morgan, R.D., Hinton, J.C.D., Blower, T.R., 2022. A widespread family of WYL-domain transcriptional regulators co-localizes with diverse phage defence systems and islands. Nucleic Acids Res 50, 5191–5207. https://doi.org/10.1093/nar/gkac334

Rajagopala, S.V., Casjens, S., Uetz, P., 2011. The protein interaction map of bacteriophage lambda. BMC Microbiology 11, 213. https://doi.org/10.1186/1471-2180-11-213

Saha, C.K., Pires, R.S., Brolin, H., Delannoy, M., Atkinson, G.C., 2021. FlaGs and webFlaGs: discovering novel biology through the analysis of gene neighbourhood conservation. Bioinformatics 37, 1312–1314. https://doi.org/10.1093/BIOINFORMATICS/BTAA788

Salamov, V.S.A., Solovyevand, A., 2011. Automatic annotation of microbial genomes and metagenomic sequences. Metagenomics and Its Applications in Agriculture, Biomedicine and Environmental Studies; Li, RW, Ed 61–78.

Sassanfar, M., Roberts, J.W., 1990. Nature of the SOS-inducing signal in Escherichia coli: The involvement of DNA replication. Journal of Molecular Biology 212, 79–96. https://doi.org/10.1016/0022-2836(90)90306-7

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, E6048–E6055. https://doi.org/10.1073/pnas.1801233115

Shevtsov, M.B., Streeter, S.D., Thresh, S.J., Swiderska, A., McGeehan, J.E., Kneale, G.G., 2015. Structural analysis of DNA binding by C.Csp231I, a member of a novel class of R-M controller proteins regulating gene expression. Acta

Crystallographica Section D: Biological Crystallography 71, 398–407. https://doi.org/10.1107/S139900471402690X

Springman, E.B., Angleton, E.L., Birkedal-Hansen, H., Van Wart, H.E., 1990. Multiple modes of activation of latent human fibroblast collagenase: Evidence for the role of a Cys73 active-site zinc complex in latency and a “cysteine switch” mechanism for activation. Proceedings of the National Academy of Sciences of the United States of America 87, 364–368. https://doi.org/10.1073/PNAS.87.1.364

Tal, N., Morehouse, B.R., Millman, A., Stokar-Avihail, A., Avraham, C., Fedorenko, T., Yirmiya, E., Herbst, E., Brandis, A., Mehlman, T., Oppenheimer-Shaanan, Y., Keszei, A.F.A., Shao, S., Amitai, G., Kranzusch, P.J., Sorek, R., 2021. Cyclic CMP and cyclic UMP mediate bacterial immunity against phages. Cell 184, 5728-5739.e16. https://doi.org/10.1016/J.CELL.2021.09.031

Tropea, J.E., Cherry, S., Waugh, D.S., 2009. Expression and purification of soluble His(6)-tagged TEV protease. Methods in molecular biology (Clifton, NJ) 498, 297–307. https://doi.org/10.1007/978-1-59745-196-3_19

Van Duyne, G.D., Standaert, R.F., Karplus, P.A., Schreiber, S.L., Clardy, J., 1993. Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. Journal of molecular biology 229, 105–124. https://doi.org/10.1006/jmbi.1993.1012

Van Wart, H.E., Birkedal-Hansen, H., 1990. The cysteine switch: A principle of regulation of metalloproteinase activity with potential applicability to the entire matrix metalloproteinase gene family. Proceedings of the National Academy of Sciences of the United States of America 87, 5578–5582. https://doi.org/10.1073/pnas.87.14.5578

Vujičić-Žagar, A., Dulermo, R., Le Gorrec, M., Vannier, F., Servant, P., Sommer, S., de Groot, A., Serre, L., 2009. Crystal Structure of the IrrE Protein, a Central Regulator of DNA Damage Repair in Deinococcaceae. Journal of Molecular Biology 386, 704–716. https://doi.org/10.1016/j.jmb.2008.12.062

Whiteley, A.T., Eaglesham, J.B., de Oliveira Mann, C.C., Morehouse, B.R., Lowey, B., Nieminen, E.A., Danilchanka, O., King, D.S., Lee, A.S.Y., Mekalanos, J.J., Kranzusch, P.J., 2019. Bacterial cGAS-like enzymes synthesize diverse nucleotide signals. Nature 567, 194–199. https://doi.org/10.1038/s41586-019-0953-5

Ye, Q., Lau, R.K., Mathews, I.T., Birkholz, E.A., Watrous, J.D., Azimi, C.S., Pogliano, J., Jain, M., Corbett, K.D., 2020. HORMA Domain Proteins and a Trip13-like ATPase Regulate Bacterial cGAS-like Enzymes to Mediate Bacteriophage

Immunity. Molecular Cell 77, 709-722.e7.
https://doi.org/10.1016/j.molcel.2019.12.009

Chapter 3, in full, is a reprint of the material as it is published in The EMBO Journal 2022. Lau, Rebecca K.; Enustun, Eray; Gu, Yajie; Nguyen, Justin V.; Corbett, Kevin D. “A Conserved Signaling Pathway Activates Bacterial CBASS Immune Signaling in Response to DNA Damage”. The dissertation/thesis author was the primary investigator and author of this paper.

CONCLUSIONS

Here, I show that bacteria possess a unique phage defense system, CBASS, which utilizes cyclic oligonucleotide signaling to enact cell death through abortive infection. In Chapter 1, I describe the components of the system and detail the mechanism of each player involved. Each CBASS system contains a CD-NTase that synthesizes a cyclic oligonucleotide second messenger when it is active. The CD-NTase in the system, CdnC, synthesizes cyclic tri-AMP (cAAA). I also show that CdnC binds both the HORMA protein, Cap7, and dsDNA, and both are required for CdnC activation in vitro. The HORMA protein must be in the closed conformation, which is stabilized by peptide binding, to bind to CdnC. However, the peptide signal that the HORMA protein binds remains unknown. *In vitro*, the HORMA protein binds a poly-His tag to stabilize this conformation, but this is not likely the actual ligand as it was artificially introduced during protein purification. Additionally, phage display revealed that the HORMA protein binds the sequence HGKILLT, but BLAST searches to this protein sequence are not a promising method to find the biologically relevant binding partner because this sequence is so short. Recently, papers from many groups have shown that phage defense systems, including Type II CBASS systems, can be activated by phage proteins. These proteins include capsid proteins, tail proteins, portal proteins, and components of phage-encoded DNA polymerases. Thus, it would not be surprising if the Type III CBASS system studied here would also bind a phage protein. Alternatively, we have not ruled out that the HORMA from CBASS could bind a bacterial protein that is specifically expressed in response to cellular stress. We also still do not

know the function of dsDNA in HORMA activation in vitro. While it is likely an in vitro artifact, we still do not know if there is some other component of this system that is required for activation within the cells.

In Chapter 2 we show that the CBASS effector nuclease NucC binds to and is activated by cAAA. After binding cAAA, NucC undergoes a conformational change that induces hexamerization. Hexamerization places the active sites of the two trimers in close proximity to each other leading to a double strand break in the target DNA. We also show that NucC is a non-specific nuclease and will cleave both host and phage DNA. In CBASS, the cAAA second messenger is synthesized by CdnC. However, components of Type III CRISPR-Cas systems have also been shown to synthesize cyclic oligoadenylates ranging in size from cA_3 – cA_6 (Niewoehner et al., 2017). Furthermore, some CRISPR-Cas systems contain NucC, such as the one from *Vibrio* shown in Chapter 2. It would be interesting to see if there is any overlap of bacterial strains containing both a Type III CRISPR system and a CBASS that has an effector that binds a cyclic oligoadenylate molecule. If there is overlap, these systems could work together to degrade viral DNA. It would also be interesting if these two systems were mutually exclusive. CRISPR-Cas systems detect foreign DNA sequences and cleave the DNA so that the phage cannot replicate, but leave the host genome intact. CBASS, on the other hand, does not discriminate between its own DNA and foreign DNA. It may not be evolutionarily advantageous to kill the cell if the CRISPR-Cas system is activated and can kill the phage without the bacteria also killing itself. In Chapter 2, we suggest that the NucC from *Vibrio* is also an indiscriminate nuclease

because it does not show sequence preference *in vitro*. A NucC effector that is found in the same operon as a Type III CRISPR system in a prophage encoded on a *Vibrio metoecus* genome is activated by cAAA synthesized when the guide RNA encounters foreign RNA (Grüschow et al., 2021). However, it is not clear whether this system also functions through abortive infection.

In Chapter 3, we find that 6.5% of CBASS systems, as well as other bacterial defense systems, are transcriptionally regulated by proteins CapH and CapP. CapH is a helix-turn-helix protein that represses CBASS expression by binding near the transcription start site. CapP is a zinc metallopeptidase that cleaves CapH upon binding single-stranded DNA. Once cleaved, CapH is unable to bind to the promoter region of CBASS, resulting in de-repression and expression of the CBASS genes. This finding leads to two questions: (1) where does the ssDNA come from and (2) what is the evolutionary advantage for some defense systems to have transcriptional regulation that is dependent on DNA damage? To answer these questions, we have two working hypotheses that are being tested in the lab by graduate student Chelsea Blankenchip. The first hypothesis is that DNA damage comes from other defense systems such as restriction-modification systems or CRISPR-Cas systems. It is well known that R-M systems create double stranded DNA breaks, often with overhangs which are single stranded. Cas proteins also cleave DNA to create some portion of single stranded DNA. Groups such as the Marraffini group have already studied interactions between R-M and CRISPR-Cas systems and shown that they are able to work together to protect against phage (Maguin et al., 2022). They show that in one instance, phage DNA is

cleaved by an R-M system and the fragments are subsequently inserted into the CRISPR array. It is possible that ssDNA generated from an R-M system or CRISPR-Cas could also activate CapP and cause an increase in CBASS expression. This way, if some of the phage are able to escape the R-M and CRISPR-Cas systems, which protect the cells without the cells themselves dying, CBASS will be primed to activate at high levels and would work as a backup system.

Our second hypothesis regarding the origin of DNA damage is also related to the question of why responding to DNA damage through activation of an anti-phage defense system may benefit a cell. We hypothesize that DNA damage may arise from an external stress. Prophages also respond to DNA damage, and will excise themselves from the genome and switch to the lytic cycle when their bacterial host is stressed. We hypothesize that by responding to DNA damage, CBASS could ensure that it is able to respond to not only lytic phages entering the cell from outside, but lysogenic phages, too. Prophages will often recognize DNA damage by binding to activated RecA, which itself binds single-stranded DNA during DNA damage repair. By binding directly to ssDNA, CapP could be activated upstream of prophage induction, which could be advantageous for the bacterial cells.

Since the discovery of CBASS, the Sorek and Kranzusch groups have identified and characterized additional phage defense systems which utilize cyclic oligonucleotides (Doron et al., 2018; Tal et al., 2021). These systems, Pycsar and Thoeris, do not contain a CD-NTase, but rather encode other enzymes that synthesize small molecule second messengers. Pycsar includes a cyclase which synthesizes cyclic

CMP and cyclic UMP. These cyclic nucleotides activate effector proteins such as transmembrane helix proteins which create leaky cell membranes or Toll/interleukin-1 receptor (TIR) proteins which deplete NAD^+ in cells, both of which ultimately result in abortive infection (Tal et al., 2021). Thoeris also activates TIR proteins, although it is thought that the small molecule second messenger in these systems is actually synthesized as a byproduct of the TIR activation, variant of cyclic adenine diphosphate ribose (v-cADPR) (Ka et al., 2020). The Sorek lab has shown that v-cADPR binds a second NADase, which further depletes NAD^+ and causes cell death (Ofir et al., 2021). CapP and CapH can be associated with Pycsar, as well as other non-cyclic nucleotide signaling systems. It will be interesting to see if these systems that share the CapH and CapP are also protective against prophage induction, as well as infection by lytic phage. Overall, this is a quickly growing field with new systems and regulatory mechanisms being discovered rapidly.

References

Doron, S., Melamed, S., Ofir, G., Leavitt, A., Lopatina, A., Keren, M., Amitai, G., Sorek, R., 2018. Systematic discovery of antiphage defense systems in the microbial pangenome. Science 359, eaar4120. https://doi.org/10.1126/science.aar4120

Grüschow, S., Adamson, C.S., White, M.F., 2021. Specificity and sensitivity of an RNA targeting type III CRISPR complex coupled with a NucC endonuclease effector. Nucleic acids research 49. https://doi.org/10.1093/nar/gkab1190

Ka, D., Oh, H., Park, E., Kim, J.-H., Bae, E., 2020. Structural and functional evidence of bacterial antiphage protection by Thoeris defense system via NAD+ degradation. Nat Commun 11, 2816. https://doi.org/10.1038/s41467-020-16703-w

Maguin, P., Varble, A., Modell, J.W., Marraffini, L.A., 2022. Cleavage of viral DNA by restriction endonucleases stimulates the type II CRISPR-Cas immune response. Molecular Cell 82, 907-919.e7. https://doi.org/10.1016/j.molcel.2022.01.012

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

Ofir, G., Herbst, E., Baroz, M., Cohen, D., Millman, A., Doron, S., Tal, N., Malheiro, D.B.A., Malitsky, S., Amitai, G., Sorek, R., 2021. Antiviral activity of bacterial TIR domains via immune signalling molecules. Nature 600, 116–120. https://doi.org/10.1038/s41586-021-04098-7

Tal, N., Morehouse, B.R., Millman, A., Stokar-Avihail, A., Avraham, C., Fedorenko, T., Yirmiya, E., Herbst, E., Brandis, A., Mehlman, T., Oppenheimer-Shaanan, Y., Keszei, A.F.A., Shao, S., Amitai, G., Kranzusch, P.J., Sorek, R., 2021. Cyclic CMP and cyclic UMP mediate bacterial immunity against phages. Cell 184, 5728-5739.e16. https://doi.org/10.1016/J.CELL.2021.09.031

APPENDIX A: Supplemental Figures and Tables for Chapter 1 HORMA Domain Proteins and a Trip13-like ATPase Regulate Bacterial cGAS-like Enzymes to Mediate Bacteriophage Immunity

Figure A.A.S24 Sequence analysis of bacterial HORMA proteins and Pch2/TRIP13-like ATPases

(A) Phylogenetic tree of bacterial HORMA domain proteins, with HORMA1 proteins colored green, HORMA2 proteins purple, and HORMA3 proteins blue. Stars indicate proteins used in this study: *Ec* HORMA, *Pa* HORMA2, and *Pa* HORMA3. Prior studies (Burroughs et al., 2015; Tromer et al., 2019) identified two families of bacterial HORMA proteins, but our analysis indicates the existence of three families. HORMA1 proteins are encoded in operons with one HORMA domain protein, while HORMA2 and HORMA3 proteins are encoded together (see panel C). **(B)** Phylogenetic tree of bacterial Pch2/TRIP13-like ATPases, with proteins from HORMA1 operons colored green, and proteins from HORMA2/3 operons colored purple. Yellow stars indicate proteins used in this study: *Ec* Trip13, *Pa* Trip13, and *Rhizobiales* (*Rh*) Trip13. The divergence of Trip13 proteins from HORMA1 and HORMA2/3 operons agrees with a prior report (Tromer et al., 2019), and suggests that HORMA and Trip13 proteins have co-evolved within each operon family. **(C)** Phylogenetic trees for 19 pairs of HORMA2 and HORMA3 proteins annotated by Burroughs et al. (Burroughs et al., 2015), showing co-evolution of the two protein families within the operon. See **Table A.A.S1** and **Table A.A.S2** for all HORMA and Trip13 sequences used to generate panels A, B and C.

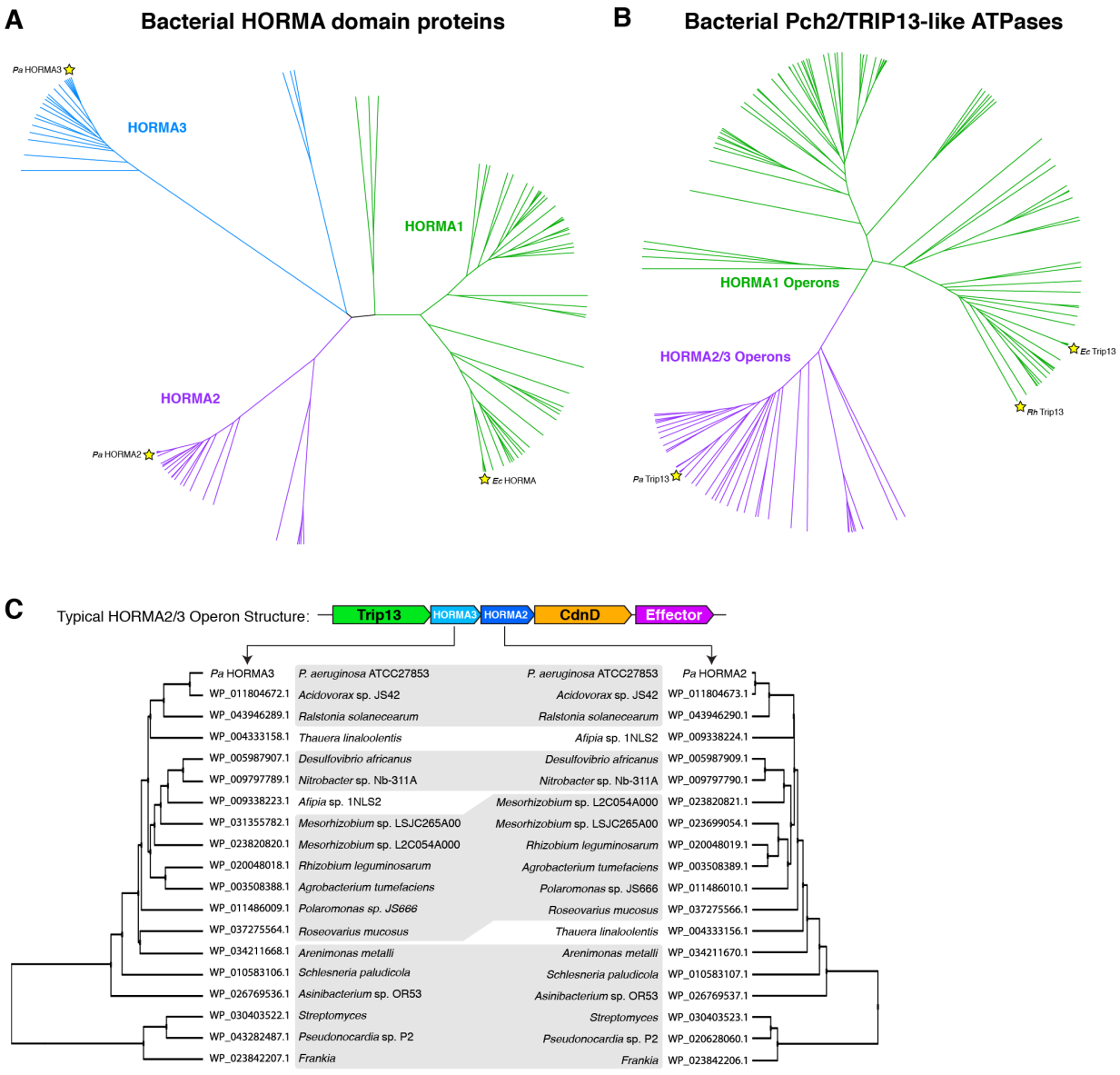
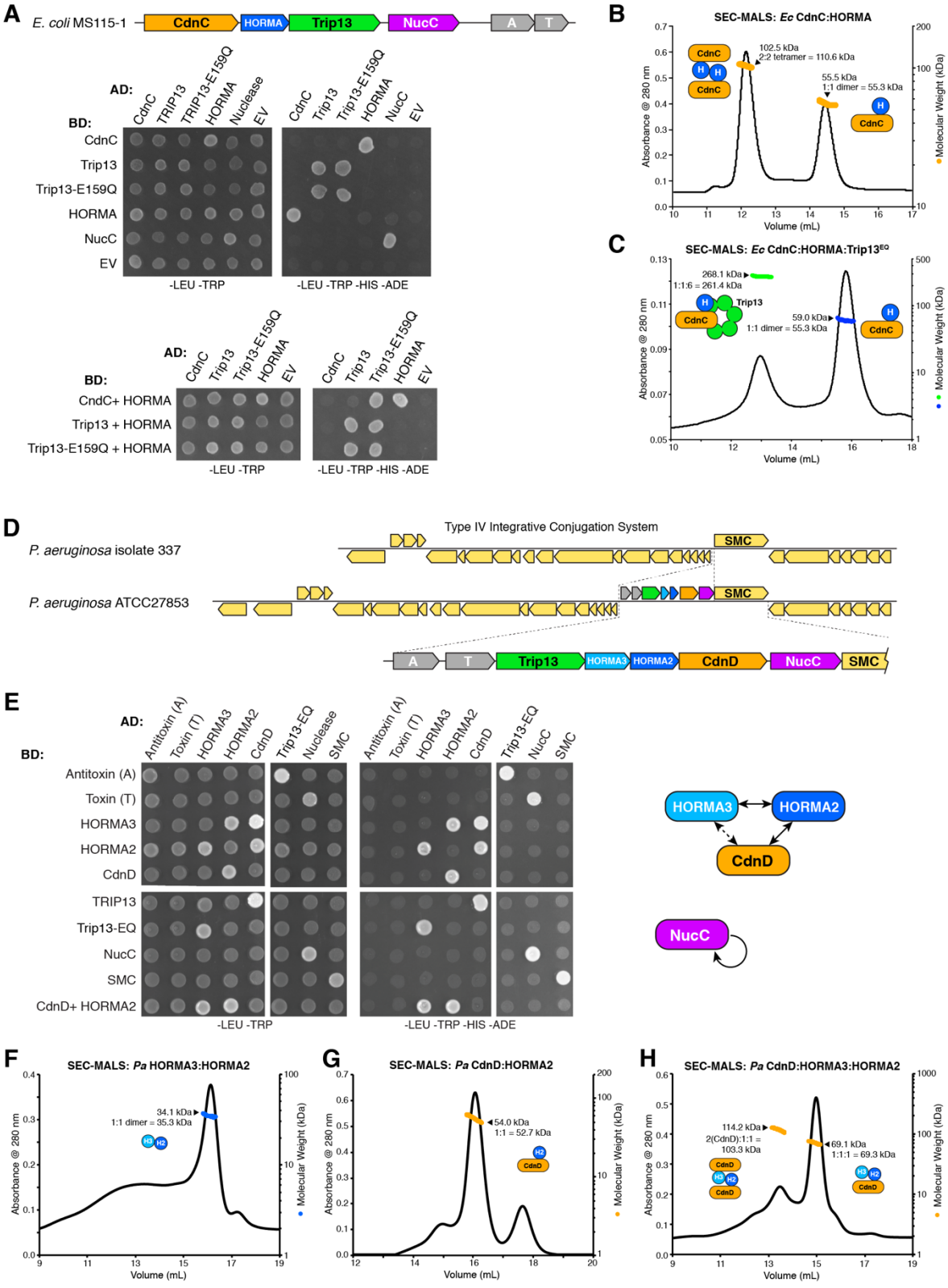

A
Bacterial HORMA domain proteins
Pa HORMA3
HORMA3
HORMA1
HORMA2
Pa HORMA2
Ec HORMA
B
Bacterial Pch2/TRIP13-like ATPases
HORMA1 Operons
HORMA2/3 Operons
Ec Trip13
Rh Trip13
Pa Trip13
C
Typical HORMA2/3 Operon Structure:
Trip13
HORMA3
HORMA2
CdnD
Effector
Pa HORMA3
WP_011804672.1
WP_043946289.1
WP_004333158.1
WP_005987907.1
WP_009797789.1
WP_009338223.1
WP_031355782.1
WP_023820820.1
WP_020048018.1
WP_003508388.1
WP_011486009.1
WP_037275564.1
WP_034211668.1
WP_010583106.1
WP_026769536.1
WP_030403522.1
WP_043282487.1
WP_023842207.1
P. aeruginosa ATCC27853
Acidovorax sp. JS42
Ralstonia solanecearum
Thauera linaloolentis
Desulfovibrio africanus
Nitrobacter sp. Nb-311A
Afipia sp. 1NLS2
Mesorhizobium sp. LSJC265A00
Mesorhizobium sp. L2C054A000
Rhizobium leguminosarum
Agrobacterium tumefaciens
Polaromonas sp. JS666
Roseovarius mucosus
Arenimonas metalli
Schlesneria paludicola
Asinibacterium sp. OR53
Streptomyces
Pseudonocardia sp. P2
Frankia
P. aeruginosa ATCC27853
Acidovorax sp. JS42
Ralstonia solanecearum
Afipia sp. 1NLS2
Desulfovibrio africanus
Nitrobacter sp. Nb-311A
Mesorhizobium sp. L2C054A000
Mesorhizobium sp. LSJC265A00
Rhizobium leguminosarum
Agrobacterium tumefaciens
Polaromonas sp. JS666
Roseovarius mucosus
Thauera linaloolentis
Arenimonas metalli
Schlesneria paludicola
Asinibacterium sp. OR53
Streptomyces
Pseudonocardia sp. P2
Frankia
Pa HORMA2
WP_011804673.1
WP_043946290.1
WP_009338224.1
WP_005987909.1
WP_009797790.1
WP_023820821.1
WP_023699054.1
WP_020048019.1
WP_003508389.1
WP_011486010.1
WP_037275566.1
WP_004333156.1
WP_034211670.1
WP_010583107.1
WP_026769537.1
WP_030403523.1
WP_020628060.1
WP_023842206.1

Figure A.A.S25 Protein-protein interactions in HORMA/Trip13-associated CBASS systems

(A) Yeast two-hybrid and three-hybrid analysis of protein-protein interactions in the *E. coli* MS115-1 CD-NTase containing operon. “-LEU -TRP” refers to non-selective growth media, and “-LEU -TRP -HIS -ADE” refers to selective growth media. This operon encodes a RES-Xre type toxin-antitoxin gene pair, labeled A/T, which was not tested for protein-protein interactions. “EV” refers to empty vector controls. **(B)** SEC-MALS of the *Ec* CdnC:HORMA complex, showing that it forms 1:1 and 2:2 complexes in solution. **(C)** SEC-MALS of *Ec* CdnC:HORMA:Trip13^{EQ}, showing that it forms a 1:1:6 complex in solution when pre-incubated with ATP. **(D)** Schematics comparing the genomic context of the P. aeruginosa ATCC27853 CBASS system to a homologous locus in *P. aeruginosa* isolate 337 (NCBI reference sequence NZ_FRQL01000018.1). The CBASS system in P. aeruginosa ATCC27853 has been incorporated into a putative Type IV integrative conjugation system (yellow) that encodes a putative SMC-like ATPase. While the CBASS system is encoded on the same strand as the putative SMC protein, this protein is not expected to function with the CBASS system. **(E)** Yeast two-hybrid and three-hybrid analysis of protein-protein interactions in the *P. aeruginosa* ATCC27853 CBASS operon. In addition to the CBASS proteins (CdnD, HORMA3, HORMA2, Trip13, and NucC), the putative toxin-antitoxin gene pair (PIN domain protein (T) and helix-turn-helix protein (A)) and SMC-like ATPase were tested. *Right:* Summary of protein-protein interactions in the *P. aeruginosa* ATCC27853 operon. The dashed line between HORMA3 and CdnD indicates that this interaction was detected with only one tag combination; nonetheless the two proteins form a complex when co-expressed in *E. coli* (not shown). **(F)** SEC-MALS analysis of *Pa* HORMA3:HORMA2, showing that the proteins form a 1:1 complex in solution. **(G)** SEC-MALS analysis of *Pa* CdnD:HORMA2, showing that the proteins form a 1:1 complex in solution. Size-exclusion chromatography of purified *Pa* CdnD:HORMA3 complex indicates a likely 1:1 stoichiometry, though aggregation of the complex precluded analysis by SEC-MALS (not shown). **(H)** SEC-MALS analysis of *Pa* CdnD:HORMA3:HORMA2, showing that the proteins form a mix of 1:1:1 and 2:1:1 stoichiometries in solution.


