HHS Public Access

Author manuscript

Nature. Author manuscript; available in PMC 2023 October 01.

Published in final edited form as:

Nature. 2023 April ; 616(7956): 319–325. doi:10.1038/s41586-022-05647-4.

An E1-E2 fusion protein primes antiviral immune signaling in bacteria

Hannah E. Ledvina[1,*], **Qiaozhen Ye**[2,*], **Yajie Gu**[2], **Ashley E. Sullivan**[1], **Yun Quan**[2], **Rebecca K. Lau**[2,3], **Huilin Zhou**[2], **Kevin D. Corbett**[2,†], **Aaron T. Whiteley**[1,†]

[1]Department of Biochemistry, University of Colorado Boulder, Boulder, Colorado, USA

[2]Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, California, USA

[3]Biomedical Sciences Graduate Program, University of California, San Diego, La Jolla, California, USA

Summary

In all organisms, innate immune pathways sense infection and rapidly activate potent immune responses while avoiding inappropriate activation (autoimmunity). In humans, the innate immune receptor cGAS detects viral infection to produce the nucleotide second messenger cGAMP, which initiates STING-dependent antiviral signaling[1]. Bacteria encode evolutionary predecessors of cGAS termed cGAS/DncV-like nucleotidyltransferases (CD-NTases)[2], which detect bacteriophage infection and produce diverse nucleotide second messengers[3]. How bacterial CD-NTase activation is controlled remains unknown. Here, we show that CD-NTase-associated protein 2 (Cap2) primes bacterial CD-NTases for activation through a ubiquitin transferase-like mechanism. A cryoelectron microscopy structure of the Cap2–CD-NTase complex reveals Cap2 as an all-in-one ubiquitin transferase-like protein, with distinct domains resembling eukaryotic E1 and E2 proteins. The structure captures a reactive-intermediate state with the CD-NTase C-terminus positioned in the Cap2 E1 active site and conjugated to AMP. We find that Cap2 conjugates the CD-NTase C-terminus to a target molecule that primes the CD-NTase for increased cGAMP production. We further demonstrate that a specific endopeptidase, Cap3, balances Cap2 activity by cleaving CD-NTase–target conjugates. Our data demonstrate that bacteria control immune signaling using an ancient, minimized ubiquitin transferase-like system and provide insight into the evolution of E1 and E2 machinery across domains of life.

† To whom correspondence should be addressed: aaron.whiteley@colorado.edu (ATW), kcorbett@ucsd.edu (KDC).
*These authors contributed equally

Author Contributions

Experiments were designed and conceived by H.E.L., Q.Y., A.T.W., and K.D.C.. Phage assays were performed be H.E.L. and A.E.S.. The *in vivo* bacterial assays, cGAMP generation assays, and bioinformatical analysis of E1, E2 and JAB domains were performed by H.E.L.. Biochemical experiments were performed by Q.Y.. Mass spectrometry experiments were performed by Q.Y., Y.Q., and H.E.L.. Structural experiments and analysis were performed by Q.Y., Y.G., and K.D.C.. Protein purification was performed by Q.Y. and R.K.L.. Bioinformatic analysis of CD-NTase and Cap2 proteins was performed by K.D.C. and H.E.L.. Figures were prepared by H.E.L., A.T.W. and K.D.C.. The manuscript was written by H.E.L., A.T.W., and K.D.C.. All authors contributed to editing the manuscript and support the conclusions.

Competing Interest

The University of Colorado Boulder and the University of California San Diego have patents pending for Cap2 and Cap3 technologies on which H.E.L., Q.Y., Y.G., K.D.C., and A.T.W are listed as inventors.

Innate immune pathways rapidly sense and respond to viral threats while limiting their activation in the absence of infection, which could otherwise lead to autoimmune disease or premature cell death. In eukaryotes, viral defense is mediated in part by the cGAS-STING (cyclic GMP–AMP Synthase-Stimulator of Interferon Genes) pathway[1]. Recent work has demonstrated that the cGAS-STING pathway originated from bacterial cyclic oligonucleotide-based antiphage signaling systems (CBASS), which serve an analogous function in the bacterial antiviral immune response[2–6].

CBASS pathways are diverse, widespread, and protect populations against phage infection by triggering programmed cell death[3,6,7]. All CBASS operons encode a cGAS/DncV-like nucleotidyltransferase (CD-NTase) that is activated upon phage infection[3] and synthesizes one of a variety of cyclic oligonucleotide second messengers[2]. Those molecules in turn activate a cell-killing effector protein[8,9] to halt phage replication, a process termed abortive infection[3,4,10]. CBASS operons are classified based on their architecture[6,7], with type I CBASS encoding only a CD-NTase and an effector protein, and types II, III, and IV encoding additional proteins with proposed regulatory roles. The mechanisms of second messenger synthesis and effector activation in CBASS have been the focus of numerous studies[9–11], but the roles of these regulatory proteins in CD-NTase activation remain largely unknown.

We focused on type II CBASS, which make up ~40% of systems[7], and selected a representative system from a pandemic strain of *Vibrio cholerae* (Fig. 1a)[12,13]. Upon phage infection, this system's CD-NTase (also termed dinucleotide cyclase in *Vibrio*, DncV) synthesizes the cyclic dinucleotide 3′,3′ cyclic GMP–AMP (cGAMP)[3,12], which activates a cell-killing phospholipase effector (CapV)[8]. This operon, like all type II CBASS, encodes two uncharacterized proteins termed CD-NTase-associated proteins 2 and 3 (Cap2 and Cap3; Fig. 1a)[7]. When expressed in *Escherichia coli*, *V. cholerae* CBASS confers broad resistance to phage infection (Fig. 1b, Extended data Fig. 1a,b). Consistent with previous reports[3], we found that the CD-NTase, CapV, and Cap2 are required for resistance, while Cap3 is dispensable (Fig. 1b, Extended data Fig. 1b).

To understand CD-NTase regulation in type II CBASS, we immunoprecipitated the CD-NTase from phage-infected bacteria (Extended data Fig. 1c,d). Mass spectrometry revealed that Cap2 copurified with the CD-NTase, suggesting that these two proteins form a complex (Fig. 1c, Supplementary Table 1). We confirmed the association between Cap2 and the CD-NTase using reciprocal immunoblots. Unexpectedly, the interaction is independent of phage infection (Fig. 1d,e, Extended data Fig. 1e–h).

Cap2–CD-NTase structure

To characterize the basis for the Cap2–CD-NTase interaction, we purified a stoichiometric Cap2–CD-NTase complex from a related CBASS found in *Enterobacter cloacae* and determined a 2.7 Å-resolution structure by cryoelectron microscopy (cryoEM; Fig. 2a–b, Extended data Fig. 1i–q, Extended Data Table 1, Supplementary Fig. 2). The structure reveals a 2:2 complex with a homodimer of Cap2 bound to two CD-NTase monomers (Fig. 2c). Cap2 adopts a modular architecture with three domains: an N-terminal E2-like domain,

a central linker domain, and a C-terminal E1-like domain (Fig. 2a). In eukaryotes, E1 and E2 domains catalyze the linkage of ubiquitin and related β-grasp fold proteins (collectively termed ubiquitin-like proteins or Ubls) to amine groups, typically lysine residues on target proteins[14]. This process begins when an E1 conjugates adenosine monophosphate (AMP) to the Ubl C-terminus (adenylation), then forms a thioester bond between the Ubl C-terminus and a catalytic cysteine in the E1. The Ubl is next shuttled to a cysteine residue on an E2, and is finally transferred to a target with the help of an E3 adapter protein[15].

In Cap2, the C-terminal adenylation/E1 domain forms a tight homodimer, similar to those observed in the bacterial E1 proteins MoeB and ThiF, which participate in sulfur metabolism[15–21]. The central linker domain of each Cap2 protomer reaches over the E1 domain of its dimer mate, positioning the N-terminal E2 domain close to the active site of its dimer-related E1 domain (Fig. 2c–d). Each Cap2 protomer is bound to a monomer of the CD-NTase via a composite interface involving one E1 domain and the nearby, dimer-related E2 domain (Fig. 2c–d). Mutations in the observed interface reduced or eliminated Cap2 binding to the CD-NTase (Extended data Fig. 2a–b). The bipartite Cap2–CD-NTase interaction appears to rigidify Cap2: a second structure from the same cryoEM dataset shows that when the Cap2 dimer is bound to only one CD-NTase, the linker and E2 domains of the unbound Cap2 protomer become flexible and are not observed in the cryoEM map (Extended data Fig. 2c–d).

The overall structure of Cap2 is strikingly similar to ATG7, a non-canonical E1 protein involved in autophagy in eukaryotes[14,15] (Extended data Fig. 3a–d). Like Cap2, ATG7 forms a homodimer through its C-terminal E1 domain. In both Cap2 and ATG7, the catalytic E1 cysteine is positioned on the “crossover loop” that extends over the E1 adenylation active site, rather than within an α-helical insertion in this loop as in canonical E1 proteins[22,23] (Extended data Fig. 3a–c). Next, the ATG7 N-terminal domain shares a common fold with the central linker domain of Cap2 and drapes over the dimer-related E1 domain in a manner similar to the Cap2 linker domain. In both Cap2 and ATG7, this domain binds and positions the E2 domain for catalysis[24–26] (Extended data Fig. 3a–d). Finally, comparison of the Cap2 E2 domain and the two noncanonical E2 proteins involved in ATG7-mediated ubiquitination (ATG3 and ATG10) reveals several features, including an incomplete UBC fold and a characteristic hairpin loop bearing the catalytic cysteine, that distinguish this group from canonical E2s (Extended data Fig. 3a,e). The unambiguous similarity of Cap2 to ATG7, plus homology between the Cap2 E2 domain and ATG3/ATG10, strongly suggest that these two systems share a common evolutionary origin distinct from canonical E1 and E2 machinery (Supplementary Discussion). Our structure suggests that Cap2 is an all-in-one transferase capable of protein ligation. Supporting this model, disruption of the Cap2 adenylation, E1, or E2 active sites eliminated the ability of *V. cholerae* CBASS to protect against phage infection (Fig. 2e, Extended data Fig. 4a,c).

Cap2 mediates CD-NTase conjugation

All known E1 enzymes use Ubls as substrates for adenylation and eventual conjugation to targets, but type II CBASS does not encode a Ubl. Our structure of the Cap2–CD-NTase complex reveals that that the extreme C-terminus of the CD-NTase (residues 375–381)

is bound to the Cap2 adenylation active site and conjugated to an AMP molecule (Fig. 2d, 3a, Extended data Fig. 2e–k, Extended Data Table 2), suggesting that the CD-NTase, rather than a ubiquitin-related β-grasp protein, is the substrate of Cap2-mediated conjugation (here, "substrate" denotes the equivalent to ubiquitin, while "target" denotes the protein/molecule to which the substrate is conjugated). The reactive-intermediate state captured in our structure closely matches prior structures of activated Ubls bound to their cognate E1 proteins[16,27,28], consistent with the CD-NTase serving as the Cap2 substrate. Further, CD-NTase enzymes in type II CBASS possess extended, disordered C-termini with a conserved C-terminal glycine or alanine residue, reminiscent of the C-terminal diglycine motif of ubiquitin (Fig. 3b, Supplementary Table 2[29,30], Extended data Fig. 5a,b).

If the CD-NTase is the substrate of Cap2, mutating or deleting the CD-NTase C-terminus or mutating the Cap2 E1 active site should destabilize the Cap2–CD-NTase complex and disrupt CBASS signaling. Accordingly, when we mutated the C-terminal glycine residue of the *V. cholerae* CD-NTase to glutamate (G436E), phage protection was lost (Fig. 3c, Extended data Fig. 4b,d). We further found that the Cap2–CD-NTase interaction in bacterial cells was compromised upon mutation of the Cap2 adenylation active site, the E1 catalytic cysteine, or the CD-NTase C-terminus (Fig. 3d, Extended data Fig. 4a,c). Importantly, inactivation of the Cap2 E2 catalytic cysteine did not disrupt the interaction with the CD-NTase (Fig. 3d). We anticipate that the E2 mutation preserves adenylation and conjugation of the CD-NTase to the E1 catalytic cysteine, trapping an intermediate state. In parallel, we established an in-cell Cap2 activity assay and found that, indeed, Cap2-mediated CD-NTase conjugation depends on the Cap2 adenylation active site, E1 and E2 catalytic cysteines, and the CD-NTase C-terminus (Fig. 3e, Extended data Fig. 4e–l).

We hypothesized that the CD-NTase C-terminus is transferred by Cap2 to target molecules similar to how ubiquitin is transferred to target molecules and tested this by probing for high molecular weight CD-NTase–target conjugates in *vivo*. Western blots demonstrated a series of novel, high molecular weight CD-NTase species in bacteria expressing the *capV-cd-ntase-cap2* operon. These species disappeared when any of the catalytic functions of Cap2 were disrupted (Fig. 3f). In total, these data demonstrate that Cap2 acts a transferase, catalyzing conjugation of the CD-NTase to an unidentified target.

Cap2 primes cGAMP synthesis

To understand the functional consequences of Cap2-mediated CD-NTase conjugation, we immunoprecipitated the CD-NTase from bacteria expressing the *capV-cd-ntase-cap2* operon and measured cGAMP synthesis by the purified protein. We found that CD-NTase purified from cells expressing wild-type Cap2 was significantly more active than CD-NTase from cells expressing Cap2 with E1 or E2 active-site mutations (Fig. 3g, Extended Data Fig 6c–f). These data demonstrate that cGAMP synthesis is enhanced by Cap2 activity. Surprisingly, the activity of purified CD-NTase was not altered by phage infection (Fig. 3h). These findings are consistent with the observation that the Cap2–CD-NTase complex is unaffected by phage infection (Fig. 1d,e). These data suggest that Cap2 covalently modifies the CD-NTase prior to phage infection to prime or license CBASS signaling, and that a further phage-mediated trigger is needed in cells to fully activate second messenger synthesis.

We next used mass spectrometry to identify the target(s) to which Cap2 conjugates the CD-NTase. We immunoprecipitated CD-NTase from bacteria expressing the *cap V-cd-ntase-cap2* operon with either wild-type or E1-mutant *cap2* alleles and quantified differentially enriched peptides (Extended data Fig. 8a–f, Supplementary Table 8). We identified 20 potential targets that were significantly enriched in samples expressing wild-type versus E1-mutant Cap2, with a strong bias toward proteins involved in metabolism (Extended data Fig. 8d, Supplementary Table 8). These data are consistent with a contemporary analysis of Cap2-mediated CD-NTase conjugation in an *E. coli* CBASS[31]. To determine if conjugation of the CD-NTase C-terminus to an arbitrary target protein is sufficient to increase CD-NTase activity, we measured cGAMP synthesis by CD-NTases fused to a C-terminal VSV-G tag, to GFP, or linked to the E1 active site of Cap2 (in a Cap2 E2 mutant). cGAMP synthesis by these proteins was equivalent to unconjugated CD-NTase (Fig. 3d,g and Extended data Fig. 5c–h, 7a,b), suggesting that CD-NTase activation depends on conjugation to a particular target protein or molecule. Future studies are required to establish the nature of the C-terminal modification and how it increases CD-NTase activity.

Cap3 antagonizes Cap2

CBASS systems that encode Cap2 invariably also encode Cap3, which is homologous to eukaryotic JAB/JAMM-family ubiquitin proteases[6,7] (Fig. 1a, Extended data Fig. 6h). We hypothesized that this protein balances CBASS activation by proteolytically cleaving CD-NTase–target conjugates. While deletion of *cap3* had no effect on CBASS-mediated phage resistance (Fig. 1b, Extended data Fig. 1b), we found that overexpression of Cap3 during infection strongly antagonized phage resistance (Fig. 4a, Extended data Fig. 6c). To directly measure Cap3 activity, we incubated *V. cholerae* and *E. cloacae* Cap3 with model substrates comprising their cognate CD-NTases fused at their C-terminus to GFP. Both Cap3 proteins precisely cleaved the CD-NTase–GFP fusions at the CD-NTase C-terminus, and this activity depended on conserved catalytic residues in the Cap3 active site and on Zn^{2+}, which is required for catalysis by JAB/JAMM family proteases (Fig. 4b,c Extended data Fig. 7, Supplementary Table 3). *V. cholerae* and *E. cloacae* Cap3 were unable to cleave substrates with mutations in the C-terminal region of their cognate CD-NTases (Extended data Fig. 7).

To further define the specificity of Cap3 *in vivo*, we overexpressed Cap3 alleles from four unrelated CBASS operons in combination with each of their cognate and non-cognate CBASS. Each Cap3 protein specifically antagonized phage protection by its cognate CBASS operon (Fig. 4d, Extended data Fig. 6d–g), demonstrating that Cap3 is exquisitely specific for its cognate CD-NTase.

Finally, we tested the ability of Cap3 to antagonize CD-NTase–target conjugates in cells. Overexpression of wild-type Cap3, but not a catalytic-dead mutant, eliminated the formation of Cap2-dependent high molecular weight CD-NTase species (Fig. 4e). Accordingly, Cap3 also antagonized cGAMP synthesis by CD-NTase immunoprecipitated from bacteria expressing Cap2 (Fig. 4f, Extended data Fig. 5c–f). These data strongly support a model in which CD-NTase activation is primed by Cap2-mediated conjugation and antagonized by Cap3-mediated cleavage of CD-NTase–target conjugates (Fig. 5a).

Bacterial E1, E2, and JABs are widespread

Antiphage systems constantly recombine and reassort into novel formulations that help bacteria gain an advantage in their conflict with phages[32–37]. We hypothesized that Cap2 and Cap3 homologs might be found in other antiphage systems and searched for these genes in CBASS-related Pycsar (pyrimidine cyclase system for antiphage resistance). Pycsar encode a phage-responsive cyclase (PycC) that generates a cyclic mononucleotide second messenger to activate an effector protein[38], analogous to type I CBASS. We identified a group of Pycsar, which we term type II Pycsar, that encode an E2-E1 fusion protein homologous to Cap2 (Pap2: PycC associated protein 2) and a protein homologous to Cap3 (Pap3; Fig. 5c). Supporting a model in which PycC is a substrate of Pap2 and Pap3, PycC proteins from type II, but not type I Pycsar have a highly conserved C-terminal extension (Fig. 5b–c) that structure predictions suggest binds to the Pap3 active site in a manner similar to CD-NTase–Cap3 binding (Extended data Fig. 8g).

