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

Graphical Abstract

CBASS OR Type III CRISPR

NucC nuclease

Cyclic triadenylate (cA₃)

Inactive trimer

Active hexamer

NucC

Genome destruction
Cell death

Highlights

- The CBASS-associated nuclease NucC kills phage-infected cells by genome destruction
- Binding of a cyclic triadenylate second messenger causes NucC homohexamer assembly
- NucC hexamer assembly juxtaposes pairs of active sites to activate DNA cleavage
- NucC is found in both CBASS systems and type III CRISPR/Cas systems

Authors
Rebecca K. Lau, Qiaozhen Ye, Erica A. Birkholz, ..., Mohit Jain, Joe Pogliano, Kevin D. Corbett

Correspondence
kcorbett@ucsd.edu

In Brief
Here, Lau et al. show that the cyclic oligonucleotide-based anti-phage signaling system (CBASS)-associated NucC nuclease causes death of phage-infected cells through genome destruction. They show how NucC is allosterically activated by a cyclic triadenylate second messenger synthesized by CBASS upon infection. NucC has also been integrated into type III CRISPR/Cas systems as an accessory nuclease.
Structure and Mechanism of a Cyclic Trinucleotide-Activated Bacterial Endonuclease Mediating Bacteriophage Immunity

Rebecca K. Lau,1,2,10 Qiaozhen Ye,1,10 Erica A. Birkholz,3 Kyle R. Berg,1,4 Lucas Patel,3 Ian T. Mathews,2,5,6 Jeramie D. Watrous,5,6 Kaori Ego,1 Aaron T. Whiteley,7,8 Brianna Lowey,7,8 John J. Mekalanos,7 Philip J. Kranzusch,7,8 Mohit Jain,5,6 Joe Pogliano,3 and Kevin D. Corbett1,4,9,11,*

1Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, CA, USA
2Biomedical Sciences Graduate Program, University of California, San Diego, La Jolla, CA, USA
3Division of Biological Sciences, University of California, San Diego, La Jolla, CA, USA
4Department of Chemistry, University of California, San Diego, La Jolla, CA, USA
5Department of Pharmacology, University of California, San Diego, La Jolla, CA, USA
6Department of Medicine, University of California, San Diego, La Jolla, CA, USA
7Department of Microbiology, Harvard Medical School, Boston, MA, USA
8Department of Cancer Immunology and Virology, Dana-Farber Cancer Institute, Boston, MA, USA
9Ludwig Institute for Cancer Research, San Diego Branch, La Jolla, CA, USA
10These authors contributed equally
11Lead Contact
*Correspondence: kcorbett@ucsd.edu
https://doi.org/10.1016/j.molcel.2019.12.010

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 synthesized by these systems’ effector complexes.

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., 2020 [this issue of Molecular Cell]). 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., 2020). 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 3-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 accesso-
sory nucleases activated by cyclic oligoadenylate second messengers.

RESULTS

NucC Activation Leads to Genome Destruction in Infected Cells

In parallel work, we showed that the CBASS system from a patient-derived strain of E. coli (MS115-1; Figure 1A) provides immunity against bacteriophage λ infection by an abortive infection mechanism (Ye et al., 2020). 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 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 min post-infection, and large-scale lysis by 80 min post-infection (Figures 1C, 1D, and S1). In cultures carrying the intact CBASS system, however, by 40 min post-infection we observe a large fraction of cells in which cellular DNA appears either partially or completely degraded (Figures 1C and 1D). These cultures show large-scale lysis by 80 min post-infection, but do not show the high levels of extracellular DNA characteristic of lysed cultures lacking CBASS (Figure 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 (Figures 1C and 1D). 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 (P). Ec NucC and Pa NucC share 77% overall sequence identity, and both proteins self-associate in yeast two-hybrid assays (Ye et al., 2020). Using size exclusion chromatography coupled to multi-angle light scattering (SEC-MALS), we found that both proteins form homotrimmers in solution (Figures S2A and S2B). 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 S1). In agreement with prior sequence analysis (Burroughs et al., 2015), the NucC monomer adopts a restriction endonuclease-like fold (Figure 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: 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 RMSD of 1.97 Å over 132 aligned Cα atoms (Figures 2A, S2C, and S2D). NucC also possesses a conserved active site motif of IDX3EAK (residues 72–106 in Ec NucC), and the conserved acidic residues Asp73 and Glu104 coordinate a Mg2+ ion in our structure (Figure 2B). Consistent with our SEC-MALS data, the structure of NucC reveals a tightly packed homotrimer with an overall triangular architecture (Figure 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 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., 2020). 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 (Figures S3 and 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 resolutions of 1.86 and 1.45 Å, respectively (Table S1; Figures