A
E. coli MS115-1
CdnC
HORMA
Trip13
NucC
A
T
AD:
BD:
CdnC
TRIP13
TRIP13-E159Q
HORMA
Nuclease
EV
Trip13
NucC
CdnC+ HORMA
Trip13 + HORMA
Trip13-E159Q + HORMA
-LEU -TRP
-LEU -TRP -HIS -ADE
B
SEC-MALS: Ec CdnC:HORMA
102.5 kDa
2:2 tetramer = 110.6 kDa
55.5 kDa
1:1 dimer = 55.3 kDa
Absorbance @ 280 nm
Molecular Weight (kDa)
Volume (mL)
C
SEC-MALS: Ec CdnC:HORMA:Trip13EQ
268.1 kDa
1:1:6 = 261.4 kDa
Trip13
59.0 kDa
1:1 dimer = 55.3 kDa
D
Type IV Integrative Conjugation System
P. aeruginosa isolate 337
P. aeruginosa ATCC27853
SMC
A
T
Trip13
HORMA3
HORMA2
CdnD
NucC
SMC
E
Antitoxin (A)
Toxin (T)
HORMA3
HORMA2
CdnD
Trip13-EQ
Nuclease
SMC
NucC
TRIP13
Trip13-EQ
CdnD+ HORMA2
-LEU -TRP
-LEU -TRP -HIS -ADE
F
SEC-MALS: Pa HORMA3:HORMA2
34.1 kDa
1:1 dimer = 35.3 kDa
G
SEC-MALS: Pa CdnD:HORMA2
54.0 kDa
1:1 = 52.7 kDa
H
SEC-MALS: Pa CdnD:HORMA3:HORMA2
114.2 kDa
2(CdnD):1:1 =
103.3 kDa
69.1 kDa
1:1:1 = 69.3 kDa

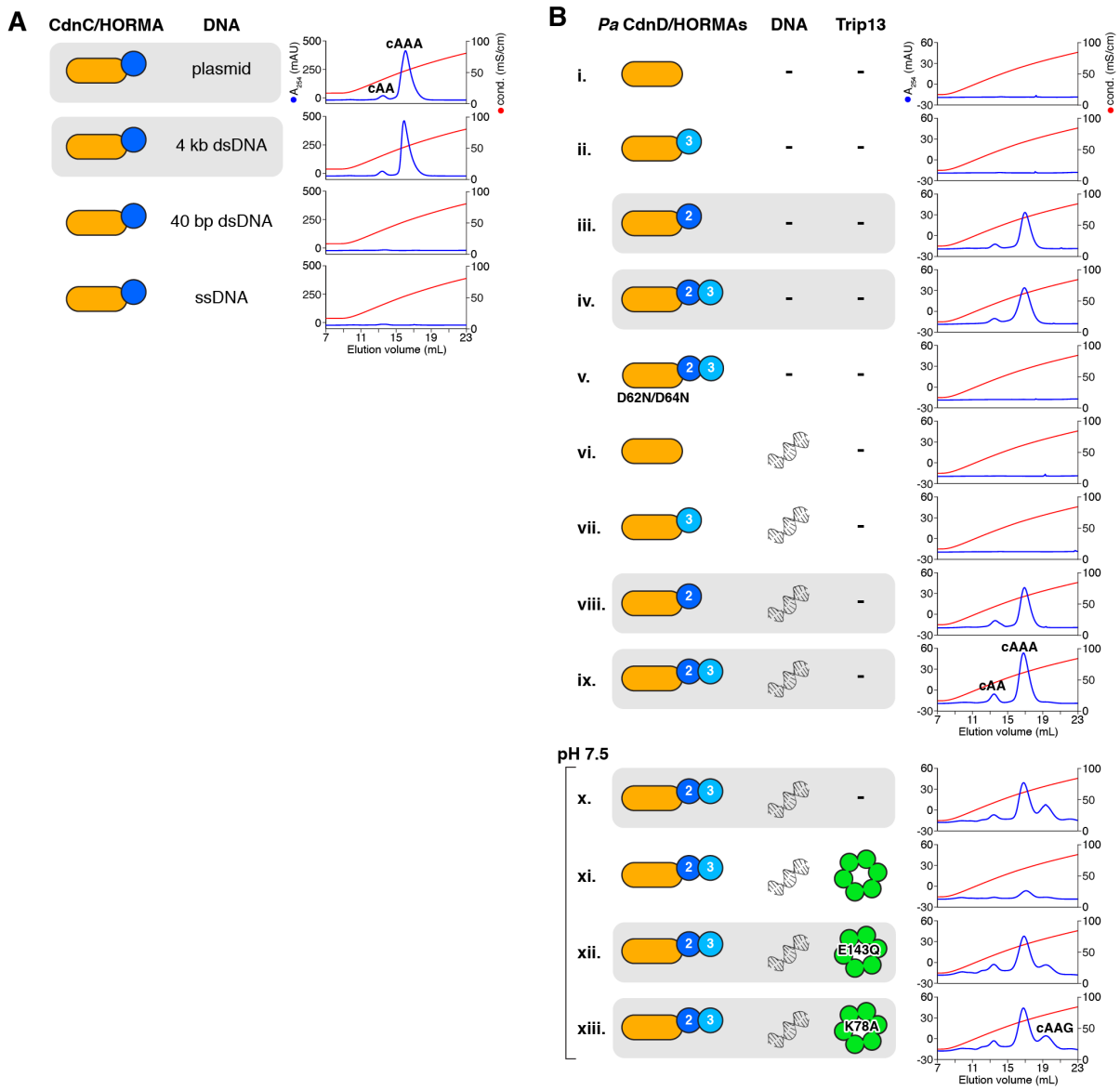
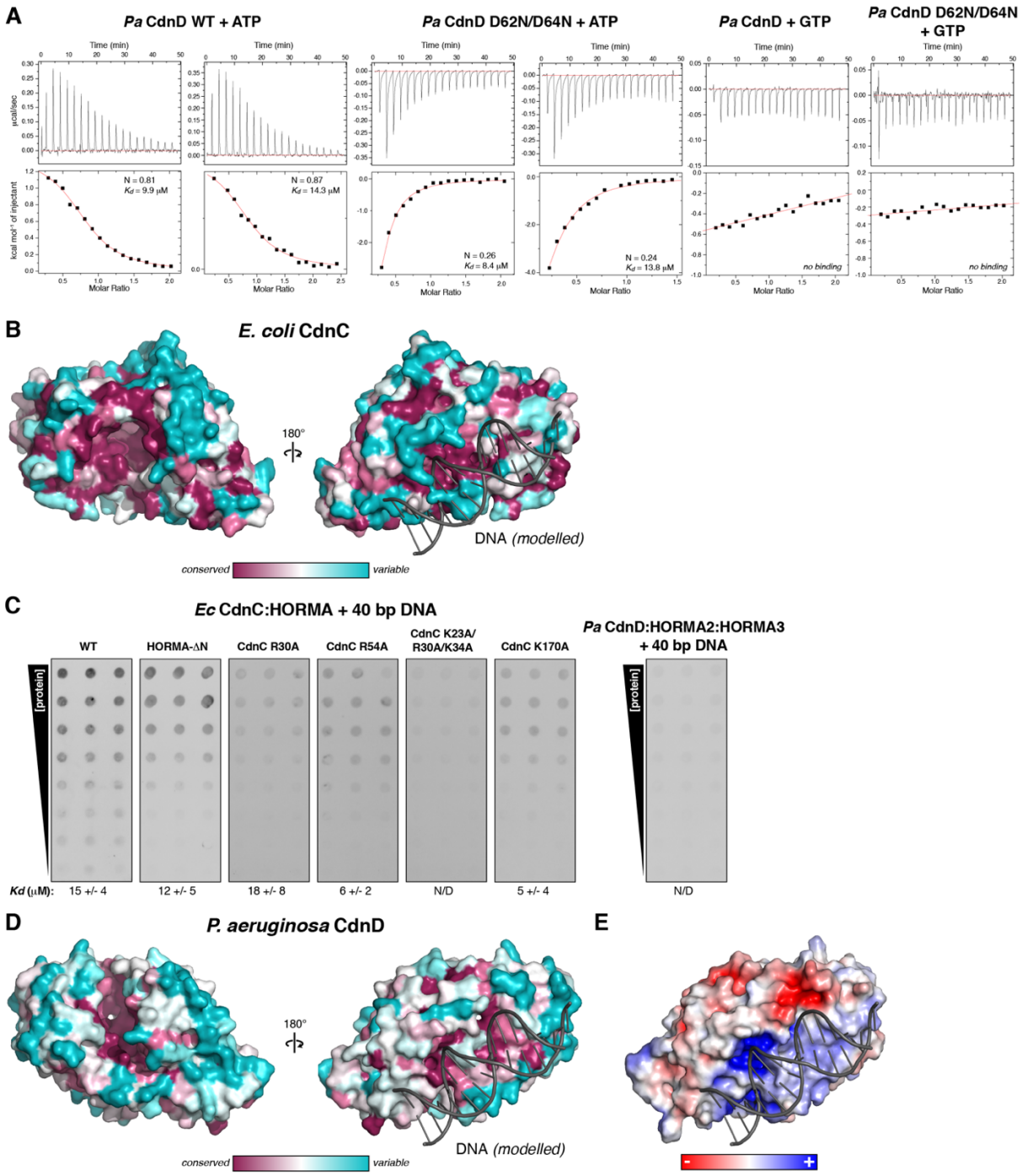


Figure A.A.S26 Second messenger synthesis by *Ec* CdnC and *Pa* CdnD

(A) Second messenger synthesis assays as in **Figure 1.1C**, with different DNAs: ~4 kb plasmid DNA (top), 4 kb linear dsDNA, 40 bp linear dsDNA, and single-stranded DNA (sheared salmon-sperm DNA). **(B)** Second messenger synthesis assays with *Pa* CdnD, HORMA3 (light blue), HORMA2 (dark blue), plasmid DNA, and Trip13. The peaks labeled “cAA” (sample ix), “cAAA” (sample ix), and “cAAG” (sample xiii) were verified by LC MS/MS. Samples x-xiii were assayed at pH 7.5 rather than 8.5 (as in samples i-ix) to accommodate *Pa* Trip13, which was insoluble at pH 8.5. *Pa* CdnD synthesizes cAAG as a minor product at pH 7.5.

Figure A.A.S27 Nucleotide and DNA binding of *Ec* CdnC and *Pa* CdnD

(A) Isothermal titration calorimetry (ITC) analysis of *Pa* CdnD binding either ATP or GTP. Both wild-type and D62N/D64N mutants bind ATP with a K_d of ~10 μM, but neither detectably binds GTP. A similar analysis of *Ec* CdnC was not possible due to its stably-bound ATP. (B) Two views of *Ec* CdnC (left panel equivalent to **Figure 1.2A**), with surface colored by conservation. Bound DNA (gray) is modelled from a superposition with *M. musculus* cGAS bound to DNA (PDB 4K9B) (Gao et al., 2013). **(C)** DNA binding dot-blot assays for *Ec* CdnC:HORMA and *Pa* CdnD:HORMA2:HORMA3 complexes. Panels show triplicate binding assays with a 5'-fluorescein-labeled 40 bp dsDNA, with protein concentration starting at 50 μM and dropping 2-fold each row (final row, 0.39 μM). *Kd* values were calculated with PRISM; N/D, not determined. **(D)** Two views of *Pa* CdnD equivalent (left panel equivalent to **Figure 1.2B**, with surface colored by conservation. **(E)** View of *Pa* CdnD colored by surface charge.


A
Pa CdnD WT + ATP
Pa CdnD D62N/D64N + ATP
Pa CdnD + GTP
Pa CdnD D62N/D64N + GTP
Time (min)
µcal/sec
kcal mol⁻¹ of injectant
Molar Ratio
N = 0.81
Kd = 9.9 µM
N = 0.87
Kd = 14.3 µM
N = 0.26
Kd = 8.4 µM
N = 0.24
Kd = 13.8 µM
no binding
no binding
B
E. coli CdnC
180°
DNA (modelled)
conserved
variable
C
Ec CdnC:HORMA + 40 bp DNA
WT
HORMA-ΔN
CdnC R30A
CdnC R54A
CdnC K23A/ R30A/K34A
CdnC K170A
Pa CdnD:HORMA2:HORMA3 + 40 bp DNA
[protein]
Kd (µM):
15 +/- 4
12 +/- 5
18 +/- 8
6 +/- 2
N/D
5 +/- 4
N/D
D
P. aeruginosa CdnD
180°
DNA (modelled)
conserved
variable
E
-
+

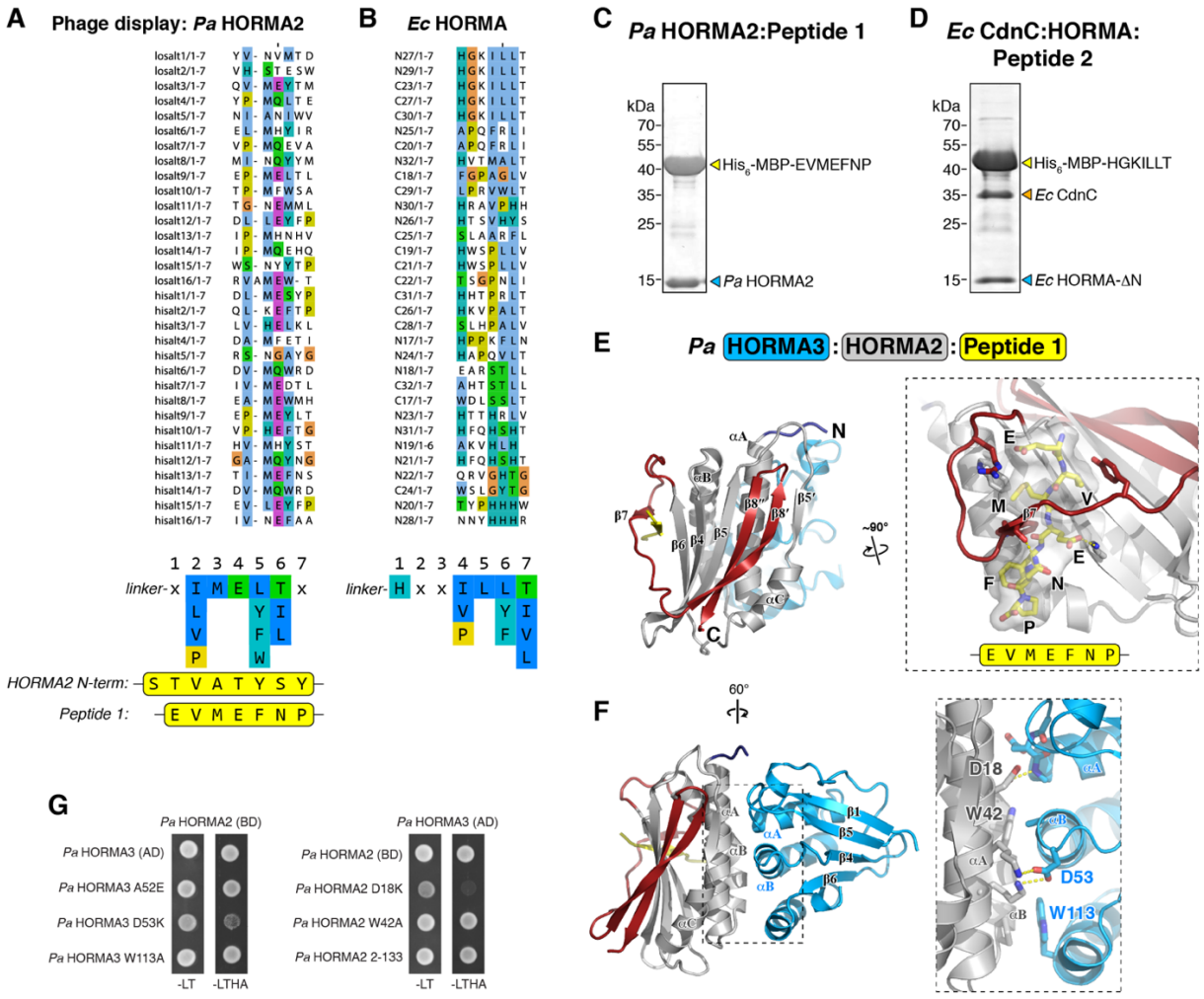


Figure A.A.S28 HORMA-closure motif and HORMA-HORMA interactions

(A) Aligned sequences from round 4 of phage display with *Pa* HORMA2 as bait. A consensus sequence is shown at bottom, aligned with the sequence of the *Pa* HORMA2 N-terminus, which binds *Pa* HORMA2 as a closure motif in our crystal structure (**Figure 1.3B**), and Peptide 1, designed from the consensus motif. **(B)** Aligned sequences from round 4 of phage display with *Ec* HORMA as bait. A consensus sequence is shown at bottom. **(C)** Ni-NTA purification of coexpressed His_6-MBP-Peptide 1 (sequence: EVMEFNP) with untagged *Pa* HORMA2. **(D)** Ni-NTA purification of coexpressed His_6-MBP-Peptide 2 (sequence: HGKILLT) with untagged *Ec* HORMA-ΔN and *Ec* CdnC. **(E)** Structure of the *Pa* HORMA3:HORMA2:Peptide 1 complex, with closeup view of HORMA2 binding Peptide 1 (yellow). **(F)** Rotated view of the *Pa* HORMA3:HORMA2:Peptide 1 complex, with closeup showing the HORMA1-HORMA2 interface. **(G)** Yeast two-hybrid analysis of the *Pa* HORMA1-HORMA2 interaction. BD: fusion to the Gal4 DNA-binding domain; AD: fusion to the Gal4 activation domain. -LT: media lacking leucine and tryptophan (non-selective); -LTHA: media lacking leucine, tryptophan, histidine, and adenine (stringent selection).

Figure A.A.S29 Structure of CD-NTase:HORMA complexes

(A) Model of a *Pa* CdnD:HORMA2:Peptide 1:HORMA3 complex, based on separate structures of CdnD:HORMA2:Peptide 1 (**Figure 1.4A**) and HORMA3:HORMA2:Peptide 1 (**Figure 1.3E**). HORMA3 and CdnD occupy different surfaces of HORMA2, suggesting they can bind simultaneously. **(B)** Overall view of the asymmetric unit in crystals of Ec CdnC:HORMA-ΔN. The N-terminal His6-tag of each HORMA-ΔN protomer mimics a closure motif in its non-crystallographic symmetry mate (see inset; two HORMA-ΔN protomers colored blue and gray), forming a symmetric 2:2 complex. **(C)** *Top:* Overall structure of *Ec* HORMA in the *Ec* CdnC:HORMA-ΔN complex. Coloring is equivalent to **Figure 1.3A-D**, with the mobile N-terminus blue and safety belt red. The protein adopts a closed, but empty, conformation. *Bottom:* Model of *Ec* HORMA bound to a consensus closure motif peptide (HGKILLT), based on a superposition of *Ec* HORMA-ΔN with the *Pa* HORMA2:Peptide 1 complex. **(D)** Superdex 200 size exclusion chromatography elution profiles for *Ec* CdnC:HORMA-ΔN with N-terminal His_6-tag on HORMA intact (dotted blue line), cleaved (solid blue line), or cleaved in the presence of Peptide 2 (HGKILLT; orange). Prior to tag cleavage, the complex adopts predominantly 2:2 stoichiometry. After cleavage, the complex adopts predominantly 1:1 stoichiometry. **(E)** Closeup views of the *Ec* CdnC-$Trip13^{EQ}$ interface in the *Ec* CdnC:HORMA:$Trip13^{EQ}$ complex, with CdnC orange and $Trip13^{EQ}$ monomers E and F shown as light blue and violet, respectively. *Top* shows closeup view equivalent to that of panel A, and *bottom* shows opposite view. **(F)** Yeast three-hybrid analysis (*Ec* Trip13 in pGADT7, *Ec* CdnC (AD fusion) and HORMA (untagged) in pBridge) showing that mutation of $Trip13^{EQ}$ residues Asp19, Phe20, Glu21, and Arg100 disrupt binding to CdnC. -LT: media lacking leucine and tryptophan (non-selective); -LTHA: media lacking leucine, tryptophan, histidine, and adenine (stringent selection). **(G)** Yeast three-hybrid analysis showing that mutation of *Ec* CdnC residue Arg68, or truncation of the HORMA N-terminus, disrupt binding to $Trip13^{EQ}$.

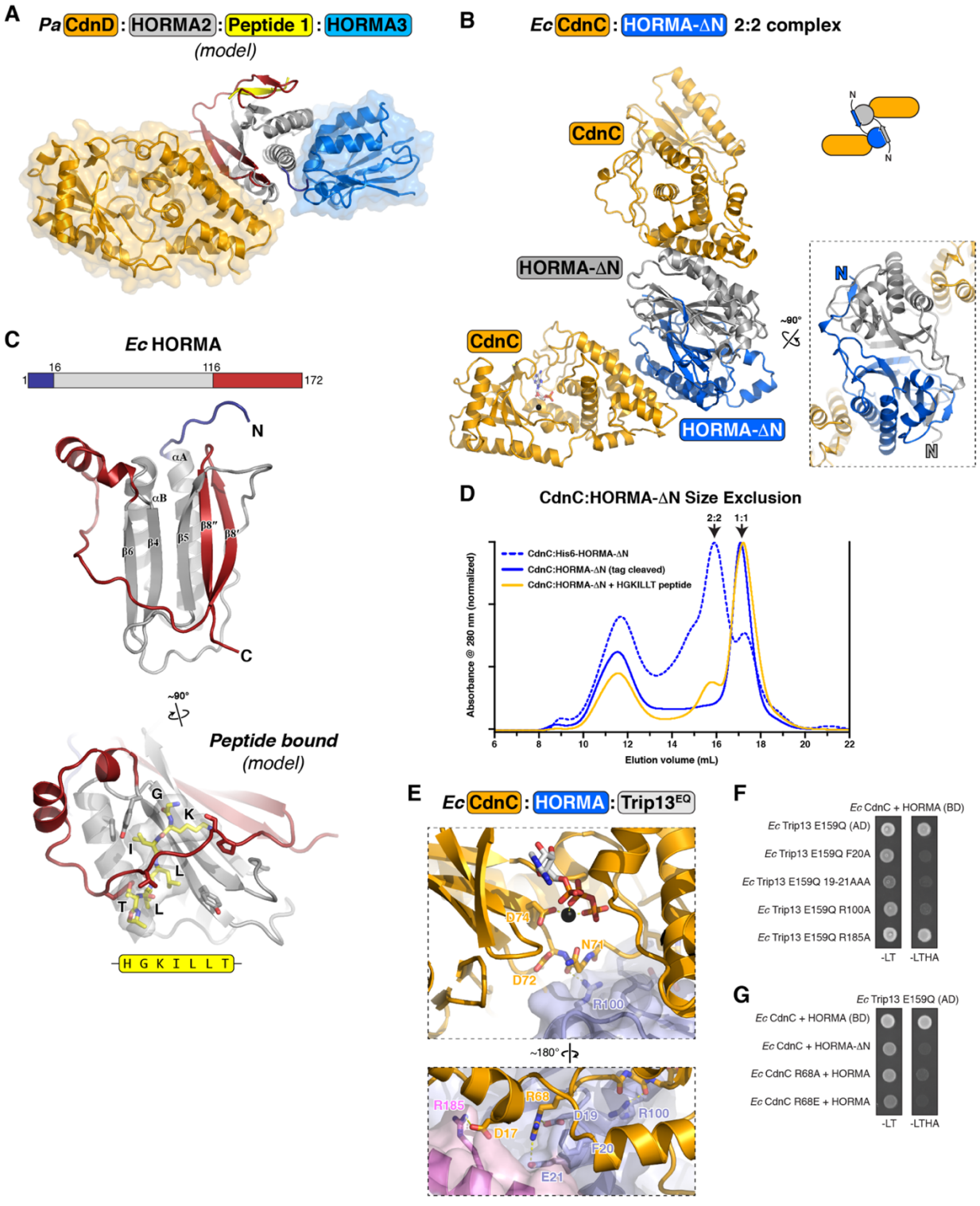

A
Pa CdnD : HORMA2 : Peptide 1 : HORMA3
(model)
B
Ec CdnC : HORMA-ΔN 2:2 complex
CdnC
HORMA-ΔN
CdnC
HORMA-ΔN
~90°
N
N
N
C
Ec HORMA
16
116
1
172
N
αA
αB
β6
β4
β5
β8″
β8′
C
~90°
Peptide bound
(model)
G
K
I
L
T
L
H G K I L L T
D
CdnC:HORMA-ΔN Size Exclusion
2:2
1:1
CdnC:His6-HORMA-ΔN
CdnC:HORMA-ΔN (tag cleaved)
CdnC:HORMA-ΔN + HGKILLT peptide
Absorbance @ 280 nm (normalized)
6
8
10
12
14
16
18
20
22
Elution volume (mL)
E
Ec CdnC : HORMA : Trip13EQ
D74
N71
D72
R100
~180°
R185
R68
D17
D19
R100
F20
E21
F
Ec CdnC + HORMA (BD)
Ec Trip13 E159Q (AD)
Ec Trip13 E159Q F20A
Ec Trip13 E159Q 19-21AAA
Ec Trip13 E159Q R100A
Ec Trip13 E159Q R185A
-LT
-LTHA
G
Ec Trip13 E159Q (AD)
Ec CdnC + HORMA (BD)
Ec CdnC + HORMA-ΔN
Ec CdnC R68A + HORMA
Ec CdnC R68E + HORMA
-LT
-LTHA

Figure A.A.S30 Structure of *Rhizobiales* Trip13

(A) Schematic of the *Rhizobiales* CBASS system, including HORMA (IMG #2619783695), Trip13 (IMG #2619783694, sequence in **Table A.A.S4**), and NucC (IMG #2619783693), compared to the *E. coli* MS115-1 operon. The sequenced contig at IMG (JGI Integrated Microbial Genomes database) ends at the putative HORMA protein, precluding identification of the inferred CD-NTase. **(B)** Overall structure of the *Rhizobiales* (*Rh*) Trip13 monomer, with N-terminal "large AAA" domain green and C-terminal "small AAA" domain red. Yellow dotted lines indicate the positions of pore loops 1 and 2, both disordered in this structure. **(C)** Closeup view of the *Rh* Trip13 active site, in the same orientation as in panel B. A SO_4^- ion is bound in the active site, surrounded by canonical Walker ATPase active site residues. Bacterial Trip13 proteins from 1-HORMA operons possess a conserved double Arg finger (R215/R216; from neighboring subunit, colored white); see **Figure A.A.S8F** for sequences. **(D)** Top and side views of the six-fold symmetric *Rh* Trip13 hexamer, formed from crystallographic packing interactions (*P*6 symmetry axis noted by black hexagon). **(C)** Sequence logos showing divergence of pore loop sequences in Trip13 proteins from one-HORMA and two-HORMA operons. **(D)** Sequence logos showing the N-termini of HORMA proteins from one-HORMA operons (top), and from the two families of HORMA proteins from two-HORMA operons (middle/bottom). The first residue of helix αA (based on crystal structures of *Ec* HORMA, *Pa* HORMA2, and *Pa* HORMA3) are indicated by blue arrows. Arrows above the sequence of *Ec* HORMA (top) indicate serine/threonine residues directly engaged by *Ec* Trip13 pore loop 2 residue His173: black arrows indicate residues engaged by pore loops in our structure, and gray arrows indicate continuation of the conserved pattern into helix αA.

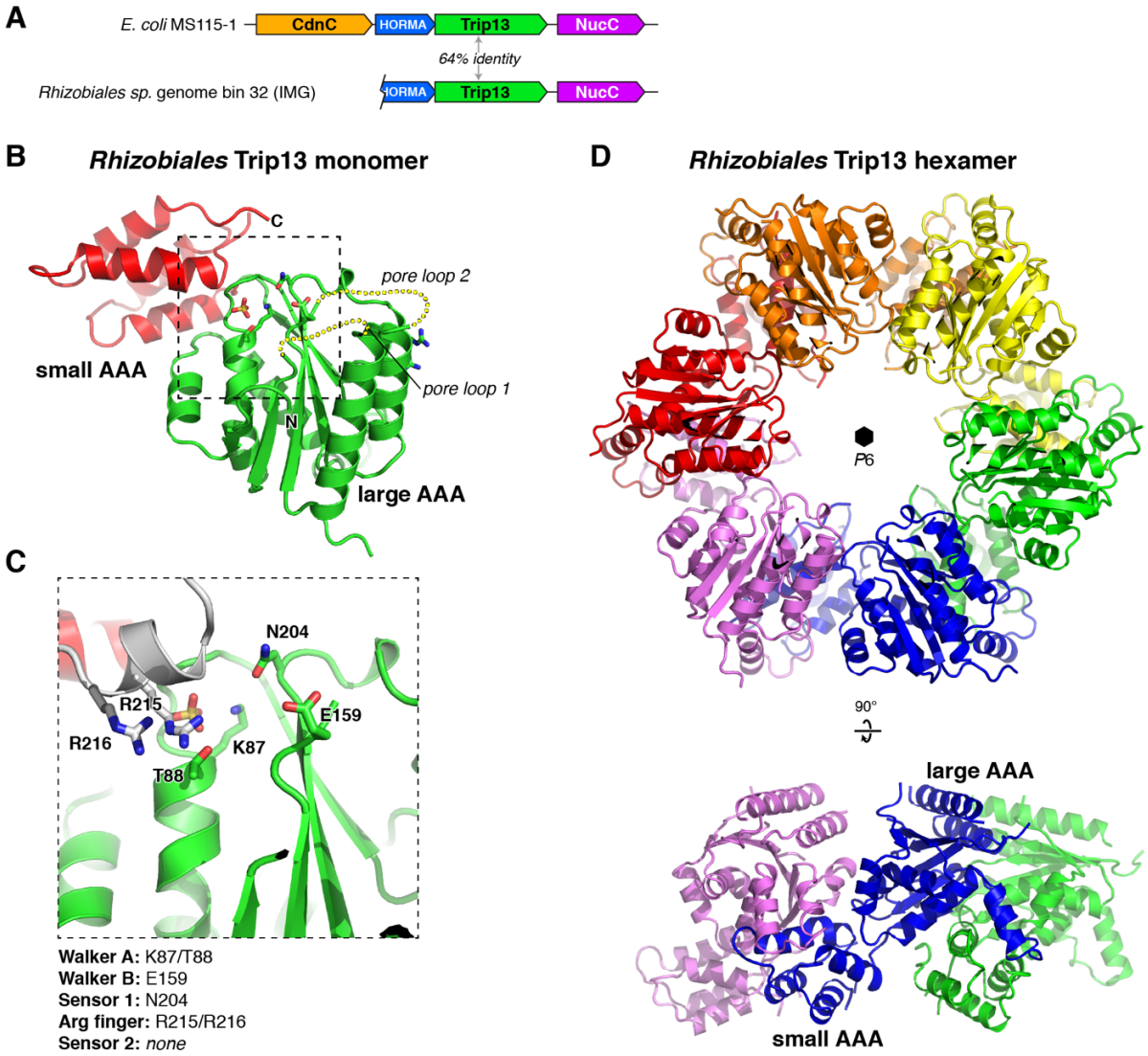
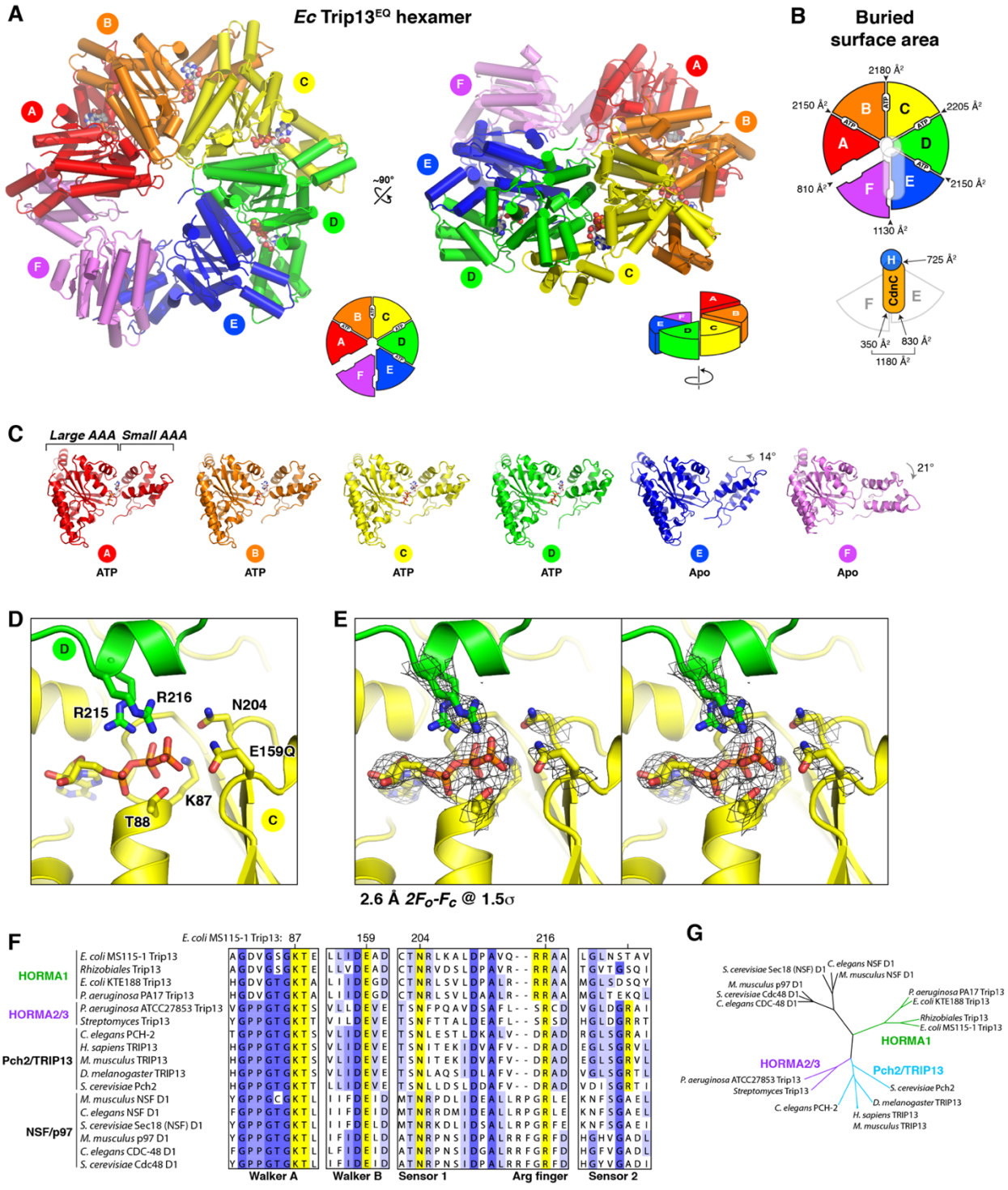

A
E. coli MS115-1
CdnC
HORMA
Trip13
NucC
64% identity
Rhizobiales sp. genome bin 32 (IMG)
HORMA
Trip13
NucC
B
Rhizobiales Trip13 monomer
C
pore loop 2
small AAA
pore loop 1
N
large AAA
C
N204
R215
E159
R216
K87
T88
Walker A: K87/T88
Walker B: E159
Sensor 1: N204
Arg finger: R215/R216
Sensor 2: none
D
Rhizobiales Trip13 hexamer
P6
90°
large AAA
small AAA
E
Bacterial Trip13 pore loop sequences
Pore loop 1
Pore loop 2
HORMA1 operons
(e.g. Ec Trip13)
Ec Trip13 residue #s:
HORMA2/3 operons
(e.g. Pa Trip13)
Pa Trip13 residue #s:
F
Bacterial HORMA N-termini
HORMA1
Ec HORMA1 residue #s:
helix αA
HORMA2
Pa HORMA2 residue #s:
helix αA
HORMA3
Pa HORMA3 residue #s:
strand β1
helix αA

Figure A.A.S31 Structure of the *Ec* CdnC:HORMA:Trip13^{EQ} complex

(A) Top and side views of the *Ec* Trip13^{EQ} hexamer in the structure of CdnC:HORMA:Trip13^{EQ}, with subunits A-F colored red, orange, yellow, green, blue, and violet, respectively. Bound ATP molecules at the A-B, B-C, C-D, and D-E interfaces are shown as spheres. **(B)** Schematics of the *Ec* CdnC:HORMA:Trip13^{EQ} complex, showing buried surface area in each subunit interface: inter-Trip13 interfaces top; Trip13-CdnC and CdnC-HORMA interfaces bottom. **(C)** Views of each individual subunit in the *Ec* Trip13^{EQ} hexamer, aligned on the large AAA domain, showing the conformational similarity of subunits A-D, and rotation of the small AAA domain in subunits E and F. **(D)** Closeup of ATP binding to *Ec* Trip13^{EQ} (subunit C), with conserved ATP-interacting residues from chain C (Walker A, Walker B, Sensor 1) in yellow and from chain D (double Arg finger) in green. **(E)** Stereo view of refined $2F_o$-F_c electron density, contoured at 1.5 σ, surrounding ATP and ATP-interacting residues in Trip13^{EQ} chain C. **(F)** Conserved AAA+ sequence motifs in bacterial Trip13, eukaryotic Pch2/TRIP13, and eukaryotic NSF and p97 proteins (D1 ATPase domain). While Trip13 proteins from HORMA2/3 operons (e.g. *P. aeruginosa* ATCC27853 Trip13) share all motifs with eukaryotic Pch2/TRIP13 proteins, Trip13 proteins from HORMA1 operons (e.g. *E. coli* MS115-1 Trip13) possess a double Arg finger (*Ec* Trip13 residues 215 and 216) and lack a sensor 2 arginine. **(G)** Phylogenetic tree showing that bacterial Trip13 proteins from HORMA2/3 operons are more closely related to eukaryotic Pch2/TRIP13 than those from HORMA1 operons.