Prior bioinformatic studies have identified at least five distinct families of bacterial operons encoding predicted E1, E2, and JAB domain proteins, one of which is now understood to be type II CBASS[20,39]. We found that a second family encodes predicted metallo-β-lactamase (MBL) alongside a Cap2-like E2-E1 protein fused to a C-terminal JAB domain (Fig. 5d, Extended data Fig. 8h,k). In this family, the MBL C-terminal region possesses a conserved glycine residue 4–6 amino acids upstream of the C-terminus (Extended data Fig. 8j), which suggests that the JAB domain processes a pro-form of the MBL prior to E2-E1–mediated conjugation; this idea is supported by structural predictions (Extended data Fig. 8i). We tested the idea that JAB domains can process pro-substrates by appending a VSV-G tag to the CD-NTase C-terminus. In our type II CBASS, *cap3* was required for phage resistance only when the C-terminal VSV-G was present, likely to remove the VSV-G tag *in vivo* and expose the native CD-NTase C-terminus for Cap2-mediated conjugation (Extended data Fig. 6a,b). These findings may also explain why a JAB domain-encoding gene from the recently described Bacterial ISG15-like system is essential for phage resistance[35].

Inspection of other operon families encoding E1, E2, and JAB domains shows that these operons also encode β-grasp fold Ubl proteins homologous to ubiquitin[20,39] (Fig. 5d, Extended data Fig. 8k). Overall, the abundance of bacterial operons encoding predicted E1, E2, and JAB-like proteins[18,20,40] along with both Ubls and other substrate proteins suggests that ubiquitin-like protein conjugation is widespread in bacteria.

The multifaceted regulation of CBASS

In this study, we show that the CBASS protein Cap2 is structurally homologous to ubiquitin transferases and conjugates a bacterial CD-NTase to an unidentified target molecule. The covalent CD-NTase adduct is primed for cGAMP synthesis and is essential for phage defense. How CD-NTase–target conjugation primes the CD-NTase for activation is unknown, but our finding that priming is independent of phage infection suggests that additional phage cues are required for full CD-NTase activation *in vivo*. CD-NTase priming can be reversed by Cap3, a sequence-specific protease (Fig. 5a). In mammals, ubiquitination is also required to prime the innate immune receptor RIG-I, which enhances signaling[41–43].

Taken together with our findings, E1 and E2-domain mediated protein conjugation may represent a conserved mechanism of immune regulation across kingdoms.

While bacterial proteins that catalyze ubiquitin conjugation have been identified[44], our findings reveal Cap2 as the first known all-in-one ATP-dependent ubiquitin transferase-like protein that combines adenylation, E1, and E2 active sites into a single polypeptide. Based on the lack of an E3 protein in CBASS operons and the apparent low specificity of Cap2-mediated conjugation, we hypothesize that target recognition is mediated directly by Cap2. The striking similarity of Cap2 to non-canonical E1 and E2 transferases from eukaryotes (ATG7 and ATG3/ATG10) suggests that these systems share a common evolutionary origin. Thus, while ancestors of canonical ubiquitin signaling can be found throughout eukaryotes and in some archaea[45,46], E1 and E2 transferases may have evolved first in bacteria, in line with prior bioinformatic observations[18,20]. Our work also reveals bacterial CD-NTases as the first known substrates of ubiquitin transferase-like systems that do not share the β-grasp fold of Ubls[47]. The all-in-one nature of Cap2, and its unique mode of substrate recognition, may enable engineering of this system to mediate customizable post-translational modifications. Similarly, a complete understanding of Cap3 specificity will enable future applications of these proteases for site-specific cleavage.

What is the selective benefit of type II CBASS encoding Cap2 and Cap3 over simpler type I CBASS that encode only a CD-NTase and effector? We hypothesize that Cap2 may increase CBASS sensitivity or license the CD-NTase to control inappropriate or spurious activation. CBASS operons with *cap2* always encode *cap3*[6], suggesting that although *cap3* is dispensable for phage resistance, it nonetheless provides a fitness advantage. The Cap2/3 signaling scheme is reminiscent of type III CBASS, which encode HORMA-like proteins (Cap7/8) required for CD-NTase activation and a TRIP13-like protein (Cap6) that disassembles activated CD-NTase–HORMA and primes HORMA proteins for peptide binding and CD-NTase activation[4,10]. The apparent dual roles of Cap6 in type III CBASS suggests that Cap3 may also play two roles: first, to limit spurious CD-NTase priming and activation, and second to disassemble nonspecific CD-NTase conjugates to recycle CD-NTase that can be specifically primed for activation. Taken together, our findings show that diverse CBASS systems use multifaceted positive and negative regulators to finely control the activation of cGAS/DncV-like enzymes and mediate broad antiphage immunity.

Methods

Bacterial strains and growth conditions

E. coli strains used in this study are listed in Supplementary Table 4. *E. coli* were cultured in LB medium (1% tryptone, 0.5% yeast extract, 0.5% NaCl) shaking at 37 °C, 220 rpm unless otherwise noted. For phage experiments and other noted assays, bacteria were grown in “MMCG” minimal medium containing M9 salts, magnesium, calcium, and glucose (47.8 mM Na_2HPO4, 22 mM KH_2PO_4, 18.7 mM NH_4Cl, 8.6 mM NaCl, 22.2 mM glucose, 2 mM $MgSO_4$, 100 μM $CaCl_2$, 3 μM Thiamine). Where applicable, media were supplemented with carbenicillin (100 μg/mL) or chloramphenicol (20 μg/mL), to ensure plasmid maintenance. When a strain with two plasmids was cultivated in MMCG medium, bacteria were cultured with 20 μg/mL carbenicillin and 4 μg/mL chloramphenicol. We defined an “overnight”

culture as 16–20hrs post-inoculation from a single colony or glycerol stock. All strains were stored in LB plus 30% glycerol at −70 °C. *E. coli* OmniPir[54] was utilized for plasmid construction and propagation whereas *E. coli* MG1655 (CGSC6300) was employed for all experimental data.

Plasmid Construction

Plasmids used in the study are listed in Supplementary Table 4. All experiments were performed with either the CBASS system from *V. cholerae* C6706 (NCBI RefSeq NZ_CP064350.1) or *E. cloacae* (NCBI RefSeq NZ_KI973084.1; see protein accession numbers in Supplementary Table 5) with the exception of the Cap3 overexpression experiments presented in Fig. 4d and Extended data Fig. 6d–g which also utilized the CBASS operons associated with CD-NTase 38, 42, and 127 (NCBI RefSeq WP_032676400, WP_000992191.1, and WP_052435251.1). For phage infections, the entire operon plus surrounding sequences were cloned into the XhoI and NotI sites of the vector pLOCO2[54]. For *in vivo cap3* expression, genes were cloned into the BamHI and NotI sites of the vector pTACxc. pTACxc (full sequence available in Supplementary Table 4) was constructed by combining the ColE1 origin of replication from pBAD24[55], chloramphenicol resistance from pBAD18cm[55], the RP4 *oriT*, *lacI*^{q} from OmniPir *E. coli*, a P_{tac} promoter[56], and *sfGFP*. For immunoprecipitation assays, a C-terminal 3×FLAG tag was added to *V. cholerae cap2* and an N-terminal VSV-G tag was added to *V. cholerae* CD-NTase. A C-terminal VSV-G tag was added to *V. cholerae* CD-NTase to test target specificity. For biochemical analysis, individual proteins were cloned into vector 2-BT (Addgene #29666; N-terminal His_6-TEV cleavage site fusion), H6-msfGFP (Addgene #29725; N-terminal His_6-TEV cleavage site fusion and C-terminal msfGFP fusion), or 2-AT (Addgene #29665; untagged).

For *E. cloacae* Cap3, sequence alignments revealed that the first 16 amino acids of the annotated gene are unlikely to be translated *in vivo*; a truncated construct comprising residues 17–180 of the annotated gene expressed at higher levels and was more soluble upon purification (for mutations, residue numbering follows the annotated gene). For *E. cloacae* Cap2-CD-NTase complex used for cryoEM, the two genes were PCR-amplified from vector 2-AT and combined to generate a polycistronic transcript, then cloned into vector 2-BT resulting in an N-terminal His_6-tag on CD-NTase and no tag on Cap2, and both catalytic cysteine residues in Cap2 (C109 and C548) were mutated to alanine. For the *E. cloacae* Cap2-CD-NTase complex used the in cell Cap2 activity assay, the two genes were cloned as above into vector 2-BT to generate a polycistronic transcript with an N-terminal His_6-tag on Cap2 and no tag on CD-NTase. For *E. cloacae* Cap2-CD-NTase complex with HA-tagged CD-NTase, the two genes were cloned as above into vector 2-AT to generate a polycistronic transcript with an N-terminal HA tag (MYPYDVPDYAGSG) fused to residue 2 of CD-NTase.

DNA sequences were cloned into destination vectors using 18–25 bp overhangs and Gibson Assembly. Point-mutations and epitope tags were cloned by mutagenic PCR and isothermal assembly. Clones were transformed either into a modified strain of OmniMax *E. coli* (Invitrogen) by electroporation, or into NovaBlue *E. coli* (Novagen) by heat-shock and plated on LB with the appropriate selection. Positive clones were verified by

Sanger Sequencing (Genewiz). Prior to use in downstream phage or immunoprecipitation experiments, sequence verified plasmids were transformed into MG1655 via heat shock and plated on LB with the appropriate selection.

Phage Amplification and Storage

Phages used in the study are listed in Supplementary Table 6. Phage lysates were generated from *E. coli* MG1655 using a modified double agar overlay plate amplification (T2) or liquid amplification (T4, T5, T6). For plate amplification, stationary phase MG1655 was infected with 10,000 plaque forming units (PFU) of phage in LB + 0.35% agar, 10 mM $MgCl_2$, 10 mM $CaCl_2$, and 100 μM $MnCl_2$. Plates were incubated overnight (16–20 hours) at 37 °C and the following day phages were harvested by adding 5 mL of SM buffer (100 mM NaCl, 8 mM $MgSO_4$, 50 mM Tris-HCl pH 7.5, 0.01% gelatin) directly to the plate, incubating for 1 hour at room temperature, then collecting and filtering the resulting liquid through a 0.2 μm Nanosep filter. For liquid amplification, early logarithmic phase MG1655 was infected at an MOI of 0.1 in 25 mL LB broth plus 10 mM $MgCl_2$, 10 mM $CaCl_2$, and 100 μM $MnCl_2$ at 37 °C with 220 rpm shaking for 2–6 hours until the culture became clear. Supernatants were then collected via centrifugation and filtration with a 0.2 μm Nanosep filter. Lysate titers were determined by spotting a serial dilution of the phage onto 0.35% LB agar plus 10 mM $MgCl_2$, 10 mM $CaCl_2$, and 100 μM $MnCl_2$ containing stationary phase MG1655. Plates were incubated overnight at 37 °C and the resulting PFU/mL was calculated. Phage stocks were stored at 4 °C in either SM buffer or LB broth.

Efficiency of plating/phage infection assays

Phage protection assays were performed using a modified double agar overlay technique[57]. Bacteria were cultivated overnight in MMCG medium, and the following day were diluted 1:10 into fresh medium and grown until mid-logarithmic phage. 400 μL of MG1655 containing the indicated vector(s) was inoculated into 3.5 mL 0.35% MMCG-agar, mixed, and poured on top of a conventional MMCG 1.6% agar plate. For the Cap3 overexpression experiments, 0, 50, or 500 μM IPTG was added to both the bacterial culture and the top agar. The plate was allowed to cool/dry for ~10 minutes after which 2 μL of phage serial dilution was spotted onto the soft agar overlay. After phage spots dried, plates were incubated at 37 °C overnight. Plates were imaged ~24 hours after infection and PFU were enumerated.

The resulting efficiency of plating for each phage was measured by quantifying PFU/mL for each phage lysate tested. PFU were enumerated for phage dilution spots with 1–30 PFU, then the dilution was used to scale PFU/mL appropriately. When individual plaques could not be counted and instead a hazy zone of clearance was observed, the lowest phage concentration at which we could detect this clearance was counted as ten plaques. When no clearance was observed, 0.9 plaques at the least dilute spot were used as the limit of detection for that assay (see Extended data Fig. 1a for an example).

The data is presented as fold protection compared to a control strain expressing GFP, which is simply the inverse of the efficiency of plating. Data is shown as the mean +/− SEM of three biological replicates. Statistical significance, as determined by an unpaired two-sided Student's *t*-test, is shown when applicable. For large and obviously significant differences

between data, such as greater than 100-fold, statistics are not indicated for clarity. All the raw data from these experiments along with the relevant P-values are found in Source Data 1.

Immunoprecipitation assays

MG1655 *E. coli* expressing the indicated vector(s) were grown to mid-logarithmic phase in MMCG. Where listed, cells were infected with the indicated phage for 30 minutes (or as noted) at an MOI of 2. Cultures were then centrifuged, and the resulting pellet was resuspended in lysis buffer (400 mM NaCl, 20 mM Tris-HCL pH 7.5, 2% glycerol, 1% triton, and 1 mM 2-mercaptoethanol). Cells were disrupted by sonication followed by centrifugation at 4 °C to remove cellular debris. Soluble lysates were then mixed with the epitope tag purification resin, as described below, overnight at 4 °C with end-over-end rotation. The following day, samples were washed five times in 1–5 mL lysis buffer and beads were processed for downstream application. For CD-NTase immunoprecipitations, lysates were incubated with either protein A magnetic beads (Pierce) containing 10 μg/mL αCD-NTase antibody or, when CD-NTase had a VSV-G tag, with agarose beads conjugated to an αVSV-G antibody (Sigma). Cap2–3×FLAG was immunoprecipitated using magnetic beads covalently linked to the αFLAG M2 antibody (Sigma).

Western blots

Rabbit αCD-NTase polyclonal antibody was generated by a commercial vendor (Genescript) using a purified, untagged CD-NTase antigen. Polyclonal αCD-NTase antibodies were further purified by antigen affinity (GenScript). Serum was used at 1:30,000 for CD-NTase immunoblot detection. αFLAG antibody (Sigma) was used at 1:10,000 to detect Cap2–3×FLAG, αVSV-G (Rockland) was used at 1:7,500 to detect VSV-G tagged CD-NTase, αRNAP (Biolegend) was used at 1:5000 for use as a loading control, and αHA (clone 3F10, Sigma-Aldrich) was used at 1:30,000 to detect HA-tagged proteins.

For whole cell lysate analysis, 5 mL of MG1655 carrying the indicated plasmid were grown to mid-logarithmic phase. Cell densities were then normalized and 5×10^9CFU were collected, centrifuged and resuspended in 50 μL of 1× LDS buffer (106 mM Tris-HCl pH7.4, 141 mM Tris Base, 2% w/v Lithium dodecyl sulfate, 10% v/v Glycerol, 0.51 mM EDTA, 0.05% Orange G). Samples were then incubated at 95 °C for 10-minutes followed by a 5-minute centrifugation at 20,000×g to remove debris. For immunoprecipitation samples, affinity purification beads were resuspended in 40 μL lysis buffer plus 40 μL 2x LDS buffer. Samples were then incubated at 95 °C followed by a 5-minute centrifugation at 20,000×g.

Samples in LDS were loaded at equal volumes to resolve using SDS-PAGE, then transferred to PVDF membranes charged in methanol. Membranes were blocked in Licor Intercept Buffer for 30 minutes at 24 °C, followed by incubation with primary antibodies diluted in Intercept buffer overnight at 4 °C. Blots were then incubated with the appropriate combination of Licor infrared (800CW/680RD) αRabbit/Mouse secondary antibodies at 1:30,000 dilution in TBS-T (0.1% Triton-X) for 45 minutes at 24 °C and visualized using the Licor Odyssey CLx. For αHA immunoblots, HRP-linked Goat αRat antibody (Pierce

31470) was used at 1:30,000 and detected with an HRP Substrate kit (Bio-Rad) and Bio-Rad ChemiDoc imager. Representative images were assembled using Adobe Illustrator CC 2022.

Mass Spectrometry analysis

Following IP enrichment as described above, samples were subjected to on bead trypsin digest followed by analysis on a Thermo Obitrap Q-Exactive HF-X using nanoLC-MS MS. Peptides were mapped to the proteome of *E. coli* MG1655 (https://www.uniprot.org/proteomes/UP000030788), the proteins composing the CBASS operon from *V. cholerae* (CapV, CD-NTase, Cap2, Cap3) and the proteome of the phage T2 (https://www.uniprot.org/proteomes/UP000503557) which was used to infect the samples. Peptides were considered significantly enriched when their LFQ score was $>10^8$ and they were more than 4-fold enriched over the $Cap2^{C522A(E1)}$ samples.

CD-NTase Enzyme Assay

6.25×10^9 CFU of MG1655 expressing the indicated plasmids were processed for IP enrichments as described above using 20 μL bead volume. Of note, the data present in Fig. 3g,h is using a vector in which *cap3* has been deleted. α-VSVG agarose beads were further washed three times in 1 mL reaction buffer (50 mM Tris-HCl pH 7.5, 50 mM KCl, 5 mM $MgCl_2$). Samples were then resuspended in 120 μL reaction buffer, split into three technical replicates and incubated with 500 μM ATP and 500 μM GTP overnight at 37 °C followed by 10 minutes at 95 °C to inactivate the CD-NTase. cGAMP levels were then quantified with the Arbor Assay 3',3'-cGAMP ELISA per manufacturer's specifications. When necessary, samples were diluted 1:5 and 1:25 in the provided assay buffer to ensure measurements were in the dynamic range. cGAMP levels were calculated using the provided standard measured in triplicate. Data shown are a representative graph from one of three independent experiments depicting the mean +/− the standard deviation of three technical replicates. To allow for accurate comparison of all the ELISA data see Extended Data Figure 6c,d.

For *in vitro* second messenger synthesis assays, 1 mL reactions with 1 μM CD-NTase or CD-NTase-GFP fusion proteins were mixed with reaction buffer containing 10 mM Tris-HCl pH 8.5, 12.5 mM NaCl, 20 mM $MgCl_2$, 1 mM DTT, 0.25 mM ATP, and 0.25mM GTP, then incubated at 37°C for 16 hours. Ten units of calf intestinal phosphatase (Quick CIP, New England Biolabs) were added and further incubated at 37°C for 2 hours. Reactions were stopped by heating at 65 °C for 20 minutes, then centrifuged at 15,000 RPM for 10 minutes to remove precipitated protein. Reaction products were separated by anion-exchange chromatography (1 mL Hitrap Q HP, Cytiva) using a gradient from 0.5 to 2M ammonium acetate, on an Akta PURE 25M FPLC (Cytiva). Products were quantified by peak-area integration from the A_{254} absorbance profile using Unicorn v. 7.3 (Cytiva).

