Control of bacterial anti-phage signaling by a WYL domain transcription factor

Chelsea L. Blankenchip[1,2], Justin V. Nguyen[2,3], Rebecca K. Lau[1,2], Qiaozhen Ye[2], Yajie Gu[2], Kevin D. Corbett[2,4]†

[1] Biomedical Sciences Graduate Program, University of California, San Diego, La Jolla CA
[2] Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla CA
[3] Division of Biological Sciences, University of California, San Diego, La Jolla CA
[4] Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla CA

†To whom correspondence should be addressed: kcorbett@ucsd.edu

Abstract

Bacteria use diverse immune systems to defend themselves from ubiquitous viruses termed bacteriophages (phages). Many anti-phage systems function by abortive infection to kill a phage-infected cell, raising the question of how these systems are regulated to avoid activation and cell killing outside the context of infection. Here, we identify a transcription factor associated with the widespread CBASS bacterial immune system, that we term CapW. CapW forms a homodimer and binds a palindromic DNA sequence in the CBASS promoter region. Two crystal structures of CapW reveal how the protein switches from a DNA binding-competent state to a ligand-bound state that cannot bind DNA due to misalignment of dimer-related DNA binding domains. We show that CapW strongly represses CBASS gene expression in uninfected cells, and that CapW disruption likely results in toxicity due to uncontrolled CBASS expression. Our results parallel recent findings with BrxR, a transcription factor associated with the BREX anti-phage system, and suggest that CapW and BrxR are the founding members of a family of universal anti-phage signaling proteins.

Keywords: Bacteriophage / Transcription Factor / CBASS / BREX

Introduction

The constant evolutionary arms race between bacteria and the viruses that infect them, called bacteriophages (phages), has resulted in the evolution of a broad array of bacterial immune systems. These include the well-characterized restriction-modification and CRISPR/Cas systems, but also less well-understood systems like BREX (Bacteriophage Exclusion) (Goldfarb *et al*, 2015; Gordeeva *et al*, 2019), CBASS (Cyclic Oligonucleotide-Based Anti-Phage Signaling System) (Burroughs *et al*, 2015; Cohen *et al*, 2019), and others (Doron *et al*, 2018; Gao *et al*, 2020). Many bacterial immune systems, including CBASS, act by a so-called abortive infection mechanism and kill an infected cell to avoid phage reproduction (Cohen *et al.*, 2019; Ye *et al*, 2020). Because of their destructive power, abortive infection systems must be exquisitely tuned to avoid activation in an uninfected cell, but activate rapidly and reliably upon infection.

CBASS immune systems are widespread and extremely diverse, with over 6,200 distinct systems identified to date across bacteria and archaea (Cohen *et al.*, 2019; Whiteley *et al*, 2019). CBASS systems show diversity in both their activation mechanisms and their cell-killing mechanisms. All CBASS systems encode an oligonucleotide cyclase related to the mammalian innate-immune sensor cGAS (cyclic GMP-AMP synthase), that synthesizes a cyclic di- or trinucleotide second messenger molecule (Govande *et al*, 2021; Whiteley *et al.*, 2019), and a second-messenger activated effector protein that mediates cell death to abort the viral infection. So-called Type I CBASS systems encode only these two proteins, suggesting that their cGAS-like enzyme has an innate infection-sensing capability (Millman *et al*, 2020). The recent discovery of Type I CBASS systems that encode both a cGAS-like enzyme and an effector related to mammalian STING (STimulator of Interferon Genes) suggests that the mammalian cGAS-STING innate immune pathway evolved from these systems (Morehouse *et al*, 2020).

The majority of bacterial CBASS systems encode putative regulators that likely provide an additional level of control over their activation. Type II CBASS systems encode enzymes related to eukaryotic ubiquitin-transfer machinery, whose roles in signaling remain unknown (Burroughs *et al.*, 2015; Millman *et al.*, 2020). In Type III CBASS systems, peptide-binding HORMA domain proteins and a AAA+ ATPase, Trip13, regulate cGAS activation through HORMA-peptide binding (Ye *et al.*, 2020). These regulators are thought to represent the evolutionary precursors of the diverse HORMA domain signaling protein family in eukaryotes, which regulate key cell-cycle checkpoints, DNA repair in both mitotic and meiotic cells, and autophagy signaling (Tromer *et al*, 2019; Ye *et al.*, 2020).

The diversity of bacterial CBASS systems suggests that additional mechanisms of phage infection sensing and signaling regulation in these systems remain to be discovered. Here, we identify a novel CBASS-associated transcription factor that we name CapW. We show that CapW is a transcriptional repressor that binds the promoter region of its cognate CBASS operon to inhibit expression of CBASS genes outside the context of infection. Two structures of CapW from different bacteria reveal a dimeric assembly that likely binds a small-molecule ligand to control DNA binding and transcriptional repression. Parallel discovery of CapW-like transcription factors in other bacterial immune systems including BREX (Luyten *et al*, 2021; Picton *et al*, 2021) suggests that CapW and its relatives make up a family of

infection-responsive transcriptional switches, which directly sense phage infection and activate expression of diverse immune systems.