A
Ec Trip13EQ hexamer
~90°
B
Buried surface area
2180 Å²
2150 Å²
2205 Å²
810 Å²
2150 Å²
1130 Å²
725 Å²
350 Å²
830 Å²
1180 Å²
CdnC
C
Large AAA Small AAA
ATP
ATP
ATP
ATP
Apo
Apo
14°
21°
D
R215
R216
N204
E159Q
K87
T88
E
2.6 Å 2Fo-Fc @ 1.5σ
F
E. coli MS115-1 Trip13: 87 159 204 216
HORMA1
HORMA2/3
Pch2/TRIP13
NSF/p97
E. coli MS115-1 Trip13
Rhizobiales Trip13
E. coli KTE188 Trip13
P. aeruginosa PA17 Trip13
P. aeruginosa ATCC27853 Trip13
Streptomyces Trip13
C. elegans PCH-2
H. sapiens TRIP13
M. musculus TRIP13
D. melanogaster TRIP13
S. cerevisiae Pch2
M. musculus NSF D1
C. elegans NSF D1
S. cerevisiae Sec18 (NSF) D1
M. musculus p97 D1
C. elegans CDC-48 D1
S. cerevisiae Cdc48 D1
Walker A
Walker B
Sensor 1
Arg finger
Sensor 2
G
S. cerevisiae Sec18 (NSF) D1
C. elegans NSF D1
M. musculus NSF D1
M. musculus p97 D1
S. cerevisiae Cdc48 D1
C. elegans CDC-48 D1
P. aeruginosa PA17 Trip13
E. coli KTE188 Trip13
Rhizobiales Trip13
E. coli MS115-1 Trip13
HORMA1
HORMA2/3
Pch2/TRIP13
P. aeruginosa ATCC27853 Trip13
Streptomyces Trip13
S. cerevisiae Pch2
C. elegans PCH-2
D. melanogaster TRIP13
H. sapiens TRIP13
M. musculus TRIP13

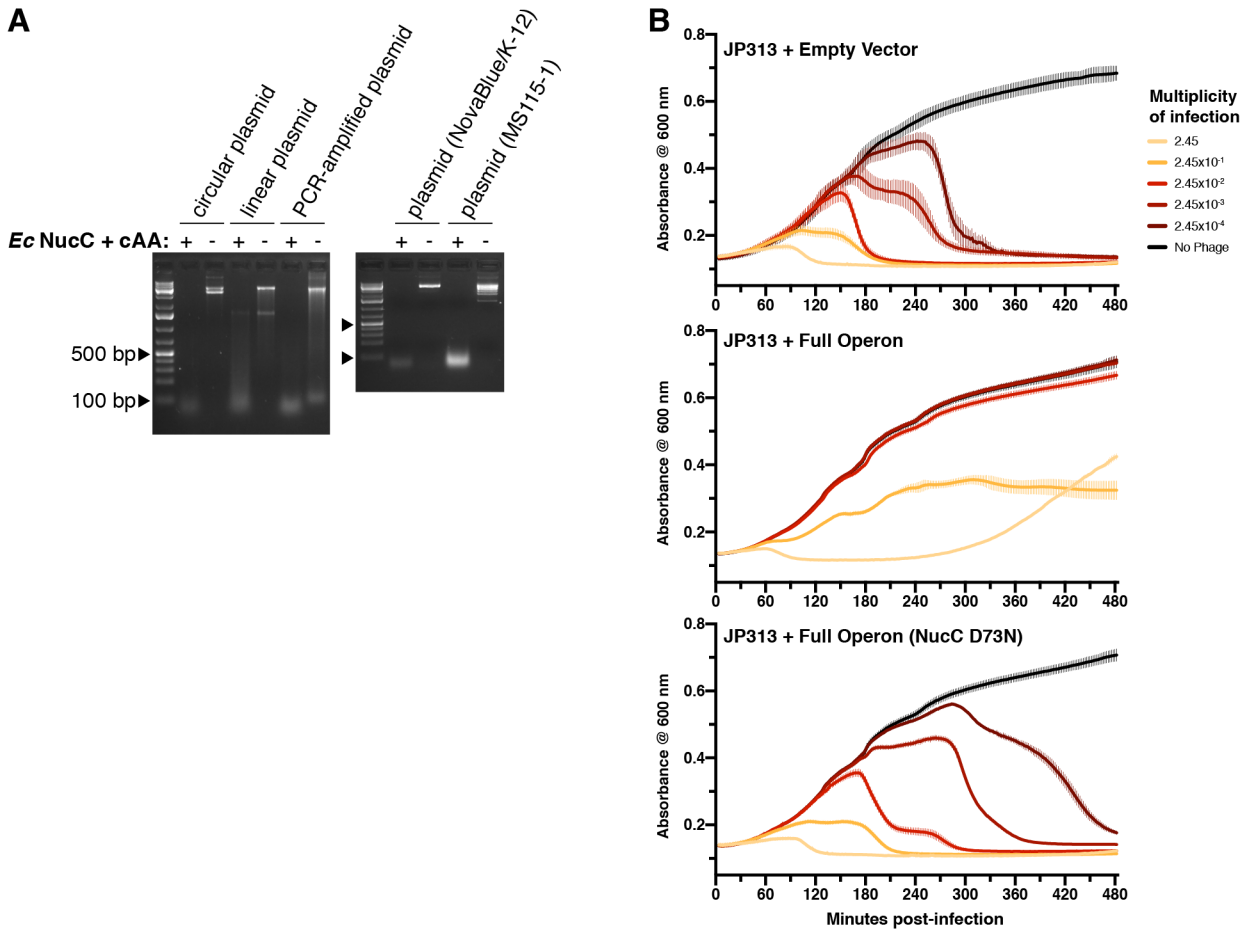


Figure A.A.S32 NucC activity and bacteriophage immunity

(A) Plasmid digestion assay with 10 nM *Ec* NucC, 1 μM cAA, and 1 μg of the indicated DNAs. **(B)** Full growth curves of *E. coli* JP313 + pLAC22 (empty vector; top), pLAC22 with the *E. coli* MS115-1 CBASS system including CdnC-HORMA-Trip13-NucC (full operon; middle), or pLAC22 with the *E. coli* MS115-1 CBASS system containing a NucC D73N mutant (bottom). Cells were infected with bacteriophage λ*cI*- at time 0, with multiplicity of infection (MOI) varying between 2.45 and 2.45 x 10^{-4}. All curves represent the mean of four replicate samples, with error bars showing standard deviation.

Methods A.A.S1 (Related to STAR Methods). Sequence of *E. coli* MS115-1 CBASS system cloned into plasmid pLAC22 for plaque assays.

/ATGTCAACAGAACATGTCGACCATAAAACTATAGCGCGATTTGCCGAAGATAAGGT
AAATCTTCCAAAAGTAAAGGCTGATGATTTCAGGGAACAGGCCAAGCGATTACAGA
ACAAACTGGAAGGGTATCTTTCTGATCATCCTGACTTTTCATTAAAGCGAATGATTC
CATCAGGTAGTCTGGCTAAAGGAACTGCTCTTCGCTCGTTAAACGATATTGATGTG
GCTGTGTATATCAGCGGCTCTGATGCACCACAGGATTTACGTGGGTTGCTTGACTA
TCTTGCTGATAGATTGCGTAAAGCATTTCCTAATTTTAGCCCTGATCAGGTTAAACC
CCAGACATACTCTGTAACGGTTTCCTTCCGGGGCTCTGGCTTAGATGTCGATATTG
TCCCTGTATTGTATTCGGGGTTACCTGACTGGCGAGGTCATTTGATAAGCCAGGAA
GATGGCTCGTTCCTTGAAACCAGTATTCCTCTGCACCTTGATTTCATCAAAGCCCG
CAAGCGTGCTGCCCCGAAGCATTTTGCTCAGGTTGTTCGTTTAGCGAAATATTGGG
CTCGTTTGATGAAGCAAGAGCGACCGAATTTTCGCTTTAAATCGTTCATGATTGAAT
TGATTCTTGCAAAATTACTGGATAATGGTGTGGATTTCTCGAATTATCCGGAAGCTT
TACAGGCATTTTTTTCCTATCTGGTGAGCACCGAATTACGTGAACGCATTGTCTTCG
AAGATAATTATCCTGCGTCAAAAATAGGCACGTTGTCAGACTTAGTGCAGATCATC
GATCCCGTTAATCCTGTTAATAATGTTGCTCGTTTATATACGCAGTCTAATGTGGAC
GCCATCATTGACGCCGCAATGGATGCCGGTGACGCTATTGATGCTGCATTCTATGC
ACCAACCAAGCAATTAACCGTAACCTATTGGCAGAAAGTTTTCGGTTCTTCATTCCA
GGGGTGA//AATCATT/ATGTCTTCTTATAGTTATACGGTAGCAGAGACGCAAACTTT
CAGCGTAACCCACGCTCGTCACATGGCTGCTAAAGTTGCAACTGACTTGCGGCGC
ATGCAACGTTTTTATGGTTACCCCAGTGATGCCGACATTGAAGCATACGAAGAAGA
ATTGGTTGTGTTTCTTAAGGCTGGATATTTGGGTGAGGTCTCTTATGGTTTTCAGAA
AAATAATAACTGGATCGAGCCGACCCTTCGATATACCGCAGGCGACTTGCTTGGCT
CAGGAACAGATGACGATCCCGGAAAAATCCGCCCAGGAAAAGATGTATCGGGTGC
ATCCTTCTACAGTTTTATGACTTATAGCTCGAAATATCTGAATGCTACTCAATCGGA
AAAAGATACTGCTTTGAAAGATCTACCATTCAAACGGGTTGGTGCCCAGTCTCCAG
GGATTAATGGCTACCTCGAAAATGATAAGACTTACTCGGCCGGTGGCCGCTCCCT
CACTCGCACTAGCGTAAGGAATTTTGT/ATGA//ACGTAAAACCATCACTTGATGAAT
TGTTTGAGCGGCGTATTAACTTCCCTGACTTTGAGCCTCAGGAGCGTCTGGCTCGT
TTAGTTGGTCTTGATGAGCACAAAGATCGATTATCAAAAATATTAGGCTTGCTTGTA
AATCCTTATGGGATTCAGGAATGGGCCAAAAAATATCATCCAGATGCTCGTGCGGC
TGTTGATACAGTACTCAGACGGCCACCTTTAGTTGTATTGGCTGGGGACGTGGGCT
CAGGTAAAACAGAGCTGGCGGAAACCATTGGTGATGCTGTAGCCCGGCAGGAAGA
TATTGATATTACGTTGTATCCTTTGAGCCTTGCTACCCGAGGGCAAGGTAGAGTTG
GCGAAATGACTCAGCTGGTTTCAGCTGCATTTGATTACACCATTGAGGCCGCAGAT
AAATTAAAAAACACGAATGGAAAAGCTCGTGGTGCGGTTTTGCTGCTGATTGATGA
AGCTGATGCATTGGCTCAGTCCCGGGAAAATGCCCAGATGCACCATGAGGATCGT
GCTGGTGTAAACGCTTTTATTCGCGGAATAGACCGGATCGCCAATCAGAAACTACC
AGCGGCAGTTCTTATGTGTACTAACCGGTTGAAAGCTCTCGATCCAGCTGTCCAGC
GTCGTGCTGCTGAAATTTTAACATTTTCCCGGCCAAATGATGAACAACGGCACTAT
TTGCTGCATTCAAAACTTACGGGATTGGGGTTAAATTCCACGGCGGTAGAAGAATT
AGTCAGGTTAACCGGTCCAAGGGACCCTAATTCTCCTGGATTTACTTTCTCAGATAT

TACGCAGCGGCTAATCCCATCAATAATTCTTGCGGCCTACCCATATAATGCAGTGT
CAGTCCATTCTGCGCTACAGGTAGTCAATAAAATGACACCAACACCCGCATTTATT
GACCGCCAGTAA//TATAGAACAGATAAAAGCCCATAAATAGTATTATGGGCCATGC
ATGCTTATGCAGCGTATAAATATTAATGAATAAAGGAATCCATT/ATGTCAGATTGGT
CGCTTTCTCAACTATTTGCATCGTTACATGAAGACATTCAGCTTCGACTGGGAACTG
CACGTAAAGCTTTCCAGCATCCCGGGGCTAAAGGGGATGCCAGTGAAGGAGTATG
GATTGAAATGCTTGATACATATCTTCCCAAACGCTATCAGGCAGCTAATGCTTTTGT
AGTTGATAGTTTGGGTAATTTTAGTGATCAGATTGATGTGGTTGTTTTTGACAGGCA
GTACTCACCATTCATTTTCAAATTTAATGAACAAATAATTGTTCCAGCAGAAAGTGT
CTATGCGGTTTTTGAAGCAAAGCAATCTGCGTCAGCAGATCTTGTTGCTTATGCAC
AGCGAAAAGTTGCAAGTGTTCGCAGACTTCATCGTACCAGCCTTCCAATTCCGCAT
GCTGGCGGGACATATCCAGCAAAACCACTCATTCCAATTTTGGGAGGGTTACTTAC
TTTTGAAAGTGACTGGAGTCCTGCATTGGGGATGTCGTTTGACAAAGCCCTCAATG
GGGATTTATCTGATGGACGACTTGATATGGGCTGTGTTGCATCGCATGGGCATTTT
TACTTCAATAACATTGATAGCAAGTTCAATTTTGAACATGGAAATAAACCAGCTACC
GCTTTTCTTTTCAGGCTTATTGCGCAACTTCAATTTAGTGGCACTGTGCCAATGATT
GACATTGATGCCTATGGAAAATGGTTGGCTAATTGA// [1]

[1]Annotations: / = ORF start; // = ORF end

Table A.A.S1 Representative bacterial HORMA1, HORMA2, and HORMA3 proteins

HORMA family	**GenBank Accession #**
HORMA1	KDS98063, WP_007179265.1, WP_011505070.1, WP_023806712.1, WP_018269866.1, WP_034464884.1, WP_037129447.1, EOY49422.1, WP_034266568.1, WP_030068533.1, WP_025144301.1, WP_043023585.1, WP_013687778.1, WP_023840430.1, WP_034736833.1, EWF90982.1, WP_020753577.1, WP_012367383.1, WP_033184441.1, WP_025760314.1, AAY49482.1, WP_044244031.1, WP_010188594.1, WP_000122586.1, WP_045114623.1, WP_020890796.1, WP_025663189.1, WP_021763012.1, WP_012071859.1, WP_023223656.1, WP_023098968.1, WP_001534690.1, AHH51003.1, WP_029855951.1, WP_004181946.1, WP_032732104.1, WP_017463403.1, WP_040602719.1, WP_028351170.1, WP_023729002.1, WP_012563549.1, WP_036046164.1, WP_037093525.1, WP_007011043.1, WP_003581386.1
HORMA2	WP_003090159, WP_005987909.1, WP_004333156.1, WP_009797790.1, WP_023699054.1, WP_026769537.1, WP_030403523.1, WP_023842206.1, WP_034211670.1, WP_003050262.1, WP_011804673.1, WP_043946290.1, WP_011486010.1, WP_037275566.1, WP_009338224.1, WP_023820821.1, WP_020048019.1, WP_003508389.1, WP_010583107.1, WP_020628060.1, WP_037146310.1
HORMA3	WP_003090160, WP_005987907.1, WP_004333158.1, WP_009797789.1, WP_031355782.1, WP_026769536.1, WP_030403522.1, WP_023842207.1, WP_034211668.1, WP_023087302.1, WP_011804672.1, WP_043946289.1, WP_011486009.1, WP_037275564.1, WP_009338223.1, WP_023820820.1, WP_020048018.1, WP_003508388.1, WP_010583106.1, WP_043282487.1, WP_052182635.1

Table A.A.S2 Representative bacterial TRIP13 protein

HORMA family	**GenBank Accession #**
HORMA1	IMG 2619783694 , EFJ98158.1, WP_042596015.1, WP_009734074.1, WP_018269867.1, WP_011505069.1, WP_034464887.1, WP_020043204.1, WP_020753578.1, WP_012367384.1, WP_024540133.1, WP_025586231.1, WP_033184386.1, EWF90983.1, WP_032742913.1, KDS98062.1, WP_036770686.1, WP_038992386.1, WP_017355006.1, WP_009807868.1, WP_007179266.1, WP_041190635.1, WP_025663190.1, WP_004777375.1, ENX61411.1, WP_023806713.1, WP_035095387.1, WP_003941252.1, WP_039798001.1, EFD67588.1, EOY49423.1, WP_003974510.1, WP_030068534.1, WP_028443308.1, WP_026454397.1, WP_011434599.1, WP_039922222.1, WP_000694216.1, WP_013687779.1, WP_031453383.1, WP_002981077.1, WP_003002482.1, WP_020890797.1, WP_012929335.1, WP_028692343.1, WP_042122382.1, WP_005764367.1, ACP17986.1, WP_007987530.1, WP_003139489.1, AHH51004.1, WP_038632721.1, WP_012071860.1, WP_000047137.1, WP_001534689.1, WP_029096108.1, AAY49483.1, WP_017463404.1, WP_011657208.1, WP_010396300.1, WP_004181947.1, WP_043123885.1, WP_042058390.1, WP_028758490.1, WP_037093526.1, WP_021244675.1, WP_028351171.1, WP_012563548.1, WP_036046183.1, WP_007011042.1, WP_010236966.1, WP_020084793.1, WP_023729003.1, WP_097572462.1, WP_003581388.1, WP_010188595.1, WP_022673617.1, WP_023223655.1, WP_008956655.1, WP_029855952.1, WP_032732106.1, WP_010454845.1, WP_038350075.1, WP_042114565.1, EDT42154.1, WP_044244032.1, WP_015056962.1, WP_002702939.1, WP_025144302.1, WP_013685199.1
HORMA2/3	WP_016852809.1, WP_005987905.1, WP_009797788.1, WP_008125046.1, WP_009338222.1, WP_018427718.1, WP_023699052.1, WP_023734942.1, WP_027533735.1, WP_023820819.1, WP_080646630.1, WP_003508386.1, WP_043156229.1, WP_021205716.1, WP_003050256.1, WP_023087301.1, WP_003105629.1, WP_043946288.1, WP_018229299.1, WP_017454944.1, WP_034211666.1, WP_029065345.1, WP_010141873.1, WP_037275562.1, WP_081492585.1, WP_004333160.1, WP_021248589.1, WP_026437714.1, WP_018988636.1, WP_011486008.1, WP_037565884.1, WP_010583105.1, WP_037360968.1, WP_018891425.1, WP_030403521.1, WP_023840428.1, WP_035910099.1, WP_020659317.1, WP_020628058.1, WP_023842208.1, WP_035734229.1, WP_037146424.1, WP_033530995.1

Table A.A.S3 Crystallographic Data

	***Ec* CdnC**	***Ec* CdnC SeMet**	***Pa* CdnD Apo SeMet**	***Pa* CdnD D62N/D64N+ATP**
Data collection				
Synchrotron/Beamline	APS 24ID-E	APS 24ID-E	APS 24ID-C	APS 24ID-E
Date collected	4/19/17	4/19/17	8/23/17	10/28/17
Resolution (Å)	20 – 1.44	20 – 1.92	100 – 2.05	100 - 1.47
Wavelength (Å)	0.97918	0.97918	0.97918	0.9793
Space Group	$P3_221$	$P3_221$	$P22_12_1$	$P2_1$
Unit Cell Dimensions (a, b, c) Å	76.18, 76.18, 100.50	76.22, 76.22, 100.21	104.59, 117.39, 147.80	46.48, 152.17, 88.25
Unit cell Angles (α,β,γ) °	90, 90, 120	90, 90, 120	90, 90, 90	90, 90.123, 90
I/s (last shell)	15.0 (0.6)	10.7 (0.7)	10.5 (0.7)	9.7 (0.97)
[a] R_{sym} (last shell)	0.029 (1.429)	0.074 (1.872)	0.14 (2.03)	0.088 (0.948)
[b] R_{meas} (last shell)	0.034 (1.776)	0.082 (2.140)	0.15 (2.23)	0.104 (1.143)
[c] $CC_{1/2}$, last shell	1 (0.263)	0.998 (0.266)	0.997 (0.336)	0.996 (0.356)
Completeness (last shell) %	98.6 (96.8)	98.1 (74.3)	99.5 (89.9)	97.4 (92.7)
[d] Number of reflections	179378	138287	753705	682978
unique	60488	25989	114327	201994
Multiplicity (last shell)	3.0 (2.7)	5.3 (3.9)	6.6 (5.3)	3.4 (3.1)
Refinement				
Resolution (Å)	20 – 1.5		100 - 2.05	88 - 1.47
[d] No. of reflections	53578		217384 (unmerged)	201964
working	50882		206483 (unmerged)	191910
free	2696		10901 (unmerged)	10054
[e] R_{work} (last shell) %	15.39 (29.38)		21.27 (36.78)	19.51 (33.34)
[e] R_{free} (last shell) %	18.89 (33.06)		24.26 (41.44)	22.87 (37.33)
Structure/Stereochemistry				
Number of atoms	5435		10437	20633
solvent	294		619	1444
ligands	39		170 (SO_4^-)	172
ions	2		0	8
hydrogen	2557		0	9505

Table A.A.S3 Crystallographic Data

r.m.s.d. bond lengths (Å)	0.006		0.007	0.009
r.m.s.d. bond angles (°)	0.974		0.825	1.116
[f] PDB ID	6P80	-	6P82	6P8J
[g] SBGrid Data Bank ID	668	669	670	671

	***Ec* CdnC**	***Ec* CdnC SeMet**	***Pa* CdnD Apo SeMet**	***Pa* CdnD D62N/D64N+ATP**
Data collection				
Synchrotron/Beamline	APS 24ID-E	APS 24ID-E	APS 24ID-C	APS 24ID-E
Date collected	4/19/17	4/19/17	8/23/17	10/28/17
Resolution (Å)	20 – 1.44	20 – 1.92	100 – 2.05	100 - 1.47
Wavelength (Å)	0.97918	0.97918	0.97918	0.9793
Space Group	$P3_221$	$P3_221$	$P22_12_1$	$P2_1$
Unit Cell Dimensions (a, b, c) Å	76.18, 76.18, 100.50	76.22, 76.22, 100.21	104.59, 117.39, 147.80	46.48, 152.17, 88.25
Unit cell Angles (α,β,γ) °	90, 90, 120	90, 90, 120	90, 90, 90	90, 90.123, 90
I/s (last shell)	15.0 (0.6)	10.7 (0.7)	10.5 (0.7)	9.7 (0.97)
[a] R_{sym} (last shell)	0.029 (1.429)	0.074 (1.872)	0.14 (2.03)	0.088 (0.948)
[b] R_{meas} (last shell)	0.034 (1.776)	0.082 (2.140)	0.15 (2.23)	0.104 (1.143)
[c] $CC_{1/2}$, last shell	1 (0.263)	0.998 (0.266)	0.997 (0.336)	0.996 (0.356)
Completeness (last shell) %	98.6 (96.8)	98.1 (74.3)	99.5 (89.9)	97.4 (92.7)
[d] Number of reflections	179378	138287	753705	682978
unique	60488	25989	114327	201994
Multiplicity (last shell)	3.0 (2.7)	5.3 (3.9)	6.6 (5.3)	3.4 (3.1)
Refinement				
Resolution (Å)	20 – 1.5		100 - 2.05	88 - 1.47
[d] No. of reflections	53578		217384 (unmerged)	201964
working	50882		206483 (unmerged)	191910
free	2696		10901 (unmerged)	10054
[e] R_{work} (last shell) %	15.39 (29.38)		21.27 (36.78)	19.51 (33.34)
[e] R_{free} (last shell) %	18.89 (33.06)		24.26 (41.44)	22.87 (37.33)

Table A.A.S3 Crystallographic Data

Structure/Stereochemistry				
Number of atoms	5435		10437	20633
solvent	294		619	1444
ligands	39		170 (SO_4^-)	172
ions	2		0	8
hydrogen	2557		0	9505
r.m.s.d. bond lengths (Å)	0.006		0.007	0.009
r.m.s.d. bond angles (°)	0.974		0.825	1.116
[f] PDB ID	6P80	-	6P82	6P8J
[g] SBGrid Data Bank ID	668	669	670	671

	***Ec* CdnC**	***Ec* CdnC SeMet**	***Pa* CdnD Apo SeMet**	***Pa* CdnD D62N/D64N+ATP**
Data collection				
Synchrotron/Beamline	APS 24ID-E	APS 24ID-E	APS 24ID-C	APS 24ID-E
Date collected	4/19/17	4/19/17	8/23/17	10/28/17
Resolution (Å)	20 – 1.44	20 – 1.92	100 – 2.05	100 - 1.47
Wavelength (Å)	0.97918	0.97918	0.97918	0.9793
Space Group	$P3_221$	$P3_221$	$P22_12_1$	$P2_1$
Unit Cell Dimensions (a, b, c) Å	76.18, 76.18, 100.50	76.22, 76.22, 100.21	104.59, 117.39, 147.80	46.48, 152.17, 88.25
Unit cell Angles (α,β,γ) °	90, 90, 120	90, 90, 120	90, 90, 90	90, 90.123, 90
I/s (last shell)	15.0 (0.6)	10.7 (0.7)	10.5 (0.7)	9.7 (0.97)
[a] R_{sym} (last shell)	0.029 (1.429)	0.074 (1.872)	0.14 (2.03)	0.088 (0.948)
[b] R_{meas} (last shell)	0.034 (1.776)	0.082 (2.140)	0.15 (2.23)	0.104 (1.143)
[c] $CC_{1/2}$, last shell	1 (0.263)	0.998 (0.266)	0.997 (0.336)	0.996 (0.356)
Completeness (last shell) %	98.6 (96.8)	98.1 (74.3)	99.5 (89.9)	97.4 (92.7)
[d] Number of reflections	179378	138287	753705	682978
unique	60488	25989	114327	201994
Multiplicity (last shell)	3.0 (2.7)	5.3 (3.9)	6.6 (5.3)	3.4 (3.1)
Refinement				
Resolution (Å)	20 – 1.5		100 - 2.05	88 - 1.47

Table A.A.S3 Crystallographic Data

[d] No. of reflections	53578		217384 (unmerged)	201964
working	50882		206483 (unmerged)	191910
free	2696		10901 (unmerged)	10054
[e] R_{work} (last shell) %	15.39 (29.38)		21.27 (36.78)	19.51 (33.34)
[e] R_{free} (last shell) %	18.89 (33.06)		24.26 (41.44)	22.87 (37.33)
Structure/Stereochemistry				
Number of atoms	5435		10437	20633
solvent	294		619	1444
ligands	39		170 (SO_4^-)	172
ions	2		0	8
hydrogen	2557		0	9505
r.m.s.d. bond lengths (Å)	0.006		0.007	0.009
r.m.s.d. bond angles (°)	0.974		0.825	1.116
[f] PDB ID	6P80	-	6P82	6P8J
[g] SBGrid Data Bank ID	668	669	670	671

	***Ec* CdnC**	***Ec* CdnC SeMet**
Data collection		
Synchrotron/Beamline	APS 24ID-E	APS 24ID-E
Date collected	4/19/17	4/19/17
Resolution (Å)	20 – 1.44	20 – 1.92
Wavelength (Å)	0.97918	0.97918
Space Group	$P3_221$	$P3_221$
Unit Cell Dimensions (a, b, c) Å	76.18, 76.18, 100.50	76.22, 76.22, 100.21
Unit cell Angles (α,β,γ) °	90, 90, 120	90, 90, 120
I/s (last shell)	15.0 (0.6)	10.7 (0.7)
[a] R_{sym} (last shell)	0.029 (1.429)	0.074 (1.872)
[b] R_{meas} (last shell)	0.034 (1.776)	0.082 (2.140)
[c] $CC_{1/2}$, last shell	1 (0.263)	0.998 (0.266)
Completeness (last shell) %	98.6 (96.8)	98.1 (74.3)

Table A.A.S3 Crystallographic Data

[d] Number of reflections	179378	138287
unique	60488	25989
Multiplicity (last shell)	3.0 (2.7)	5.3 (3.9)
Refinement		
Resolution (Å)	20 – 1.5	
[d] No. of reflections	53578	
working	50882	
free	2696	
[e] R_{work} (last shell) %	15.39 (29.38)	
[e] R_{free} (last shell) %	18.89 (33.06)	
Structure/Stereochemistry		
Number of atoms	5435	
solvent	294	
ligands	39	
ions	2	
hydrogen	2557	
r.m.s.d. bond lengths (Å)	0.006	
r.m.s.d. bond angles (°)	0.974	
[f] PDB ID	6P80	-
[g] SBGrid Data Bank ID	668	669

[a] R_{sym} = åå$_j$| I_j – áIñ|/åI_j, where I_j is the intensity measurement for reflection j and áIñ is the mean intensity for multiply recorded reflections.
[b] R_{meas} = å$_h$ [√(n/(n-1)) å$_j$ [I_{hj} - áI_hñ] / å$_{hj}$ áI_hñ, where I_{hj} is a single intensity measurement for reflection h, áI_hñ is the average intensity measurement for multiply recorded reflections, and n is the number of observations of reflection h.
[c] $CC_{1/2}$ is the Pearson correlation coefficient between the average measured intensities of two randomly-assigned half-sets of the measurements of each unique reflection.
[d] Reflection counts in scaling and refinement are unmerged (F(+) and F(-) treated as separate reflections).
[e] $R_{work, free}$ = å||F_{obs}| – |F_{calc}||/|F_{obs}|, where the working and free R-factors are calculated using the working and free reflection sets, respectively.
[f] Coordinates and structure factors have been deposited in the RCSB Protein Data Bank (www.pdb.org) with the noted accession numbers.