Protein Alignments

Cap2 and Cap3 protein alignments were generated with the MUSCLE algorithm[58] within Geneious software, then adjusted by hand based on structure superpositions performed using the PDBeFold server (https://www.ebi.ac.uk/pdbe/). Sequence logos were generated using WebLogo (https://weblogo.berkeley.edu/logo.cgi).

Protein Expression and Purification

Protein expression vectors used in the study are listed in Supplementary Table 4. For protein purification, expression vectors were transformed into *E. coli* Rosetta2 pLysS (EMD Millipore) or LOBSTR (Kerafast), grown at 37 °C in 2×YT media to an OD_{600} of 0.6, then protein expression was induced by the addition of 0.25 mM IPTG (isopropylthio-β-galactoside). Cultures were shifted to 20 °C for 16 hours, then cells were harvested by centrifugation. Cells were resuspended in binding buffer (25 mM Tris-HCl pH 8.5, 5 mM imidazole, 300 mM NaCl, 5 mM $MgCl_2$, 10% glycerol, and 5 mM 2-mercaptoethanol), lysed by sonication, and centrifuged (20,000 ×g for 30 minutes) to remove cell debris. Clarified lysate was passed over a Ni^{2+} affinity column (Ni-NTA Superflow, Qiagen) and eluted in a buffer with 250 mM imidazole. For cleavage of His_6-tags, proteins were buffer-exchanged to binding buffer, then incubated 48 hours at 4 °C with His_6-tagged TEV protease[59]. Cleavage reactions were passed through a Ni^{2+} affinity column again to remove uncleaved protein, His_6-tags, and TEV protease. Flow-through fractions were passed over a size-exclusion chromatography column (Superdex 200; Cytiva) in gel filtration buffer (25 mM Tris-HCl pH 8.5, 300 mM NaCl, 5 mM $MgCl_2$, 10% glycerol, 1 mM DTT). Gel filtration buffer without glycerol was used for samples for cryoelectron microscopy. Purified proteins were concentrated and stored at −80 °C for analysis or 4 °C for crystallization.

Cryoelectron Microscopy

For grid preparation, freshly purified *E. cloacae* Cap2–CD-NTase complex was collected from size-exclusion chromatography and diluted to 8 μM. Immediately prior to use, Quantifoil Cu 1.2/1.3 300 grids were glow-discharged for 10 sec in a pre-set program using a Solarus II plasma cleaner (Gatan). Sample was applied to a grid as a 3.5 μL drop in the environmental chamber of a Vitrobot Mark IV (Thermo Fisher Scientific) held at 4 °C and 100% humidity. After a 1-minute incubation, the grid was blotted with filter paper for 5 seconds prior to plunging into liquid ethane cooled by liquid nitrogen. Grids were mounted into standard AutoGrids (Thermo Fisher Scientific) for imaging.

All samples were imaged using a Titan Krios G3 transmission electron microscope (Thermo Fisher Scientific) operated at 300 kV configured for fringe-free illumination and equipped with a K2 direct electron detector (Gatan) mounted post Quantum 968 LS imaging filter (Gatan). The microscope was operated in EFTEM mode with a slit-width of 20 eV and using a 100 μm objective aperture. Automated data acquisition was performed using EPU (Thermo Fisher Scientific) and all images were collected using the K2 in counting mode. Ten-second movies were collected at a magnification of 165,000x and a pixel size of 0.84 Å, with a total dose of 64.8 $e^-/Å^2$ distributed uniformly over 40 frames. In total, 2437 movies were acquired with a realized defocus range of −0.5 to −2.5 μm.

CryoEM data analysis was performed in cryoSPARC version 3.2[60] (Extended data Fig. 1i–q, Extended Data Table 1). Movies were motion-corrected using patch motion correction (multi) and CTF-estimated using patch CTF estimation (multi)[61], and a 200-image subset was used for initial particle picking using the blob picker. Initial picks were subjected to 2D classification, and two classes were picked as templates for template-based particle picking of the entire 2437-image dataset. 2D classification of the resulting ~2.5M particles

revealed relatively few high-quality classes, likely a result of particle shape irregularity and high density on the grids. A ~1M particle subset was used for initial ab initio 3D reconstruction, then the entire ~2.5M particle dataset was used for heterogeneous refinement against these models, resulting in a 663,199-particle set that gave a ~5.5 Å resolution map with recognizable protein features. This particle set was subjected to a further round of ab initio reconstruction and heterogeneous refinement to separate 2:2 Cap2:CD-NTase complexes from 2:1 complexes. The two separate particle sets were cleaned with another round of 3D classifications, then re-extracted with a 440-pixel (370 Å) box size and refined using the Non-Uniform Refinement NEW job type in cryoSPARC with the following options enabled: Maximize over-particle scale; Optimize per-particle defocus; Optimize per-group CTF params. For the 2:2 complex, C2 symmetry was applied during refinement. The resulting reconstructions showed resolution values of 2.74 Å (2:2 complex) and 2.91 Å (2:1 complex) using the 0.143-cutoff criterion of the Fourier shell correlations between masked independently refined half-maps. Resolution anisotropy for both reconstructions was assessed using the 3DFSC web server[48].

An initial model for *E. cloacae* Cap2 was generated by AlphaFold2[52]. This model and the crystal structure of ATP-bound *E. cloacae* CD-NTase (PDB ID 7LJL[30]) were manually docked into the final 2:2 complex cryoEM map using UCSF Chimera[62] and rebuilt in COOT[63]. For the E1 domain of Cap2 and for CD-NTase, high-resolution crystal structures were used to verify the accuracy of the resulting model. The final rebuilt model was real-space refined in phenix.refine[64]. This model was then docked into the 2:1 complex map, disordered regions were deleted, and the final model was real-space refined in phenix.refine[72]. Structure validation was performed with MoProbity[65] and EMRinger[66]. Structures were visualized in ChimeraX[62] and PyMOL (Schrödinger, LLC).

Crystallography

To determine a crystal structure of the Cap2 E1 domain bound to the CD-NTase C-terminus in the apo state, we cloned and purified a fusion construct with *E. cloacae* Cap2 residues 374–600 (C548A mutant) fused at its C-terminus to a flexible linker and residues 370–381 of CD-NTase (sequence: GSGKPAEPQKTGRFA). Purified protein was exchanged into a buffer containing 25 mM Tris-HCl pH 8.5, 200 mM NaCl, 5 mM $MgCl_2$ and 1 mM TCEP, then concentrated to 30 mg/mL. Small rod-shaped crystals grew in hanging drop format by mixing 1:1 of protein with well solution containing 0.1 M Tris-HCl pH 8.5, 0.8 M LiCl, and 25% PEG 3350. Crystals were transferred to a cryoprotectant containing an additional 10% glycerol, then flash-frozen in liquid nitrogen. We collected a 1.77 Å resolution diffraction dataset at NE-CAT beamline 24ID-C at the Advanced Photon Source at Argonne National Laboratory (Extended Data Table 2). Data were processed with the RAPD pipeline, which uses XDS[67] for data indexing and reduction, AIMLESS[68] for scaling, and TRUNCATE[69] for conversion to structure factors. We determined the structure by molecular replacement in PHASER[70] using the refined Cap2 E1 domain structure from our cryoEM model of Cap2-CD-NTase. The model was rebuilt in COOT[63], followed by refinement in phenix.refine[71] using positional, individual B-factor, and TLS refinement (statistics in Extended Data Table 2).

To determine a crystal structure of the Cap2 E1 domain bound to the CD-NTase C-terminus in the AMP-bound reactive intermediate state, we cloned and purified a fusion construct with *E. cloacae* Cap2 residues 363–600 (C548A mutant) fused at its C-terminus to a flexible linker and residues 370–381 of CD-NTase (sequence: GSGKPAEPQKTGRFA). Purified protein was exchanged into crystallization buffer and concentrated to 30 mg/mL. A final concentration of 2.5 mM ATP was added to the protein and incubated overnight at 4 °C. Needle crystals grew in hanging drop format by mixing 1:1 of protein with well solution containing 0.1 M Tris-HCl pH 8.5, 0.2 M $MgCl_2$, and 30% PEG 3350). Crystals were looped directly from the drop and flash-frozen in liquid nitrogen. A 2.11 Å resolution diffraction dataset was collected at NE-CAT beamline 24ID-E at Advanced Photon Source at Argonne National Laboratory and processed as above.

Cap2/3 Biochemical Assays

For Cap2 activity assays, the indicated combinations of *E. cloacae* His_6-Cap2 and untagged CD-NTase (wild-type or mutant) were coexpressed in Rosetta2 pLys *E. coli* cells, then purified as above using a Ni^{2+} affinity column. Samples were analyzed by SDS-PAGE with Coomassie staining. For quantitation, experiments were run in triplicate and Coomassie blue-stained bands quantified using Fiji software[72]. For Cap3 activity assays, model substrates comprising *E. cloacae* or *V. cholerae* His_6-CD-NTase (wild-type or mutant) fused at their C-terminus to GFP were cloned and purified as above. Model substrates (4.5 μg) were incubated with Cap3 (1.5 μg) in a reaction buffer with 20 mM HEPES pH 7.5, 100 mM NaCl, 20 mM $MgCl_2$, 20 μM $ZnCl_2$, and 1 mM DTT (20 μL total reaction volume). Reactions were incubated 30 minutes at 37 °C, then analyzed by SDS-PAGE with Coomassie blue staining.

Trypsin Mass Spectrometry

For trypsin mass spectrometry of purified proteins (HA-CD-NTase and Cap2-GFP), in-gel digestion was performed according to a previously described method[73]. Briefly, proteins in diced gel bands were reduced by 100 μL of 10 mM DTT (Dithiothreitol) for 30 minutes at 37 °C and then alkylated by 6 μL of 0.5 M iodoacetamide in water for 20 minutes at room temperature in the dark. To digest proteins, 25–30 μL of 10 ng/μL Trypsin (Promega, V511A) in 50 mM ammonium bicarbonate (pH 8) was added to cover the gel pieces and incubated on ice for 30 minutes until fully swollen. An additional 10–20 μL of ammonium bicarbonate buffer was added and the sample was incubated overnight at 37 °C. The next day, trypsin digested peptides were extracted from the gel via multiple solvent extractions, dried under vacuum and then resuspended in 5 μL of 0.6% acetic acid. The digested peptides were analyzed by a Thermo Fisher Scientific Orbitrap Fusion LUMOS Tribrid mass spectrometer using standard LC-MS/MS method[73].

MS data analysis was performed using the Trans-Proteomic Pipeline (TPP, Seattle Proteome Center). Briefly, MS data were searched using the search engine COMET against a composite *E. coli* database that additionally contained protein sequences for *E. cloacae* CD-NTase and Cap2, plus common contaminants. Variable modifications include possible oxidation of methionine (15.9949 Da) and expected FA remnant of the CD-NTase C-terminus (218.10552 Da); and a static modification of cysteine by IAA (57.021464 Da)

was included. The COMET search results were further analyzed with PeptideProphet and ProteinProphet[74]. Peptides with a probability of >0.9 and mass accuracy of <10 ppm were subjected to further manual inspection of the MS/MS spectra to confirm major fragment ions are accounted for.

Bioinformatic analyses

The CD-NTase alignments and tree in Fig. 3b and Extended Data Fig. 5 were adapted from previously published datasets[2,6,35] (Supplementary Table 10). CD-NTase clades with >~75% of CD-NTases encoded adjacent to Cap2 and Cap3 homologs were deemed type II systems and highlighted. For each subset of CD-NTases indicated, C-terminal residues were extracted and aligned to create a sequence logo using WebLogo (https://weblogo.berkeley.edu/logo.cgi). Sequence logos in Extended Data Fig. 5a use data from Milman et al., while sequence logos in Figure 3b and Extended Data Fig. 5b used data from Whiteley et al. and Burroughs et al.

To identify Pycsar systems that contain E1, E2, and JAB domains we initially searched the Integrated Microbial Genomes (IMG) database for homologs of the cyclase, *pycC* from *Pseudomonas aeruginosa* (2736613764). All available hits for this gene were downloaded along with 10,000 bp upstream and downstream of the gene. The operon predictor Glimmer was used on each region of DNA and all identified genes were extracted and translated. Interpro was used to predict the protein domains found within each extracted protein sequence. In order to identify E1/E2 proteins, we searched for the protein domain ThiF, which is an E1 domain. We then confirmed that these sequences also encode for an E2 and JAB domain containing protein. All *pycC* that were associated with these domains were then extract, translated and the last 9 amino acids were aligned to generate a sequence logo.

We then broadened our search to include E1 and E2 domains that were previously reported[20,39]. We expanded upon their analysis and IMG was utilized to identify homologs of the genes encoding these proteins and a representative 500 genes +/− 10,000 bp were extracted. We again utilized Glimmer and Interpro to identify protein domains associated with E1 and E2 domains. From this analysis we identified numerous operons that could be divided into 4 broad classes, those that contain an MBL-domain, those with a CEHH domain, αhelical- domain containing operons, and finally those that contain a DUF6527 domain. Representatives of each operon architecture (Fig. 5d. Extended data Fig. 8k) were identified. Data can be found in Supplementary Table 9.

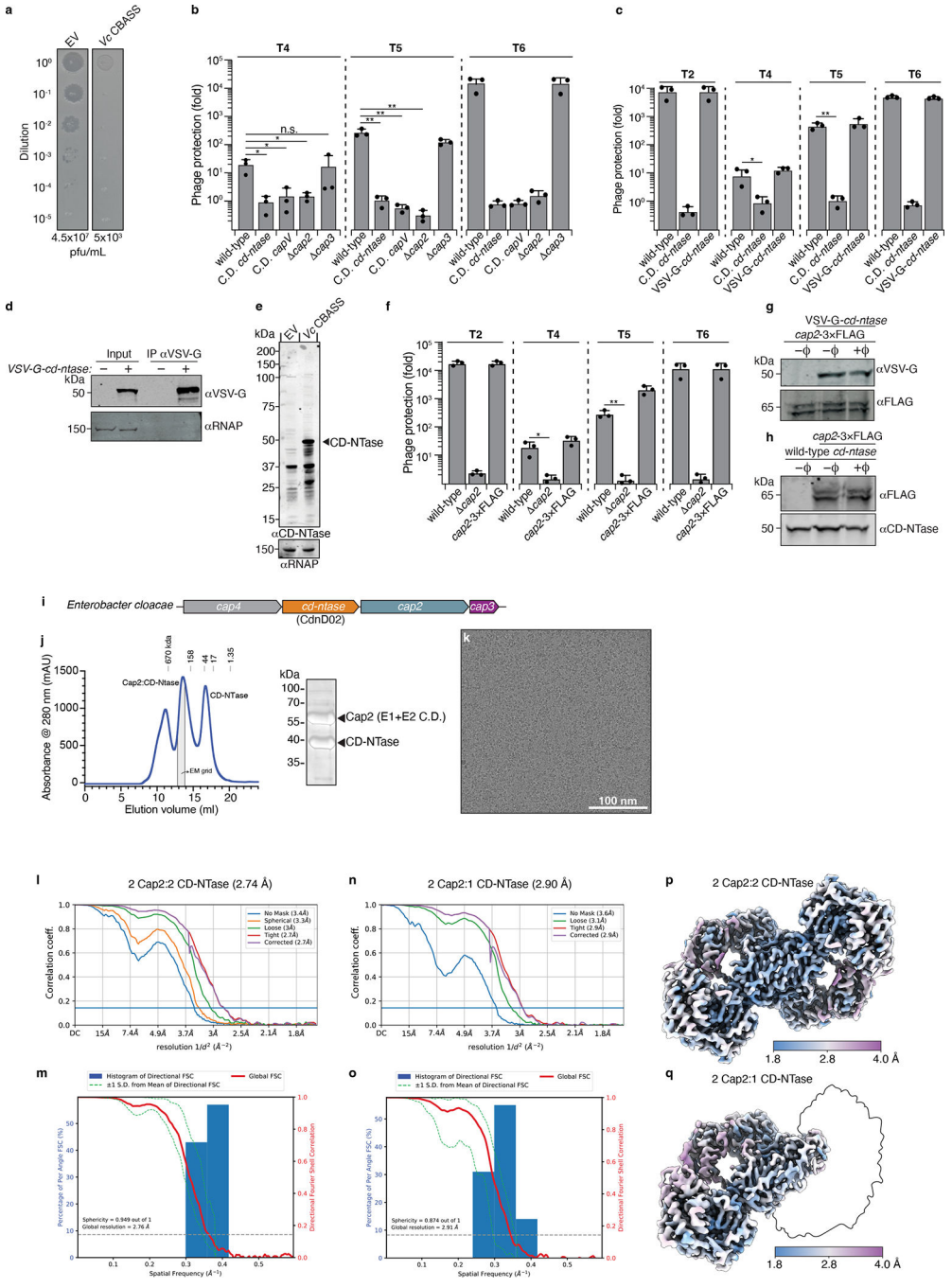
For structure/model predictions, we used a local installation of ColabFold (https://github.com/YoshitakaMo/localcolabfold)[75], which implements the AlphaFold[52] and AlphaFold-Multimer[76] algorithms.

Statistics and Reproducibility

All efficiency of plating phage assays performed within the paper N=3 independent biological replicates observed on different days. Data is presented as the mean +/− SEM and a two-sided Student's t-test was used to calculate significance; n.s., $p>0.05$; *, $p<0.05$; **, $p<0.001$. The actual p-values are listed in Source Data 1.

All western blots and Coomassie analysis presented are representative of N=3 independent biological replicates. This includes Fig. 1d, 1e, 3d, 3f, 4b, 4e and Extended data Fig. 1d, 1e, 1g, 1h, 3b, 5c–e, 5g, 5i, 5l, 6e, 6f, 7b, 8c, 8e, 8f, 9a, and 9b. Mass spectrometry analysis was performed twice on two independent biological replicates for each experiment conducted.

Extended Data



Extended Data Figure 1. Phage protection assays, inputs for IPs, and CD-NTase antibody verification along with CryoEM information.