Results

Identification of a WYL-family transcription factor associated with CBASS systems

To identify novel regulators of anti-phage signaling in CBASS systems, we manually inspected a set of Type III CBASS systems encoded by different *E. coli* isolates. We identified an uncharacterized gene in a CBASS system from *E. coli* strain upec-117, which is predicted to possess both a winged-helix (winged helix-turn-helix; wHTH) DNA binding domain and a WYL domain, named for a conserved three-amino acid motif (tryptophan-tyrosine-leucine) in this domain (Makarova *et al*, 2014). We named this gene *capW* (CBASS-associated protein with WYL domain) (**Fig. 1a**). To identify other CBASS systems with *capW* genes, we searched for homologs of *E. coli* upec-117 *capW* within 10 kb of 6,233 previously-identified CBASS systems in diverse bacteria and archaea (Cohen *et al.*, 2019), and identified 160 CBASS systems encoding this protein (**Table S1**). In these systems, the *capW* gene is consistently found upstream of the core CBASS genes and encoded on the opposite strand (**Fig. 1a**). CapW is associated with all major CBASS types, including Type I (40 systems), Type II (59 systems), and Type III (56 systems) (**Fig. 1b**). These systems encode diverse putative effector proteins, including the phospholipase CapV, endonucleases Cap4 and NucC, and transmembrane proteins (Cap14 and Cap15) (**Fig. 1a**). The *E. coli* upec-117 CBASS system encodes an uncharacterized predicted effector similar to MTA/SAH family nucleoside phosphorylases (Millman *et al.*, 2020), which we named Cap17, and an uncharacterized predicted 3'-5' exonuclease that we named Cap18 (**Table S2**).

The WYL domain adopts an SH3 β-barrel fold and has been proposed to function in ligand binding in multiple contexts, including in the regulation of bacterial immunity as part of CRISPR/Cas systems (Makarova *et al.*, 2014). This family includes WYL1, a dimeric WYL domain-containing protein that binds single-stranded RNA and positively regulates Cas13d in a Type VI-D CRISPR-Cas system domain (Yan *et al*, 2018; Zhang *et al*, 2019). The largest family of bacterial WYL domain-containing proteins possess CapW's domain structure with an N-terminal wHTH domain, a central WYL domain, and a conserved C-terminal domain termed WCX (WYL C-terminal extension). The structural mechanisms of these proteins, and their roles in bacterial signaling, are largely unknown. The most well-characterized members of this family are PafB and PafC, which together regulate the DNA damage response in mycobacteria (Fudrini Olivencia *et al*, 2017; Muller *et al*, 2018). A recent structure of a naturally-fused PafBC protein from *Arthrobacter aurescens* in the absence of bound ligand or DNA (Müller *et al*, 2019) revealed an asymmetric overall structure, leaving unanswered the question of how these proteins' DNA binding propensity may be regulated by ligand binding. Overall, our bioinformatics data suggest that CapW is a ligand-responsive transcription factor that may regulate expression of its associated CBASS operon in response to phage infection. Moreover, CapW represents a large family of uncharacterized bacterial transcription factors involved in diverse signaling pathways.

CapW specifically binds the CBASS promoter region

To determine whether CapW is a transcription factor that controls CBASS expression, we first purified the protein and tested its binding to the shared promoter region between the CBASS core genes and CapW. We purified three CapW proteins, from *E. coli* upec-117 (*Ec* CapW), *Stenotrophomonas maltophilia* C11 (*Sm* CapW; 41% identical to *Ec* CapW), and *Pseudomonas aeruginosa* PA17 (*Pa* CapW; 64% identical to *Ec* CapW). All three proteins are associated with Type III CBASS systems (**Fig. 1a, S1a**). By size exclusion chromatography coupled to multi-angle light scattering (SEC-MALS), both *Sm* CapW and *Pa* CapW form homodimers in solution (**Fig S1b,c**).