[g] Diffraction data have been deposited with the SBGrid Data Bank (https://data.sbgrid.org) with the noted accession numbers.

Table A.A.S4 Protein Sequences

Protein	NCBI Accession	Sequence
***E. coli* MS115-1**		
[1]CdnC	EFJ98156	MSSGLDRVKTSSEDEMSTEHVDHKTIARFAEDKVNLPKVKADDFR EQAKRLQNKLEGYLSDHPDFSLKRMIPSGSLAKGTALRSLNDIDVA VYISGSDAPQDLRGLLDYLADRLRKAFPNFSPDQVKPQTYSVTVS FRGSGLDVDIVPVLYSGLPDWRGHLISQEDGSFLETSIPLHLDFIKA RKRAAPKHFAQVVRLAKYWARLMKQERPNFRFKSFMIELILAKLLD NGVDFSNYPEALQAFFSYLVSTELRERIVFEDNYPASKIGTLSDLV QIIDPVNPVNNVARLYTQSNVDAIIDAAMDAGDAIDAAFYAPTKQLT VTYWQKVFGSSFQG
[2]HORMA	KDS98063	MSSYSYTVAETQTFSVTHARHMAAKVATDLRRMQRFYGYPSDAD IEAYEEELVVFLKAGYLGEVSYGFQKNNNWIEPTLRYTAGDLLGSG TDDDPGKIRPGKDVSGASFYSFMTYSSKYLNATQSEKDTALKDLP FKRVGAQSPGINGYLENDKTYSAGGRSLTRTSVRNFV
Trip13	EFJ98158	MNVKPSLDELFERRINFPDFEPQERLARLVGLDEHKDRLSKILGLL VNPYGIQEWAKKYHPDARAAVDTVLRRPPLVVLAGDVGSGKTELA ETIGDAVARQEDIDITLYPLSLATRGQGRVGEMTQLVSAAFDYTIEA ADKLKNTNGKARGAVLLLIDEADALAQSRENAQMHHEDRAGVNAF IRGIDRIANQKLPAAVLMCTNRLKALDPAVQRRAAEILTFSRPNDEQ RHYLLHSKLTGLGLNSTAVEELVRLTGPRDPNSPGFTFSDITQRLIP SIILAAYPYNAVSVHSALQVVNKMTPTPAFIDRQ
NucC	EFJ98159	MSDWSLSQLFASLHEDIQLRLGTARKAFQHPGAKGDASEGVWIE MLDTYLPKRYQAANAFVVDSLGNFSDQIDVVVFDRQYSPFIFKFNE QIIVPAESVYAVFEAKQSASADLVAYAQRKVASVRRLHRTSLPIPH AGGTYPAKPLIPILGGLLTFESDWSPALGMSFDKALNGDLSDGRLD MGCVASHGHFYFNNIDSKFNFEHGNKPATAFLFRLIAQLQFSGTV PMIDIDAYGKWLAN
***P. aeruginosa* ATCC27853**		
Antitoxin (helix-turn-helix protein)	WP_003050245	MTTATAQSKMTLPAAGEVKAAVQGQRALAAYLATQFETQHIQIFD DHKQAHQVELPTSALRLLVDILAELADGNAVKVVPVHAELTTQEAA DLLNVSRPHFVKLLEDGVLAFHRTGKHRRVRFADLMQYKEARERA SEQAMAELAQQSQELGMGYE
Toxin (PIN domain containing protein	WP_003153636	MRHSPFTAVYDACVLYPAPLRDFLMWLGLSGRFRARWSQAIHEE WKRNLLINRPDLTRVQVDRTSDLMDRAIPDGLVEGYEALVAGLTLP DPNDRHVLAAAIRCGASVIVTFNERDFPNDLLAPYGIESQHPDEFV DNLLDLDAAAVVSAAQRQRAQLKHPPIDVDRYLEILLRQGLVQTTK VLATYRTIL

Table A.A.S4 Protein Sequences

Trip13	WP_016852809	MTKNPSSDATLPKGIHRSWKLPDKSLGDLWDSIVMDEAIKKQLLSQAIVNFTVRPKVERTVLPLHGVILLVGPPGTGKTSLARGLAHRVAESFSSAKFRLLEVEPHTLTSSAMGKTQRAVADLFSQSIAESAAAGPTIVLLDEVETLAADRAKLSLEANPVDVHRATDAVLVQLDMLAERNPHLLFVATSNFPQAVDSAFLSRCDMVMEVPLPGKDACKQILVDCLNGLAKTFPGIGKLSSAHQFDVCAGECVGLDGRAIRKVVANALAADPQVAIDPNKLSVEHLRSAIRQAKQMRLQGGKQK
HORMA3	WP_003090160	MTTVVSRTFRSSPHRDALQTWDAIVELLTQGKDGTARSELRAVTGVAASLIADQAPKSAPIVATCDGPRTRIYCLFDEDAIDGDDANEEVLGFEPLKGDWGVSLPCPKEQLGWVQSALKKHSSRIIARDLSQGIATQAQADAGQALSLDLGGFLKS
HORMA2	WP_003090159	MSTVATYSYTHSVTYVTDNILKSLKDIILLSGLDPEHFADRWESNTRAIKTWLGTGDLRKVILEIYNPATDKLVTRWDIDIVYGWSDGDGSFWTDTEQLKYAIKKAGLLPSQAKYKLMLDTKPGRPDVEGWSKGSYRSTDGMVKQSLGSTVEHSGLAGQAGYWRQR
CdnD	WP_003090158	MLSIDEAFRKFKSRLELNEREQKNASQRQNEVRDYLQTKFGIARSFLTGSYARYTKTKPLKDIDIFFVLKDSEKHYHGKAASVVLDDFHSALVEKYGSAAVRKQARSINVDFGVHIDAEDNTDYRVVSVDAVPAFDTGDQYEIPDTASGKWIKTDPEIHKDKATAAHQAYANEWKGLVRMVKYWNNNPKHGDLKPVKPSFLIEVMALECLYGGWGGSFDREIQSFFATLADRVHDEWPDPAGLGPAISNDMDAARKQRAQQLLFQASQDASIAIDHARRGRNIEALRAWRALFGPKFPLS
NucC	WP_003050273	MSQWSLSQLLSSLHEDIQQRLSVVRKTFGHPGTKGDASENVWIDMLDTYLPKRYQAAKAHVVDSLGNFSQQIDVVVFDRQYSPFIFTYENETIIPAESVYAVFEAKQTADAGLVAYAQEKVASVRRLHRTSLPIPHAGGTYPAKPLIPILGGLLTFESEWSPALGPSMDKALNANLTEGRLDIGCVAAHGHFFYDQASGAYSYTNENKPATAFLFKLIAQLQFSGTVPMIDVEAYGQWLTK
SMC	WP_003050276	MASIRSEYHRFLAHLAQRHVHDDVRRLAHLVLDHLQPLAEVGAARRGRSTRLAPLAIAHLAQMPVAYDGDARGPENGPALGRLHQLEVGPFRGFMRQETFDLSHDITLVYGANGTGKSSFCEALEVAMLGSISEAQAKRVDQRTYCNNARLRRHVAPVLSSTAAGEAQAVQPDEAEYRFCFIEKNRLDDFARIAARTPSDQRQLIATLFGVDQFSEFVRGFNPSLDQDLMLAGVQAAQLAQRRLRLASSEQTIAAYPQKIAAVEGLEQALAQRMSPGATYQACVDWLLGTPQQQGRLPYVQAQLDANPPAIHEVTQARLQALLAEAYRVQGLWQASSAQLAARAGEVSYAKLYEAVQALADGATACPACGTGLAAVAQDPFARARMGLEQLAQLAVLQQQEAGHRTQLSEAVRALWDEMRRVVAAAGVACPAESQAAGLPLLPPTSAGNWLGGWVNGDQRAWQALLRIAQIIEGFDAQARDVHAQRGAMAQERDRLQQHQLEIERLRTMRTTADQELAAARQTVAQFDDANRGLIQAATDEMPVVVHHQRVKAAYDGFLPEIQAYLTALPGVLLQGLGDQARHLYNAFNRADPPGDLLHALWLPVAENGKIEVEFAGEPGVRYDALIVFSEGHIKCLGLAILLAKNLAQGCPVVIFDDVVNAIDDDHRDGIWRTFFEDGLLHGKQVILTSHAEEFLHRIQQELGVRRAAAIKRYKFLPHQGEHELRVDSDPPAKNYVLLAQQALAADEKREALRQARPALESLTDRLWTWLGRRADGRIDIKLSGPRAPWELNNKCTKLRSAVERIAAQHAGAPDAVGALVRLLNVSGTSIEWGYLNSGVHDAQRDHEFDRATVRTVVEAVTALDAALDTLQNR

Table A.A.S4 Protein Sequences

Rhizobiales Trip13	2619783694 (see footnote 2)	MSMTPDLAGLFEGASTYPDVDARERLNNLVGLDTHKSRLSKMLAV LVNPDGLSAWAKKHHPAAEALVKNVIRRPPLIVLAGDVGSGKTELA ETIGDDVARRESIRITLLPLSLSARGQGRVGEMTQLISAAFEHTVSE ARKLKSSGTKARGAVILLVDEADALAQSREAAQMHHEDRAGVNAF IRGIDRLGNGALPAAVIMCTNRVDSLDPAVRRRAAEIITFDRPNDAQ RRAVITTTLQGTGVTGSQIEGLVAATGPARPRDYGFTFSDLTQRLI PSIVLDAYPDTSINPARALAIAQAMAPTAPFQDKSR

[1]Sequence underlined is annotated, but not part of the protein

[2]Accession number is from the Joint Genome Institute (JGI) Integrated Microbial Genomes & Microbiomes database (https://img.jgi.doe.gov)

References

Burroughs, A.M., Zhang, D., Schäffer, D.E., Iyer, L.M., Aravind, L., 2015. Comparative genomic analyses reveal a vast, novel network of nucleotide-centric systems in biological conflicts, immunity and signaling. Nucleic acids research 43, gkv1267--10654. https://doi.org/10.1093/nar/gkv1267

Gao, P., Ascano, M., Wu, Y., Barchet, W., Gaffney, B.L., Zillinger, T., Serganov, A.A., Liu, Y., Jones, R.A., Hartmann, G., Tuschl, T., Patel, D.J., 2013. Cyclic [G(2',5')pA(3',5')p] is the metazoan second messenger produced by DNA-activated cyclic GMP-AMP synthase. Cell 153, 1094–1107. https://doi.org/10.1016/j.cell.2013.04.046

Tromer, E.C., van Hooff, J.J.E., Kops, G.J.P.L., Snel, B., 2019. Mosaic origin of the eukaryotic kinetochore. Proceedings of the National Academy of Sciences of the United States of America 116, 12873–12882. https://doi.org/10.1073/pnas.1821945116

APPENDIX B: Supplemental Figures and Tables for Chapter 2 Structure and Mechanism of a Cyclic Trinucleotide-Activated Bacterial Endonuclease Mediating Bacteriophage Immunity

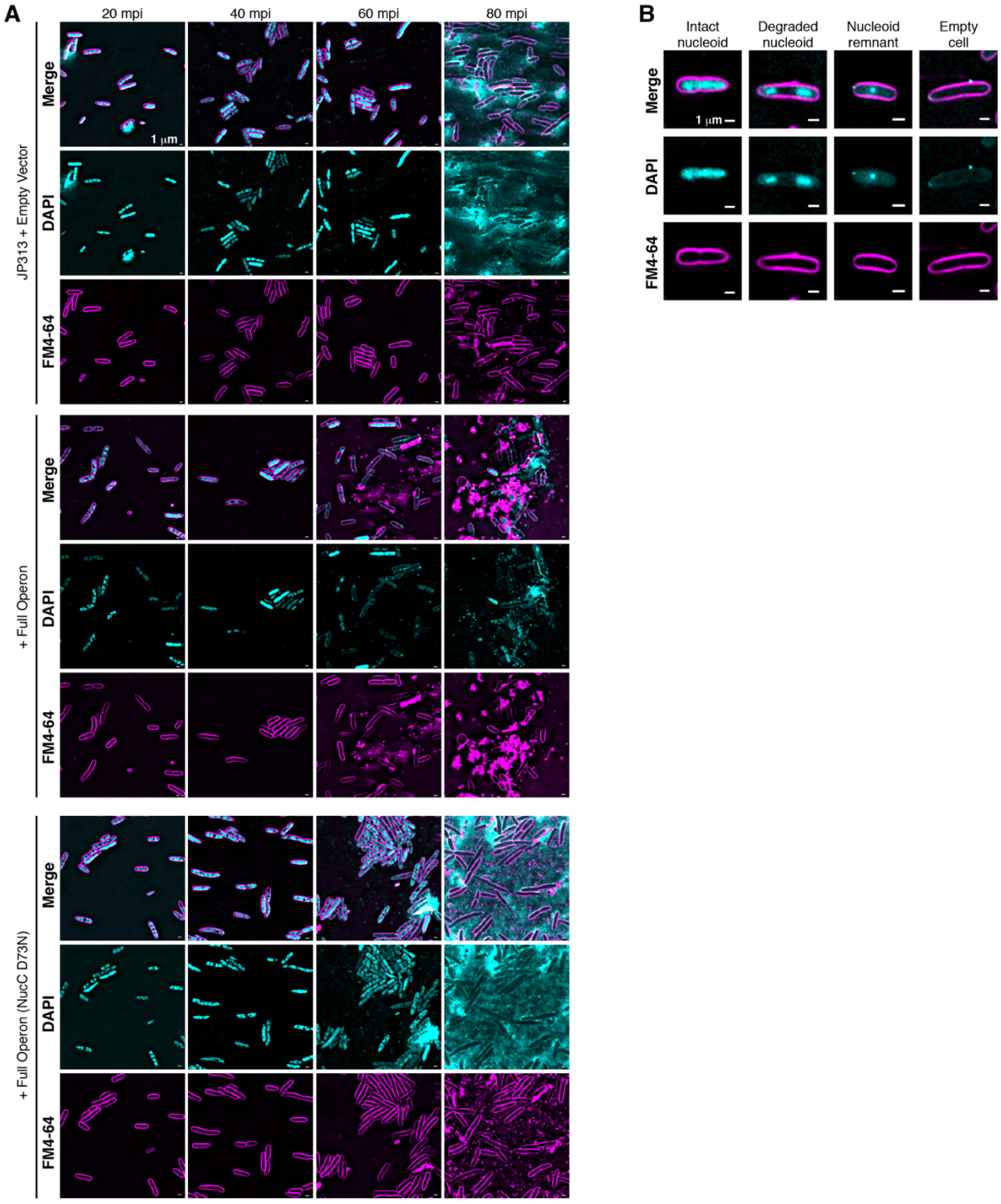


Figure A.B.S33 NucC activation leads to chromosome destruction in infected cells

(A) Separated channels for images shown in Figure 1A. All scale bars = 1 μm. **(B)** Separated channels for images shown in Figure 1A. All scale bars = 1 μm**.**

Figure A.B.S34 NucC sequence and structure

(A) Size exclusion chromatography coupled to multi-angle light scattering (SEC-MALS) of purified *Ec* NucC (theoretical trimer = 80.1 kDa). **(B)** SEC-MALS of purified *Pa* NucC (theoretical trimer = 80.0 kDa). **(C)** Closeup view of the NucC active site, with conserved active-site residues shown as sticks and Mg^{2+} shown as black sphere. *Inset* shows active-site consensus motif. **(D)** Closeup view of the active site of restriction endonuclease SgrAI, from a DNA-bound structure (PDB ID 3DW9) (Dunten et al., 2008). Bound DNA, which coordinates one Mn^{2+} ion with D188 and the main-chain carbonyl group of F241, is not shown. *Inset* shows active-site consensus motif. **(E)** Sequence alignment of *E. coli* MS115-1 (*Ec*) and *P. aeruginosa* ATCC27853 (*Pa*) NucC with four NucC homologs associated with Type III CRISPR systems (all from cluster 3 in **Figure 2.7A**). Secondary structure shown is that of *Ec* NucC, with elements of the core nuclease fold in violet and embellishments in gray (**Figure 2.2A**). Orange highlights indicate the active-site consensus motif ($IDx_{30}EAK$) and yellow highlights indicate the gate loop latch residues H136 and Y141. Green and cyan underlines indicate residues involved in the two interfaces that mediate NucC trimer assembly, calculated by PISA (Krissinel and Henrick, 2007).

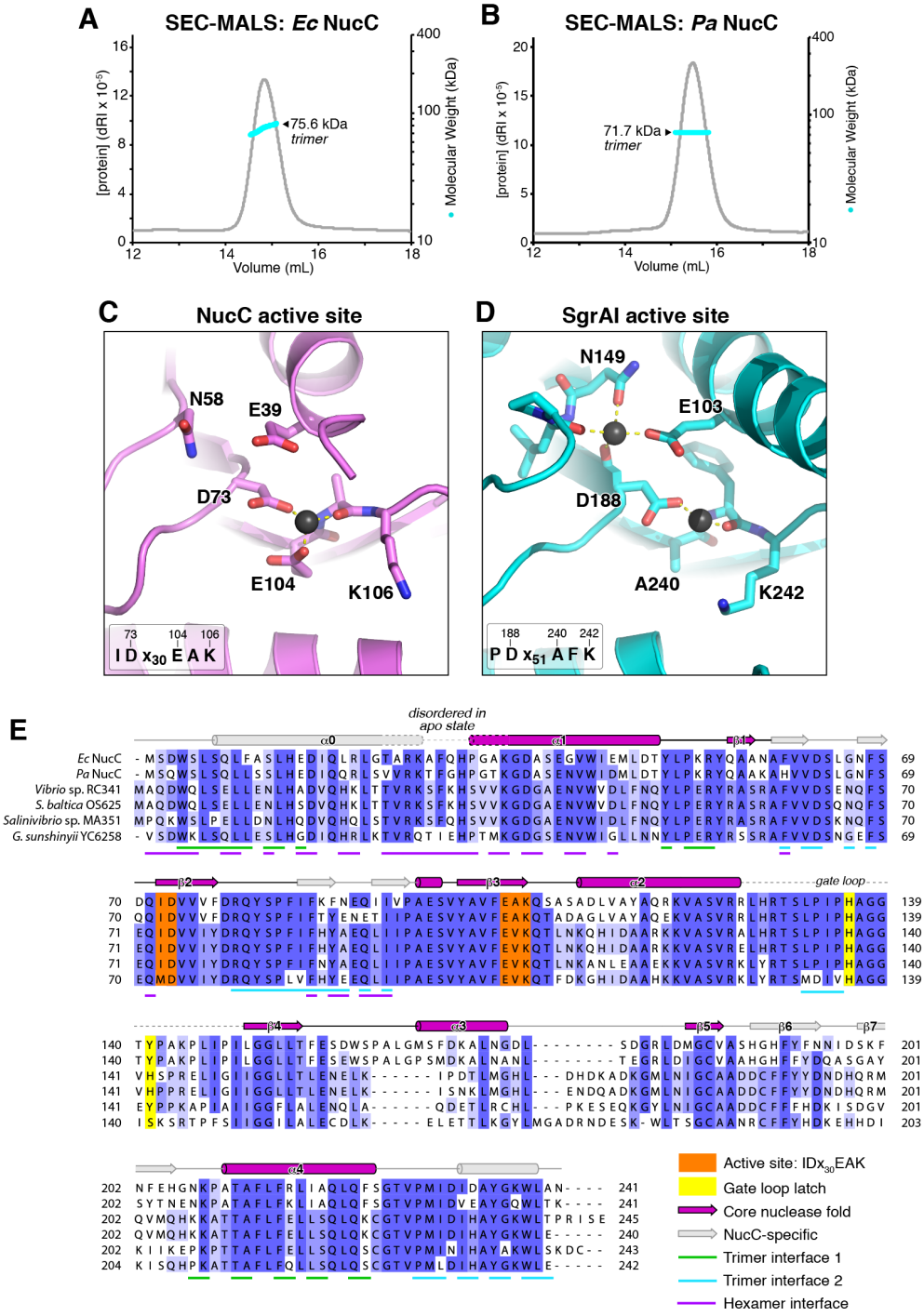
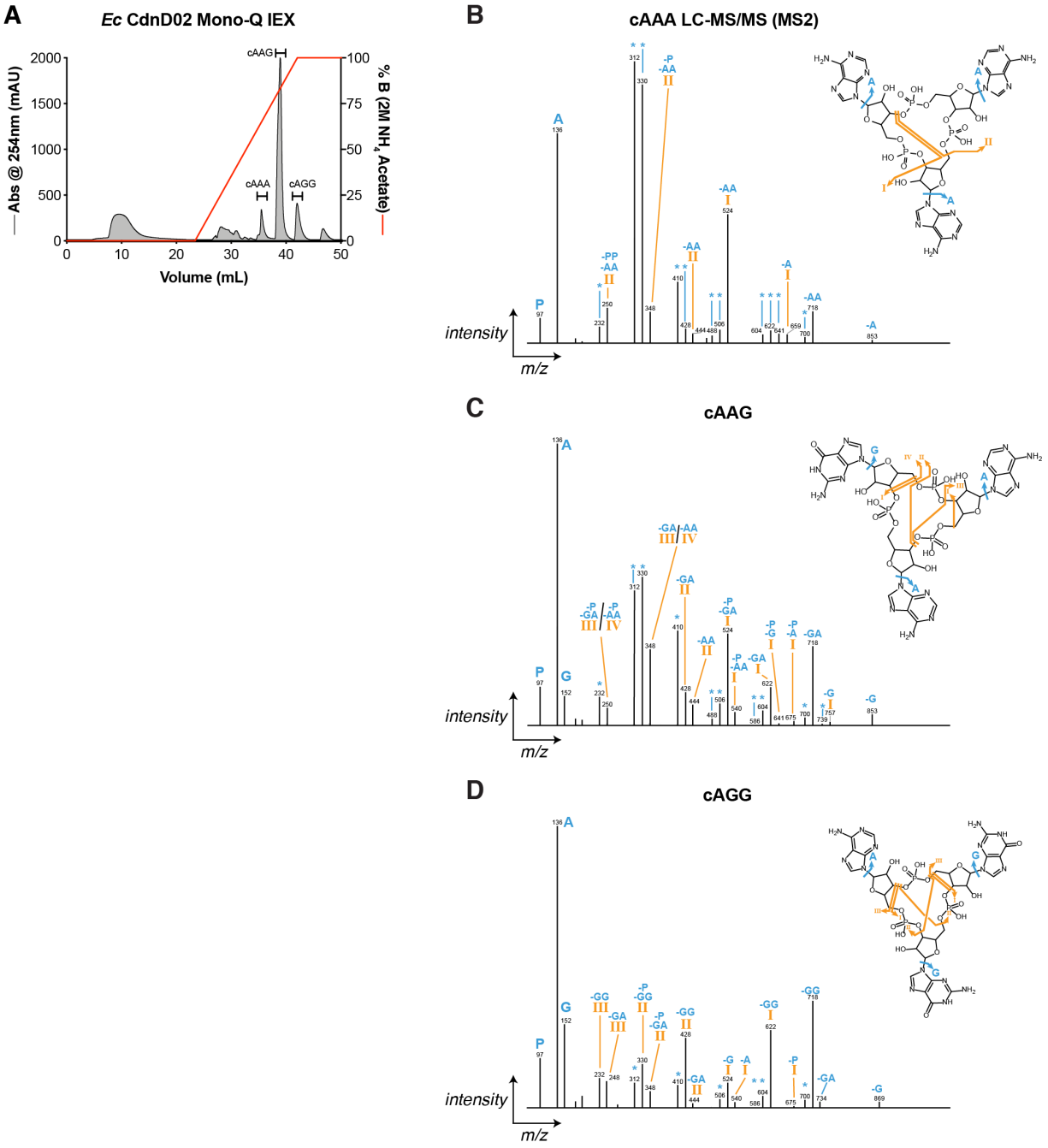

A
SEC-MALS: Ec NucC
[protein] (dRI x 10-5)
Molecular Weight (kDa)
75.6 kDa
trimer
Volume (mL)
B
SEC-MALS: Pa NucC
71.7 kDa
trimer
C
NucC active site
N58
E39
D73
E104
K106
I D x30 E A K
D
SgrAI active site
N149
E103
D188
A240
K242
P D x51 A F K
E
disordered in apo state
Ec NucC
Pa NucC
Vibrio sp. RC341
S. baltica OS625
Salinivibrio sp. MA351
G. sunshinyii YC6258
gate loop
Active site: IDx30EAK
Gate loop latch
Core nuclease fold
NucC-specific
Trimer interface 1
Trimer interface 2
Hexamer interface

Figure A.B.S35 Cyclic trinucleotide synthesis by *Enterobacter cloacae* CdnD02

(A) Mono-Q ion-exchange chromatography elution profile of reaction products generated by *Enterobacter cloacae* CdnD02 when supplied with GTP and ATP. The most prominent peak (cAAG) was previously characterized by NMR (Whiteley et al., 2019). The cAAG peak was verified, and two minor products identified as cAAA and cAGG, by LC-MS/MS mass spectrometry (panels B-D). **(B-D)** Mass fragmentation pattern of Mono-Q peak 1 (cAAA, panel B), peak 2 (cAAG , panel C), and peak 3 (cAGG, panel D). Roman numerals indicate associated bond cleavage, with asterisk (*) indicating additional loss of water. “-A”, “-G”, and “-P” indicate loss of adenine, guanine, and phosphate, respectively.