(a) Image of double agar overlay phage infection assay used to measure efficiency of plating for a lysate of phage T2. *E. coli* MG1655 expressing the indicated vectors is shown. Zones of clearance (plaques) represent successful phage infection and replication. Apparent plaque forming units (PFU) per mL is calculated for the lysate infecting each bacterial genotype. Fold protection is the PFU per mL of empty vector divided by *Vc* CBASS, ~10^4 in this assay.

(b) Efficiency of plating of the indicated phage when infecting *E. coli* expressing CBASS with the indicated genotype. Data plotted as in Fig. 1b. C.D. CD-NTase: DID131AIA.; C.D. *capV*: C62A.

(c) Efficiency of plating of the indicated phage when infecting *E. coli* expressing *V. cholerae* CBASS with the indicated genotypes. Data plotted as in Fig. 1b.

(d) Western blot analysis of cell lysates (inputs) and αVSV-G immunoprecipitation of *E. coli* expressing CBASS with the indicated genotypes. These samples correspond to the mass spectrometry in Fig. 1c. αRNAP western blot serves as a loading control for bacterial cells. (−): CBASS operon, CD-NTase without VSV-G; (+): CBASS operon, CD-NTase with N-terminal VSV-G.

(e) Whole cell western blot analysis of *E. coli* expressing either an empty vector (EV) or CBASS (wild-type). αCD-NTase Western blot used a custom CD-NTase antibody; arrow indicates monomeric CD-NTase at the expected molecular weight. αRNAP western blot serves as a loading control for bacterial cells.

(f) Efficiency of plating of the indicated phage when infecting *E. coli* expressing CBASS with the indicated genotypes. Data plotted as in Fig. 1b.

(g) Whole cell western blot analysis of *E. coli* expressing the indicated genotypes of CBASS. Data are the input for the immunoprecipitation presented in Fig. 1d.

(h) Whole cell western blot analysis of *E. coli* expressing the indicated genotypes of CBASS. Data are the input for the immunoprecipitation presented in Fig. 1e. For (f),(g) and (h) ±ϕ indicates phage T2 at an MOI of 2.

(i) Operon structure of CBASS from *E. cloacae*. See Supplementary Table 5 for relevant accession numbers.

(j) Size exclusion chromatography elution profile (Superdex 200 Increase 10/300 GL) and SDS-PAGE analysis of *E. cloacae* Cap2–CD-NTase. The fraction used for cryoEM analysis is shaded in gray. C.D. Cap2: C109A/C548A

(k) Representative electron micrograph of *E. cloacae* Cap2–CD-NTase.

(l) Fourier Shell Correlation (FSC) curve for the final refinement of the 2:2 Cap2–CD-NTase complex.

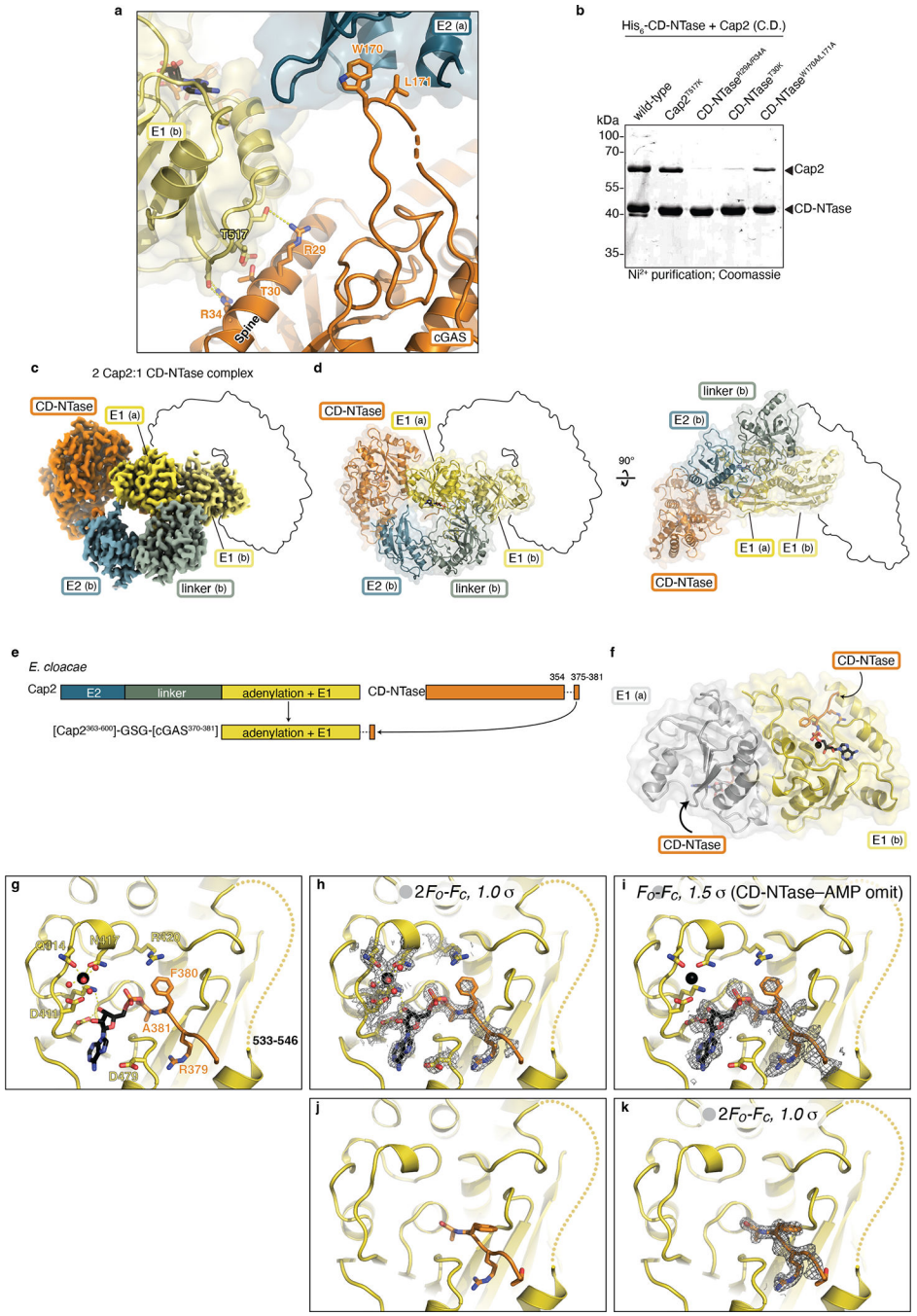
(m) 3D FSC analysis[48] for the 2:2 Cap2–CD-NTase complex.

(n) Fourier Shell Correlation (FSC) curve for the final refinement of the 2:1 Cap2–CD-NTase complex.

(o) 3D FSC analysis for the 2:1 Cap2–CD-NTase complex.

(p) Local resolution of the final refined map for the 2:2 Cap2–CD-NTase complex, colored from blue (≤1.8 Å) to magenta (≥4.00 Å).

(q) Local resolution of the final refined map for the 2:1 Cap2–CD-NTase complex colored from blue (≤1.8 Å) to magenta (≥4.00 Å). Outline indicates the areas of missing density compared to the 2:2 Cap2–CD-NTase complex.



Extended Data Figure 2. The CD-NTase rigidifies Cap2 through a bipartite interaction and crystal structures of an *E. cloacae* Cap2 E1-CD-NTase fusion.

(a) Closeup view of the interaction between Cap2 (yellow/blue) and CD-NTase (orange), with key residues shown as sticks and labeled.

(b) Coomassie stained SDS-PAGE of proteins that purified by Ni^{2+}-affinity chromatography from *E. coli* co-expressing *E. cloacae* 6×His-tagged *cd-ntase* (His_6-CD-NTase) and catalytically inactivated *cap2* (C548A/C109A; C.D.) with the indicated genotype.

(c) CryoEM density for 2:1 Cap2–CD-NTase complex, with domains labeled and colored as in Fig. 2b. Outline indicates the areas of missing density compared to the 2:2 Cap2–CD-NTase complex, including one protomer of CD-NTase and the E2 and linker domains for the (a) protomer of Cap2.

(d) Two views of the 2:1 Cap2–CD-NTase complex, with domains labeled and colored as in panel (c). Outlines indicate the areas of missing density compared to the 2:2 Cap2–CD-NTase complex.

(e) Design of a fusion between the C-terminal E1 domain of *E. cloacae* Cap2 (residues 363–600) and the C-terminus of CD-NTase (residues 370–381), with a three-residue GSG linker.

(f) 2.1 Å resolution crystal structure of the *E. cloacae* Cap2–CD-NTase fusion crystallized in the presence of ATP, with two Cap2 E1 domains colored yellow and gray, and the two CD-NTase C-termini colored orange. See also Extended Data Table 2.

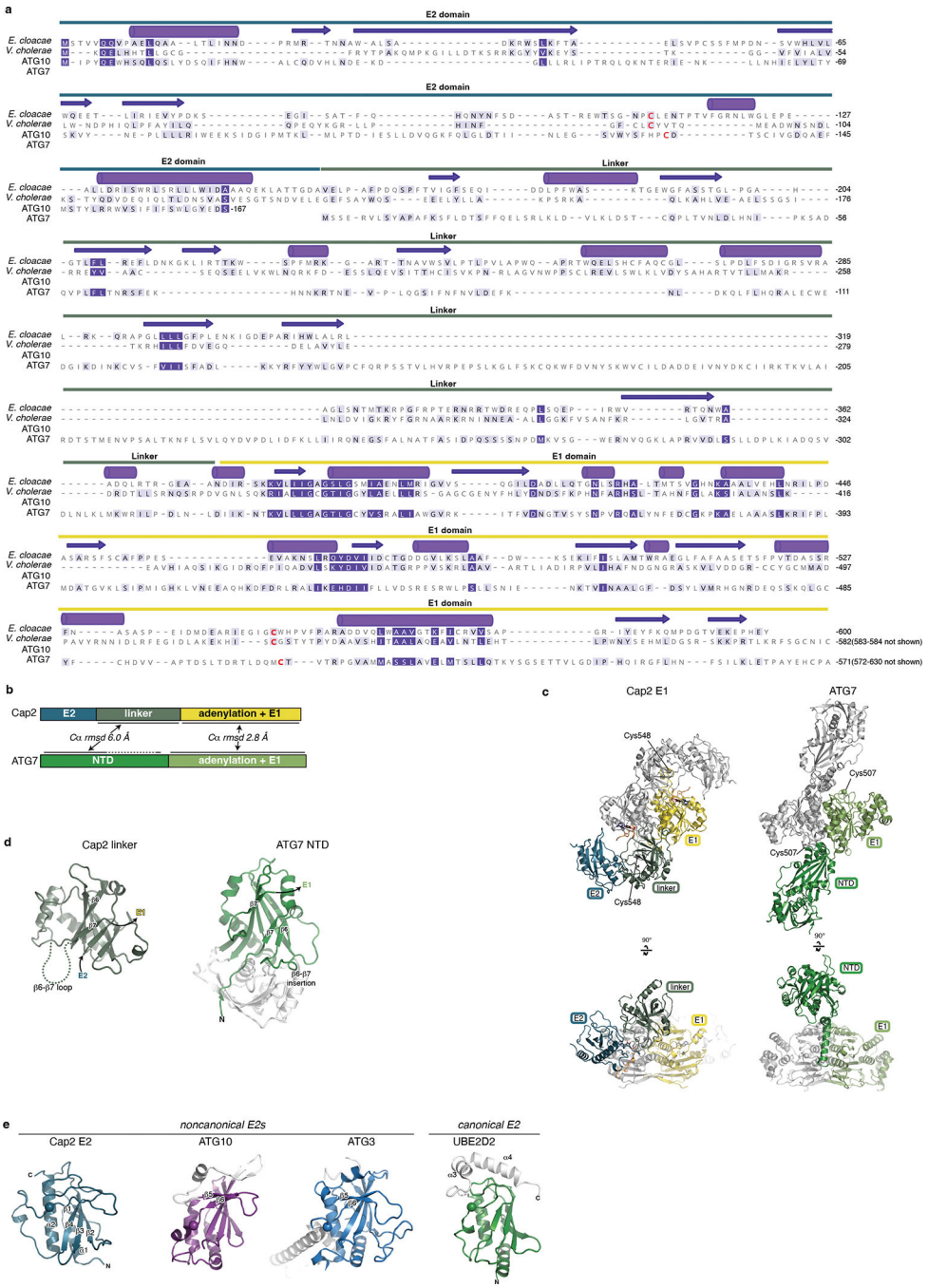
(g) Closeup of the Cap2 adenylation active site, showing the CD-NTase–AMP conjugate and active site residues. Residues 533–546 are disordered and represented by a dotted line. Bound Mg^{2+} ion is shown in black.

(h) View as in (g), with $2F_O$-F_C electron density contoured at 1.0 σ around the CD-NTase–AMP conjugate and active site residues.

(i) View as in (g), with F_O-F_C omit map density contoured at 1.5 σ around the CD-NTase–AMP conjugate.

(j) Closeup of the Cap2 adenylation active site in a 1.8 Å-resolution structure of the Cap2-CD-NTase fusion crystallized in the absence of added nucleotide (apo state).

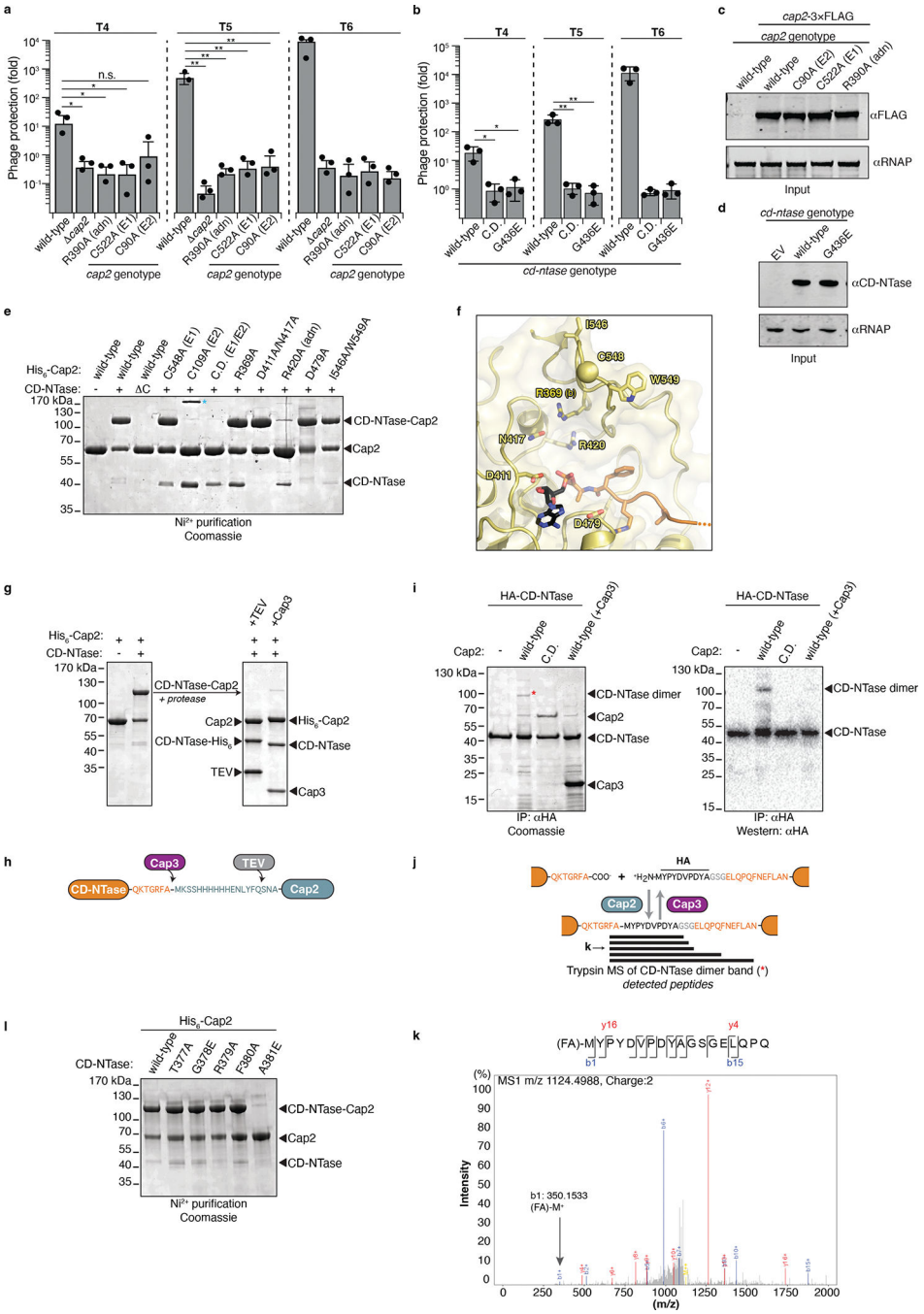
(k) View as in (j), with 2*Fo-Fc* electron density contoured at 1.0 σ around the CD-NTase C-terminus.



Extended Data Figure 3. Cap2 is related to autophagy E1 and E2 proteins.

(a) Protein alignment of Cap2 from *E. cloacae*, Cap2 from *V. cholerae*, ATG10 from *S. cerevisiae* (4EBR[49]), and ATG7 from *S. cerevisiae* (3T7H[24]). Domains are indicated above the alignment with colors corresponding to Fig. 2. The secondary structure of Cap2 from *E. cloacae* is indicated in purple with alpha helices depicted as cylinders and beta sheets as arrows. The catalytic cysteines found in the E2 and E1 domains are highlighted in red. See Supplementary Table 5 for relevant accession numbers.

(b) Domain schematic of *E. cloacae* Cap2 and *S. cerevisiae* ATG7, with approximate root-mean-squared distance (Cα r.m.s.d.) values for the linker/NTD and E1 domains noted.
(c) Structures of *E. cloacae* Cap2 (left), compared to *S. cerevisiae* ATG7 (right; PDB ID 4GSK[22]), with one protomer colored as in panel (a) and the dimer mate colored gray. For each protein, the E1 active-site cysteine residue (C548 for Cap2, C507 for ATG7) is shown as a sphere and labeled.
(d) Structures of the *E. cloacae* Cap2 linker domain (left), compared to the *S. cerevisiae* ATG7 NTD (right; PDB ID 4GSK[22]). ATG7 features a second subdomain (residues 147–268, shown in white) inserted into the loop separating β-strands 6 and 7 (labeled) where Cap2 has a partially disordered loop (residues 319–356).
(e) Structure of the Cap2 E2 domain (active-site C109 shown as a sphere), compared to *Kluyveromyces marxianus* ATG10 (PDB ID 3VX7[23]), *S. cerevisiae* ATG3 (PDB ID 2DYT[50]), and *Homo sapiens* UBE2D2 (PDB ID 4DDG[51]). Structural features not shared are shown in white. The active-site cysteine of each protein is shown as a sphere.