Please cite this article in press as: Lau et al., Structure and Mechanism of a Cyclic Trinucleotide-Activated Bacterial Endonuclease Mediating Bacteriophage Immunity, Molecular Cell (2019), https://doi.org/10.1016/j.molcel.2019.12.010
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 S1; Figures S5D–S5F). The structures reveal that a single second messenger molecule binds in a conserved, 3-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 (Figures 3A–3C). 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 (Figures 3D and 3E). In addition, NucC residues Arg53 and Thr226 form a hydrogen-bond network with each phosphate group of the cyclic trinucleotide (Figure 3D). 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 3F). 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 3-fold axis of the NucC trimer. Mutation of gate loop residues His136 or Tyr141, or cAAA-binding residues...
ArgS3, Tyr81, or Thr226, completely disrupts Ec NucC’s cAAA-activated endonuclease activity (Figure 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 (Figures 3B and 3C). Within the NucC trimer, each subunit undergoes a rotation of ~3° “upward” away from the cAAA binding site compared to the unbound structure (Figure 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 (Figures 4A–4C and S5A). The trimer-trimer interface involves a symmetric interaction of the x0 helices of facing NucC protomers, plus a hydrophobic cluster assembled from three different NucC protomers. In the x0 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 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 4C). The docking of Phe28 into this hydrophobic pocket causes residues 23–34 linking x0 and x1, 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 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 (Figures 4D and 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 x0 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 (Figures 4E and S6A–S6F) and their cAAA-stimulated DNA cleavage activity (Figures 4F and S6G). Four of the six mutants (L19D, A27E, A27K, and F28A) strongly disrupt both hexamer formation and DNA cleavage activity (Figure 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 4H).

To better understand why hexamer formation activates NucC, we modeled 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 5A), and we could model a continuous bent DNA duplex that contacts a pair of juxtaposed NucC-active sites (Figure 5B). In this model, the putative scissile phosphate groups are positioned two base pairs apart on opposite DNA strands (Figure 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 Mg2+ conditions to enrich for singly cut plasmid in a plasmid digestion assay (Figure 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 2-fold symmetry in the preferred cleavage sites, with a strong preference for “T” at the 5’0 end (+1 base) of the cleaved fragment, and an equally strong preference for “A” at the -3 position (Figure 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 2-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 (Figures 5D and 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 5G). 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, which are encoded not within CBASS systems but rather within type III CRISPR/Cas systems (Figures 6A–6C). 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 6A; Table S4). These proteins are found predominantly in Gammaproteobacteria and Firmicutes, but are also found in Alpha-, Beta-, and Deltaproteobacteria (Table 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 6A; Table 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 several NucC proteins and members of a second, likely unrelated nuclease family (Table S4; Shah et al., 2019).

Type III CRISPR/Cas systems encode a multi-subunit effector complex, which when activated by target binding synthesizes a range of cyclic oligoadenylate second messengers with 2–6 bases (cA2–cA6), often with a high proportion of cAAA/cA 3 (Gruschow 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).
et al., 2017; Niewoehner et al., 2017). More recently, a cA4-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 metoe cus, a close relative of V. cholerae (Kirchberger et al., 2014). We found that V. metoe cus (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 (Figures 7A and 7B; Ye et al., 2020). 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 7D).

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 7C). Vm NucC shares a common active-site structure with Ec NucC (Figures S7A and S7B), and the Vm NucC trimer overlays with the Ec NucC trimer with an overall Cα RMSD 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

Figure 5. 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: 3DW9; Dunten et al., 2008). The two modeled DNA can be joined by a bent-two-base-pair linker (dotted lines) to generate a contiguous double-stranded DNA (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 Mg2+. 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 2-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.
and Pa NucC (Figure 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 (Figures S7C and S7D). 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., 2020). 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., 2020). 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.

**Figure 6. NucC as 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 S4 for a complete list of type III CRISPR-associated NucC homologs and Table 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 (A). Core type III genes are colored as in (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 (SMODS-associated and fused to various effector domains) domain-containing proteins (Burroughs et al., 2015) are shown in white. The operon for *Acetivibrio cellulolyticus* CD2 was poorly annotated, and the gene names shown are based on the closest match from BLAST (Basic Local Alignment Search Tool) searches. RT, predicted reverse transcriptase.
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 \( \frac{C24}{50} \) 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 (Rostøl and Marraffini, 2019), 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.
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.