A
Ec CdnD02 Mono-Q IEX
Abs @ 254nm (mAU)
% B (2M NH4 Acetate)
Volume (mL)
cAAG
cAAA
cAGG
B
cAAA LC-MS/MS (MS2)
intensity
m/z
C
cAAG
intensity
m/z
D
cAGG
intensity
m/z

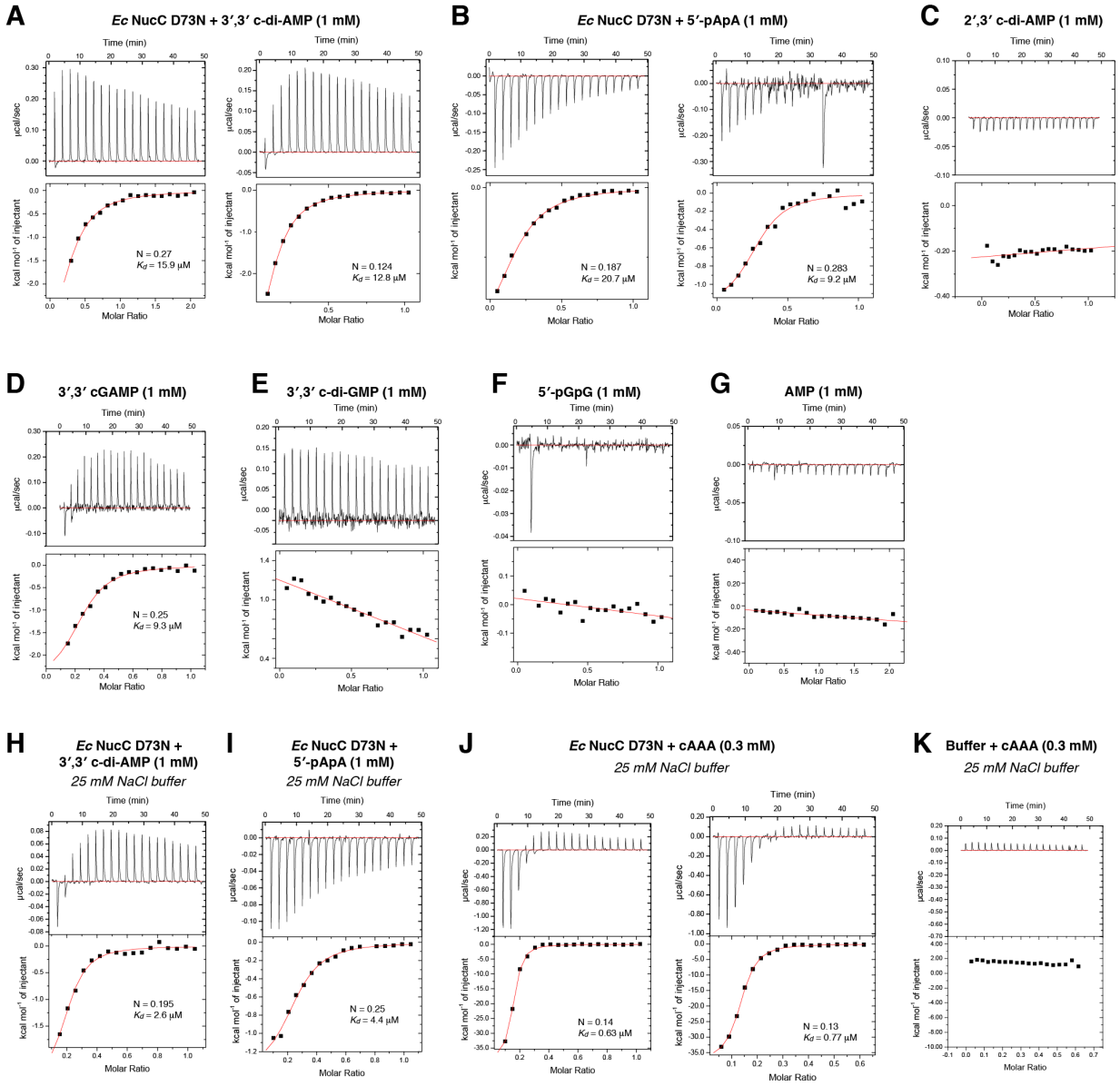
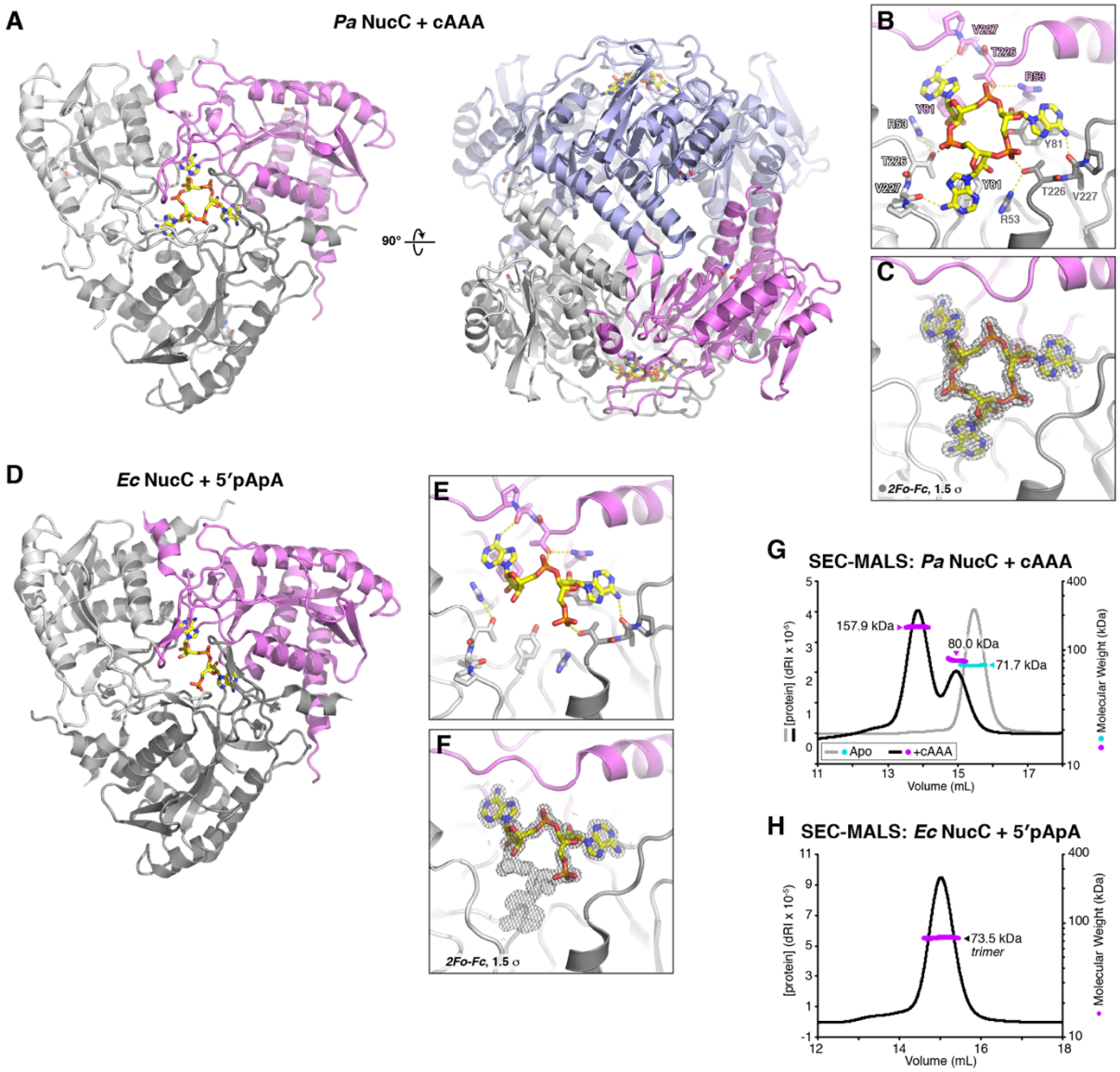


Figure A.B.S36 Binding of NucC to second messengers

(A) Binding of *Ec* NucC D73N to 3′,3′ cyclic di-AMP in high-salt buffer by isothermal titration calorimetry. See **Table A.B.S2** for affinity and stoichiometry of binding for panels A-G. **(B)** Binding of *Ec* NucC D73N to 5′-pApA (linear di-AMP). **(C)** Binding of *Ec* NucC D73N to 2′,3′ cyclic di-AMP. **(D)** Binding of *Ec* NucC D73N to 3′,3′ cyclic GMP-AMP (cGAMP). **(E)** Binding of *Ec* NucC D73N to 3′,3′ cyclic di-GMP. **(F)** Binding of *Ec* NucC D73N to 5′-pGpG (linear di-GMP). **(G)** Binding of *Ec* NucC D73N to AMP. **(H)** Binding of *Ec* NucC D73N to 3′,3′ cyclic di-AMP in low-salt buffer. See **Table S3** for affinity and stoichiometry of binding for panels H-J. **(I)** Binding of *Ec* NucC D73N to 5′-pApA (linear di-AMP) in low-salt buffer. **(J)** Binding of *Ec* NucC D73N to cAAA in low-salt buffer. **(K)** Negative control titration of low-salt buffer with cAAA.

Figure A.B.S37 Structures of Pa NucC+cAAA and Ec NucC + 5'-pApA

(A) *Left:* Bottom view of a *Pa* NucC trimer (chains colored violet, gray, and white) with bound cAAA shown as sticks. The *Pa* NucC trimer overlays with the cAAA-bound *Ec* NucC trimer with an overall Cα r.m.s.d. of 0.51 Å (649 Cα pairs). *Right:* Side view of a *Pa* NucC hexamer in the cAAA-bound state (one trimer colored as in panel A, and the second trimer colored light blue). Active site residues Asn73 (D73N mutant) and Glu104 are shown as sticks. **(B)** Closeup view of cAAA binding to *Pa* NucC. **(C)** 2*Fo-Fc* electron density for cAAA at 1.45 Å resolution, contoured at 1.5 σ. **(D)** Bottom view of an *Ec* NucC trimer (chains colored violet, gray, and white) with bound 5′-pApA shown as sticks. The *Ec* NucC+5′-pApA complex crystallizes equivalently to the *Ec* NucC+cAAA complex, forming a hexamer equivalent to that shown in **Figure 2.4A**. **(E)** Closeup view of 5′-pApA binding to *Ec* NucC. **(F)** 2*Fo-Fc* electron density for 5′-pApA at 1.66 Å resolution, contoured at 1.5 σ. The density is three-fold symmetric due to random orientation of 5′-pApA in the NucC allosteric pocket. **(G)** Size exclusion chromatography coupled to multi-angle light scattering (SEC-MALS) for *Pa* NucC in the absence of cAAA (gray line with molecular weight measurement in cyan) and in the presence of cAAA (black line with molecular weight measurement in violet). The molar mass of a *Pa* NucC homotrimer is 80.0 kDa, and a homohexamer is 160.0 kDa. **(H)** SEC-MALS analysis of *Ec* NucC in the presence of 5′-pApA. Despite crystallizing equivalently to *Ec* NucC+cAAA, 5′-pApA is unable to mediate NucC hexamer assembly in solution.


A
Pa NucC + cAAA
90°
B
V227
T226
R53
Y81
R53
Y81
T226
V227
Y81
T226
V227
R53
C
2Fo-Fc, 1.5 σ
D
Ec NucC + 5′pApA
E
F
2Fo-Fc, 1.5 σ
G
SEC-MALS: Pa NucC + cAAA
157.9 kDa
80.0 kDa
71.7 kDa
[protein] (dRI x 10⁻⁵)
Molecular Weight (kDa)
Apo
+cAAA
Volume (mL)
H
SEC-MALS: Ec NucC + 5′pApA
73.5 kDa
trimer
[protein] (dRI x 10⁻⁵)
Molecular Weight (kDa)
Volume (mL)

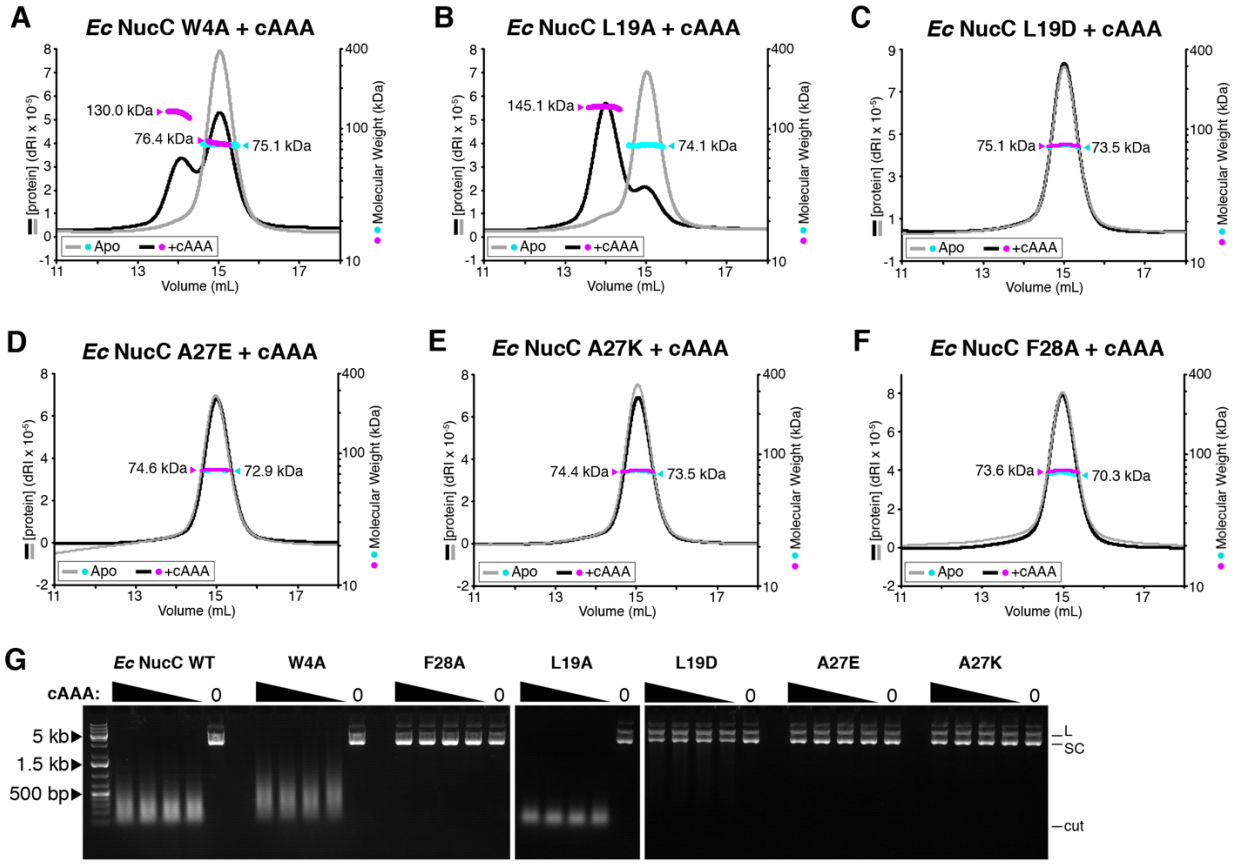


Figure A.B.S38 SEC-MALS analysis of Ec NucC hexamer interface mutants

(A) Size exclusion chromatography coupled to multi-angle light scattering (SEC-MALS) for *Ec* NucC W4A mutant (with additional D73N active-site mutant) in the absence of cAAA (gray line with molecular weight measurement in cyan) and in the presence of cAAA (black line with molecular weight measurement in violet). The molar mass of an *Ec* NucC homotrimer is 80.1 kDa, and a homohexamer is 160.3 kDa. **(B)** SEC-MALS analysis of *Ec* NucC L19A mutant. **(C)** SEC-MALS analysis of *Ec* NucC L19D mutant. **(D)** SEC-MALS analysis of *Ec* NucC A27E mutant **(E)** SEC-MALS analysis of *Ec* NucC A27K mutant. **(F)** SEC-MALS analysis of *Ec* NucC F28A mutant. **(G)** Plasmid digestion assay with wild-type and hexamer-interface mutant *Ec* NucC (10 nM), in the presence of cAAA (for each set: 400, 100, 25, 6.25, and 0 nM cAAA). “L” denotes linear plasmid, “SC” denotes closed-circular supercoiled plasmid, and “cut” denotes fully-digested DNA. Results from all panels are summarized in **Figure 2.4H**.

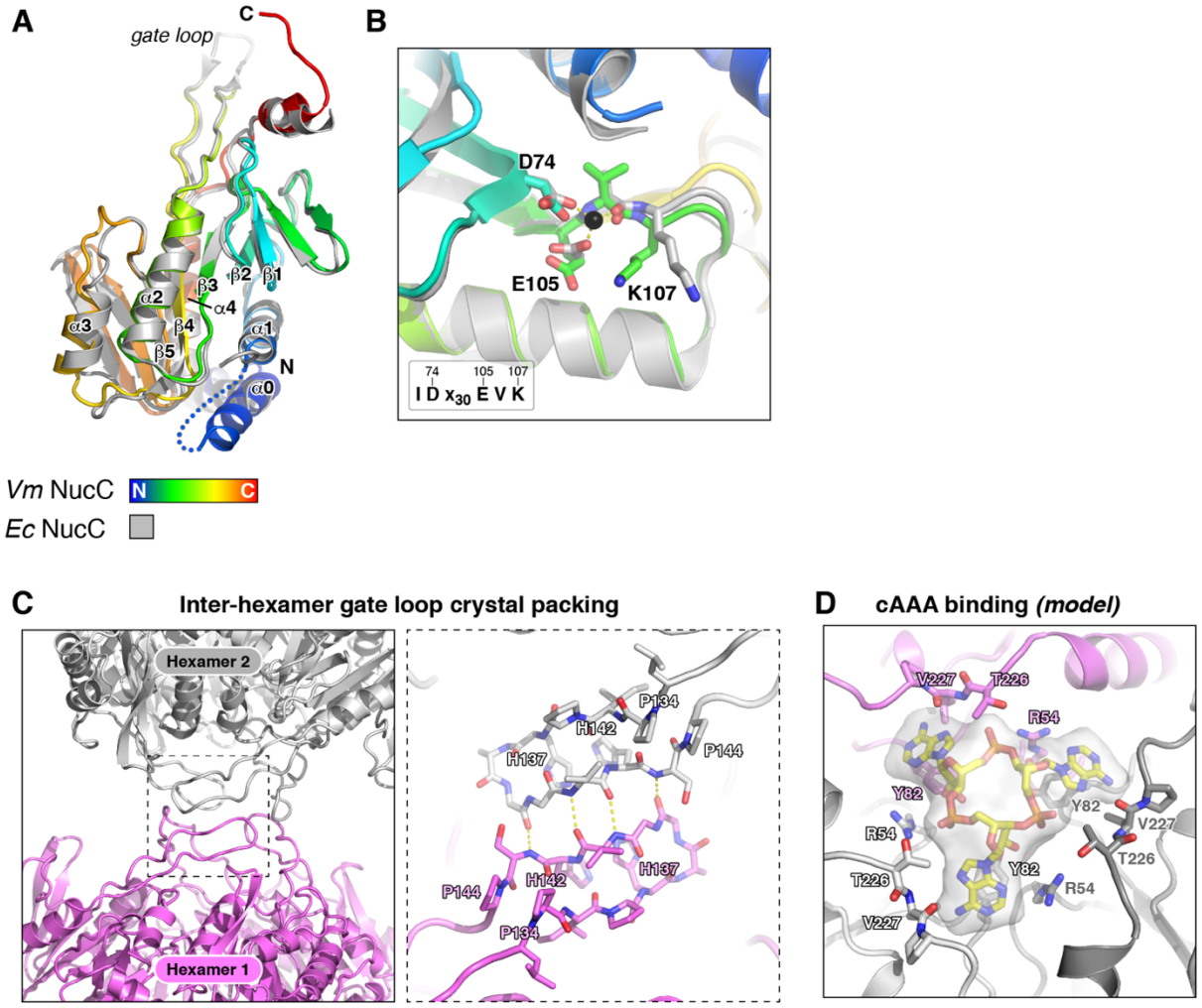


Figure A.B.S39 Structure of *V. metoecus* NucC

(A) Structural superposition of a single NucC monomer from apo-state trimer structures of *V. metoecus* NucC (colored as a rainbow from N-terminus to C-terminus) and *E. coli* MS115-1 NucC (gray). The two proteins overlay with an overall r.m.s.d. of 0.77 Å over 170 equivalent Cα atoms. **(B)** Close-up view of the NucC active site, with *Vm* NucC and *Ec* NucC colored as in panel (A). Active site residues for *Vm* NucC (D74, E105, and K107) are labeled. The Mg^{2+} ion bound in the *Ec* NucC active site is shown in black. **(C)** Overall view (left) and closeup view (right) of inter-hexamer packing in the structure of the *Vm* NucC hexamer. The gate loops of two adjacent hexamers interact symmetrically, pulling the gate loops into an open state and likely allowing release of bound cAAA. **(D)** Model of cAAA (sticks with semi-transparent surface) binding in the allosteric pocket of *Vm* NucC, showing that the pocket can accommodate cAAA.

Methods A.B.S1 (Related to STAR Methods). Sequence of UC Berkeley Macrolab plasmid 2AT (Addgene #29665), used for NucC plasmid digestion assays and high-throughput sequencing.

TTCTTGAAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGAT
AATAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACC
CCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAAC
CCTGATAAATGCTTCAATAACATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCC
GTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAG
AAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTA
CATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAAC
GTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTG
TTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTT
GGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGA
GAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCT
GACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGAT
CATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACG
ACGAGCGTGACACCACGATGCCTGCAGCAATGGCAACAACGTTGCGCAAACTATT
AACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGG
CGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTAT
TGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTG
GGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGG
CAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAG
CATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTC
ATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAAT
CCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAG
GATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAAC
CACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCG
AAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCC
GTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGC
TAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTG
GACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGT
TCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTAC
AGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGT
ATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGG
GGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCG
TCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAAC
GCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCT
GCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATAC
CGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGG
AAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGC
ATATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTAT
ACACTCCGCTATCGCTACGTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAAC
ACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAA
GCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGA

AACGCGCGAGGCAGCTGCGGTAAAGCTCATCAGCGTGGTCGTGAAGCGATTCACA
GATGTCTGCCTGTTCATCCGCGTCCAGCTCGTTGAGTTTCTCCAGAAGCGTTAATG
TCTGGCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTTTTTTCCTGTTTGGTCACT
GATGCCTCCGTGTAAGGGGGATTTCTGTTCATGGGGGTAATGATACCGATGAAAC
GAGAGAGGATGCTCACGATACGGGTTACTGATGATGAACATGCCCGGTTACTGGA
ACGTTGTGAGGGTAAACAACTGGCGGTATGGATGCGGCGGGACCAGAGAAAAATC
ACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGGTGTTCCACAGGGTA
GCCAGCAGCATCCTGCGATGCAGATCCGGAACATAATGGTGCAGGGCGCTGACTT
CCGCGTTTCCAGACTTTACGAAACACGGAAACCGAAGACCATTCATGTTGTTGCTC
AGGTCGCAGACGTTTTGCAGCAGCAGTCGCTTCACGTTCGCTCGCGTATCGGTGA
TTCATTCTGCTAACCAGTAAGGCAACCCCGCCAGCCTAGCCGGGTCCTCAACGAC
AGGAGCACGATCATGCGCACCCGTGGCCAGGACCCAACGCTGCCCGAGATGCGC
CGCGTGCGGCTGCTGGAGATGGCGGACGCGATGGATATGTTCTGCCAAGGGTTG
GTTTGCGCATTCACAGTTCTCCGCAAGAATTGATTGGCTCCAATTCTTGGAGTGGT
GAATCCGTTAGCGAGGTGCCGCCGGCTTCCATTCAGGTCGAGGTGGCCCGGCTC
CATGCACCGCGACGCAACGCGGGGAGGCAGACAAGGTATAGGGCGGCGCCTACA
ATCCATGCCAACCCGTTCCATGTGCTCGCCGAGGCGGCATAAATCGCCGTGACGA
TCAGCGGTCCAGTGATCGAAGTTAGGCTGGTAAGAGCCGCGAGCGATCCTTGAAG
CTGTCCCTGATGGTCGTCATCTACCTGCCTGGACAGCATGGCCTGCAACGCGGGC
ATCCCGATGCCGCCGGAAGCGAGAAGAATCATAATGGGGAAGGCCATCCAGCCTC
GCGTCGCGAACGCCAGCAAGACGTAGCCCAGCGCGTCGGCCGCCATGCCGGCG
ATAATGGCCTGCTTCTCGCCGAAACGTTTGGTGGCGGGACCAGTGACGAAGGCTT
GAGCGAGGGCGTGCAAGATTCCGAATACCGCAAGCGACAGGCCGATCATCGTCG
CGCTCCAGCGAAAGCGGTCCTCGCCGAAAATGACCCAGAGCGCTGCCGGCACCT
GTCCTACGAGTTGCATGATAAAGAAGACAGTCATAAGTGCGGCGACGATAGTCAT
GCCCCGCGCCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCTCAAGGGCATCGG
TCGACGCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTTGA
GGCCGTTGAGCACCGCCGCCGCAAGGAATGGTGCATGCAAGGAGATGGCGCCCA
ACAGTCCCCCGGCCACGGGGCCTGCCACCATACCCACGCCGAAACAAGCGCTCA
TGAGCCCGAAGTGGCGAGCCCGATCTTCCCCATCGGTGATGTCGGCGATATAGGC
GCCAGCAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGCGTCCGGCGTA
GAGGATCGAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGAGACCAC
AACGGTTTCCCTCTAGTGCCGGCTCCGGAGAGCTCTTTAATTAAGCGGCCGCCCT
GCAGGACTCGAGTTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATAGATAT
CCCAACTCCATAAGGATCCGCGATCGCGGCGCGCCACCTGGTGGCCGGCCGGTA
CCACGCGTGCGCGCTGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTG
GCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGG
TCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGACATCCACAGGACGG
GTGTGGTCGCCATGATCGCGTAGTCGATAGTGGCTCCAAGTAGCGAAGCGAGCAG
GACTGGGCGGCGGCCAAAGCGGTCGGACAGTGCTCCGAGAACGGGTGCGCATA
GAAATTGCATCAACGCATATAGCGCTAGCAGCACGCCATAGTGACTGGCGATGCT
GTCGGAATGGACGACATCCCGCAAGAGGCCCGGCAGTACCGGCATAACCAAGCC
TATGCCTACAGCATCCAGGGTGACGGTGCCGAGGATGACGATGAGCGCATTGTTA

GATTTCATACACGGTGCCTGACTGCGTTAGCAATTTAACTGTGATAAACTACCGCA
TTAAAGCTTATCGATGATAAGCTGTCAAACATGAGAA

Table A.B.S5 Crystallographic Data

	***Ec* NucC Apo**	***Ec* NucC Apo SeMet**	***Ec* NucC D73N + 5¢-pApA**	***Ec* NucC D73N + cAAA**
Data collection				
Synchrotron/Beamline	SSRL 14-1	SSRL 14-1	APS 24ID-E	APS 24ID-E
Date collected	1/12/17	1/12/17	8/1/18	2/26/19
Resolution (Å)	40 – 1.75	40 – 1.82	104 - 1.66	104 – 1.66
Wavelength (Å)	1.19499	0.9792	0.97918	0.97918
Space Group	$F4_132$	$F4_132$	$H3_2$	$H3_2$
Unit Cell Dimensions (a, b, c) Å	208.39, 208.39, 208.39	208.79, 208.79, 208.79	131.78, 131.78, 252.84	131.59, 131.59, 252.11
Unit cell Angles (α,β,γ) °	90, 90, 90	90, 90, 90	90, 90, 120	90, 90, 120
I/s (last shell)	16.5 (1.2)	12.4 (1.0)	9.7 (1.7)	20.5 (3.5)
[a] R_{sym} (last shell)	0.065 (1.517)	0.125 (2.059)	0.138 (0.830)	0.093 (0.601)
[b] R_{meas} (last shell)	0.068 (1.633)	0.129 (2.169)	0.154 (0.924)	0.098 (0.636)
[c] $CC_{1/2}$, last shell	0.998 (0.530)	0.998 (0.460)	0.994 (0.617)	0.999 (0.890)
Completeness (last shell) %	99.9 (97.8)	99.3 (89.8)	99.2 (97.6)	99.2 (84.2)
[d] Number of reflections	369692	486245	491090	977367
unique	39582	35180	98697	97528
Multiplicity (last shell)	9.3 (7.2)	13.8 (9.5)	5.0 (4.9)	10.0 (9.3)
Refinement				
Resolution (Å)	37.0 – 1.75	-	56 – 1.66	66 – 1.66
[d] No. of reflections	39517	-	98601	97502
working	37284	-	93761	92714
free	2233	-	4840	4788
[e] R_{work} (last shell) %	17.08 (28.47)	-	15.88 (26.21)	14.19 (20.03)
[e] R_{free} (last shell) %	19.56 (30.86)	-	18.46 (29.09)	16.21 (24.51)
Structure/Stereochemistry				
Number of atoms	3752	-	12374	12416
solvent	173	-	936	861
ligands	70	-	135	198
ions	2	-	1	1
hydrogen	1771	-	5614	5656

Table A.B.S1 Crystallographic Data

r.m.s.d. bond lengths (Å)	0.019	-	0.012	0.008
r.m.s.d. bond angles (°)	1.576	-	1.025	0.968
[f] PDB ID	6P7O	-	6P7Q	6P7P
[g] SBGrid Data Bank ID	664	665	667	666

	***Ec* NucC Apo**	***Ec* NucC Apo SeMet**	***Ec* NucC D73N + 5¢-pApA**
Data collection			
Synchrotron/Beamline	SSRL 14-1	SSRL 14-1	APS 24ID-E
Date collected	1/12/17	1/12/17	8/1/18
Resolution (Å)	40 – 1.75	40 – 1.82	104 - 1.66
Wavelength (Å)	1.19499	0.9792	0.97918
Space Group	$F4_132$	$F4_132$	$H3_2$
Unit Cell Dimensions (a, b, c) Å	208.39, 208.39, 208.39	208.79, 208.79, 208.79	131.78, 131.78, 252.84
Unit cell Angles (α,β,γ) °	90, 90, 90	90, 90, 90	90, 90, 120
I/s (last shell)	16.5 (1.2)	12.4 (1.0)	9.7 (1.7)
[a] R_{sym} (last shell)	0.065 (1.517)	0.125 (2.059)	0.138 (0.830)
[b] R_{meas} (last shell)	0.068 (1.633)	0.129 (2.169)	0.154 (0.924)
[c] $CC_{1/2}$, last shell	0.998 (0.530)	0.998 (0.460)	0.994 (0.617)
Completeness (last shell) %	99.9 (97.8)	99.3 (89.8)	99.2 (97.6)
[d] Number of reflections	369692	486245	491090
unique	39582	35180	98697
Multiplicity (last shell)	9.3 (7.2)	13.8 (9.5)	5.0 (4.9)
Refinement			
Resolution (Å)	37.0 – 1.75	-	56 – 1.66
[d] No. of reflections	39517	-	98601
working	37284	-	93761
free	2233	-	4840
[e] R_{work} (last shell) %	17.08 (28.47)	-	15.88 (26.21)
[e] R_{free} (last shell) %	19.56 (30.86)	-	18.46 (29.09)

Table A.B.S1 Crystallographic Data

Structure/Stereochemistry			
Number of atoms	3752	-	12374
solvent	173	-	936
ligands	70	-	135
ions	2	-	1
hydrogen	1771	-	5614
r.m.s.d. bond lengths (Å)	0.019	-	0.012
r.m.s.d. bond angles (°)	1.576	-	1.025
[f] PDB ID	6P7O	-	6P7Q
[g] SBGrid Data Bank ID	664	665	667

[a] R_{sym} = åå$_j$| I_j – áIñ|/åI_j, where I_j is the intensity measurement for reflection j and áIñ is the mean intensity for multiply recorded reflections.
[b] R_{meas} = å$_h$ [√(n/(n-1)) å$_j$ [I_{hj} - áI_hñ] / å$_{hj}$ áI_hñ, where I_{hj} is a single intensity measurement for reflection h, áI_hñ is the average intensity measurement for multiply recorded reflections, and n is the number of observations of reflection h.
[c] $CC_{1/2}$ is the Pearson correlation coefficient between the average measured intensities of two randomly-assigned half-sets of the measurements of each unique reflection.
[d] Reflection counts in scaling and refinement are unmerged (F(+) and F(-) treated as separate reflections).
[e] $R_{work,\ free}$ = å||F_{obs}| – |F_{calc}||/|F_{obs}|, where the working and free R-factors are calculated using the working and free reflection sets, respectively.
[f] Coordinates and structure factors have been deposited in the RCSB Protein Data Bank (www.pdb.org) with the noted accession numbers.
[g] Diffraction data have been deposited with the SBGrid Data Bank (https://data.sbgrid.org) with the noted accession numbers.

Table A.B.S6 NucC binding second messenger molecules in high-salt conditions

Second messenger	K_d (μM)	*N* per NucC trimer[a]
3′,3′ cyclic di-AMP	14.4	0.60
5′-pApA	15.0	0.72
2′,3′ cyclic di-AMP	*no binding*	-
3′,3′ cyclic di-GMP	*no binding*	-
5′-pGpG	*no binding*	-
3′,3′ cyclic GMP-AMP (cGAMP)	9.3	0.75
AMP	*no binding*	-

[a]*N* refers to the stoichiometry of second messenger binding to the NucC homotrimer.

Table A.B.S7 NucC binding second messenger molecules in low-salt conditions

Second messenger	K_d (μM)	*N* per NucC trimer[a]
3′,3′ cyclic di-AMP	2.6	0.59
5′-pApA	4.4	0.75
cAAA	0.7	0.41

[a]*N* refers to the stoichiometry of second messenger binding to the NucC homotrimer.