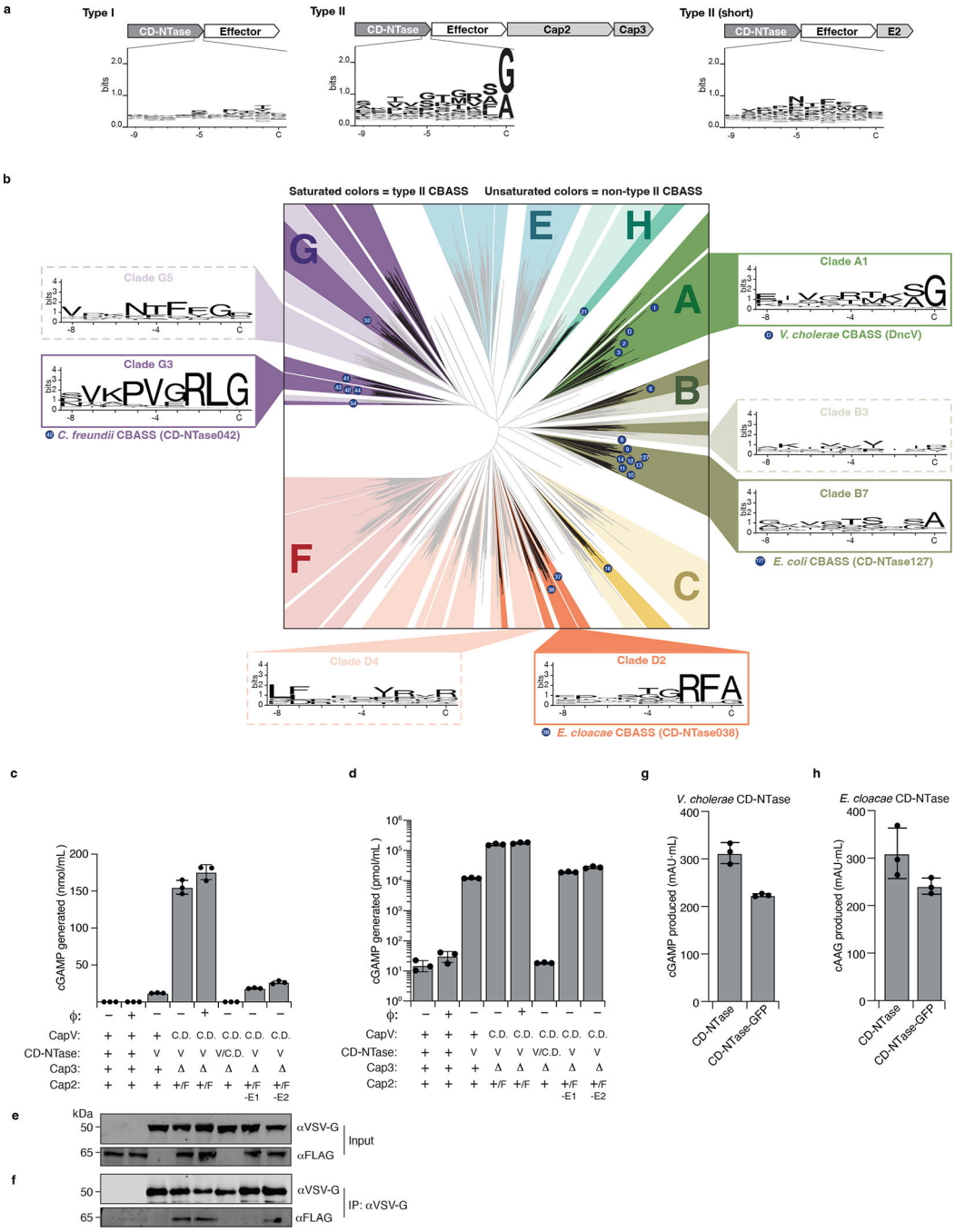
Extended Data 4. Analysis of Cap2 mutants and epitope-tagged CD-NTase and evidence that Cap2 conjugates the CD- NTase C-terminus to a target.

(a) Efficiency of plating of the indicated phage when infecting *E. coli* expressing CBASS with the indicated genotype. Data plotted as in Fig. 1b.

(b) Efficiency of plating of the indicated phage when infecting *E. coli* expressing CBASS with the indicated genotype. Data plotted as in Fig. 1b.

(c) Western blot analysis of cell lysates from *E. coli* expressing CBASS with the indicated genotypes demonstrating that the mutations do not affect expression levels.

(d) Western blot analysis of cell lysates from *E. coli* expressing CBASS with the indicated genotypes demonstrating that the mutations do not affect protein expression levels.
(e) SDS-PAGE analysis of *E. cloacae* Cap2 activity assay. The indicated genotypes of His_6-Cap2 and the CD-NTase were expressed from a single plasmid and the formation of a CD-NTase–His_6-Cap2 conjugate was used as an indicator of Cap2 activity. (−): no CD-NTase; (+): wild-type CD-NTase; (ΔC): CD-NTase lacking its C-terminal 19 residues; C.D. Cap2: C548A/C109A. Blue asterisk indicates a putative intermediate with CD-NTase thioester-linked to the Cap2 E1 catalytic cysteine (C548). The formation of a CD-NTase–Cap2 conjugate in the absence of a functional E1 catalytic cysteine (C548A) indicates that *in vitro*, this residue is dispensable for catalysis and the nearby E2 catalytic cysteine (C109) can function in instead.
(f) Cap2 E1 active site (yellow) in Cap2–CD-NTase cryoEM structure with the residues mutated in (a) indicated and the E1 active-site cysteine residue (C548 for Cap2) shown as a sphere and labeled. The CD-NTase C-terminus (orange) conjugated to AMP (black).
(g) Left: SDS-PAGE analysis of Ni^{2+}-purified *E. cloacae* His_6-Cap2, expressed either alone or with full-length CD-NTase. Right: Protease treatment (TEV or Cap3) of the CD-NTase–His_6-Cap2 conjugate.
(h) Schematic of the inferred CD-NTase–His_6-Cap2 conjugate formed upon coexpression of *E. cloacae* His_6-Cap2 and CD-NTase, with cleavage sites for Cap3 and TEV protease indicated.
(i) SDS-PAGE analysis with detection by coomassie (left) or αHA western blot (right) of αHA immunoprecipitated *E. cloacae* Cap2 coexpressed with HA-CD-NTase. C.D. Cap2: C548A/C109A. Red asterisk indicates band used for tryptic mass spectrometry analysis in (f-g).
(j) Peptides detected in tryptic mass spectrometry of the marked band in €, showing conjugation of CD-NTase to the N-terminus of a second HA-CD-NTase molecule. See Supplementary Table 7 for mass spectrometry data.
(k) Collision-induced dissociation mass spectrum of the peptide indicated in (f), with b1 peak indicated (mass of 350.1533 is that of Met+(H+)+(Phe-Ala)).
(l) SDS-PAGE analysis of Ni^{2+}-purified *E. cloacae* His_6-Cap2 with CD-NTase with the indicated genotype.



Extended Data Figure 5. The C-terminus of the CD-NTase is conserved in type II CBASS systems and the quantification of CD-NTase-mediated second messenger generation.

(a) Sequence logos of C-terminal 10 residues of the CD-NTase in type I (2284 sequences), type II (1556 sequences), and type II (short) (593 sequences) CBASS systems[7]. Type II (short) CBASS systems encode an E2 ubiquitin transferase-like enzyme without a linked E1 domain, and do not encode a JAB isopeptidase.

(b) Phylogenetic tree adapted from Whiteley et al.[2], with sequence logos of the C-terminus for CD-NTase clades analyzed in Cap3 experiments (Fig. 4d, Extended data Fig. 6d–g)

shown. Saturated colors bordered with solid lines depict branches of the tree that contain type II systems, whereas the de-saturated colors bordered with dashed lines depict clades with non-type II systems. The CD-NTases used in this study are listed below each sequence logo. Blue circles with numbers represent CD-NTase numbers as reported previously[2].

(c) cGAMP generated by αVSV-G immunoprecipitation from *E. coli* expressing CBASS operons with the indicated genotypes. Western blots of input are in (e) and immunoprecipitation in (f). N=3 technical replicates representative of three independent biological replicates. Data is presented as the mean values ± the SEM. ϕ (−): no infection; ϕ (+): phage T2 at an MOI of 2; CapV (+): wild-type; CapV (C.D.): S62A; CD-NTase (+): wild-type; CD-NTase (C.D.): DID131AIA; CD-NTase (V): N-terminal VSV-G epitope tagged CD-NTase; Cap3 (+): wild-type; Cap3 (Δ): genetically deleted *cap3*; Cap2 (+): wild-type; Cap2 (F): C-terminal 3x-FLAG epitope tagged Cap2; Cap2 (E1): C522A; Cap2 (E2): C90A.

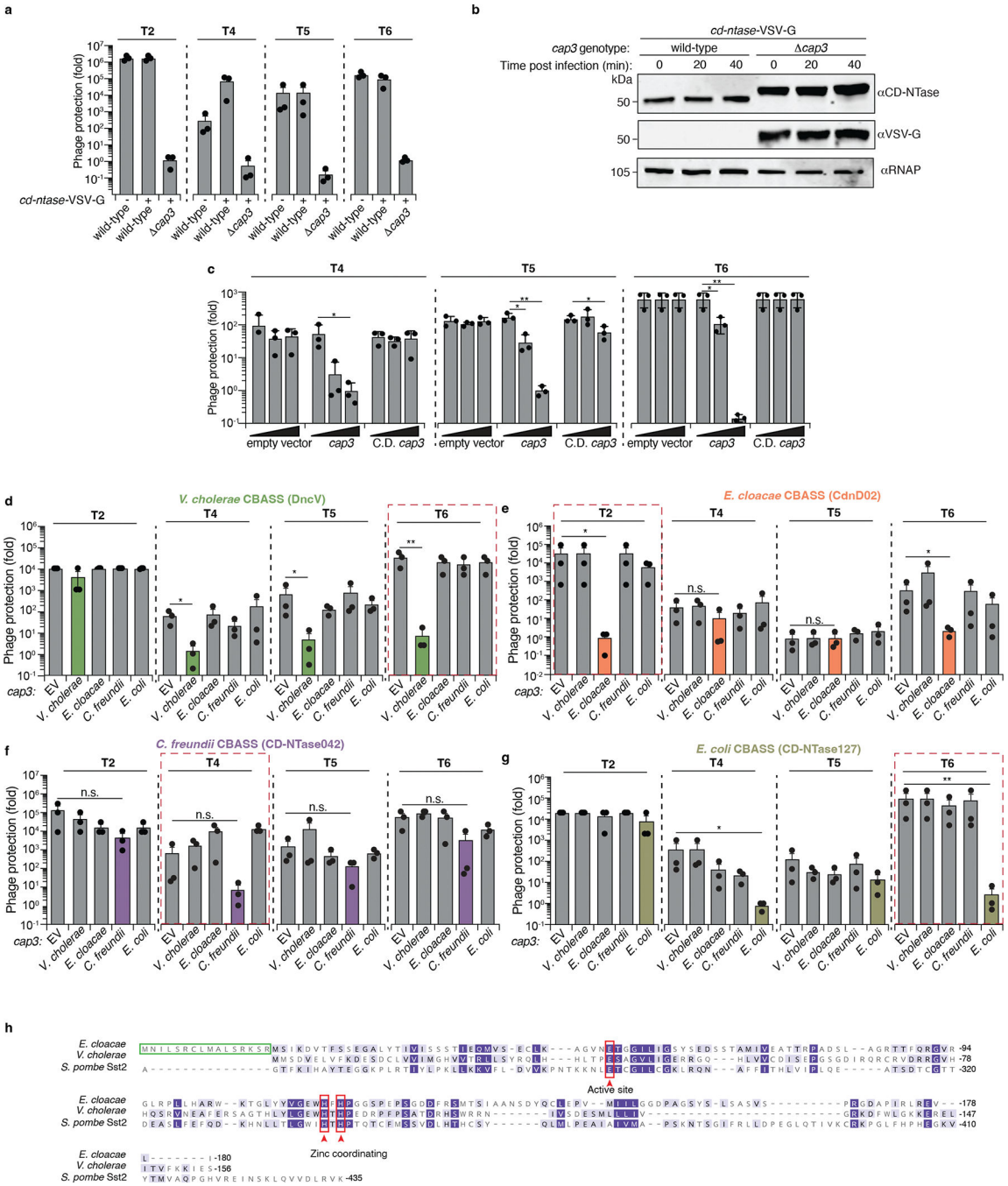
(d) Data in (c) presented on a $\log_{10}$ scale.

(e) Whole cell western blot analysis of *E. coli* expressing CBASS with the indicated genotype. Data corresponds to the input for the immunoprecipitation shown in (a) and (b).

(f) Western blot analysis of αVSV-G immunoprecipitation of *E. coli* expressing CBASS with the indicated genotypes. These data correspond to the samples used to measure cGAMP synthesis in (c) and (d).

(g) Quantification of the cGAMP produced by the *V. cholerae* CD-NTase with, or without, a C-terminal GFP fusion.

(h) Quantification of the cAAG produced by the *E. cloacae* CD-NTase with, or without, a C-terminal GFP fusion. For (g) and (h) N=3 independent biological replicates and the data is presented as the mean ± the SD.



Extended Data Figure 6. Cap3 overexpression inhibits phage protection by cognate CBASS.

(a) Efficiency of plating of the indicated phage when infecting *E. coli* expressing CBASS with the indicated genotype. *cd-ntase*-VSV-G (−): wild-type CD-NTase; *cd-ntase*-VSV-G (+): C-terminal VSV-G epitope tagged CD-NTase; wild-type indicates otherwise a full CBASS operon; Δ*cap3*: CBASS operon with only *cap3* deletion. Data plotted as in Fig. 1b. **(b)** Western blot analysis of cell lysates from *E. coli* expressing CBASS with the indicated genotypes, abbreviated as in (a). Cells were infected with phage T5 at an MOI of 2 for the indicated time prior to harvesting for analysis.

(c) Efficiency of plating of the indicated phage when infecting *E. coli* expressing CBASS Δ*cap3* in the absence or presence of overexpressed *cap3* with the indicated genotype. Data plotted as in Fig. 4a. C.D. *cap3*: HTH101ATA. A Two-sided Student's t-test was used to calculate significance; n.s., $p>0.05$; *, $p<0.05$; **, $p<0.001$. See Extended data Fig. 6h for protein alignment.

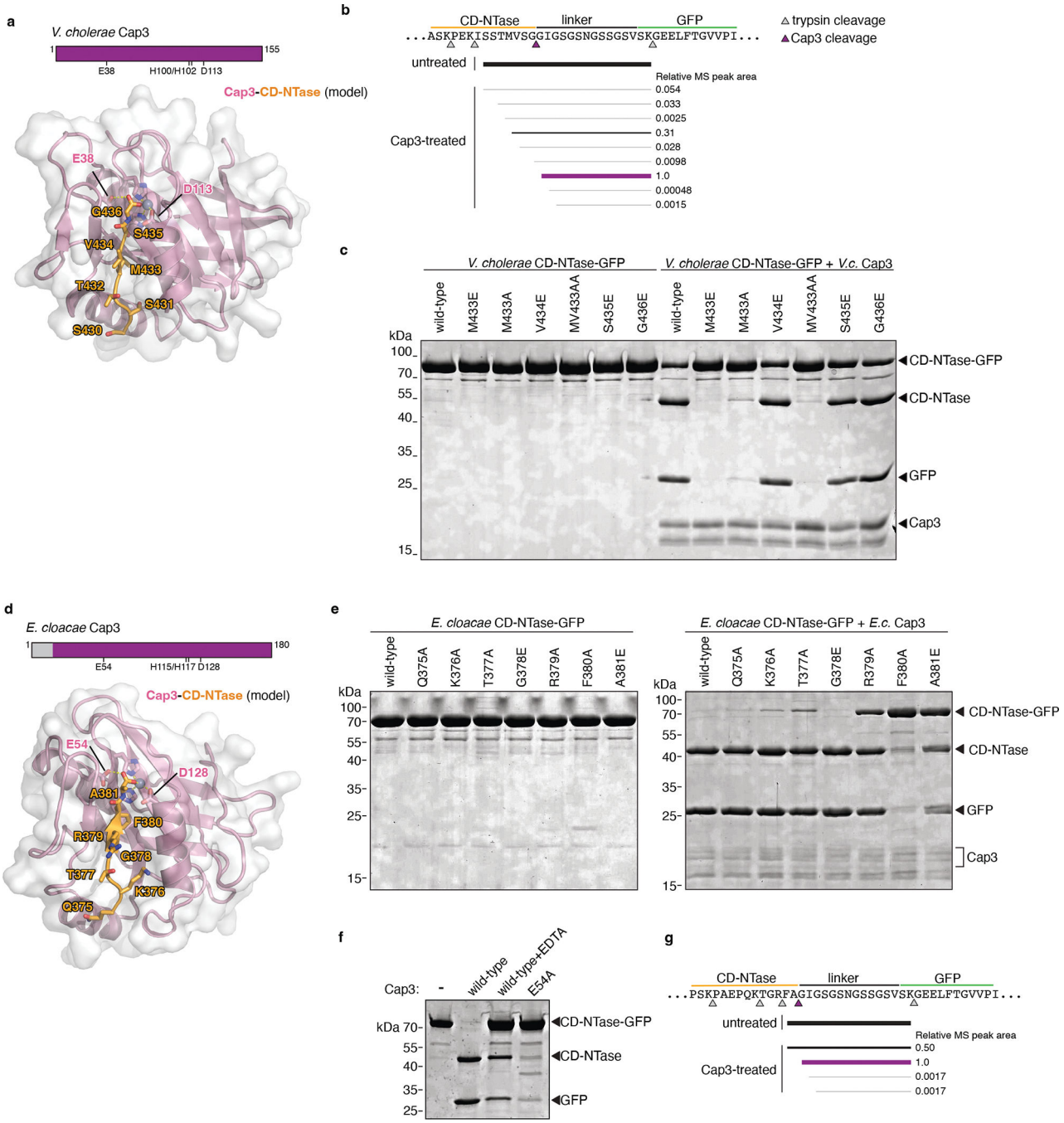
(d) Efficiency of plating of the indicated phage when infecting *E. coli* expressing a full CBASS operon from *V. cholerae* in the absence or presence of overexpressed *cap3* from another CBASS system, indicated on the *x*-axis. Data plotted as in Fig. 4a.