STAR METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **LEAD CONTACT AND MATERIALS AVAILABILITY**
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
- **METHOD DETAILS**
  - Bacterial Growth and Microscopy
  - Protein Expression, Purification, and Characterization
  - Second Messenger Molecules
  - Mass Spectrometry
  - Isothermal Titration Calorimetry
  - Crystallization and Structure Determination
  - Synchrotron Support Statements
  - Nuclease Assays
  - High-Throughput Sequencing and Data Analysis
  - Identification of CRISPR-Associated NucC Homologs
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
- **DATA AND CODE AVAILABILITY**

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.molcel.2019.12.010.

ACKNOWLEDGMENTS

The authors thank the staff of the Stanford Synchrotron Radiation Lightsource, the Advanced Photon Source NE-CAT beamlines, and the Advanced Light Source beamline 6.3.1 for assistance with crystallographic data collection; A. Bobkov (Sanford Burnham Prebys Medical Discovery Institute, Protein Expression Core) for assistance with isothermal titration calorimetry; and A. Desai (Sanford Burnham Prebys Medical Discovery Institute, Protein Expression, Purification, and Characterization) for critical reading and helpful suggestions. R.K.L. was supported by the UC San Diego Quantitative and Integrative Biology training grant (NIH T32 GM127235). I.T.M. was supported by NIH grants K01 DK116917 and P30 DK063491. M.J. was supported by NIH grants S10 OD020025 and R01 ES027595. A.T.W. was supported as a fellow of the Jane Coffin Childs Memorial Fund for Medical Research. B.L. was supported by NIH grants R01 AI081845 and R01 A1026289. P.J.K. was supported by the Richard and Susan Smith Foundation and the Parker Institute for Cancer Immunotherapy. J.P. was supported by NIH grant R01 GM129245. K.D.C. was supported by the Ludwig Institute for Cancer Research and the University of California, San Diego.

AUTHOR CONTRIBUTIONS


DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: July 6, 2019
Revised: October 4, 2019
Accepted: December 12, 2019
Published: January 10, 2020