Table A.B.S8 Type III CRISPR/Cas-associated NucC homologs

Database Identifier	Database (1)	Length	Shah et al. cluster (2)	Species	Bacterial Group	% identity to prior sequence
2513842568	IMG	250		Teredinibacter turnerae T8602	Gamma-proteobacteria	
WP_028881082.1	GenBank	250		Teredinibacter turnerae	Gamma-proteobacteria	96.4%
2526117325	IMG	247		Teredinibacter sp. 991H.S.0a.06	Gamma-proteobacteria	95.6%
647193063	**IMG**	**245**		**Vibrio metoecus sp. RC341**	Gamma-proteobacteria	**52.7%**
WP_076024803.1	GenBank	245		Vibrio cholerae	Gamma-proteobacteria	99.6%
WP_148499653.1	GenBank	245		Vibrio cholerae	Gamma-proteobacteria	97.6%
WP_085602491.1	GenBank	245		Vibrio cholerae	Gamma-proteobacteria	99.6%
WP_071170739.1	GenBank	245		Vibrio cholerae	Gamma-proteobacteria	99.2%
2507444198	IMG	240		Shewanella baltica OS625	Gamma-proteobacteria	95.4%
2837149145	IMG	240		Marinomonas arctica 328	Gamma-proteobacteria	82.1%
2792187373	IMG	243		Salinivibrio sp. MA351	Gamma-proteobacteria	75.7%
2632356247	**IMG**	**242**		**Gynuella sunshinyii YC6258**	Gamma-proteobacteria	**63.9%**
WP_044616068.1	GenBank	242		Gynuella sunshinyii	Gamma-proteobacteria	99.6%
WP_077298232.1	GenBank	310		Aquaspirillum sp. LM1	Beta-proteobacteria	40.8%
WP_110389129.1	GenBank	255		Rivicola pingtungensis	Beta-proteobacteria	60.1%

Table A.B.S4 Type III CRISPR/Cas-associated NucC homologs

WP_012885876.1	GenBank	307	PD-DExK (3)	Dickeya zeae (strain Ech586)	Gamma-proteobacteria	32.0%
WP_064665190.1	GenBank	292		Pseudoalteromonas prydzensis	Gamma-proteobacteria	41.4%
WP_015841417.1	GenBank	250	PD-DExK	Pectobacterium carotovorum subsp. carotovorum (strain PC1)	Gamma-proteobacteria	44.4%
WP_021014788.1	GenBank	250	PD-DExK	Serratia ATCC 39006	Gamma-proteobacteria	92.4%
WP_013659859.1	GenBank	256	PD-DExK	Marinomonas mediterranea strain ATCC 700492	Gamma-proteobacteria	47.4%
2831805761	IMG	258		Pokkaliibacter plantistimulans L1E11	Gamma-proteobacteria	54.8%
2779856863	IMG	248		Deltaproteobacteria bacterium NP119	Delta-proteobacteria	49.4%
2808182764	IMG	239		Faecalicatena contorta 2789STDY5834876	Firmicutes	37.9%
2789927606	IMG	275		Anaeromassilibacillus sp. An250	Firmicutes	66.5%
WP_050698056.1	GenBank	245		Anaeromassilibacillus senegalensis	Firmicutes	72.4%
2667999241	IMG	231		Cellulosilyticum ruminicola JCM 14822	Firmicutes	63.6%
WP_015279331.1	Genbank	241	PD-DExK	Thioflavicoccus mobilis 8321	Gamma-proteobacteria	32.6%
2806389584	IMG	254		Nitrospirillum amazonense BR 11880	Alpha-proteobacteria	48.7%
2510777369	IMG	243		Acetivibrio cellulolyticus CD2	Firmicutes	38.6%
2653800354	IMG	245		Tindallia californiensis APO	Firmicutes	51.7%
2595103576	IMG	242		Methylocaldum sp. 175	Gamma-proteobacteria	41.3%

Table A.B.S4 Type III CRISPR/Cas-associated NucC homologs

Bold type indicates proteins selected for in vitro expression and purification
Underlines denote the 10 separate clusters of NucC homologs associated with Type III CRISPR systems (see Figure 2.5)
(1) Databases include Genbank (https://www.ncbi.nlm.nih.gov) and IMG (https://img.jgi.doe.gov)
(2) See Shah et al., "Comprehensive search for accessory proteins encoded with archaeal and bacterial type III CRISPR-cas gene cassettes reveals 39 new cas gene families" (2019) *RNA Biology* **16**(4):530-542
(3) As originally classified, cluster PD-DExK included nine proteins. Four members of the cluster (Uniprot IDs HPL003_18785, Desru_3523, ERIC2_c04110, and Sgly_1266) contain a likely nuclease active site but do not show clear hallmarks of NucC proteins (e.g. the gate loop), and were excluded from further analysis

Table A.B.S9 NucC proteins used for sequence alignment

Database	Database Identifier
Genbank	EFJ98159.1, WP_003050273.1, WP_012885876.1, WP_013659859.1, WP_015841417.1, WP_021014788.1, WP_015279331.1, WP_117109623.1, WP_072039935.1, WP_075201405.1, EBT1697807.1, WP_126521968.1, WP_080865893.1, WP_053287146.1, WP_087760536.1, WP_125297157.1, WP_088126339.1, WP_072252788.1, WP_058799372.1, EAT3274191.1, WP_065684574.1, WP_129753977.1, WP_117110152.1, WP_094962242.1, WP_105575604.1, WP_036770685.1, WP_048638343.1, WP_054989155.1, WP_110818094.1, WP_058427461.1, WP_122262520.1, WP_103367126.1, WP_102823321.1, WP_132727533.1, WP_132884923.1, WP_132607396.1, WP_033895614.1, WP_066412676.1, WP_058139025.1, WP_005408888.1, WP_023087304.1, WP_003120003.1, WP_049253691.1, WP_134560360.1, WP_043555297.1, WP_046720530.1, WP_122048403.1, WP_134310892.1, WP_061198537.1, WP_061193800.1, WP_124102865.1, AAN62156.1, VTP97482.1, CDH75061.1, WP_024067897.1, WP_043156220.1, WP_021205712.1, WP_100773617.1, WP_070490265.1, WP_065502002.1, WP_096453106.1, WP_011829959.1, WP_050157971.1, WP_007514452.1, WP_076401812.1, WP_018427714.1, OYX76705.1, OYY88408.1, WP_140899275.1, WP_008125037.1, WP_104952816.1, WP_113240871.1, WP_099555781.1, WP_134490779.1, WP_073135241.1, WP_099095839.1, WP_009338226.1, WP_011663883.1, WP_012110740.1, WP_043946292.1, WP_068071346.1, WP_046716131.1, WP_026070607.1, WP_027533731.1, WP_076845471.1, WP_020048021.1, WP_094087287.1, WP_131705022.1, WP_131591324.1, WP_113473263.1, WP_003508391.1, WP_121557654.1, WP_110737489.1, WP_037275570.1, WP_093993975.1, WP_074839947.1, WP_064225022.1, WP_099249422.1, WP_013592364.1, KND60647.1, WP_111461857.1, WP_075935152.1, WP_017454940.1, VFJ74888.1, WP_137391929.1, WP_152293612.1, WP_054534569.1, WP_089937202.1, ETX06256.1, WP_036274931.1, WP_084353930.1, WP_106572897.1, WP_014705157.1, WP_023820823.1, WP_035095385.1, WP_053505562.1, WP_024540132.1, WP_045579177.1, WP_012367385.1, WP_056321645.1, TAG75758.1, WP_066918475.1, WP_025586683.1, WP_057063717.1, WP_020753579.1, WP_151814670.1, OYV47967.1, WP_122401238.1, THD49852.1, WP_111545628.1, WP_140659809.1, WP_140905777.1, WP_097523110.1, WP_131549119.1, WP_124964925.1, TMQ11038.1, WP_010141868.1, PPC99544.1, EEF26502.1, WP_069053887.1, PZQ87067.1, WP_031552623.1, WP_114280403.1, HBR51527.1, AKH70559.1, PHR31563.1, WP_108019640.1, MAZ87931.1, WP_066110810.1, WP_087652341.1, WP_067733011.1, WP_137680558.1, WP_138400860.1, OYZ77835.1, WP_033184387.1, SSC70240.1, WP_064832767.1, WP_112522133.1, WP_102761945.1, WP_071204881.1, WP_065662025.1, MAM37104.1, WP_138328097.1, WP_147076126.1, GEO97840.1, WP_027316147.1, WP_141855040.1, WP_015860162.1, WP_046194712.1, WP_132883904.1, BBC71262.1, WP_017500488.1, WP_109858633.1, WP_044585648.1, WP_068028174.1, WP_092914429.1, WP_090473014.1, WP_126652207.1, WP_116218477.1, MBD77336.1, PHR75222.1, WP_009807867.1, WP_150170293.1, PZO22609.1, WP_121164080.1, WP_056205236.1, WP_085856867.1, KOX56122.1, WP_116530112.1, WP_114947284.1, WP_113586875.1, WP_091862449.1,

Table A.B.S5 NucC proteins used for sequence alignment

Genbank	SDH65907.1, WP_089174987.1, WP_091880658.1, WP_024342650.1, WP_130030035.1, WP_147221866.1, WP_039289906.1, OJX19609.1, RYE07397.1, WP_023795328.1, WP_025429978.1, AZO59673.1, WP_064990655.1, AXA43707.1, WP_082849659.1, KQY50733.1, WP_082546164.1, WP_081159716.1, WP_136508923.1, WP_121692443.1, WP_021585828.1, WP_093002118.1, WP_088202460.1, WP_057214036.1, EHS52301.1, WP_113381622.1, WP_080579622.1, KQW42171.1, WP_082555673.1, WP_059978167.1, WP_018309834.1, WP_008343067.1, WP_054532754.1, WP_084413134.1, WP_013214703.1, WP_085163840.1, WP_116930240.1, WP_022951181.1, WP_077409906.1, RYG99173.1, WP_029934543.1, WP_013569417.1, WP_091303795.1, WP_149140697.1, WP_028881082.1, THD31264.1, WP_127670994.1, WP_130854256.1, WP_090198348.1, TAN48999.1, WP_015279331.1, WP_146388053.1, WP_145762146.1, WP_081268939.1, WP_039587461.1, WP_065628717.1, WP_115579840.1, WP_046984188.1, WP_078997919.1, WP_066608956.1, WP_150411994.1, WP_041043285.1, WP_130160209.1, WP_020818215.1, WP_037484142.1, WP_116086422.1, RYE92681.1, WP_135195579.1, WP_096225140.1, OYU37529.1, WP_101754818.1, WP_064697328.1, WP_130831990.1, WP_132526865.1, WP_034211676.1, WP_076024803.1, WP_071170739.1, WP_085602491.1, WP_148499653.1, WP_044616068.1, TIM26514.1, STN25082.1, WP_068024502.1, WP_110389129.1, WP_077298232.1, OUM83853.1, WP_076243885.1, WP_127451361.1, WP_056832935.1, WP_150459777.1, WP_120464927.1, WP_090647638.1, WP_059040663.1, WP_093373009.1, WP_069328609.1, WP_143033195.1, WP_053477638.1, WP_069195342.1, WP_034601308.1, WP_040037072.1, WP_096686516.1, OFY43238.1, WP_070865900.1, WP_070613181.1, WP_070538717.1, WP_097115301.1, WP_115006206.1, WP_070826115.1, WP_128581797.1, WP_111279514.1, WP_003663249.1, WP_050837778.1, WP_115323103.1, WP_013745861.1, WP_148323777.1, WP_148327772.1, WP_147489074.1, WP_072281338.1, WP_107813391.1, WP_003776236.1, WP_003742867.1, WP_049228838.1, WP_049357872.1, WP_094490573.1, KRU22285.1, WP_136438553.1, EEI92810.1, WP_095920276.1, WP_076503085.1, WP_062672312.1, WP_047433492.1, HCY42412.1, PKP47452.1, HCC51546.1, WP_009642756.1, WP_018612328.1, WP_013687740.1, OIQ21200.1, WP_139066763.1, MAP53886.1, WP_127136010.1, SHK73382.1, WP_078796289.1, WP_059323545.1, WP_089732682.1, KUJ54641.1, WP_036243807.1, WP_007395196.1, POR18699.1, WP_144214993.1, WP_086896082.1, WP_110305702.1, WP_089478536.1, WP_125011916.1, WP_111394778.1, WP_114006080.1, WP_095704069.1, RZJ91194.1, WP_114758269.1, WP_047433805.1, WP_089053279.1, MAC94214.1, WP_134203216.1, SJL33907.1, WP_065319358.1, RYY87994.1, WP_006799226.1, EXY19392.1, HBR21520.1, WP_140997585.1, AWM14340.1, WP_035558331.1, WP_070729753.1, WP_074649703.1, WP_097777949.1, WP_095175828.1, WP_104805224.1, WP_092966780.1, WP_126792910.1, WP_094552149.1, WP_145214961.1, WP_148074277.1, WP_106199822.1, WP_118712708.1, WP_074926419.1, PWL74508.1, WP_115683850.1, WP_144192059.1, WP_133667404.1, PIQ24505.1, EKD88011.1, WP_102242633.1, OGK64504.1, WP_145126747.1, AQA06654.1,

Table A.B.S5 NucC proteins used for sequence alignment

Genbank	WP_083739558.1, WP_117429583.1, WP_099289419.1, WP_138746413.1, WP_133058134.1, WP_079630835.1, TDR68883.1, WP_133759739.1, WP_145796011.1, WP_143025364.1, WP_116104466.1, WP_013884197.1, WP_110107914.1, WP_144690991.1, KAB1645257.1, WP_052682315.1, WP_144827708.1, WP_052112058.1, WP_123574903.1, WP_073969933.1, WP_056343452.1, WP_051629117.1, WP_068200280.1, WP_143015727.1, WP_076709335.1, WP_005054591.1, MAL27734.1, MAY50569.1, WP_055935121.1, WP_127012855.1, WP_091480116.1, WP_137159429.1, WP_104126704.1, WP_127435240.1, WP_123294490.1, WP_056755513.1, WP_094656857.1, WP_064078136.1, WP_139623557.1, WP_141805374.1, WP_084125985.1, WP_123813513.1, WP_119283361.1, WP_132165409.1, MBS42855.1, AZP70722.1, WP_078469341.1, WP_108237914.1, WP_059400005.1, WP_058170113.1, AXA56675.1, WP_122439601.1, WP_081235227.1, WP_103387598.1, WP_131184733.1, WP_058142645.1, WP_018229295.1, WP_119304441.1, SDR44551.1, WP_129134627.1, WP_052575195.1, WP_140885097.1, WP_124605499.1, WP_011486012.1, WP_018988632.1, WP_071615533.1, WP_118909630.1, WP_150618621.1, WP_102503002.1, WP_140063057.1, WP_148553794.1, MAF58692.1, WP_063457668.1, WP_115016765.1, WP_048387300.1, WP_140311641.1, WP_025627806.1, WP_050482130.1, WP_109848362.1, WP_107799270.1, OYY21385.1, WP_144307966.1, MBI42370.1, OGP49126.1, CBX31253.1, WP_096783435.1, TNC82801.1, HAF00665.1, WP_151723430.1, WP_015841417.1, WP_021014788.1, WP_038903548.1, MAT73457.1, WP_010583109.1, WP_024336406.1, TDU69446.1, PWU09498.1, TRZ49248.1, MAY36593.1, WP_012885876.1, ELS02931.1, WP_040897781.1, WP_144863505.1, WP_013659859.1, WP_041330525.1, HBH72950.1, WP_064665190.1, OGA35668.1, OFV91899.1, WP_139308183.1, KQY06463.1, WP_141215942.1, WP_143546124.1, PYX72524.1, WP_131984287.1, HCH74539.1, WP_016178418.1, EAG0986839.1, WP_044955721.1, WP_018597323.1, WP_050698056.1, KFI52303.1, WP_078686534.1, ORV42121.1, WP_139825224.1, TEX45487.1, HAL29488.1, WP_114594477.1
IMG	647193063, 2632356247, 2831805761, 2779856863, 2595103576, 2513842568, 2526117325, 2507444198, 2837149145, 2792187373, 2806389584, 2653800354, 2510777369, 2808182764, 2667999241, 2789927606

(1) Databases include Genbank (https://www.ncbi.nlm.nih.gov) and IMG (https://img.jgi.doe.gov)
(2) See Shah et al., "Comprehensive search for accessory proteins encoded with archaeal and bacterial type III CRISPR-cas gene cassettes reveals 39 new cas gene families" (2019) *RNA Biology* **16**(4):530-542

References

Dunten, P.W., Little, E.J., Gregory, M.T., Manohar, V.M., Dalton, M., Hough, D., Bitinaite, J., Horton, N.C., 2008. The structure of SgrAI bound to DNA; recognition of an 8 base pair target. Nucleic Acids Research 36, 5405–5416. https://doi.org/10.1093/nar/gkn510

Krissinel, E., Henrick, K., 2007. Inference of Macromolecular Assemblies from Crystalline State. Journal of Molecular Biology 372, 774–797. https://doi.org/10.1016/j.jmb.2007.05.022

Whiteley, A.T., Eaglesham, J.B., de Oliveira Mann, C.C., Morehouse, B.R., Lowey, B., Nieminen, E.A., Danilchanka, O., King, D.S., Lee, A.S.Y., Mekalanos, J.J., Kranzusch, P.J., 2019. Bacterial cGAS-like enzymes synthesize diverse nucleotide signals. Nature 567, 194–199. https://doi.org/10.1038/s41586-019-0953-5

APPENDIX C: Supplemental figures and tables for Chapter 3 A conserved signaling pathway activates bacterial CBASS immune signaling in response to DNA damage

Figure A.C.S40 CBASS expression reporter systems

(A) Full western blot for GFP reporter assay. RNAP: RNA Polymerase loading control.

(B) Full western blot for FLAG-NucC expression reporter assay. RNAP: RNA Polymerase loading control.

(C) Quantitative plaque assay showing infectivity of λ cI- against cells containing no CBASS system (EV: empty vector), the wild-type *E. coli* MS115-1 CBASS system (WT), or mutant systems with catalytic-dead CdnC (D72N/D74N) or NucC (D73N). Data is shown as average and standard deviation of plaque forming units per mL of purified phage (PFU/mL), from triplicate experiments. For NucC D73N, all three plates showed equal plaques (52 plaques at the tested dilution).

(D) Size of phage plaques for λ cI- infecting cells containing no CBASS system (EV: empty vector), the wild-type *E. coli* MS115-1 CBASS system (WT), or mutant systems with catalytic-dead CdnC (D72N/D74N) or NucC (D73N).

(E) qRT-PCR for *cdnC* using JP313 cells containing the *E. coli* MS115-1 CBASS operon on a plasmid under exponential growth conditions (black circles), after phage λ infection at MOI 10 (green sqares), or after addition of zeocin at 100 μg/mL (orange triangles).

(F) Minimum inhibitory concentration (MIC) analysis of JP313 cells with either empty vector (black circles), E. coli MS115-1 CBASS (green squares), or a mutant CBASS system with catalytic-dead NucC (D73N; orange triangles). Data shown are the average and standard deviation of three measurements (see **Methods**).

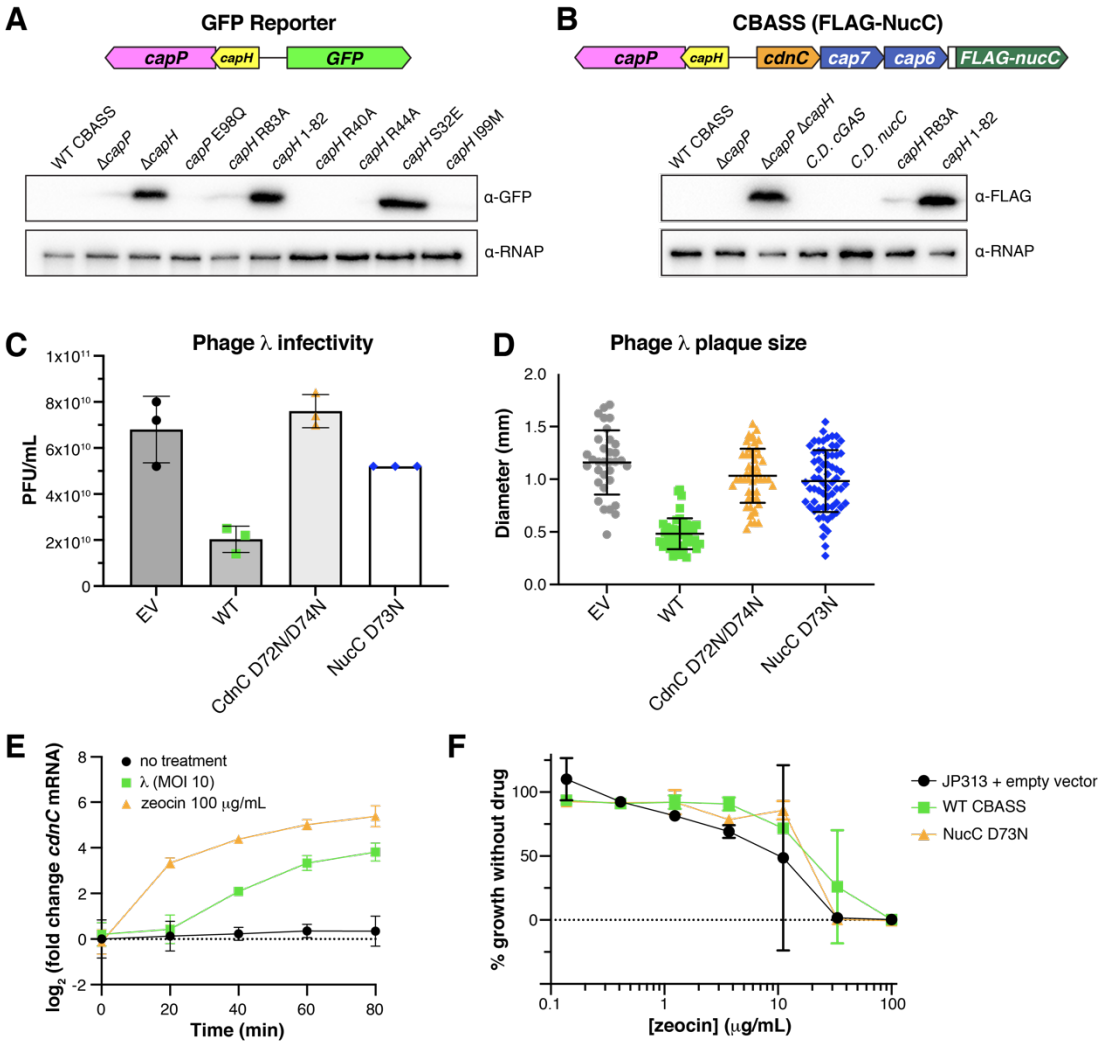

A
GFP Reporter
capP
capH
GFP
WT CBASS
ΔcapP
ΔcapH
capP E98Q
capH R83A
capH 1-82
capH R40A
capH R44A
capH S32E
capH I99M
α-GFP
α-RNAP
B
CBASS (FLAG-NucC)
capP
capH
cdnC
cap7
cap6
FLAG-nucC
WT CBASS
ΔcapP
ΔcapP ΔcapH
C.D. cGAS
C.D. nucC
capH R83A
capH 1-82
α-FLAG
α-RNAP
C
Phage λ infectivity
PFU/mL
1x10¹¹
8x10¹⁰
6x10¹⁰
4x10¹⁰
2x10¹⁰
0
EV
WT
CdnC D72N/D74N
NucC D73N
D
Phage λ plaque size
Diameter (mm)
2.0
1.5
1.0
0.5
0.0
EV
WT
CdnC D72N/D74N
NucC D73N
E
log2 (fold change cdnC mRNA)
no treatment
λ (MOI 10)
zeocin 100 μg/mL
8
6
4
2
0
-2
0
20
40
60
80
Time (min)
F
% growth without drug
100
50
0
0.1
1
10
100
[zeocin] (μg/mL)
JP313 + empty vector
WT CBASS
NucC D73N

Figure A.C.S41 Phage infection of cells with E. coli MS115-1 CBASS

(A) Live-cell fluorescence microscopy of λ cI- infecting cells containing no CBASS system (EV: empty vector), the wild-type *E. coli* MS115-1 CBASS system (WT), or a mutant system lacking *capP* (Δ*capP*). DNA (DAPI) is colored cyan, and cell membranes (FM4-64) are colored magenta.

(B) Quantification of fluorescence microscopy of λ cI- infecting cells containing no CBASS system (EV: empty vector), the wild-type *E. coli* MS115-1 CBASS system (WT), or a mutant system lacking *capP* (Δ*capP).* Cells were manually quantified and assigned categories based on examples shown at top (n ≥ 30 cells for each strain and time point).

(C) Growth curves of JP313 cells containing plasmids with no insert (Empty vector; gray), the core four CBASS genes under a lac-inducible promoter (Core CBASS; blue) (Lau et al., 2020), the six-gene CBASS system (WT CBASS; green), or the six-gene CBASS system with capP deleted (CBASS Δ*capP*; orange). All strains grew comparably when not infected with phage (Uninfected, top) and showed differing levels of protection when infected with phage λ at different multiplicity of infection (MOI). Data shown is the average and standard deviation of 3 independent measurements.

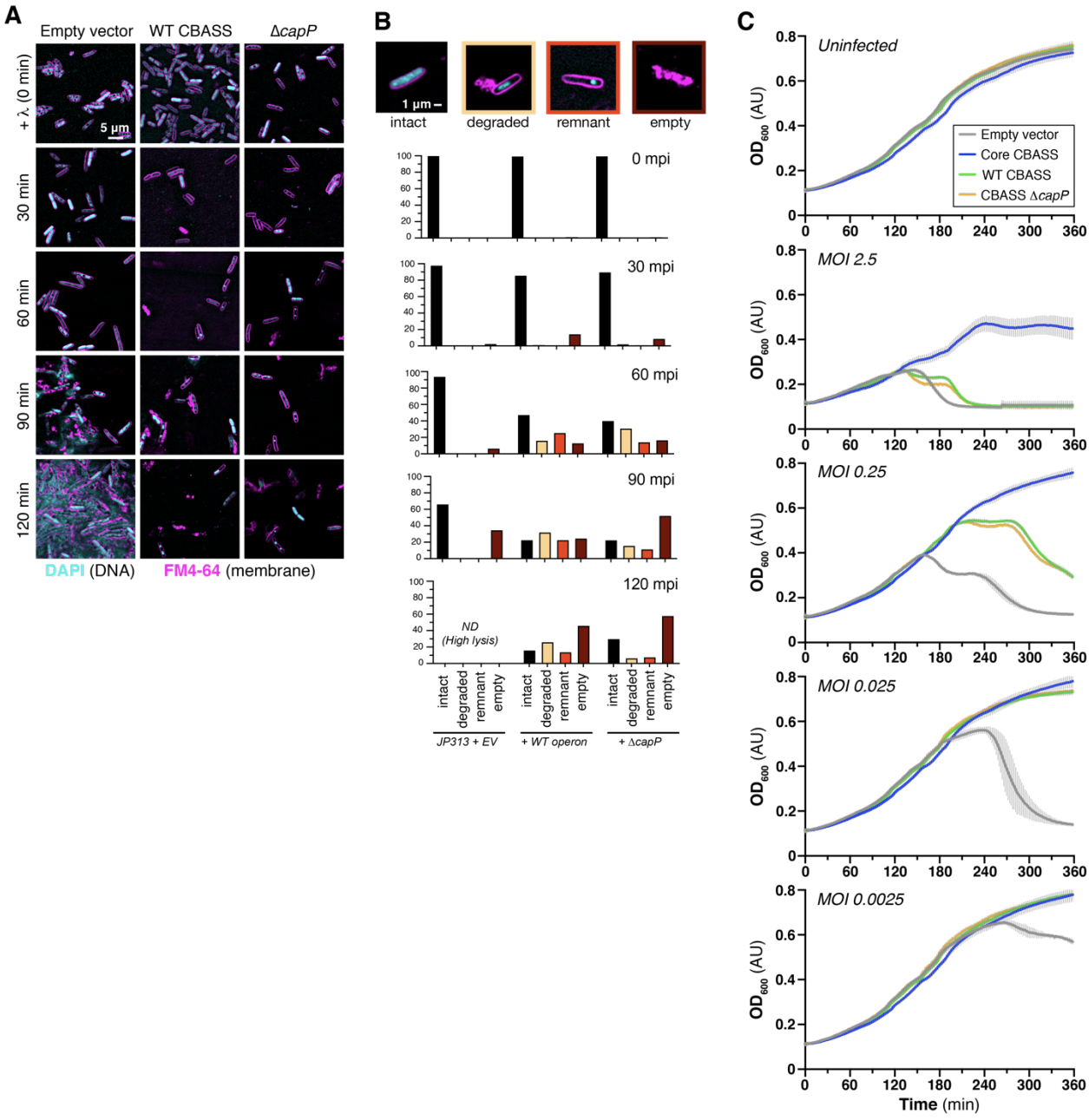

A
Empty vector
WT CBASS
ΔcapP
+ λ (0 min)
30 min
60 min
90 min
120 min
5 μm
DAPI (DNA)
FM4-64 (membrane)
B
1 μm
intact
degraded
remnant
empty
0 mpi
30 mpi
60 mpi
90 mpi
120 mpi
ND
(High lysis)
JP313 + EV
+ WT operon
+ ΔcapP
C
Uninfected
MOI 2.5
MOI 0.25
MOI 0.025
MOI 0.0025
Empty vector
Core CBASS
WT CBASS
CBASS ΔcapP
OD600 (AU)
Time (min)

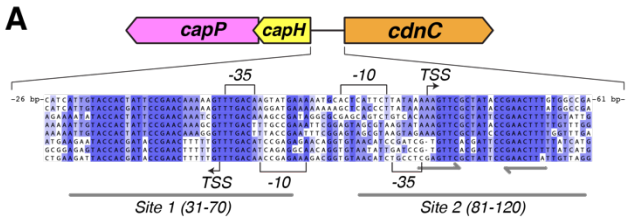
Figure A.C.S42 CapH binds MS115-1 CBASS intergenic region

(A) Sequence alignment of CBASS promoter regions in *E. coli* MS115-1 (NCBI RefSeq GG771785.1), *Cronobacter sakazakii* strain cro3915C2 (NZ_NRJY01000012.1), *Pseudomonas stutzeri* strain KC NODE_1_length_951488_cov_16.453 (NZ_POUN01000001.1), *Pseudomonas* sp. RIT 412 RIT412_S3_7 (NZ_QBJA02000007.1), *Pseudomonas* sp. MF4836 (NZ_MVOL01000002.1), *Burkholderia pseudomallei* strain MSHR4301 (NZ_LXCN01000015.1), *Ralstonia insidiosa* strain WCHRI065162 (NZ_PKPC01000011.1), *Thauera* sp. K11 plasmid pTX1 (NZ_CP023440.1). Promoter sequences (-35, -10, and TSS) were predicted by the BPROM server (Salamov and Solovyevand, 2011). Identified CapH-binding Sites 1 and 2 are denoted by gray underlines, and the palindrome within Site 2 is denoted by gray arrows.

(B) DNA binding affinities (K_d) and cooperativity (Hill coefficient) from fluorescence polarization measurements of CapH binding overlapping 40-41 bp DNAs spanning the *E. coli* MS115-1 CBAS promoter region (**Fig 3.2A**).

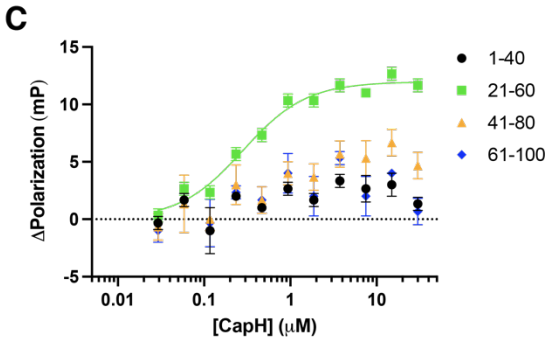
(C-F) Fluorescence polarization assays showing binding of *E. coli* MS115-1 CapH (His_6-MBP tagged) to 40-41 bp DNAs spanning the *E. coli* MS115-1 CBAS promoter region. Fit K_d and Hill coefficient for each DNA is shown in panel B.

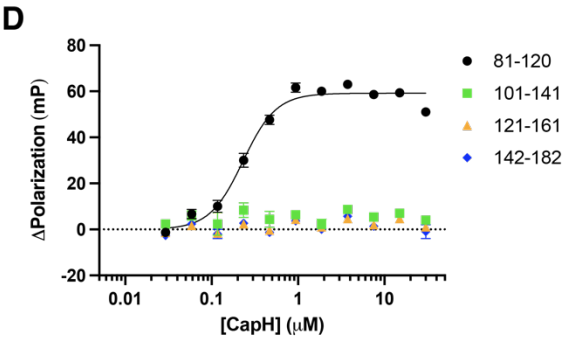
(G) Schematic of the CapH NTD with sequence logo indicating the conservation of each residue in an alignment of 56 unique CBASS-associated CapH proteins.