(e) Efficiency of plating of the indicated phage when infecting *E. coli* expressing a full CBASS operon from *E. cloacae* in the absence or presence of overexpressed *cap3* from another CBASS system, indicated on the *x*-axis. Data plotted as in Fig. 4a.

(f) Efficiency of plating of the indicated phage when infecting *E. coli* expressing a full CBASS operon from *C. freundii* in the absence or presence of overexpressed *cap3* from another CBASS system, indicated on the *x*-axis. Data plotted as in Fig. 4a.

(g) Efficiency of plating of the indicated phage when infecting *E. coli* expressing a full CBASS operon from *E. coli* in the absence or presence of overexpressed *cap3* from another CBASS system, indicated on the *x*-axis. Data plotted as in Fig. 4a. For (d-g) the red dashed boxes indicated the data utilized in Fig. 4d, See Supplementary Table 5 for relevant accession numbers.

(h) Protein alignment of the JAMM/JAB protease Sst2 from *S. pombe* (Uniprot ID Q9P371; residues 235–435), Cap3 from *E. cloacae*, and Cap3 from *V. cholerae*. The active site glutamate, as well as two zinc-coordinating histidine residues, are noted. For experiments using Cap3 from *E. cloacae,* the first 16 annotated amino acids (green box) were removed as we found the translation start site is likely misannotated for this gene. See Supplementary Table 5 for relevant accession numbers.

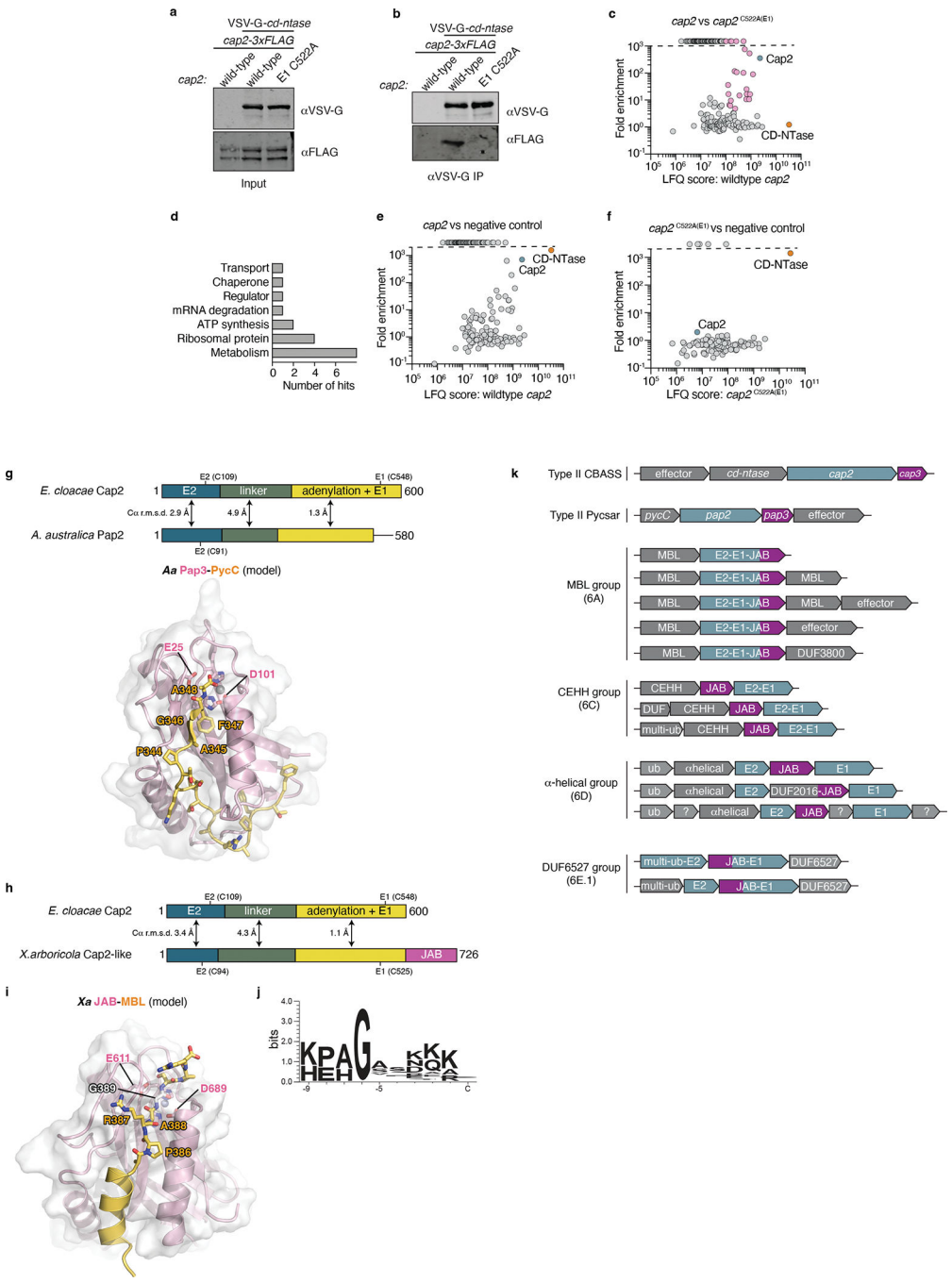


Extended Data Figure 7. Cap3 cleavage of a CD-NTase model substrate.

(a) Domain schematic and predicted structure/model of the *V. cholerae* Cap3-CD-NTase complex[52] and with the CD-NTase C-terminus and Zn^{2+} ion manually modeled from an overlay with a structure of *S. pombe* Sst2 bound to ubiquitin[53] (PDB ID 4K1R).

(b) Summary of tryptic digest mass spectrometry analysis of the *V. cholerae* Cap3-treated CD-NTase bands as in Fig. 4b. Pink arrow indicates the inferred Cap3 cleavage site; gray arrows indicate trypsin cleavage sites. See Supplementary Table 3 for data.

(c) Coomassie stained SDS-PAGE of a *V. cholerae* model substrate (CD-NTase-GFP fusion protein) with the indicated mutations in the CD-NTase C-terminus, with and without incubation with *V. cholerae* Cap3.
(d) Domain schematic and predicted structure/model of the *E. cloacae* Cap3-CD-NTase complex[52] with the CD-NTase C-terminus and Zn^{2+} ion manually modeled from an overlay with a structure of *S. pombe* Sst2 bound to ubiquitin[53] (PDB ID 4K1R).
(e) Coomassie stained SDS-PAGE of an *E. cloacae* model substrate (CD-NTase-GFP fusion protein) with the indicated mutations in the CD-NTase C-terminus, with and without incubation with *E. cloacae* Cap3.
(f) Coomassie stained SDS-PAGE of an *E. cloacae* model substrate (CD-NTase-GFP fusion protein) incubated with *E. cloacae* Cap3 with the indicated reaction condition/genotype.
(g) Summary of tryptic digest mass spectrometry analysis of the *E. cloacae* Cap3-treated CD-NTase bands as in (f), showing the putative Cap3 cleavage site. Pink arrow indicates the inferred Cap3 cleavage site; gray arrows indicate trypsin cleavage sites. See Supplementary Table 3 for data.



Extended Data Figure 8. -NTase immunoprecipitation reveals numerous potential protein targets and Cap2 homologs are found in other bacteria.

CD (a) Western blot analysis of cell lysates generated from *E. coli* expressing CBASS Δ*cap3* with the additional indicated genotypes.

(b) Western blot analysis of αVSV-G immunoprecipitations generated from *E. coli* expressing CBASS Δ*cap3* with the additional indicated genotypes. These samples were used in the mass spectrometry analysis displayed in (c-f).

(c) Mass spectrometry of immunoprecipitated VSV-G-CD-NTase as shown in (b). Data are label free quantitation (LFQ) score and fold enrichment comparing immunoprecipitations from bacteria expressing CBASS Δ*cap3* to a strain expressing CBASS Δ*cap3 cap2*$^{C522A(E1)}$. Both strains encode an N-terminally VSV-G tagged CD-NTase. Cap2 and CD-NTase are represented as colored circles corresponding to Fig. 1a and are labeled. Proteins which we determined were significantly enriched (LFQ> 10^8 and a fold enrichment >4) are colored in pink. Circles above the dotted line are proteins with peptides identified only in the sample listed on the *x*-axis. See Supplementary Table 8 for data.

(d) Characterization of the predicted functions of the proteins that were significantly enriched in (c).

(e) Mass spectrometry of immunoprecipitated VSV-G-CD-NTase as in (c) comparing immunoprecipitations from bacteria expressing CBASS Δ*cap3* where CD-NTase has an N-terminal VSV-G tag to a strain expressing CBASS lacking a VSV-G tag (negative control).

(f) Mass spectrometry of immunoprecipitated VSV-G-CD-NTase as in (c) comparing immunoprecipitations from bacteria expressing CBASS Δ*cap3 cap2*$^{C522A(E1)}$ where CD-NTase has an N-terminal VSV-G tag to a strain expressing CBASS lacking a VSV-G tag (negative control).

(g) Top: Structural comparison between *E. cloacae* Cap2 and a predicted structure/model of *Azohydromonas australica* Pap2. Cα r.m.s.d. values are reported for superposition of individual domains: E2 domain (52 Cα atoms overlaid), linker (90 Cα atoms), and E1 (133 Cα atoms). Predicted catalytic cysteine residues are noted for each protein. Bottom: Structural prediction of a Pap3 (pink) and the C-terminus of PycC (yellow) from *A. australica.* Predicted active site residues of Pap3 are shown as sticks with a zinc ion (gray) modeled from a structure of *S. pombe* Sst2 bound to ubiquitin (PDB ID 4K1R).

(h) Structural comparison between *E. cloacae* Cap2 and a predicted structure/model of the Cap2-like protein from the *Xanthamonas arboricola* MBL-group operon. Cα r.m.s.d. values are reported for superposition of individual domains: E2 domain (94 Cα atoms overlaid), linker (50 Cα atoms), and E1 (154 Cα atoms). Predicted catalytic cysteine residues are noted for each protein.

(i) Predicted structure/model of a complex between *X. arboricola* JAB domain (pink) and the C-terminus of MBL (yellow). Predicted active site residues of the JAB domain are shown in sticks, with a zinc ion (gray) modeled from a structure of *S. pombe* Sst2 bound to ubiquitin[53] (PDB ID 4K1R). The conserved glycine residue of MBL (white) is positioned for cleavage.

(j) Sequence logo for the C-terminal 9 residues of 268 MBL encoded within MBL-group operons. See also Supplementary Table 9.

(k) Operon structure of previously described and proposed phage defense systems that contain E1, E2 and JAB domain containing proteins along with operons of unknown function that contain these domains. Operons are grouped by conserved protein domains. The E1-superfamily these groups is also indicated in paratheses[20,39]. Genes are colored by domain type; E1 and E2 domains, blue; JAB domains, purple; all other domains, grey. Metallo-β-lactamase (MBL); metal binding domain (CEHH); tandem β-grasp fold domain containing protein (multi-ub); single β-grasp fold domain containing protein (ub); Domains of unknown function (DUF); genes with no discernable domains (?). See also Supplementary Table 9.

Extended Data Table 1.

Summary of Cryo-EM data collection, refinement and validation statistics.

Table containing all the information related to Cryo-EM data collection and downstream processing. 3D FSC values were calculated by the Remote 3DFSC Processing Server (https://3dfsc.salk.edu)[50]. EMRinger scores were calculated by EMRinger[67]. Coordinates have been deposited in the RCSB Protein Data Bank (http://www.rcsb.org). EM maps have been deposited with the Electron Microscopy Data Bank (https://www.ebi.ac.uk/emdb/).

	Cap2-CD-NTase 2:2 complex	**Cap2-CD-NTase 2:1 complex**
Data collection and processing		
Magnification	165,000	165,000
Voltage (kV)	300	300
Electron exposure (e–/Å^2)	65	65
Defocus range (μm)	− 0.5 to −2.5	− 0.5 to −2.5
Pixel size (Å)	0.84	0.84
Symmetry imposed	C2	C1
Initial particle images (no.)	2,554,650	2,554,650
Final particle images (no.)	138,202	147,709
Map resolution (Å)	2.74	2.9
FSC threshold	0.143	0.143
Map resolution range (Å)	2.58–5.47	2.71–5.7
Refinement		
Initial model used (PDB code)	*de novo*	7TO3
Model resolution (Å)	2.7	2.9
FSC threshold	0.143	0.143
Map sharpening *B* factor (Å^2)	80	77.6
Model composition		
Non-hydrogen atoms	14764	8690
Protein residues	1850	1096
Ligands	4 Mg^{2+}, 2 AMP, 2 ADP, 2 ATP	2 Mg^{2+},1 AMP, 1 ADP, 1 ATP
B factors (Å^2)		
Protein	99.97	116.24
Ligand	90.14	110.7
R.m.s. deviations		
Bond lengths (Å)	0.002	0.003
Bond angles (°)	0.544	0.519
Validation		
MolProbity score	1.11	1.14
Clashscore	3.2	3.15
Poor rotamers (%)	0	0

	Cap2-CD-NTase 2:2 complex	Cap2-CD-NTase 2:1 complex
Ramachandran plot		
Favored (%)	98.14	97.87
Allowed (%)	1.86	2.13
Disallowed (%)	0	0
PDB ID	7TO3	7TQD
EMDB ID	26028	26066

Extended Data Table 2.
Summary of Crystallographic data collection and refinement statistics

A single crystal was used for each dataset.

	Cap2 E1-CD-NTase Apo (PDB 7TSX)	Cap2 E1-CD-NTase:AMP (PDB 7TSQ)
Data collection		
Space group	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions		
a, *b*, *c* (Å)	50.26, 76.55, 127.63	49.69, 76.35, 126.90
α, β, γ (°)	90, 90, 90	90, 90, 90
Resolution (Å)	128–1.77 (1.80–1.77) *	127–2.11 (2.18–2.11)
R_{merge}	0.087 (1.315)	0.146 (0.860)
$I / \sigma I$	14.1 (1.3)	15.3 (2.5)
Completeness (%)	99.9 (98.7)	99.6 (97.0)
Redundancy	6.7 (6.5)	6.5 (5.8)
Refinement		
Resolution (Å)	65.65–1.77	65.42–2.11
No. reflections	48,900	28.241
R_{work} / R_{free}	16.95 / 19.15	20.22 / 22.93
No. atoms	6970	6830
Protein	6585	6576
Ligand/ion	0	36
Water	385	218
B-factors	35.68	47.26
Protein	34.8	46.77
Ligand/ion		73.59
Water	43.22	48.96
R.m.s. deviations		
Bond lengths (Å)	0.004	0.006
Bond angles (°)	0.64	0.78

* A single crystal was used for each dataset.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors thank R. Parker and A. Desai for critical reading of the manuscript; M. Herzik for advice on cryoEM data collection and processing; C. Ebmeirer and the Mass Spectrometry Facility at the University of Colorado Boulder (RRID:SCR_018992) for assistance with sample preparation, experimental details, and data analysis (NIH S10-OD025267) which was funded by the Core Voucher from the CU Boulder Financial Futures Program (A.T.W); the CU Boulder Department of Biochemistry Shared Instruments Pool core facility (RRID:SCR_018986) and its staff; J. Kralj at CU Boulder for use of his laboratory's plate reader; and members of the Corbett and Whiteley labs for advice and helpful discussion. The authors acknowledge the facilities of the cryoEM facility at UC San Diego, and technical assistance of R. Ashley on cryoEM sample preparation and data collection. This work was funded by NIH R35 GM144121 (K.D.C.), the NIH Office of the Director DP2 AT012346 and a Mallinckrodt Foundation Grant (A.T.W.), and NIH R01 GM116897 and S10 OD023498 (H.Z.). H.E.L. is supported as a fellow of the Jane Coffin Childs Memorial Fund for Medical Research and R.K.L. is supported by NIH F31GM137600. X-ray diffraction data were collected 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). The Eiger 16M detector on 24-ID-E 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.

Data Availability

All data supporting the findings of this study are available within the Article and associated Supplementary Information. For the E. cloacae 2:2 Cap2–CD-NTase complex, coordinates are available at the RCSB Protein Data Bank (PDB; http://www.rcsb.org) under accession code 7TO3, and EM maps are available at the Electron Microscopy Data Bank (EMDB; https://www.ebi.ac.uk/emdb/) under accession code 26028. For the E. cloacae 2:1 Cap2–CD-NTase complex, coordinates are available under PDB accession code 7TQD, and EM maps are available at the Electron Microscopy Data Bank (EMDB; https://www.ebi.ac.uk/emdb/) under accession code 26066. For the E. cloacae Cap2 E1 domain-CD-NTase fusion in the Apo state, coordinates and crystallographic structure factors are available under PDB accession code 7TSX, and diffraction data are available at the SBGrid Data Bank (http://data.sbgrid.org) under dataset number 878. For the E. cloacae Cap2 E1 domain-CD-NTase fusion in the AMP state, coordinates and crystallographic structure factors are available under PDB accession code 7TSQ, and diffraction data are available at the SBGrid Data Bank under dataset number 877.