REFERENCES


# STAR METHODS

## KEY RESOURCES TABLE

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bacterial and Virus Strains</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E. coli Rosetta2(DE3) pLysS</td>
<td>EMD Millipore</td>
<td>#71403</td>
</tr>
<tr>
<td>E. coli JP313</td>
<td>Joe Pogliano, UCSD; Economou et al., 1995</td>
<td>N/A</td>
</tr>
<tr>
<td>E. coli MS115-1</td>
<td>BEI Resources</td>
<td>HM-344</td>
</tr>
<tr>
<td>Bacteriophage lambda cI</td>
<td>Joe Pogliano, UCSD</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Biological Samples</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Genomic DNA for P. aeruginosa ATCC27853</td>
<td>American Type Culture Collection (ATCC)</td>
<td>#27853</td>
</tr>
<tr>
<td><strong>Chemicals, Peptides, and Recombinant Proteins</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5'-pApA</td>
<td>Invivogen</td>
<td>#tlrl-napapa</td>
</tr>
<tr>
<td>3',3' cyclic di-AMP</td>
<td>Invivogen</td>
<td>#tlrl-nacda</td>
</tr>
<tr>
<td>2',3' cyclic di-AMP</td>
<td>Invivogen</td>
<td>#tlrl-nacda23</td>
</tr>
<tr>
<td>3',3' cyclic GMP-AMP (cGAMP)</td>
<td>Invivogen</td>
<td>#tlrl-nacga</td>
</tr>
<tr>
<td>5'pGpG</td>
<td>Invivogen</td>
<td>#tlrl-napapa</td>
</tr>
<tr>
<td>3',3' cyclic di-GMP</td>
<td>Invivogen</td>
<td>#tlrl-nacdg</td>
</tr>
<tr>
<td>3',3',3' cyclic tri-AMP (cAAA)</td>
<td>This work; Whiteley et al., 2019</td>
<td>N/A</td>
</tr>
<tr>
<td>3',3',3' cyclic AMP-AMP-GMP (cAAG)</td>
<td>This work; Whiteley et al., 2019</td>
<td>N/A</td>
</tr>
<tr>
<td>Recombinant TEV protease (homemade)</td>
<td>Tropea et al., 2009</td>
<td>N/A</td>
</tr>
<tr>
<td>Enterobacter cloacae CdnD (homemade)</td>
<td>Whiteley et al., 2019</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Critical Commercial Assays</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accel-NGS 1S Plus DNA Library Kit</td>
<td>Swift Biosciences</td>
<td>#10024</td>
</tr>
<tr>
<td><strong>Deposited Data</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Structure of E. coli NucC (Apo)</td>
<td>This work</td>
<td>Protein Data Bank (<a href="http://www.rcsb.org">http://www.rcsb.org</a>) 6P7Q <a href="https://doi.org/10.2210/pdb6P7Q/pdb">https://doi.org/10.2210/pdb6P7Q/pdb</a></td>
</tr>
<tr>
<td>Diffraction data for E. coli NucC (Apo) (Native)</td>
<td>This work</td>
<td>SBGrid Data Bank (<a href="https://data.sbgrid.org">https://data.sbgrid.org</a>) 664 <a href="https://doi.org/10.15785/SBGRID/664">https://doi.org/10.15785/SBGRID/664</a></td>
</tr>
<tr>
<td>Diffraction data for E. coli NucC (Apo) (SeMet)</td>
<td>This work</td>
<td>SBGrid-DB 665 <a href="https://doi.org/10.15785/SBGRID/665">https://doi.org/10.15785/SBGRID/665</a></td>
</tr>
<tr>
<td>Structure of E. coli NucC (5'-pApA)</td>
<td>This work</td>
<td>PDB 6P7Q <a href="https://doi.org/10.2210/pdb6P7Q/pdb">https://doi.org/10.2210/pdb6P7Q/pdb</a></td>
</tr>
<tr>
<td>Diffraction data for E. coli NucC (5'pApA)</td>
<td>This work</td>
<td>SBGrid-DB 667 <a href="https://doi.org/10.15785/SBGRID/667">https://doi.org/10.15785/SBGRID/667</a></td>
</tr>
<tr>
<td>Structure of E. coli NucC (cAAA)</td>
<td>This work</td>
<td>PDB 6P7P <a href="https://doi.org/10.2210/pdb6P7P/pdb">https://doi.org/10.2210/pdb6P7P/pdb</a></td>
</tr>
<tr>
<td>Diffraction data for E. coli NucC (cAAA)</td>
<td>This work</td>
<td>SBGrid-DB 666 <a href="https://doi.org/10.15785/SBGRID/666">https://doi.org/10.15785/SBGRID/666</a></td>
</tr>
<tr>
<td>Structure of P. aeruginosa NucC (cAAA)</td>
<td>This work</td>
<td>PDB 6Q1H <a href="https://doi.org/10.2210/pdb6Q1H/pdb">https://doi.org/10.2210/pdb6Q1H/pdb</a></td>
</tr>
<tr>
<td>Diffraction data for P. aeruginosa NucC (cAAA)</td>
<td>This work</td>
<td>SBGrid-DB 732 <a href="https://doi.org/10.15785/SBGRID/732">https://doi.org/10.15785/SBGRID/732</a></td>
</tr>
<tr>
<td>Structure of V. metoecus NucC (trimer)</td>
<td>This work</td>
<td>PDB 6UXG <a href="https://doi.org/10.2210/pdb6UXG/pdb">https://doi.org/10.2210/pdb6UXG/pdb</a></td>
</tr>
<tr>
<td>Diffraction data for V. metoecus NucC (trimer)</td>
<td>This work</td>
<td>SBGrid-DB 734 <a href="https://doi.org/10.15785/SBGRID/734">https://doi.org/10.15785/SBGRID/734</a></td>
</tr>
<tr>
<td>Structure of V. metoecus NucC (hexamer)</td>
<td>This work</td>
<td>PDB 6UXF <a href="https://doi.org/10.2210/pdb6UXF/pdb">https://doi.org/10.2210/pdb6UXF/pdb</a></td>
</tr>
<tr>
<td>Diffraction data for V. metoecus NucC (hexamer)</td>
<td>This work</td>
<td>SBGrid-DB 735 <a href="https://doi.org/10.15785/SBGRID/735">https://doi.org/10.15785/SBGRID/735</a></td>
</tr>
<tr>
<td>Original imaging data for Figures 1 and S1</td>
<td>This work</td>
<td>Mendeley Data <a href="https://doi.org/10.17632/dr8mjtwh2d.1">https://doi.org/10.17632/dr8mjtwh2d.1</a></td>
</tr>
<tr>
<td><strong>Recombinant DNA</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bacterial expression plasmid for Enterobacter cloacae CdnD</td>
<td>Whiteley et al., 2019</td>
<td>N/A</td>
</tr>
<tr>
<td>UC Berkeley Macrolab Vector 1B</td>
<td>Addgene</td>
<td>#29653</td>
</tr>
<tr>
<td>UC Berkeley Macrolab Vector 2AT</td>
<td>Addgene</td>
<td>#29665</td>
</tr>
<tr>
<td>UC Berkeley Macrolab Vector 2BT</td>
<td>Addgene</td>
<td>#29666</td>
</tr>
</tbody>
</table>