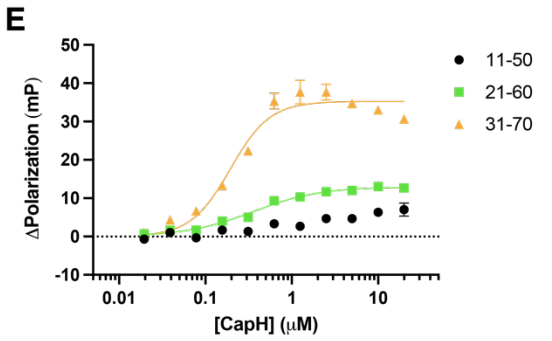


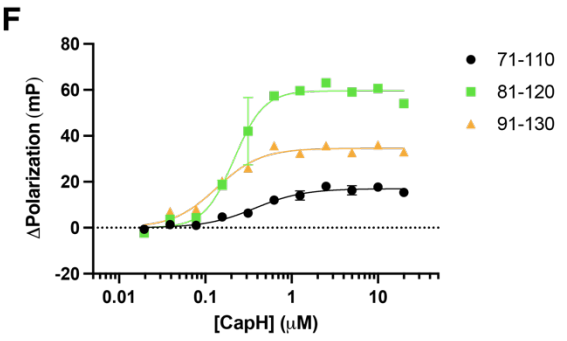
B

DNA	K_d	Hill coeff.
1-40	N/A	N/A
11-50	N/A	N/A
21-60	0.36 +/- 0.06	1.1 +/- 0.2
31-70	0.19 +/- 0.03	1.9 +/- 0.5
41-80	N/A	N/A
61-100	N/A	N/A
71-110	0.37 +/- 0.08	1.5 +/- 0.2
81-120	0.22 +/- 0.02	2.5 +/- 0.6
91-130	0.13 +/- 0.02	1.7 +/- 0.3
101-141	N/A	N/A
121-161	N/A	N/A
142-182	N/A	N/A









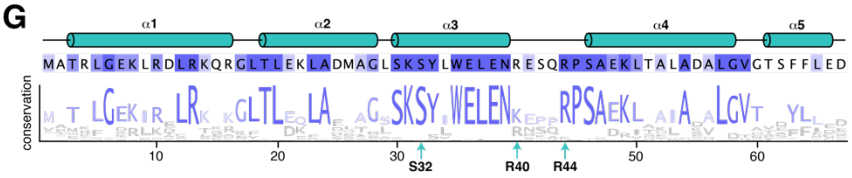


Figure A.C.S43 CBASS-associated CapP contains an internal cystein switch

(A) Operon schematic of the *Thauera* sp. K11 CBASS system, compared to the *E. coli* MS115-1 system, with sequence identity between the two systems' CapH and CapP proteins indicated.

(B) Structure of *Thauera* sp. K11 CapP, with closeup of its Zn^{2+} metallopeptidase domain (pink) with internal cysteine switch loop (blue) and cysteine switch residue (C113).

(C) Structure of human matrix metalloprotease MMP9 (PDB IF 1L6J; (Elkins et al., 2002)), with closeup of its Zn^{2+} metallopeptidase domain (pink) and N-terminal cysteine switch domain (blue) and cysteine switch residue (C99). The orientation of the Zn2+ metallopeptidase domains in panels B and C is identical.

(D) Evolutionary tree of 408 CBASS-associated CapP proteins, colored by the presence (black) or absence (blue) of the internal cysteine switch.

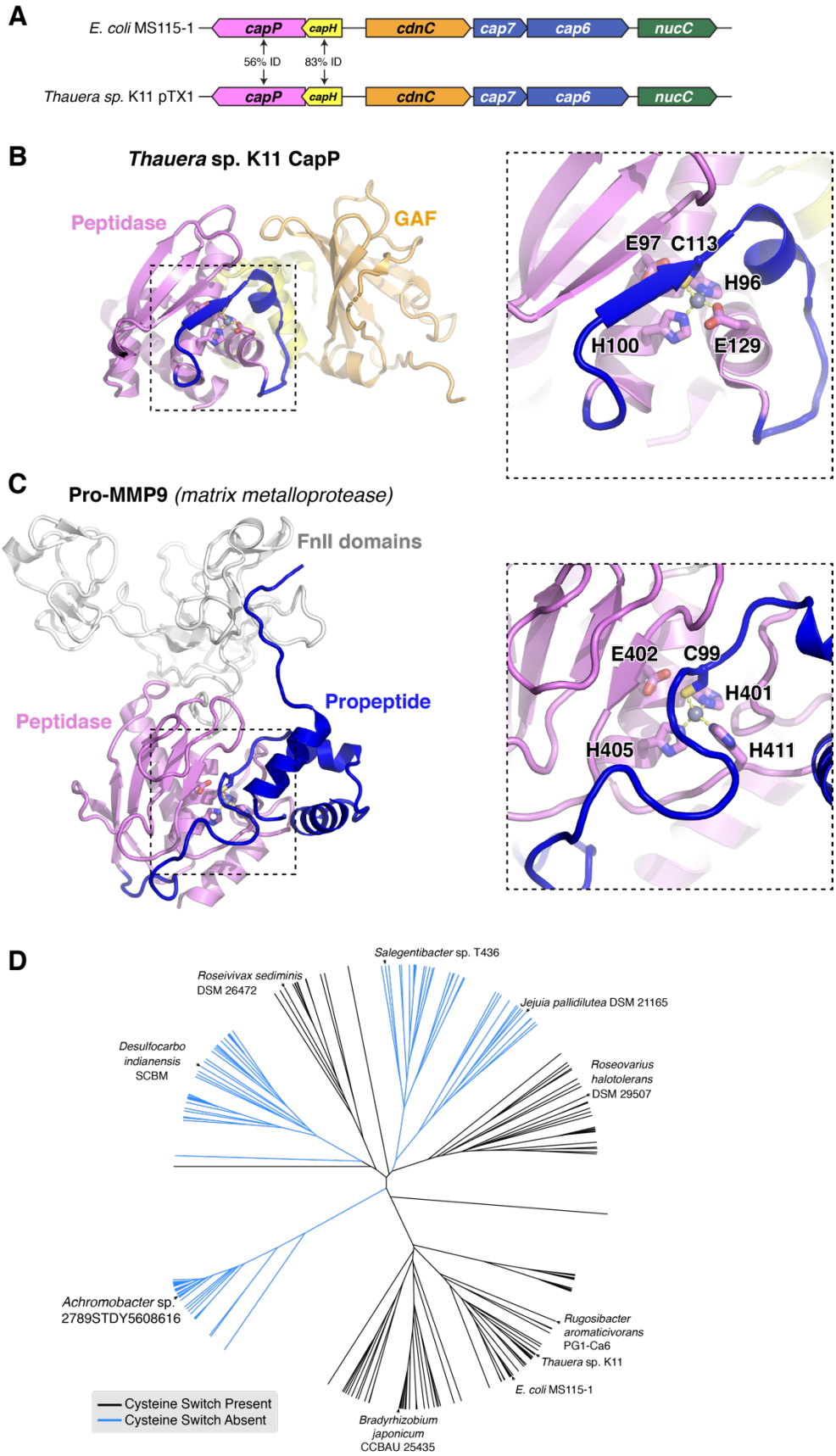

A
E. coli MS115-1
capP
capH
cdnC
cap7
cap6
nucC
56% ID
83% ID
Thauera sp. K11 pTX1
B
Thauera sp. K11 CapP
Peptidase
GAF
E97
C113
H96
H100
E129
C
Pro-MMP9 (matrix metalloprotease)
FnII domains
Propeptide
Peptidase
E402
C99
H401
H405
H411
D
Salegentibacter sp. T436
Roseivivax sediminis DSM 26472
Jejuia pallidilutea DSM 21165
Desulfocarbo indianensis SCBM
Roseovarius halotolerans DSM 29507
Achromobacter sp. 2789STDY5608616
Rugosibacter aromaticivorans PG1-Ca6
Thauera sp. K11
E. coli MS115-1
Bradyrhizobium japonicum CCBAU 25435
Cysteine Switch Present
Cysteine Switch Absent

Figure A.C.S44 DNA-mediated activation of CapP

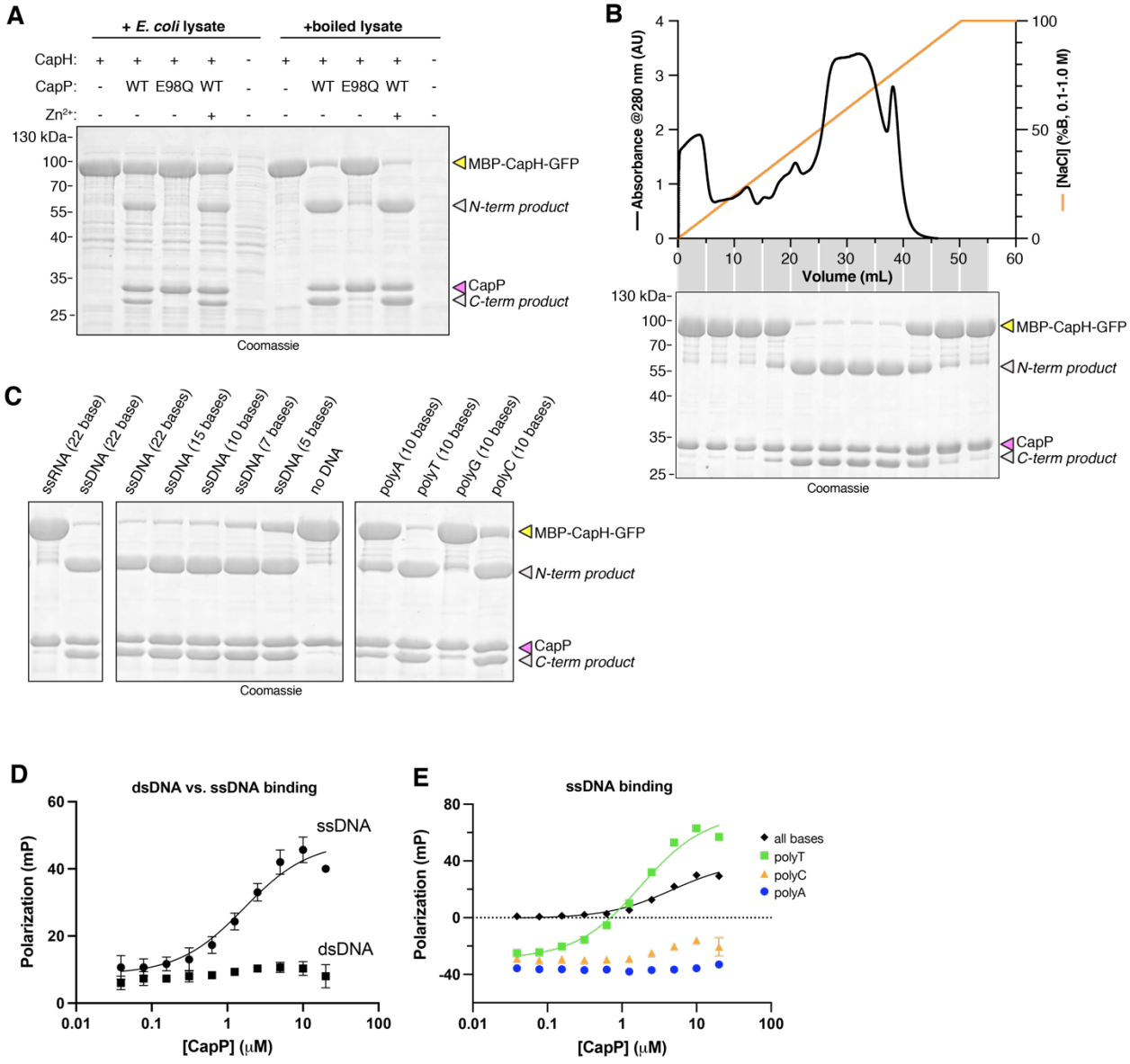
(A) In vitro cleavage assay with purified *E. coli* MS115-1 CapP (wild-type or catalytic-dead E98Q mutant), MBP-CapH-GFP, and *E. coli* cell lysate.

(B) *Top:* UV absorbance (280 nm) of boiled *E. coli* cell lysate separated by HiTrap Q column. Gradient (orange line) goes from 100 mM NaCl (0% B) to 1 M NaCl (100% B). *Bottom:* Coomassie-stained SDS-PAGE gel of cleavage assay with purified *E. coli* MS115-1 CapP, MBP-CapH-GFP, and combined fractions from HiTrap Q column. A similar assay with a Superdex 75 gel filtration column showed high activity across the entire separation range (not shown).

(C) In vitro cleavage assay with purified *E. coli* MS115-1 CapP, MBP-CapH-GFP, and the indicated nucleic acids.

(D) Fluorescence polarization DNA binding assay for *E. coli* MS115-1 CapP and either single-stranded DNA (circles and solid line; K_d=1.7 +/- 0.6 μM) or double-stranded DNA (squares; no binding detected).

(E) Fluorescence polarization DNA binding assay for *E. coli* MS115-1 CapP and single-stranded DNAs including a random sequence with all four bases (black diamonds; K_d=4.7 +/- 1.2 μM), poly-T (green squares; K_d=1.8 +/- 0.3 μM), poly-C (orange triangles; no binding detected), or poly-A (blue circles; no binding detected).


A
+ E. coli lysate
+boiled lysate
CapH:
CapP:
Zn2+:
WT E98Q WT
130 kDa
100
70
55
40
35
25
MBP-CapH-GFP
N-term product
CapP
C-term product
Coomassie
B
Absorbance @280 nm (AU)
[NaCl] (%B, 0.1-1.0 M)
Volume (mL)
130 kDa
MBP-CapH-GFP
N-term product
CapP
C-term product
Coomassie
C
ssRNA (22 base)
ssDNA (22 base)
ssDNA (22 bases)
ssDNA (15 bases)
ssDNA (10 bases)
ssDNA (7 bases)
ssDNA (5 bases)
no DNA
polyA (10 bases)
polyT (10 bases)
polyG (10 bases)
polyC (10 bases)
MBP-CapH-GFP
N-term product
CapP
C-term product
Coomassie
D
dsDNA vs. ssDNA binding
Polarization (mP)
ssDNA
dsDNA
[CapP] (μM)
E
ssDNA binding
Polarization (mP)
all bases
polyT
polyC
polyA
[CapP] (μM)

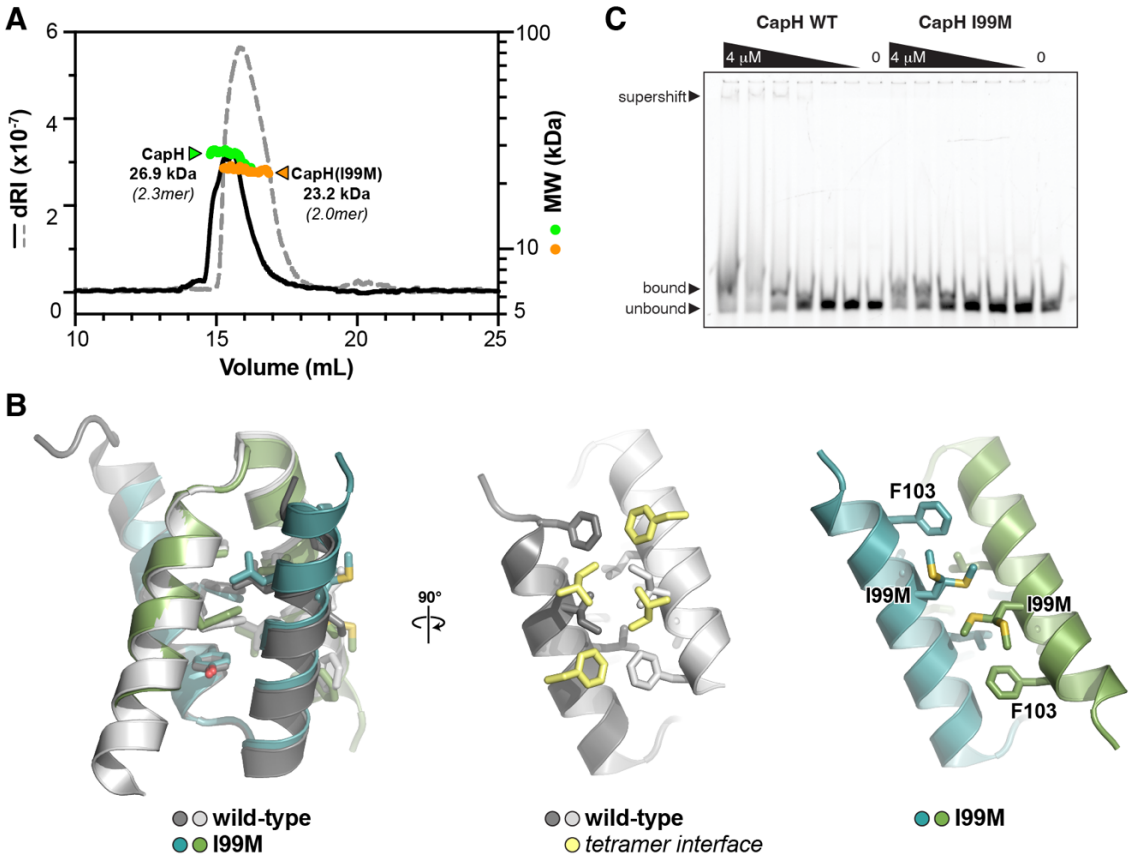


Figure A.C.S45 The CapH I99M mutation disrupts tetramer formation

(A) *Left:* Cartoon view of the *E. coli* MS115-1 CapH^{CTD} (I99M) structure, with the two protomers colored blue and green. Overlaid is the structure of wild-type CapH^{CTD} (dark gray/light gray). *Center:* Views of the wild-type CapH^{CTD} tetramerization interface, with the dimer shown in dark gray/light gray and interacting residues from the opposite dimer in light yellow. Right: View of the CapH^{CTD} (I99M) tetramerization interface, showing the position of the I99M mutation that disrupts tetramer formation. In the electron density maps, the M99 residues showed partial occupancy for two rotamers, and were modeled with each rotamer at 50% occupancy (both shown as sticks).

(B) SEC-MALS analysis of MBP-fused CapH (I99M) mutant (monomer molecular weight 57.0 kDa), showing that it forms a homodimer.

(C) Fluorescence polarization DNA binding assay showing that the CapH I99M mutation reduces, but does not eliminate, the DNA binding ability of CapH (all proteins used were His_6-MBP tagged).

(D) Electrophoretic mobility shift of 5'-FAM labeled DNA with a sequence comprised of residues 81-120 of the intergenic region of MS115-1 CBASS after incubation with either wild-type or I99M CapH. Protein concentrations are 4 uM and two-fold dilutions thereof.

(E) Size exclusion chromatography coupled to multiangle light scattering chromatogram of CapH and CapH I99M proteins without tags.

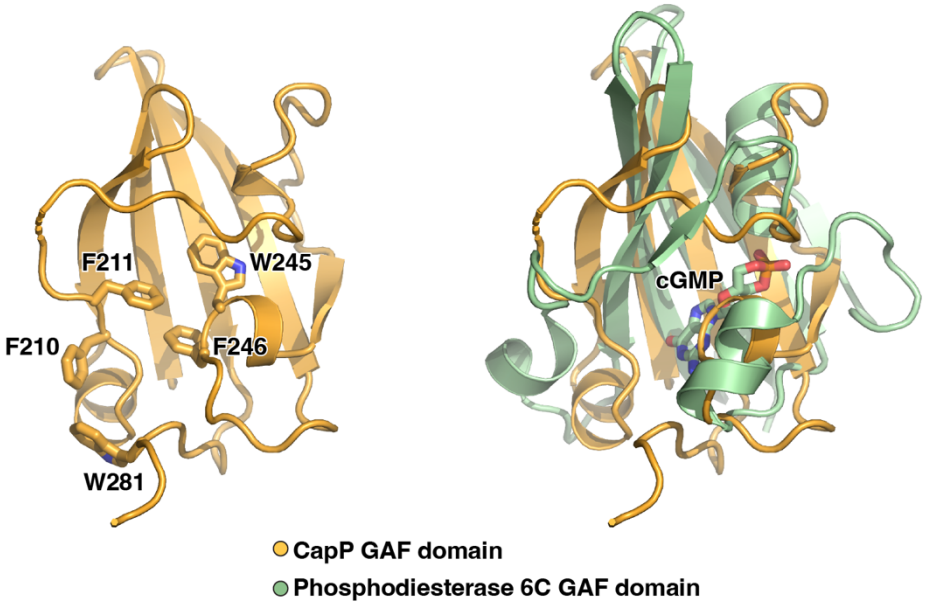


Figure A.C.S46 Structure of the CapP GAF domain

Structural overlay between the *Thauera* sp. K11 CapP GAF domain (orange) and the GAF domain of phosphodiesterase 6C bound to cyclic GMP (green; PDB ID 3DBA; (Martinez et al., 2008)). Surface-exposed aromatic residues potentially involved in CapP-ligand binding are shown as sticks in the left panel.

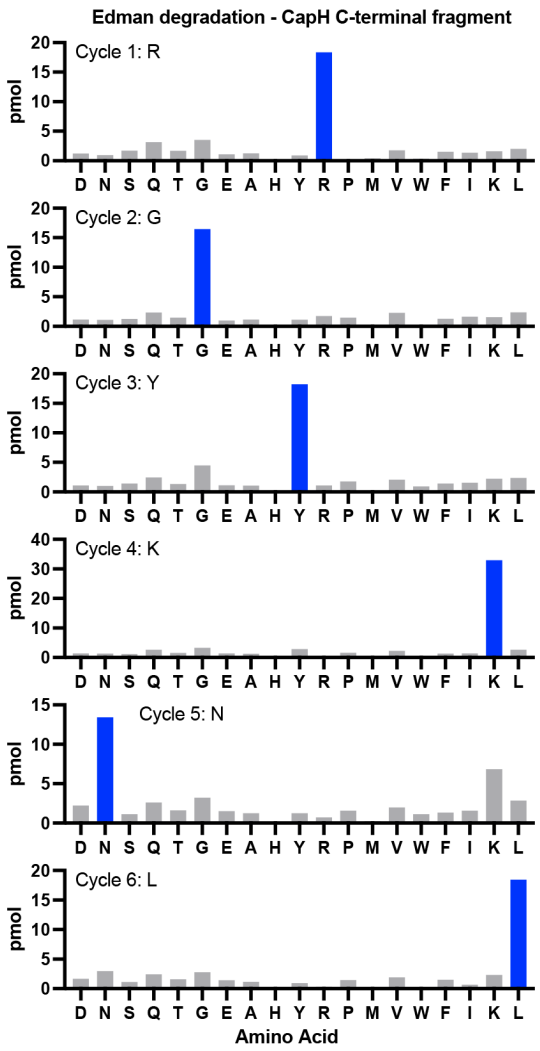


Figure A.C.S47 Edman degradation of CapP cleavage product

Bar graphs showing picomoles (pmol) of each amino acid detected by Edman degradation of the C-terminal cleavage product of MBP-CapH-GFP (denoted by red asterisk in **Figure 3.5B**). The dominant amino acid for each cycle is shown as a blue bar. The inferred N-terminal sequence (RGYKNL) matches CapH residues 83-88 (see **Figure 3.5C**).

Table A.C.S10 CapH and CapP regulators in CBASS systems

CD-NTase (IMG Accession #)	capH (IMG Acc#)	capP (IMG Acc#)	Cap cysteine switch present?	CBASS type	Effector type	Notes
2507506249	2507506251	2507506252	Yes	Type II short	phospholipase	*Sigma-70 present*
2665779205	2665779203	2665779202	Yes	Type II short	phospholipase	*Sigma-70 present*
2558280643	2558280645	2558280644		Type II short	endonuclease	
2738940440	2738940439	2738940438	Yes	Type II short	endonuclease	
2516899368	2516899369	2516899370	Yes	Type II short	endonuclease	
2600377306	2600377307	2600377308	Yes	Type II short	endonuclease	
2585553962	2585553961	2585553960	Yes	Type II short	endonuclease	
2585563704	2585563703	2585563702	Yes	Type II short	endonuclease	
2756163455	2756163454	2756163453	Yes	Type II short	endonuclease	
2690689082	2690689081	2690689080	Yes	Type II short	phospholipase	
2712152206	2712152203	2712152202	Yes	Type II short	phospholipase	
2509145587	2509145589	2509145590		Type II short	phospholipase	
2644290320	2644290317	2644290316		Type II short	phospholipase	
2524877765	2524877762	2524877761		Type II short	phospholipase	
2524870332	2524870335	2524870336		Type II short	phospholipase	
2686558267	2686558270	2686558271		Type II short	phospholipase	
2611863577	2611863578	2611863579		Type II short	phospholipase	
2585409678	2585409677	2585409676		Type II short	phospholipase	

Table A.C.S1 CapH and CapP regulators in CBASS

2633634880	2633634882	2633634883	Yes	Type II short	2TM	
2558203594	2558203596	2558203597	Yes	Type II short	2TM	
2509383948	2509383952	2509383953		Type II short	2TM	
2710633767	2710633770	2710633771		Type II short	2TM	
2668215717	2668215722	2668215723		Type II short	2TM	*Sigma-70 present*
2623290238	2623290233	2623290232		Type II short	2TM	
2690431750	2690431754	2690431755		Type II short	2TM	
2590612567	2590612563	2590612562		Type II short	2TM	*Sigma-70 present*
2667962606	2667962602	2667962601		Type II short	2TM	
2729850490	2729850486	2729850485		Type II short	2TM	
2595015100	2595015097	2595015096	Yes	Type II	endonuclease	*Shown in Fig. 1A*
2655195476	2655195474	2655195473	Yes	Type II	endonuclease	
2644292418	2644292416	2644292415	Yes	Type II	endonuclease	
2519814111	2519814113	2519814114		Type II	endonuclease	
2689753568	2689753570	2689753571	Yes	Type II	endonuclease	
2690477417	2690477415	2690477414	Yes	Type II	endonuclease	
2624310701	2624310703	2624310704	Yes	Type II	endonuclease	
2513139536	2513139533	2513139532		Type II short	endonuclease	
2657176113	2657176116	2657176117		Type II short	endonuclease	

Table A.C.S1 CapH and CapP regulators in CBASS

2513127817	2513127814	2513127813		Type II short	endonuclease	
2621283147	2621283144	2621283143		Type II short	endonuclease	
2596295743	2596295740	2596295739		Type II short	endonuclease	*Sigma-70 present*
2719648670	2719648674	2719648675		Type II short	phospholipase	*Sigma-70 present*
2623266198	2623266203	2623266204		Type II short	phospholipase	
2729910573	2729910570	2729910569		Type II short	phospholipase	
2744676524	2744676526	2744676527	Yes	Type I	phospholipase	*Shown in Fig. 1A*
2641367376	-	2641367378		Type II short	phospholipase	*CapH not annotated*
2617838939	-	2617838941		Type II short	phospholipase	*CapH-CapP fusion*
2701402228	-	2701402225		Type II short	phospholipase	*CapH-CapP fusion*
2639679470	-	2639679468		Type II short	phospholipase	*CapH-CapP fusion*
2739332132	-	2739332134		Type II short	phospholipase	*CapH-CapP fusion*
2513842423	2513842419	2513842418	Yes	Type II short	endonuclease	
2752845634	2752845632	2752845631		Type II short	phospholipase	
2730460220	2730460222	2730460223	Yes	Type II short	endonuclease	
2596846691	2596846689	2596846688		Type II short	endonuclease	*Minimal CapP*
2513651720	2513651718	2513651717		Type II short	endonuclease	*Minimal CapP*
2669546849	2669546851	2669546852		Type II short	endonuclease	*Minimal CapP*
2669696928	2669696926	2669696925		Type II short	endonuclease	*Minimal CapP*

Table A.C.S1 CapH and CapP regulators in CBASS

2721434241	2721434243	2721434244		Type II short	endonuclease	*Minimal CapP*
2647284086	2647284088	2647284089		Type II short	endonuclease	*Minimal CapP*
2683167510	2683167508	2683167507		Type II short	endonuclease	*Minimal CapP*
2694075569	2694075571	2694075572		Type II short	endonuclease	*Minimal CapP*
2596982008	2596982010	2596982011		Type II short	endonuclease	*Minimal CapP*
2702579069	2702579071	2702579072		Type II short	endonuclease	*Minimal CapP*
2645931479	2645931481	2645931482		Type II short	endonuclease	*Minimal CapP*
2719791480	2719791482	2719791483		Type II short	endonuclease	*Minimal CapP*
2513669909	2513669907	2513669906		Type II short	endonuclease	*Minimal CapP*
2513946128	2513946130	2513946131		Type II short	endonuclease	*Minimal CapP*
2655482197	2655482199	2655482200		Type II short	endonuclease	*Minimal CapP*
2688586993	2688586995	2688586996		Type II short	endonuclease	*Minimal CapP*
2696273992	2696273990	2696273989		Type II short	endonuclease	*Minimal CapP*
2513879207	2513879205	2513879204		Type II short	endonuclease	*Minimal CapP*
2517889406	2517889408	2517889409		Type II short	endonuclease	*Minimal CapP*
2665775671	2665775669	2665775668		Type II short	endonuclease	*Minimal CapP*
2683158154	2683158152	2683158151		Type II short	endonuclease	*Minimal CapP*
2517106502	2517106500	2517106499		Type II short	endonuclease	*Minimal CapP*
2509373550	2509373552	2509373553		Type II short	endonuclease	*Minimal CapP*

Table A.C.S1 CapH and CapP regulators in CBASS

2645136499	2645136501	2645136502	Yes	Type II short	phosph olipase	*Sigma-70 present*
2675846562	2675846559	2675846558	Yes	Type II short	phosph olipase	
2685259843	2685259845	2685259846	Yes	Type II short	phosph olipase	
2674658948	2674658945	2674658944	Yes	Type II short	phosph olipase	
2678929850	2678929852	2678929853	Yes	Type II short	phosph olipase	
2619896652	2619896654	2619896655	Yes	Type II short	phosph olipase	
2702575790	2702575791	2702575792	Yes	Type II short	phosph olipase	*Shown in Fig. 1A*
2721435309	2721435310	2721435311	Yes	Type II short	phosph olipase	
2509154264	2509154263	2509154262	Yes	Type II short	phosph olipase	
2619129804	2619129805	2619129806	Yes	Type II short	phosph olipase	
2597440108	2597440109	2597440110	Yes	Type II short	phosph olipase	
2671917814	2671917813	2671917812	Yes	Type II short	phosph olipase	
2574522140	2574522139	2574522138	Yes	Standalone cyclase		
2654613574	2654613575	2654613576	Yes	Type II short	phosph olipase	
2597812841	2597812840	2597812839	Yes	Type II short	phosph olipase	
2515445369	2515445370	2515445371	Yes	Type II short	phosph olipase	
2644651528	2644651527	2644651526	Yes	Type II short	phosph olipase	
2644208277	2644208276	2644208275	Yes	Type II short	phosph olipase	
2739693679	2739693680	2739693681	Yes	Type II short	phosph olipase	