Main text references

1. Ni G, Ma Z & Damania B cGAS and STING: At the intersection of DNA and RNA virus-sensing networks. PLOS Pathog. 14, e1007148 (2018). [PubMed: 30114241]
2. Whiteley AT et al. Bacterial cGAS-like enzymes synthesize diverse nucleotide signals. Nature 567, 194–199 (2019). [PubMed: 30787435]
3. Cohen D et al. Cyclic GMP-AMP signalling protects bacteria against viral infection. Nature 574, 691–695 (2019). [PubMed: 31533127]
4. Ye Q et al. HORMA Domain Proteins and a Trip13-like ATPase Regulate Bacterial cGAS-like Enzymes to Mediate Bacteriophage Immunity. Mol. Cell 77, 709–722.e7 (2020). [PubMed: 31932165]
5. Morehouse BR et al. STING cyclic dinucleotide sensing originated in bacteria. Nature 586, 429–433 (2020). [PubMed: 32877915]

6. Burroughs AM, Zhang D, Schäffer DE, Iyer LM & Aravind L Comparative genomic analyses reveal a vast, novel network of nucleotide-centric systems in biological conflicts, immunity and signaling. Nucleic Acids Res. 43, 10633–10654 (2015). [PubMed: 26590262]
7. Millman A, Melamed S, Amitai G & Sorek R Diversity and classification of cyclic-oligonucleotide-based anti-phage signalling systems. Nat. Microbiol. 5, 1608–1615 (2020). [PubMed: 32839535]
8. Severin GB et al. Direct activation of a phospholipase by cyclic GMP-AMP in El Tor *Vibrio cholerae*. Proc. Natl. Acad. Sci. 115, E6048–E6055 (2018). [PubMed: 29891656]
9. Lowey B et al. 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 (2020). [PubMed: 32544385]
10. Lau RK et al. Structure and Mechanism of a Cyclic Trinucleotide-Activated Bacterial Endonuclease Mediating Bacteriophage Immunity. Mol. Cell 77, 723–733.e6 (2020). [PubMed: 31932164]
11. Duncan-Lowey B, McNamara-Bordewick NK, Tal N, Sorek R & Kranzusch PJ Effector-mediated membrane disruption controls cell death in CBASS antiphage defense. Mol. Cell 81, 5039–5051.e5 (2021). [PubMed: 34784509]
12. Davies BW, Bogard RW, Young TS & Mekalanos JJ Coordinated Regulation of Accessory Genetic Elements Produces Cyclic Di-Nucleotides for V. cholerae Virulence. Cell 149, 358–370 (2012). [PubMed: 22500802]
13. Dziejman M et al. Comparative genomic analysis of Vibrio cholerae: genes that correlate with cholera endemic and pandemic disease. Proc. Natl. Acad. Sci. U. S. A. 99, 1556–1561 (2002). [PubMed: 11818571]
14. Schulman BA & Harper JW Ubiquitin-like protein activation by E1 enzymes: the apex for downstream signalling pathways. Nat. Rev. Mol. Cell Biol. 10, 319–331 (2009). [PubMed: 19352404]
15. Cappadocia L & Lima CD Ubiquitin-like Protein Conjugation: Structures, Chemistry, and Mechanism. Chem. Rev 118, 889–918 (2018). [PubMed: 28234446]
16. Lake MW, Wuebbens MM, Rajagopalan KV & Schindelin H Mechanism of ubiquitin activation revealed by the structure of a bacterial MoeB-MoaD complex. Nature 414, 325–329 (2001). [PubMed: 11713534]
17. Xu X, Wang T, Niu Y, Liang K & Yang Y The ubiquitin-like modification by ThiS and ThiF in Escherichia coli. Int. J. Biol. Macromol. 141, 351–357 (2019). [PubMed: 31442507]
18. Burroughs AM, Iyer LM & Aravind L The natural history of ubiquitin and ubiquitin-related domains. Front. Biosci. Landmark Ed. 17, 1433–1460 (2012). [PubMed: 22201813]
19. Nandi D, Tahiliani P, Kumar A & Chandu D The ubiquitin-proteasome system. J. Biosci. 31, 137–155 (2006). [PubMed: 16595883]
20. Burroughs AM, Iyer LM & Aravind L Natural history of the E1-like superfamily: Implication for adenylation, sulfur transfer, and ubiquitin conjugation. Proteins Struct. Funct. Bioinforma. 75, 895–910 (2009).
21. Lehmann C, Begley TP & Ealick SE Structure of the Escherichia coli ThiS-ThiF complex, a key component of the sulfur transfer system in thiamin biosynthesis. Biochemistry 45, 11–19 (2006). [PubMed: 16388576]
22. Kaiser SE et al. Noncanonical E2 recruitment by the autophagy E1 revealed by Atg7-Atg3 and Atg7-Atg10 structures. Nat. Struct. Mol. Biol. 19, 1242–1249 (2012). [PubMed: 23142976]
23. Yamaguchi M et al. Noncanonical recognition and UBL loading of distinct E2s by autophagy-essential Atg7. Nat. Struct. Mol. Biol. 19, 1250–1256 (2012). [PubMed: 23142983]
24. Taherbhoy AM et al. Atg8 transfer from Atg7 to Atg3: a distinctive E1-E2 architecture and mechanism in the autophagy pathway. Mol. Cell 44, 451–461 (2011). [PubMed: 22055190]
25. Noda NN et al. Structural basis of Atg8 activation by a homodimeric E1, Atg7. Mol. Cell 44, 462–475 (2011). [PubMed: 22055191]
26. Hong SB et al. Insights into noncanonical E1 enzyme activation from the structure of autophagic E1 Atg7 with Atg8. Nat. Struct. Mol. Biol. 18, 1323–1330 (2011). [PubMed: 22056771]

27. Schäfer A, Kuhn M & Schindelin H Structure of the ubiquitin-activating enzyme loaded with two ubiquitin molecules. Acta Crystallogr. D Biol. Crystallogr. 70, 1311–1320 (2014). [PubMed: 24816100]
28. Olsen SK, Capili AD, Lu X, Tan DS & Lima CD Active site remodelling accompanies thioester bond formation in the SUMO E1. Nature 463, 906–912 (2010). [PubMed: 20164921]
29. Kranzusch PJ et al. Structure-guided reprogramming of human cGAS dinucleotide linkage specificity. Cell 158, 1011–1021 (2014). [PubMed: 25131990]
30. Govande AA, Duncan-Lowey B, Eaglesham JB, Whiteley AT & Kranzusch PJ Molecular basis of CD-NTase nucleotide selection in CBASS anti-phage defense. Cell Rep. 35, 109206 (2021). [PubMed: 34077735]
31. Chen Z. (James), Jenson J & Li T Ubiquitin-Like Conjugation by Bacterial cGAS Enhances Anti-Phage Defense. https://www.researchsquare.com/article/rs-1578554/v1 (2022) doi:10.21203/rs.3.rs-1578554/v1.
32. Burroughs AM & Aravind L Identification of Uncharacterized Components of Prokaryotic Immune Systems and Their Diverse Eukaryotic Reformulations. J. Bacteriol. (2020) doi:10.1128/JB.00365-20.
33. Gao L et al. Diverse enzymatic activities mediate antiviral immunity in prokaryotes. 9 (2020).
34. Doron S et al. Systematic discovery of antiphage defense systems in the microbial pangenome. Science 359, eaar4120 (2018). [PubMed: 29371424]
35. Millman A et al. An expanding arsenal of immune systems that protect bacteria from phages. 10.1101/2022.05.11.491447 (2022) doi:10.1101/2022.05.11.491447.
36. Rousset F, Dowding J, Bernheim A, Rocha EPC & Bikard D Prophage-encoded hotspots of bacterial immune systems. 10.1101/2021.01.21.427644 (2021) doi:10.1101/2021.01.21.427644.
37. Vassallo C, Doering C, Littlehale ML, Teodoro G & Laub MT Mapping the landscape of anti-phage defense mechanisms in the *E. coli* pangenome. 10.1101/2022.05.12.491691 (2022) doi:10.1101/2022.05.12.491691.
38. Tal N et al. Cyclic CMP and cyclic UMP mediate bacterial immunity against phages. Cell 184, 5728–5739.e16 (2021). [PubMed: 34644530]
39. Iyer LM, Burroughs AM & Aravind L The prokaryotic antecedents of the ubiquitin-signaling system and the early evolution of ubiquitin-like beta-grasp domains. Genome Biol. 7, R60 (2006). [PubMed: 16859499]
40. Aravind L, Iyer LM & Burroughs AM Discovering Biological Conflict Systems Through Genome Analysis: Evolutionary Principles and Biochemical Novelty. Annu. Rev. Biomed. Data Sci. 5, annurev-biodatasci-122220–101119 (2022).
41. Oudshoorn D, Versteeg GA & Kikkert M Regulation of the innate immune system by ubiquitin and ubiquitin-like modifiers. Cytokine Growth Factor Rev. 23, 273–282 (2012). [PubMed: 22964110]
42. Zinngrebe J, Montinaro A, Peltzer N & Walczak H Ubiquitin in the immune system. EMBO Rep. 15, 28–45 (2014). [PubMed: 24375678]
43. Hu H & Sun S-C Ubiquitin signaling in immune responses. Cell Res. 26, 457–483 (2016). [PubMed: 27012466]
44. Qiu J et al. Ubiquitination independent of E1 and E2 enzymes by bacterial effectors. Nature 533, 120–124 (2016). [PubMed: 27049943]
45. Grau-Bové X, Sebé-Pedrós A & Ruiz-Trillo I The eukaryotic ancestor had a complex ubiquitin signaling system of archaeal origin. Mol. Biol. Evol. 32, 726–739 (2015). [PubMed: 25525215]
46. Hennell James R et al. Functional reconstruction of a eukaryotic-like E1/E2/(RING) E3 ubiquitylation cascade from an uncultured archaeon. Nat. Commun. 8, 1120 (2017). [PubMed: 29066714]
47. Iyer LM, Burroughs AM & Aravind L The prokaryotic antecedents of the ubiquitin-signaling system and the early evolution of ubiquitin-like beta-grasp domains. Genome Biol. 7, R60 (2006). [PubMed: 16859499]
48. Tan YZ et al. Addressing preferred specimen orientation in single-particle cryo-EM through tilting. Nat. Methods 14, 793–796 (2017). [PubMed: 28671674]

49. Hong SB, Kim B-W, Kim JH & Song HK Structure of the autophagic E2 enzyme Atg10. Acta Crystallogr. D Biol. Crystallogr 68, 1409–1417 (2012). [PubMed: 22993095]
50. Yamada R et al. Cell-autonomous involvement of Mab21l1 is essential for lens placode development. Dev. Camb. Engl. 130, 1759–1770 (2003).
51. Juang Y-C et al. OTUB1 co-opts Lys48-linked ubiquitin recognition to suppress E2 enzyme function. Mol. Cell 45, 384–397 (2012). [PubMed: 22325355]
52. Jumper J et al. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589 (2021). [PubMed: 34265844]
53. Shrestha RK et al. Insights into the mechanism of deubiquitination by JAMM deubiquitinases from cocrystal structures of the enzyme with the substrate and product. Biochemistry 53, 3199–3217 (2014). [PubMed: 24787148]
54. Kibby EM et al. Bacterial NLR-related proteins protect against phage. 10.1101/2022.07.19.500537 (2022) doi:10.1101/2022.07.19.500537.
55. Guzman LM, Belin D, Carson MJ & Beckwith J Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. J. Bacteriol. 177, 4121–4130 (1995). [PubMed: 7608087]
56. de Boer HA, Comstock LJ & Vasser M The tac promoter: a functional hybrid derived from the trp and lac promoters. Proc. Natl. Acad. Sci. U. S. A. 80, 21–25 (1983). [PubMed: 6337371]
57. Kropinski AM, Mazzocco A, Waddell TE, Lingohr E & Johnson RP Enumeration of bacteriophages by double agar overlay plaque assay. Methods Mol. Biol. Clifton NJ 501, 69–76 (2009).
58. Edgar RC MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797 (2004). [PubMed: 15034147]
59. Raran-Kurussi S, Cherry S, Zhang D & Waugh DS Removal of Affinity Tags with TEV Protease. Methods Mol. Biol. Clifton NJ 1586, 221–230 (2017).
60. Punjani A, Rubinstein JL, Fleet DJ & Brubaker MA cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat. Methods 14, 290–296 (2017). [PubMed: 28165473]
61. Zheng SQ et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. Nat. Methods 14, 331–332 (2017). [PubMed: 28250466]
62. Pettersen EF et al. UCSF Chimera--a visualization system for exploratory research and analysis. J. Comput. Chem. 25, 1605–1612 (2004). [PubMed: 15264254]
63. Emsley P, Lohkamp B, Scott WG & Cowtan K Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr 66, 486–501 (2010). [PubMed: 20383002]
64. Afonine PV et al. Real-space refinement in PHENIX for cryo-EM and crystallography. Acta Crystallogr. Sect. Struct. Biol. 74, 531–544 (2018).
65. Williams CJ et al. MolProbity: More and better reference data for improved all-atom structure validation. Protein Sci. Publ. Protein Soc. 27, 293–315 (2018).
66. Barad BA et al. EMRinger: side chain-directed model and map validation for 3D cryo-electron microscopy. Nat. Methods 12, 943–946 (2015). [PubMed: 26280328]
67. Kabsch W XDS. Acta Crystallogr. D Biol. Crystallogr 66, 125–132 (2010). [PubMed: 20124692]
68. Evans PR & Murshudov GN How good are my data and what is the resolution? Acta Crystallogr. D Biol. Crystallogr 69, 1204–1214 (2013). [PubMed: 23793146]
69. Evans P Scaling and assessment of data quality. Acta Crystallogr. D Biol. Crystallogr. 62, 72–82 (2006). [PubMed: 16369096]
70. McCoy AJ et al. Phaser crystallographic software. J. Appl. Crystallogr. 40, 658–674 (2007). [PubMed: 19461840]
71. Afonine PV et al. Towards automated crystallographic structure refinement with phenix.refine. Acta Crystallogr. D Biol. Crystallogr. 68, 352–367 (2012). [PubMed: 22505256]
72. Schindelin J et al. Fiji: an open-source platform for biological-image analysis. Nat. Methods 9, 676–682 (2012). [PubMed: 22743772]

73. Zhou W, Ryan JJ & Zhou H Global analyses of sumoylated proteins in Saccharomyces cerevisiae. Induction of protein sumoylation by cellular stresses. J. Biol. Chem. 279, 32262–32268 (2004). [PubMed: 15166219]

74. Ma K, Vitek O & Nesvizhskii AI A statistical model-building perspective to identification of MS/MS spectra with PeptideProphet. BMC Bioinformatics 13 Suppl 16, S1 (2012).

75. Mirdita M et al. ColabFold: making protein folding accessible to all. Nat. Methods 19, 679–682 (2022). [PubMed: 35637307]

76. Evans R et al. Protein complex prediction with AlphaFold-Multimer. 10.1101/2021.10.04.463034 (2021) doi:10.1101/2021.10.04.463034.

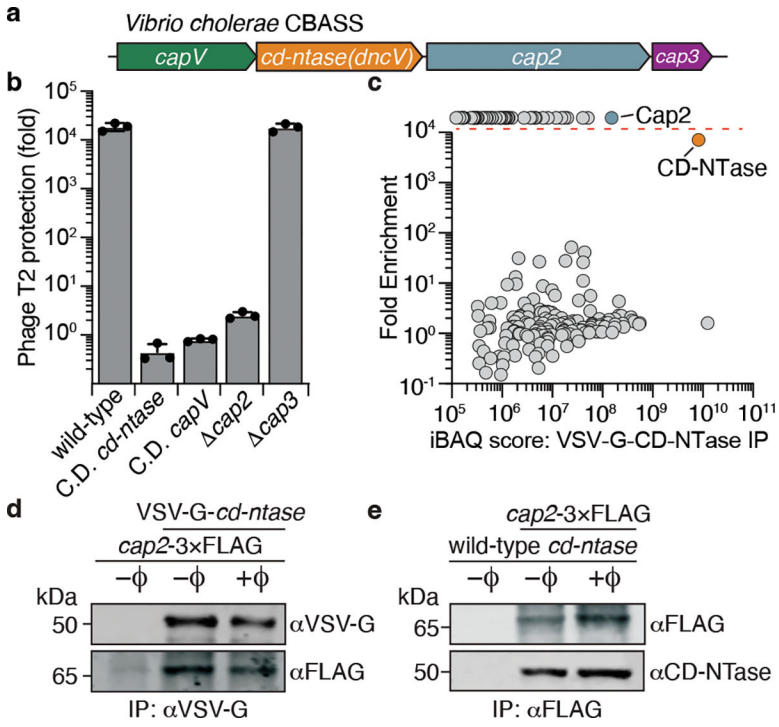


Figure 1: Cap2 is essential for CBASS function and directly interacts with the CD-NTase.

(a) Operon structure of CBASS from *V. cholerae*. See Supplementary Table 5 for relevant accession numbers.

(b) Efficiency of plating of phage T2 when infecting *E. coli* expressing *V. cholerae* CBASS with the indicated genotype. Data represent fold decrease in plaque forming units compared to bacteria expressing an empty vector. See Extended data Fig. 1b for infections with phages T4, T5, and T6. Catalytically dead (C.D.) *cd-ntase*: DID131AIA.; C.D. *capV*: S62A.

(c) Mass spectrometry of immunoprecipitated VSV-G-CD-NTase. Data are iBAQ quantitation score and fold enrichment comparing anti-VSV-G (αVSV-G) immunoprecipitations from bacteria expressing wild-type CBASS where the CD-NTase has an N-terminal VSV-G tag to a strain expressing the CBASS operon without a VSV-G tag. Cap2 and CD-NTase are represented as colored circles corresponding to (a) and are labeled. Circles above the dotted line are proteins with peptides only identified in the VSV-G-tagged samples and not untagged control.

(d) Western blot analysis of αVSV-G immunoprecipitation from *E. coli* expressing CBASS with the indicated genotypes. ±ϕ indicates phage T2 at a multiplicity of infection (MOI) of 2. See Extended data Fig. 1g for pre-IP sample analysis.

(e) Western blot analysis of αFLAG immunoprecipitation from *E. coli* expressing CBASS with the indicated genotype. ±ϕ indicates phage T2 at an MOI of 2. See Extended data Fig. 1h for pre-IP sample analysis.