(Continued on next page)
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 MODEL 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., 2020). 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 OD600 of 0.1 in a reservoir containing LB with 100ug/ml ampicillin and 0.2mM IPTG. Bacteriophage λcI diluted in phage buffer with 1mM CaCl2 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 strain 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 His6-tag. Mutants of Ec NucC were generated by PCR mutagenesis.

Please cite this article in press as: Lau et al., Structure and Mechanism of a Cyclic Trinucleotide-Activated Bacterial Endonuclease Mediating Bacteriophage Immunity, Molecular Cell (2019), https://doi.org/10.1016/j.molcel.2019.12.010
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 centrifuged 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 an Ni²⁺ 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 Ni²⁺ 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 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 (Figures 4 and 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 mM 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 MgCl₂, 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 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 S2A) were pooled, evaporated using a speedvac, then resuspended in water and analyzed by LC-MS/MS. cAAA synthesized by the Ec-CdnD02 and 0.25 mM each ATP and GTP (ATP alone forming with protein also containing the D73N active-site mutation.

**Mass Spectrometry**

Whiteley et al. (2019) previously identified the major product of Ec-CdnD02 as cAAG by NMR. To further characterize the products of Ec-CdnD02, we performed liquid chromatography-tandem mass spectrometry (LC-MS/MS) (Figures S2B–S2D). 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 1×106, 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 1×105; 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 MgCl₂, 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 (Figures S3A–S3G; Table 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 MgCl2, and 1 mM TCEP (Figures S3H–S3K; Table 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 MgCl2, 1 mM TCEP) 1:1 with well solution containing 27% PEG 400, 400 mM MgCl2, 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 molecular replacement using a 1.82 Å resolution dataset from selenomethione-derivatized protein. We identified heavy-atom sites using hkl2map (Pape et al., 2004) (implemented model building (Terwilliger, 2003b). 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 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 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 S2E), we used the PISA structure-analysis server (Krusin 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, then flash-froze crystals 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 S1).

**V. metoecus NucC**

For *V. metoecus* 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-froze 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 P6322 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 S1).

For *V. metoecus* Apo trimer crystals, Vm NucC (20 mg/mL) in crystallization buffer (25 mM Tris-HCl, 0.2 M NaCl, 5 mM MgCl2, 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-froze 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 P43212 by XDS. Molecular replacement was performed with PHASER using a single chain from the *V. metoecus* 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 GM103393). The Eger 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), Plexonik 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-BstEll 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 MgCl₂, 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 Adaptor 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 adaptor ligation; (3) ligates the Adaptor 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 Adaptor 1 (full sequence: 5’ AATGATACGGCGACCAGGAGATCTCAGCTTCCTCCTACACGACGCTCTTCCGATCT 3’) in R1, searching for the sequence “AGATCGGA.” We calculated the length of each NucC fragment as: (length of RT prior to AGATCGGA) minus 8 (average length of adaptase-added low-complexity sequence). The graph in Figure 5E represents the lengths of 343 million NucC fragments.

For NucC cleavage site identification (Figure 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 IDx3EAK 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/qbench/) 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 S4 for the complete list of Type III CRISPR associated NucC proteins, and Table S5 for the complete list of bacterial NucC proteins used for the phylogenetic tree in Figure 6A.
QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis of primary diffraction datasets was performed with standard measures including $R_{\text{sym}}$, $R_{\text{meas}}$, and $CC_{1/2}$ to determine data quality and resolution cutoffs. Statistical analysis of refined crystal structures used standard measures including $R$ and $R_{\text{free}}$ 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 S1 and the Key Resources Table (see Methods). 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 S1 and the Key Resources Table (see Methods). The original imaging data for Figures 1 and S1 is available at Mendeley Data (https://data.mendeley.com) under https://doi.org/10.17632/dr8mjtwh2d.1. All high-throughput sequencing data used for Figure 5 are available from the Lead Contact without restriction.