Table A.C.S1 CapH and CapP regulators in CBASS

2645151397	2645151398	2645151399	Yes	Type II short	phosph olipase	
2645594096	2645594095	2645594094	Yes	Type II short	phosph olipase	
2645603717	2645603716	2645603715	Yes	Type II short	phosph olipase	
2741196916	2741196917	2741196918	Yes	Type II short	phosph olipase	
2508005295	2508005296	2508005297	Yes	Type II short	phosph olipase	
2730980924	2730980925	2730980926	Yes	Type II short	phosph olipase	
2600194924	2600194921	2600194920	Yes	Type II short	phosph olipase	*Sigma-70 present*
2631785771	2631785768	2631785767	Yes	Type II short	phosph olipase	*Sigma-70 present*
2745475457	2745475459	2745475460	Yes	Type II short	phosph olipase	*Sigma-70 present*
2647117907	2647117906	2647117905	Yes	Type II short	phosph olipase	
2620456171	2620456172	2620456173	Yes	Type II short	phosph olipase	
2599099869	-	2599099868	Yes	Type II short	phosph olipase	*CapH not annotated*
2579752047	2579752048	2579752049	Yes	Type II short	phosph olipase	
2518559491	2518559492	2518559493		Type II short	phosph olipase	
2622920052	2622920053	2622920054		Type II short	phosph olipase	
2602595621	-	2602595620		Type II short	phosph olipase	*CapH not annotated*
2637055429	2637055428	2637055427		Type II short	phosph olipase	
2646929475	-	2646929476		Type II short	phosph olipase	*CapH not annotated*
2651090307	2651090306	2651090305		Type II short	phosph olipase	

Table A.C.S1 CapH and CapP regulators in CBASS

2682264007	2682264006	2682264005		Type II short	phosph olipase	
2656510512	2656510510	2656510509	Yes	Type II short	phosph olipase	
2723788314	2723788311	2723788310	Yes	Type II short	phosph olipase	*Sigma-70 present*
2512957821	2512957818	2512957817	Yes	Type II short	phosph olipase	*Sigma-70 present*
2688820243	2688820240	2688820239	Yes	Type II short	phosph olipase	*Sigma-70 present*
2714951169	2714951166	2714951165	Yes	Type II short	phosph olipase	*Sigma-70 present*
2687438464	2687438463	2687438462		Type II short	phosph olipase	
2635168335	2635168334	2635168333	Yes	Type II short	phosph olipase	
2660231444	2660231445	2660231446		Type II short	1TM	*CapP truncated*
2507213650	2507213651	2507213652	Yes	Type III (HORMA1)	endonu clease	
2652622416	2652622415	2652622414	Yes	Type III (HORMA1)	endonu clease	
2673404074	2673404073	2673404072	Yes	Type III (HORMA1)	genom e record truncat ed	
2660430062	2660430063	2660430064	Yes	Type III (HORMA1)	endonu clease	
2753898904	2753898906	2753898907	Yes	Type III (HORMA1)	endonu clease	
2699670736	2699670735	2699670734	Yes	Type III (HORMA1)	endonu clease	
2721244544	2721244546	2721244547	Yes	Type III (HORMA1)	endonu clease	
2743170184	2743170183	2743170182	Yes	Type III (HORMA1)	endonu clease	
2730133759	2730133761	2730133762	Yes	Type III (HORMA1)	endonu clease	

Table A.C.S1 CapH and CapP regulators in CBASS

2733819243	2733819242	2733819241	Yes	Type III (HORMA1)	endonuclease	
2744296961	2744296959	2744296958	Yes	Type III (HORMA1)	endonuclease	
2662639312	2662639314	2662639315	Yes	Type III (HORMA1)	endonuclease	*Identical to E. coli MS115-1 CBASS system; Shown in Fig. 1A*
2704038509	2704038508	2704038507	Yes	Type III (HORMA1)	endonuclease	
2699633080	2699633079	2699633078	Yes	Type III (HORMA1)	endonuclease	
2746146147	2746146146	2746146145	Yes	Type III (HORMA1)	endonuclease	
2714191489	2714191488	2714191487	Yes	Type III (HORMA1)	endonuclease	
2566040933	2566040934	2566040935	Yes	Type III (HORMA1)	endonuclease	
2695589792	2695589791	2695589790	Yes	Type III (HORMA1)	endonuclease	
2648087047	2648087048	2648087049	Yes	Type III (HORMA1)	endonuclease	
2691523049	2691523048	2691523047	Yes	Type III (HORMA1)	endonuclease	
2738620088	2738620089	2738620090	Yes	Type III (HORMA1)	endonuclease	
2752282199	2752282198	2752282197	Yes	Type III (HORMA1)	endonuclease	
2518945505	2518945506	2518945507	Yes	Type III (HORMA1)	endonuclease	
2669116360	2669116359	2669116358	Yes	Type III (HORMA1)	endonuclease	
2651800495	2651800496	2651800497	Yes	Type III (HORMA1)	endonuclease	
2508682644	2508682643	2508682642	Yes	Type III (HORMA1)	endonuclease	
2503167229	2503167230	2503167231	Yes	Type III (HORMA1)	endonuclease	

Table A.C.S1 CapH and CapP regulators in CBASS

2691906884	2691906885	2691906886	Yes	Type III (HORMA1)	endonuclease	
2508536762	2508536763	2508536764	Yes	Type III (HORMA1)	endonuclease	
2717844112	2717844113	2717844114		Type III (HORMA1)	endonuclease	
2748061513	2748061515	2748061516		Type III (HORMA1)	endonuclease (partial)	
2624401413	2624401414	2624401415		Type III (HORMA1)	endonuclease	
2513957181	-	2513957182		Type III (HORMA1)	not detected	*CapH not annotated*
2723993301	2723993300	2723993299		Type III (HORMA1)	endonuclease	
2528820621	2528820622	2528820623	Yes	Type III (HORMA1)	endonuclease	
2636886035	2636886036	2636886037		Type III (HORMA1)	endonuclease	
2726800886	2726800885	2726800884		Type III (HORMA1)	endonuclease	
2669843165	2669843166	2669843167		Type III (HORMA1)	endonuclease	
2726828623	2726828622	2726828621		Type III (HORMA1)	endonuclease	
2715118952	2715118953	2715118954		Type III (HORMA1)	endonuclease	
2754383797	2754383796	2754383795	Yes	Type III (HORMA1)	endonuclease	
2655270820	2655270819	2655270817		Type III (HORMA1)	endonuclease	
2727725939	2727725940	2727725941	Yes	Type III (HORMA1)	endonuclease	
2751157772	2751157769	2751157768		Type III (HORMA1)	endonuclease	
2690503223	2690503221	2690503220		Type III (HORMA1)	endonuclease	*Sigma-70 present*
2730470228	2730470231	2730470232		Type III (HORMA1)	endonuclease	*Sigma-70 present*

Table A.C.S1 CapH and CapP regulators in CBASS

2619783553	2619783554	2619783555		Type III (HORMA1)	genome record truncated	
2718698219	2718698220	2718698221		Type III (HORMA1)	endonuclease	
2676870477	-	2676870478		Type III (HORMA1)	endonuclease	*CapH-CapP fusion*
2643694093	2643694096	2643694097		Type III (HORMA1)	2TM	
2700132864	2700132867	2700132868		Type III (HORMA1)	2TM	
2529462631	2529462628	2529462627		Type III (HORMA1)	2TM	
2587746750	2587746753	2587746754		Type III (HORMA1)	2TM	
2684366124	2684366121	2684366120		Type III (HORMA1)	2TM	
2611140343	2611140345	2611140346		Type III (HORMA1)	2TM	
2624321938	2624321936	2624321935		Type III (HORMA1)	2TM	
2596256426	2596256429	2596256430		Type III (HORMA1)	2TM	
2689737943	2689737945	2689737946	Yes	Type II	endonuclease	*Sigma-70 present*
2644992264	2644992266	2644992267	Yes	Type II	endonuclease	*Sigma-70 present*
2653997566	2653997567	2653997568		Type II	endonuclease	*Sigma-70 present*
2745656233	-	2745656232		Type II	endonuclease	*Sigma-70 present; CapH not annotated*
2715924822	2715924824	2715924825	Yes	Type II short	4TM	
2598269393	2598269391	2598269390	Yes	Type II short	4TM	
2618503394	2618503392	2618503391	Yes	Type II short	4TM	

Table A.C.S1 CapH and CapP regulators in CBASS

2598390702	2598390700	2598390699	Yes	Type II short	4TM	
2521671244	2521671246	2521671247	Yes	Type II short	4TM	
2546983264	2546983262	2546983261	Yes	Type II short	4TM	
2632632472	2632632474	2632632475	Yes	Type II short	4TM	
2633337825	2633337827	2633337828	Yes	Type II short	4TM	
2689678230	2689678232	2689678233	Yes	Type II short	4TM	
2732579425	2732579426	2732579427		Type II short	4TM	
2512560318	2512560317	2512560316		Type II short	2TM	
2671594882	2671594879	2671594878	Yes	Type I	2TM	
2690183553	2690183558	2690183559	Yes	Type I	2TM	
2524257494	2524257489	2524257498		Type I	2TM	
2596978699	2596978704	2596978705	Yes	Type I	4TM	
2517105985	2517105980	2517105979	Yes	Type I	4TM	
2610036861	2610036859	2610036858	Yes			
2612570398	2612570397	2612570396	Yes			
2612569470	2612569469	2612569468	Yes			
2695394772	2695394771	2695394770	Yes	Standalone cyclase		
2656910963	2656910960	2656910959	Yes	Type I	endonuclease	
2509372087	2509372090	2509372091		Type II short	phospholipase	
2525851215	2525851217	2525851218		Standalone cyclase	not detected	
2698167474	2698167476	2698167477		Type I	2TM	
2690689086	2690689081	2690689080	Yes	Type I	2TM	
2635135361	-	2635135357		Type III (HORMA2/3)	4TM	*CapH-CapP fusion*

Table A.C.S1 CapH and CapP regulators in CBASS

2546948199	-	2546948195		Type III (HORMA2/3)	4TM	*CapH-CapP fusion*
2639795727	-	2639795723		Type III (HORMA2/3)	4TM	*CapH-CapP fusion*
2578300842	-	2578300838		Type III (HORMA2/3)	4TM	*CapH-CapP fusion*
2694064494	-	2694064490		Type III (HORMA2/3)	4TM	*CapH-CapP fusion*
2518083082	-	2518083078		Type III (HORMA2/3)	4TM	*CapH-CapP fusion*
2642329704	-	2642329700		Type III (HORMA2/3)	4TM	*CapH-CapP fusion*
2660133629	-	2660133633		Type III (HORMA2/3)	4TM	*CapH-CapP fusion*
2517368101	-	2517368105		Type III (HORMA2/3)	4TM	*CapH-CapP fusion*
2663671870	2663671874	2663671875		Type III (HORMA2/3)	endonuclease	
2663674314	2663674318	2663674319	Yes	Type III (HORMA2/3)	endonuclease	
2513642674	2513642679	2513642680	Yes	Type III (HORMA2/3)	endonuclease	
2707329536	2707329541	2707329542	Yes	Type III (HORMA2/3)	endonuclease	
2707240832	2707240837	2707240838	Yes	Type III (HORMA2/3)	endonuclease	
2674659172	2674659167	2674659166	Yes	Type III (HORMA2/3)	endonuclease	

Table A.C.S1 CapH and CapP regulators in CBASS

2511390540	2511390545	2511390546		Type III (HORMA2/ 3)	endonu clease	
2596885468	2596885476	2596885477	Yes	Type I	4TM	
2502683541	2502683544	2502683545		Type II short	2TM	
2515273650	2515273648	2515273647		Type II short	4TM	*Sigma-70 present*
2563801038	2563801036	2563801035		Type II short	4TM	
2671579105	2671579104	2671579103		Type I	TIR	
2602096539	-	2602096537		Type II	2TM	*CapH not annotated*
2650343826	2650343829	2650343830		Type II	2TM	
2698174438	2698174441	2698174442		Type II	endonu clease	
2656896813	2656896811	2656896810		Type II short	endonu clease	
2619745356	2619745353	2619745352	Yes	Type II short	endonu clease	
2697558710	2697558711	2697558712	Yes	Type II	endonu clease	
2662726368	2662726369	2662726370	Yes	Type II	endonu clease	
2665390329	2665390330	2665390331	Yes	Type II	endonu clease	
2696037028	2696037027	2696037026	Yes	Type II	endonu clease	
2664494764	2664494765	2664494766	Yes	Type II	endonu clease	
2657780097	2657780099	2657780100		Type II short	phosph olipase	
2687359245	2687359244	2687359243		Type III (HORMA1)	endonu clease	
2687354813	2687354812	2687354811		Type III (HORMA1)	endonu clease	
2647255930	2647255929	2647255928		Type I	4TM	
2596978701	2596978704	2596978705		Type I	4TM	

Table A.C.S1 CapH and CapP regulators in CBASS

2517105983	2517105980	2517105979		Type I	4TM	
2657054369	2657054366	2657054365	Yes	Type I	DUF2235	
2711700095	2711700093	2711700092		Type I	4TM	
2712705672	2712705667	2712705666		Type II short	4TM	
2713070507	2713070512	2713070513		Type II short	4TM	
2712804583	2712804588	2712804589		Type II short	4TM	
2710834855	2710834860	2710834861		Type II short	4TM	
2641262536	2641262541	2641262542		Type II short	4TM	
2528323693	2528323688	2528323687		Type II short	4TM	
2668349933	2668349938	2668349939		Type II short	4TM	
2732808049	2732808044	2732808043		Type II short	4TM	
2626510053	2626510058	2626510059/2626510060	Yes	Type II short	4TM	
2740524163	-	2740524161		Type I	endonuclease	*CapH-CapP fusion*
2698131277	2698131274	2698131273		Type II	2TM	*SIgma-70 present*
2728057638	2728057641	2728057642		Type II	2TM	
2684397613	2684397609	2684397608		Type II	2TM	
2753669380	2753669384	2753669385		Type II	phospholipase	
2558234065	2558234068	2558234069		Type II	phospholipase	
2523995709	2523995705	2523995704		Type II	phospholipase	
2654896397	-	2654896400		Type II short	not detected	*CapH not annotated*

Table A.C.S1 CapH and CapP regulators in CBASS

2645173595	2645173597	2645173598		Type II	phosph olipase	
2598492051	2598492049	2598492048		Type II	phosph olipase	
2739079492	2739079494	2739079495		Type II	phosph olipase	
2738954077	2738954079	2738954080		Type II	phosph olipase	
2686655639	2686655637	2686655636		Type II	phosph olipase	
2529237622	2529237624	2529237625	Yes	Type II	phosph olipase	
2689729383	2689729387	2689729388	Yes	Type II	phosph olipase	
2609801512	2609801508	2609801507	Yes	Type II	phosph olipase	
2732550022	2732550019	2732550018		Type II	phosph olipase	
2524883671	2524883674	2524883675	Yes	Type I	endonu clease	
2524876148	2524876145	2524876144	Yes	Type I	endonu clease	
2626514909	2626514912	2626514913	Yes	Type II	phosph olipase	*Sigma-70 present*
2626513340	2626513337	2626513336	Yes	Type II	phosph olipase	*Sigma-70 present*
2611046455	2611046452	2611046451	Yes	Type II	phosph olipase	
2599209493	2599209496	2599209497		Type II	phosph olipase	
2720975408	2720975411	2720975412		Type II	phosph olipase	
2599213912	2599213915	2599213916		Type II	phosph olipase	
2618837821	2618837818	2618837817		Type II	phosph olipase	
2629725376	2629725373	2629725372		Type II	phosph olipase	

Table A.C.S1 CapH and CapP regulators in CBASS

2631778469	2631778472	2631778473		Type II	phosph olipase	
2631936835	2631936832	2631936831		Type II	phosph olipase	
2637230762	2637230765	2637230766		Type II	phosph olipase	
2634050317	2634050314	2634050313		Type II	phosph olipase	
2640861876	2640861879	2640861880		Type II	phosph olipase	
2629803928	2629803925	2629803924		Type II	phosph olipase	
2630903411	2630903414	2630903415		Type II	phosph olipase	
2628459832	2628459835	2628459836		Type II	phosph olipase	
2647085408	2647085405	2647085404		Type II	phosph olipase	
2651755067	2651755064	2651755063		Type II	phosph olipase	
2652650346	2652650343	2652650342		Type II	phosph olipase	
2723801827	-	2723801824		Type II	phosph olipase	*CapH not annotated*
2668599363	2668599360	2668599359		Type II	phosph olipase	
2516838873	2516838871	2516838870		Type II	phosph olipase	
2630205316	2630205319	2630205320		Type II	phosph olipase	
2678852886	2678852883	2678852882		Type II	phosph olipase	
2662062718	2662062721	2662062722		Type II	phosph olipase	
2682747150	2682747153	2682747154		Type II	phosph olipase	
2686968366	2686968369	2686968370		Type II	phosph olipase	

Table A.C.S1 CapH and CapP regulators in CBASS

2693582531	2693582534	2693582535		Type II	phosph olipase	
2722614184	2722614187	2722614188		Type II	phosph olipase	
2752970036	2752970039	2752970040		Type II	phosph olipase	
2633653043	2633653046	2633653047		Type II	phosph olipase	
2666217522	2666217519	2666217518		Type II	phosph olipase	
2695721296	2695721293	2695721292		Type II	phosph olipase	
2638390396	2638390399	2638390400		Type II	phosph olipase	
2661876269	2661876272	2661876273		Type II	phosph olipase	
2699460238	2699460241	2699460242		Type II	phosph olipase	
2641747950	2641747953	2641747954		Type II	phosph olipase	
2662608527	2662608524	2662608523		Type II	phosph olipase	
2672584262	2672584259	2672584258		Type II	phosph olipase	
2719176830	2719176833	2719176834		Type II	phosph olipase	
2519167982	2519167985	2519167986		Type II	phosph olipase	
2523626522	2523626519	2523626518		Type II	phosph olipase	
2668936802	2668936805	2668936806		Type II	phosph olipase	
2638368062	2638368065	2638368066		Type II	phosph olipase	
2668921521	2668921518	2668921517		Type II	phosph olipase	
2676330811	2676330808	2676330807		Type II	phosph olipase	

Table A.C.S1 CapH and CapP regulators in CBASS

2702355012	2702355009	2702355008		Type II	phosph olipase	
2662713317	2662713320	2662713321		Type II	phosph olipase	
2674088326	2674088323	2674088322		Type II	phosph olipase	
2677649745	2677649748	2677649749		Type II	phosph olipase	
2669058946	2669058949	2669058950		Type II	phosph olipase	
2732718620	2732718623	2732718624		Type II	phosph olipase	
2671892771	2671892768	2671892767		Type II	phosph olipase	
2635346377	2635346380	2635346381		Type II	phosph olipase	
2508686093	2508686090	2508686089		Type II	phosph olipase	
2669502069	-	2669502067		Type II	phosph olipase	*CapH not annotated*
2617852874	-	2617852877		Type II	phosph olipase	*CapH not annotated*
2502333864	2502333867	2502333868		Type II	phosph olipase	
2632721008	2632721011	2632721012		Type II	phosph olipase	
2630333406	-	2630333409		Type II	phosph olipase	*CapH not annotated*
2665511197	-	2665511200		Type II	phosph olipase	*CapH not annotated*
2506665658	2506665661	2506665662		Type II	phosph olipase	
2547106029	2547106026	2547106025		Type II	phosph olipase	
2647740546	2647740543	2647740542		Type II	phosph olipase	
2670512700	-	2670512697		Type II	phosph olipase	*CapH not annotated*

Table A.C.S1 CapH and CapP regulators in CBASS

2675537097	2675537094	2675537093		Type II	phosph olipase	
2704068073	2704068070	2704068069		Type II	phosph olipase	
2673771445	2673771442	2673771441		Type II	phosph olipase	
2699554687	2699554684	2699554683		Type II	phosph olipase	
2673632610	2673632613	2673632614		Type II	phosph olipase	
2674805878	2674805875	2674805874		Type II	phosph olipase	
2704039980	2704039983	2704039984		Type II	phosph olipase	
2674038584	2674038581	2674038580		Type II	phosph olipase	
2683541550	2683541553	2683541554		Type II	phosph olipase	
2703948844	2703948847	2703948848		Type II	phosph olipase	
2641572094	-	2641572096		Type II	phosph olipase	*CapH not annotated*
2597804387	2597804390	2597804391		Type II	phosph olipase	
2736560602	-	2736560599		Type II	phosph olipase	*CapH not annotated*
2736568852	-	2736568849		Type II	phosph olipase	*CapH not annotated*
2736553705	-	2736553702		Type II	phosph olipase	*CapH not annotated*
2652628329	2652628332	2652628333		Type II	phosph olipase	
2688089839	-	2688089842		Type II	phosph olipase	*CapH not annotated*
2689085657	-	2689085660		Type II	phosph olipase	*CapH not annotated*
2688716432	-	2688716435		Type II	phosph olipase	*CapH not annotated*

Table A.C.S1 CapH and CapP regulators in CBASS

2685603808	-	2685603811		Type II	phosph olipase	*CapH not annotated*
2673097265	-	2673097262		Type II	phosph olipase	*CapH not annotated*
2674340297	-	2674340294		Type II	phosph olipase	*CapH not annotated*
2677724982	-	2677724985		Type II	phosph olipase	*CapH not annotated*
2692697646	-	2692697649		Type II	phosph olipase	*CapH not annotated*
2696356224	-	2696356227		Type II	phosph olipase	*CapH not annotated*
2697820223	-	2697820220		Type II	phosph olipase	*CapH not annotated*
2673272645	-	2673272642		Type II	phosph olipase	*CapH not annotated*
2674273657	-	2674273654		Type II	phosph olipase	*CapH not annotated*
2675662305	-	2675662308		Type II	phosph olipase	*CapH not annotated*
2683901263	-	2683901260		Type II	phosph olipase	*CapH not annotated*
2692831768	-	2692831771		Type II	phosph olipase	*CapH not annotated; confirmed present in genome sequence*
2692848326	-	2692848323		Type II	phosph olipase	*CapH not annotated*
2696882989	-	2696882992		Type II	phosph olipase	*CapH not annotated*
2697292476	-	2697292473		Type II	phosph olipase	*CapH not annotated*
2697515814	-	2697515811		Type II	phosph olipase	*CapH not annotated*
2677623137	-	2677623140		Type II	phosph olipase	*CapH not annotated*
2683889884	-	2683889881		Type II	phosph olipase	*CapH not annotated*

Table A.C.S1 CapH and CapP regulators in CBASS

2683894738	-	2683894735		Type II	phosph olipase	*CapH not annotated*
2696336386	-	2696336383		Type II	phosph olipase	*CapH not annotated*
2697141496	-	2697141493		Type II	phosph olipase	*CapH not annotated*
2697756677	-	2697756674		Type II	phosph olipase	*CapH not annotated*
2699055511	-	2699055508		Type II	phosph olipase	*CapH not annotated*
2672803174	-	2672803171		Type II	phosph olipase	*CapH not annotated*
2678020133	-	2678020130		Type II	phosph olipase	*CapH not annotated*
2693132540	-	2693132537		Type II	phosph olipase	*CapH not annotated*
2696049324	-	2696049321		Type II	phosph olipase	*CapH not annotated*
2697510876	-	2697510873		Type II	phosph olipase	*CapH not annotated*
2698955961	-	2698955958		Type II	phosph olipase	*CapH not annotated*
2699313178	-	2699313175		Type II	phosph olipase	*CapH not annotated*
2673880498	-	2673880495		Type II	phosph olipase	*CapH not annotated*
2675444124	-	2675444121		Type II	phosph olipase	*CapH not annotated*
2695870517	-	2695870514		Type II	phosph olipase	*CapH not annotated*
2697150719	-	2697150716		Type II	phosph olipase	*CapH not annotated*
2697777604	-	2697777601		Type II	phosph olipase	*CapH not annotated*
2697998223	-	2697998220		Type II	phosph olipase	*CapH not annotated*
2698013633	-	2698013636		Type II	phosph olipase	*CapH not annotated*

Table A.C.S1 CapH and CapP regulators in CBASS

2678519127	-	2678519124		Type II	phosph olipase	*CapH not annotated*
2692412387	-	2692412390		Type II	phosph olipase	*CapH not annotated*
2695813501	-	2695813504		Type II	phosph olipase	*CapH not annotated*
2675712932	-	2675712929		Type II	phosph olipase	*CapH not annotated*
2693109836	-	2693109833		Type II	phosph olipase	*CapH not annotated*
2693225902	-	2693225899		Type II	phosph olipase	*CapH not annotated*
2693950232	-	2693950229		Type II	phosph olipase	*CapH not annotated*
2694562759	-	2694562756		Type II	phosph olipase	*CapH not annotated*
2697465810	-	2697465807		Type II	phosph olipase	*CapH not annotated*
2698005973	-	2698005976		Type II	phosph olipase	*CapH not annotated*
2672966463	-	2672966466		Type II	phosph olipase	*CapH not annotated*
2698961466	-	2698961463		Type II	phosph olipase	*CapH not annotated*

Table A.C.S11 Crystallographic data and refinement

	***Thauera* sp. K11 CapP SeMet**	***Thauera* sp. K11 CapP Native**	***E. coli* MS115-1 CapH NTD**	***E. coli* MS115-1 CapH CTD**	***E. coli* MS115-1 CapH CTD (I99M)**
Data collection					
Synchrotron/Beamline	SSRL 9-2	SSRL 9-2	ALS 5.0.2	ALS 5.0.2	APS 24ID-C
Date collected	02/14/2020	02/14/2020	04/25/2021	5/14/21	07/11/21
Resolution (Å)	40-1.6	40-1.35	40-1.02	100-1.75	59-1.26
Wavelength (Å)	0.97892	0.97892	1.00003	1.00004	0.97918
Space Group	$P4_22_12$	$P4_22_12$	$P2_12_12_1$	$P2_1$	$P4_32_12$
Unit Cell Dimensions (a, b, c) Å	95.28, 95.28, 103.95	95.31, 95.31, 104.92	32.38, 39.68, 47.24	44.02, 39.86, 46.26	38.38, 38.38, 59.43
Unit cell Angles (α,β,γ) °	90, 90, 90	90, 90, 90	90, 90, 90	90, 98.73, 90	90, 90, 90
I/σ (last shell)	21.2 (0.9)	18.7 (0.8)	12.5 (2.0)	26.2 (1.1)	21.1 (2.0)
[1]R_{sym} (last shell)	0.082 (2.61)	0.058 (2.884)	0.068 (0.424)	0.059 (0.637)	0.039 (0.732)
[2]R_{meas} (last shell)	0.085 (2.723)	0.06 (3.012)	0.074 (0.516)	0.070 (0.784)	0.043 (0.809)
[3]$CC_{1/2}$ (last shell)	1 (0.529)	0.999 (0.443)	0.997 (0.814)	1 (0.631)	0.999 (0.830)
Completeness (last shell) %	100.0 (99.8)	99.8 (98.1)	80.8 (11.8)	99.5 (98.8)	99.8 (98.3)
Number of reflections	839884	1388162	144010	50378	74958
unique	63659	106052	25495	15936	12610
Multiplicity (last shell)	13.2 (12.4)	13.1 (11.7)	5.6 (3.0)	3.2 (2.8)	5.9 (5.5)
Refinement					
Resolution (Å)	-	40-1.35	30.4-1.02	45.7-1.75	32.2-1.26
No. of reflections	-	105839	25448	15917	12560
working	-	102934	24251	15208	11966
free	-	2905	1197	709	594
[4]R_{work} (last shell) (%)	-	16.21 (42.68)	16.30 (17.77)	20.98 (28.61)	19.13 (22.50)
[4]R_{free} (last shell) (%)	-	16.79 (43.92)	17.64 (17.00)	23.70 (31.23)	21.36 (24.19)
Structure/Stereochemistry					
No. of atoms	-	4741	1166	2531	642
solvent	-	343	87	60	46
ligand	-	0	10	0	0
hydrogen	-	2159	536	1212	297

Table A.C.S2 Crystallographic data and refinement

r.m.s.d. bond lengths (Å)	-	0.008	0.008	0.010	0.011
r.m.s.d. bond angles (°)	-	0.94	0.993	1.000	0.896
Ramachandran favored/allowed (%)	-	98.2%/100%	100%/100%	99.3%/100%	100%/100%
Molprobity score	-	0.82	0.78	0.64	1.31
[5]SBGrid Data Bank ID	865	864	866	867	868
[6]Protein Data Bank ID	-	7T5T	7T5U	7T5W	7T5V

[1]$R_{sym} = \sum\sum_j |I_j - \langle I \rangle| / \sum I_j$, where I_j is the intensity measurement for reflection j and $\langle I \rangle$ is the mean intensity for multiply recorded reflections.
[2]$R_{meas} = \sum_h [\sqrt{(n/(n-1))} \sum_j [I_{hj} - \langle I_h \rangle]] / \sum_{hj} \langle I_h \rangle$
where I_{hj} is a single intensity measurement for reflection h, $\langle I_h \rangle$ is the average intensity measurement for multiply recorded reflections, and n is the number of observations of reflection h.
[3]$CC_{1/2}$ is the Pearson correlation coefficient between the average measured intensities of two randomly assigned half-sets of the measurements of each unique reflection. $CC_{1/2}$ is considered significant above a value of ~0.15.
[4]$R_{work, free} = \sum||F_{obs}| - |F_{calc}|| / |F_{obs}|$, where the working and free R-factors are calculated using the working and free reflection sets, respectively.
[5]Diffraction data for each structure have been deposited with the SBGrid Data Bank (https://data.sbgrid.org) with the noted accession codes.
[6]Coordinates and structure factors for each structure have been deposited with the Protein Data Bank (http://www.rcsb.org) with the noted accession codes.

References

Elkins, P.A., Ho, Y.S., Smith, W.W., Janson, C.A., D'Alessio, K.J., McQueney, M.S., Cummings, M.D., Romanic, A.M., 2002. Structure of the C-terminally truncated human ProMMP9, a gelatin-binding matrix metalloproteinase. Acta Crystallographica Section D 58, 1182–1192. https://doi.org/10.1107/S0907444902007849

Lau, R.K., Ye, Q., Birkholz, E.A., Berg, K.R., Patel, L., Mathews, I.T., Watrous, J.D., Ego, K., Whiteley, A.T., Lowey, B., Mekalanos, J.J., Kranzusch, P.J., Jain, M., Pogliano, J., Corbett, K.D., 2020. Structure and Mechanism of a Cyclic Trinucleotide-Activated Bacterial Endonuclease Mediating Bacteriophage Immunity. Molecular Cell 77, 723-733.e6. https://doi.org/10.1016/j.molcel.2019.12.010

Martinez, S.E., Heikaus, C.C., Klevit, R.E., Beavo, J.A., 2008. The structure of the GAF a domain from phosphodiesterase 6C reveals determinants of cGMP binding, a conserved binding surface, and a large cGMP-dependent conformational change. Journal of Biological Chemistry 283, 25913–25919. https://doi.org/10.1074/jbc.M802891200

Salamov, V.S.A., Solovyevand, A., 2011. Automatic annotation of microbial genomes and metagenomic sequences. Metagenomics and Its Applications in Agriculture, Biomedicine and Environmental Studies; Li, RW, Ed 61–78.