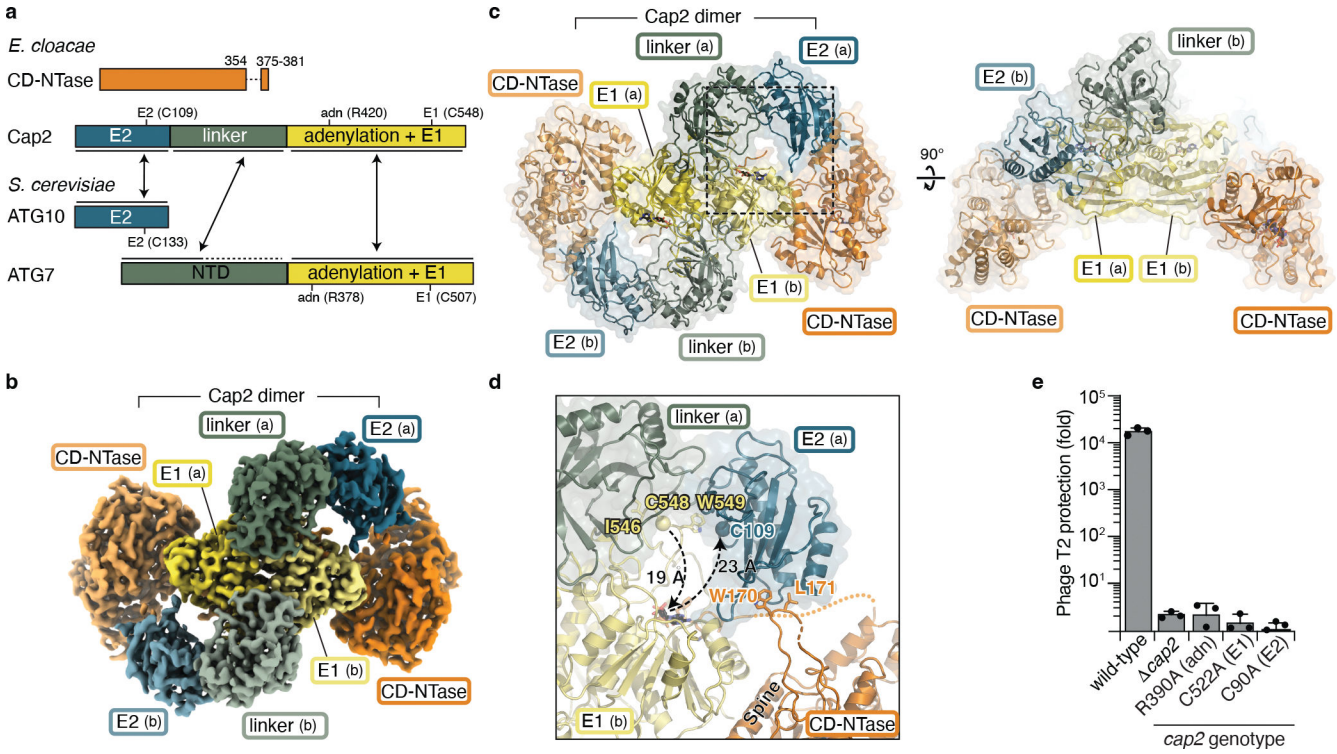


Figure 2: Cryoelectron microscopy structure of a Cap2–CD-NTase complex.

(a) Domain schematic of CD-NTase and Cap2 from *Enterobacter cloacae* and ATG10 and ATG7 from *Saccharomyces cerevisiae*, with domains colored and labeled to represent similarity. See also Extended data Fig. 3.

(b) 2.74 Å resolution cryoEM density for the *E. cloacae* Cap2–CD-NTase (C.D. Cap2: C109A/C548A) complex, with domains colored as in (a). See also Extended Data Table 1, Extended data Fig. 1i–q, and Supplementary Fig. 2.

(c) Two views of the 2:2 heterotetrameric Cap2–CD-NTase complex, with domains colored as in (a).

(d) View of one set of active sites in the Cap2–CD-NTase complex, with the E1 and E2 active-site residues (C548 and C109, respectively; both mutated to alanine in this structure) shown as spheres.

(e) Efficiency of plating of phage T2 when infecting *E. coli* expressing CBASS with the indicated genotype. Adenylation residue (adn). Data plotted as in Fig. 1b. See Extended data Fig. 4a for infections with phages T4, T5, and T6.

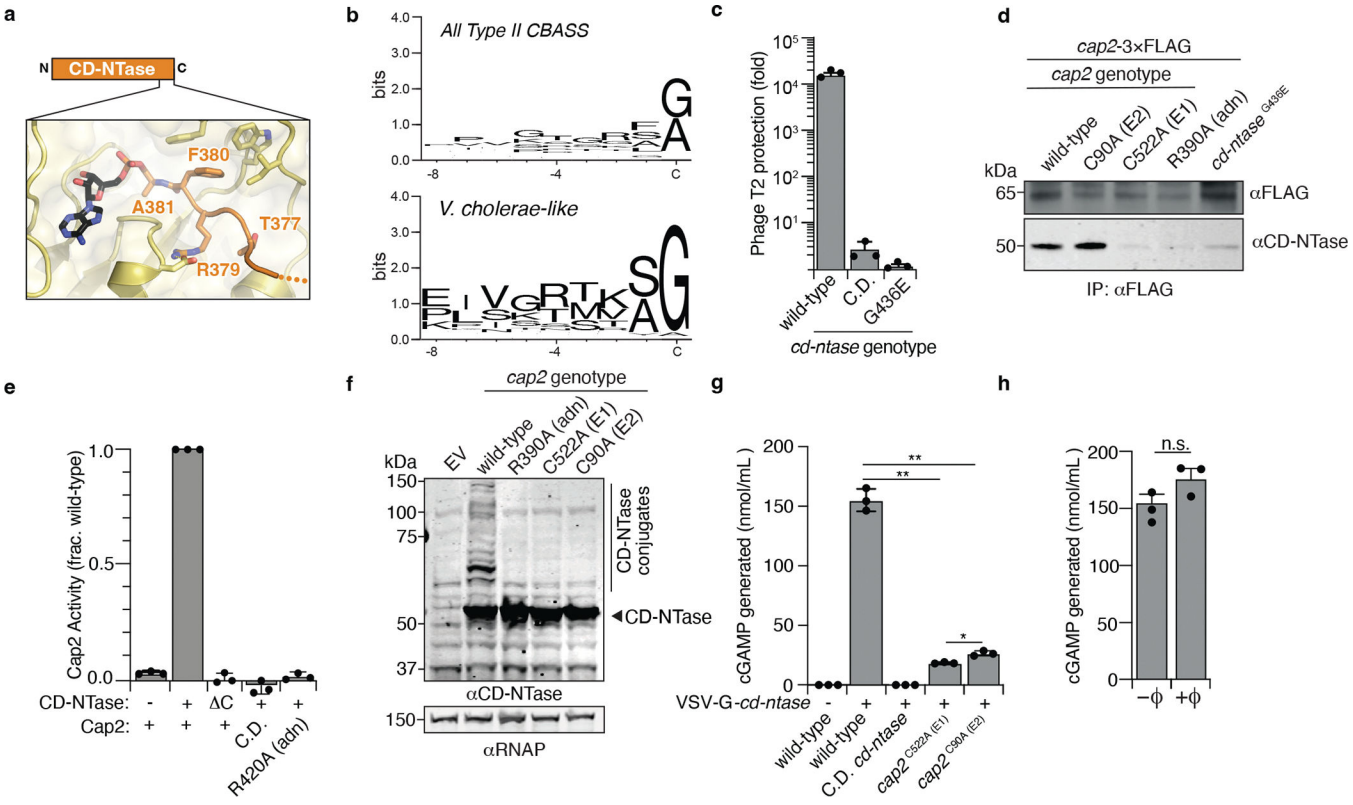


Figure 3: The CD-NTase is the substrate of Cap2.

(a) Cap2 adenylation active site in Cap2–CD-NTase cryoEM structure. The CD-NTase C-terminus (orange) is conjugated to AMP (black). See also Extended data Fig. 2e–k.

(b) Sequence logos for the C-terminal 9 residues of 1556 CD-NTase enzymes from diverse type II CBASS or CD-NTases from only the *V. cholerae*-like group (Clade A1; Extended data Fig. 5d). Data are depicted as bits and signified by the height of each residue.

(c) Efficiency of plating by phage T2 when infecting *E. coli* expressing CBASS with the indicated genotype. Data plotted as in Fig. 1b, see Extended data Fig. 4b for infections with phages T4, T5, and T6. C.D. CD-NTase: DID131AIA.

(d) Western blot analysis of αFLAG immunoprecipitation from *E. coli* expressing CBASS with the indicated genotype.

(e) *E. cloacae* Cap2 activity assay, representing Cap2-mediated catalysis as a fraction of wild-type. The indicated genotypes of Cap2 and CD-NTase were expressed from a single plasmid and the formation of a CD-NTase conjugate was measured (in this assay, CD-NTase is conjugated to the flexible N-terminus of His_6-Cap2; see Extended data Fig. 4e–k for details). N=3 independent biological replicates and data are presented as the mean values ± the SD. (−): no protein; (+): wild-type protein; (ΔC): CD-NTase lacking its C-terminal 19 residues; C.D. Cap2: C548A/C109A. See Extended data Fig. 3a for Cap2 protein alignment.

(f) Western blot analysis of cell lysates from *E. coli* expressing empty vector (EV) or *cap V-cd-ntase-cap2* (CBASS Δ*cap3*) with the indicated genotype. For (d), (e), and (f) adenylation residue (adn).

(g) cGAMP generated by αVSV-G immunoprecipitation from *E. coli* expressing CBASS Δ*cap3* with the indicated genotype. (−): CD-NTase without VSV-G; (+): CD-NTase with N-terminal VSV-G: C.D. CD-NTase: DID131AIA. See also Extended data Fig. 5c–f.
(h) cGAMP generated by αVSV-G immunoprecipitation from *E. coli* expressing *capV-(vsv-g-cd-ntase)-cap2*. ±ϕ indicates phage T2 at an MOI of 2. For (g) and (h) N=3 technical replicates representative of three independent biological replicates. Data is presented as the mean values ± the SEM. Two-sided student's t-test was used to calculate significance; n.s., $p>0.05$; *, $p<0.05$ (p-value = 0.0028); **, $p<0.001$ (p-value<0.0001).

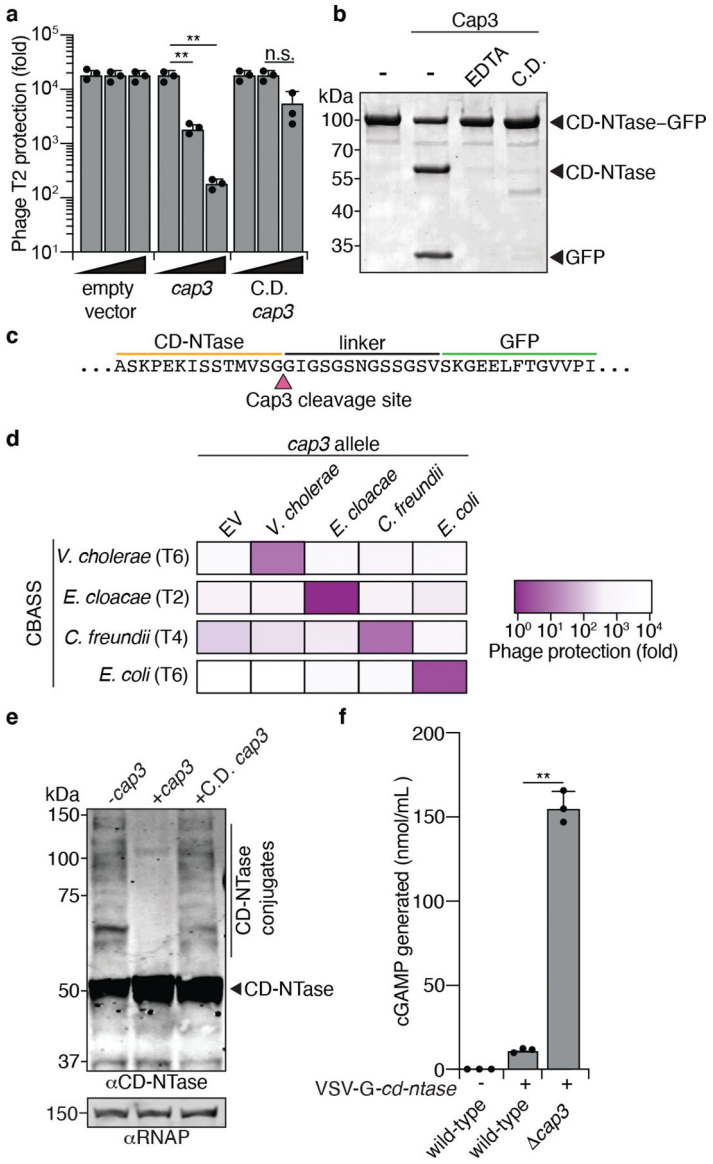


Figure 4: Cap3 antagonizes CBASS phage defense

(a) Efficiency of plating of phage T2 when infecting *E. coli* expressing CBASS Δ*cap3* in the absence or presence of overexpressed *cap3* with the indicated genotype. Data plotted as in Fig. 1b. C.D. *cap3*: HTH101ATA. See Extended data Fig. 6h for protein alignment and Extended data Fig. 6c for infections with phages T4, T5, and T6.

(b) Coomassie stained SDS-PAGE of a *V. cholerae* model substrate (CD-NTase–GFP fusion protein) incubated with *V. cholerae* Cap3 with the indicated reaction condition/genotype. See Extended data Fig. 7 for cleavage of CD-NTase mutants and for activity assays with *E. cloacae* Cap3.

(c) Summary of tryptic digest mass spectrometry analysis of the Cap3-treated CD-NTase bands as in **(b)**, showing the putative *V. cholerae* Cap3 cleavage site. See also Extended data Fig. 7b and Supplementary Table 3.

(d) Efficiency of plating of the indicated phage when infecting *E. coli* expressing CBASS and *cap3* from the indicated system. All samples contained 500 μM IPTG to induce expression of *cap3.* See also Extended data Fig. 6d–g.
(e) Western blot analysis of cell lysates from *E. coli* expressing CBASS Δ*cap3* plus a second vector expressing *cap3* with the indicated genotype. C.D. *cap3*: HTH101ATA.
(f) cGAMP generated by αVSV-G immunoprecipitation from *E. coli* expressing CBASS with the indicated genotype. (−): CD-NTase without VSV-G; (+): CD-NTase with N-terminal VSV-G. See also Extended data Fig. 5c–f. N=3 technical replicates representative of three independent biological replicates. Data is presented as the mean values ± the SEM. Two-sided Student's t-test was used to calculate significance; n.s., $p>0.05$; *, $p<0.05$; **, $p<0.001$ ($p\text{-value}<0.0001$).

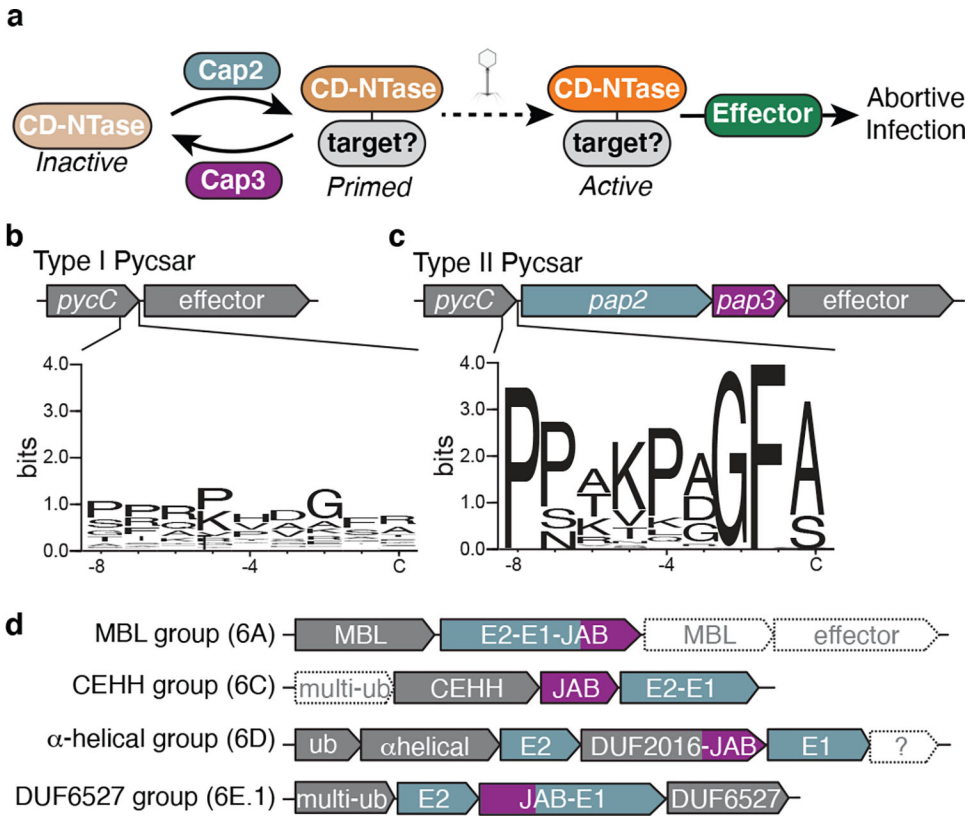


Figure 5: Proposed mechanism for the role of Cap2 and Cap3 in CBASS signaling

(a) Model depicting the role of Cap2 and Cap3 in CBASS regulation. Briefly, Cap2 conjugates the CD-NTase to an unknown target via an E1-E2 ubiquitin transferase-like mechanism. CD-NTase conjugation primes the CD-NTase for activation by phage infection. Upon infection, the CD-NTase becomes enzymatically active and generates a cyclic oligonucleotide second messenger which then activates an effector protein. Effector protein activity leads to cell death, inhibiting phage by abortive infection. This process is antagonized by Cap3 protease activity which removes the CD-NTase from target proteins, thereby limiting priming.

(b) General operon structure of type I Pycsar systems[38]. Below is a sequence logo for the C-terminal 9 residues of 550 PycC proteins encoded within type I Pycsar. See also Supplementary Table 9.

(c) General operon structure of identified type II Pycsar systems. Below is a sequence logo for the C-terminal 9 residues of 55 PycC encoded within type II Pycsar. See also Supplementary Table 9.

(d) General structures of operons that encode E1, E2 and JAB domain containing proteins. Genes are colored by domain type; E1 and E2 domains, blue; JAB domains, purple; all other domains, grey. Genes in dashed boxes are not always found within the operons. Operons are grouped by conserved protein domains and the E1-superfamily these groups is also indicated in paratheses[20]. Metallo-β-lactamase (MBL); metal binding domain (CEHH); tandem β-grasp fold domain containing protein (multi-ub); single β-grasp fold domain containing protein (ub). See also Extended data Fig. 8k and Supplementary Table 9.