Using an electrophoretic mobility shift assay (EMSA), we found that *Sm* CapW robustly binds to its cognate promoter region (**Fig. 2a**). As *Sm* CapW is a homodimer, we searched its promoter for palindromic sequences that would represent a likely binding site for a two-fold symmetric CapW dimer. We identified two palindromic sequences within the *S. maltophilia* C11 CBASS promoter region (**Fig. 2b**), and used a fluorescence polarization assay to demonstrate specific binding to a 24-bp imperfect palindrome that overlaps the promoter's -10 site (**Fig. 2e**). Next, we searched the promoter regions of *E. coli* upec-117 and *P. aeruginosa* PA17 CBASS for similarly-positioned palindromic sequences. We identified a 21-bp imperfect palindrome in the promoter of *E. coli* upec-117 CBASS, and a 19-bp imperfect palindrome in the promoter of *P. aeruginosa* PA17 CBASS, both of which overlap their promoters' -10 sites (**Fig. 2c, 2d**). Using fluorescence polarization, we found that *Ec* CapW and *Pa* CapW specifically bind these sequences (**Fig. 2f, 2g**). Based on these data, we conclude that CBASS-associated CapW proteins bind palindromic sequences that overlap the -10 sites within the promoters of their cognate CBASS operons. Because the -10 site represents the binding site for RNA polymerase within the promoter, this finding suggests that CapW acts as a transcriptional repressor by inhibiting RNA polymerase-promoter binding.

Structure and DNA binding mechanism of CapW

We next crystallized and determined the structure of *Sm* CapW to 1.89 Å resolution (**Table S3**). As expected from our SEC-MALS analysis (**Fig. S2b**), CapW crystallized as a homodimer with each chain possessing an N-terminal winged-helix (wHTH) domain, a central WYL domain, and a C-terminal WCX domain (**Fig. 3a**). The CapW homodimer adopts a distinctive domain-swapped overall architecture with the N-terminal wHTH domains adjacent to one another, followed by extended linkers reaching across the dimer such that each protomer's WYL domain interacts primarily with the wHTH domain of the opposite protomer (**Fig. 3a**). The C-terminal WCX domain adopts an extended α-β fold, and each protomer's WCX domain reaches back across the top of the dimer to interact with the WYL domain of the opposite protomer (**Fig. 3a**). The two WCX domains define a groove across the top of the CapW dimer that extends between the putative ligand-binding sites of each WYL domain (see next section).

The overall architecture of CapW is very similar to that of another recently-discovered bacterial immune system associated transcription factor, BrxR. Two contemporary studies report the discovery of BrxR as a regulator of BREX anti-phage systems, and determine structures of the protein from two different bacteria (Luyten *et al.*, 2021; Picton *et al.*, 2021). These structures reveal that BrxR and CapW share a common domain organization and overall domain-swapped architecture, with wHTH domains on

one face of the dimer, and a putative ligand-binding surface on the opposite face that is made up of WYL and WCX domains.

In DNA-free structures of both CapW and BrxR, the wHTH domains of the two protomers are positioned adjacent to one another, with their DNA-binding surfaces aligned on one face of the dimer. We could model DNA binding to each wHTH domain of CapW based on known wHTH-DNA complex structures, revealing that the two wHTH domains in the CapW dimer are nearly aligned to bind a single palindromic DNA sequence 16-20 base pairs in length (**Fig. 3b**), close to the length of palindromic sequences we identified as CapW binding sites (**Fig. 2b**). Supporting this model, a structure of DNA-bound BrxR reveals that the two wHTH domains bind ~11 bp apart on a single 24-bp imperfect palindrome DNA (Luyten *et al.*, 2021).

Based on our model of DNA-bound CapW, we designed two mutants to disrupt DNA binding: a single Arg32 to alanine mutant (*Sm* CapW^{R32A}), and a triple mutant with Ser42, Gln45, and Ser47 all mutated to alanine (*Sm* CapW$^{SQS\text{-}AAA}$) (**Fig. 3c**). While Arg32 is highly conserved in CapW but not in BrxR, Ser42/Gln45/Ser47 are conserved in BrxR and all three residues are directly involved in DNA binding (Luyten *et al.*, 2021). In *Sm* CapW, both CapW^{R32A} and CapW$^{SQS\text{-}AAA}$ eliminated detectable binding of the protein to its binding site in the CBASS promoter (**Fig. 3d**). Together, these data suggest that our structure of *Sm* CapW likely represents a DNA-binding competent conformation of CapW.

The structure of *P. aeruginosa* PA17 CapW reveals a non-DNA binding conformation

The central WYL domain of CapW adopts an Sm-type SH3 β-barrel fold similar to bacterial Hfq (host factor for RNA bacteriophage Qβ replication) proteins, which bind small RNAs (Link *et al*, 2009; Schumacher *et al*, 2002). Other WYL domain containing proteins have been shown to bind single-stranded RNA (Zhang *et al.*, 2019) or DNA (Andis *et al*, 2018), suggesting that the CapW WYL domain may also bind nucleic acids and/or a small molecule ligand. The WYL domain is named for a set of three highly-conserved amino acids, tryptophan-tyrosine-leucine, located on one of the domain's β-strands. In CapW, the tryptophan residue is highly conserved, while the typical tyrosine residue in WYL domains is replaced by a highly conserved histidine (His171; **Fig. S2b**). His171 is solvent-exposed on the top face of the CapW dimer, and is surrounded by a cluster of highly-conserved hydrophobic and polar residues including Tyr145, Ser147, Trp156, Arg169, Arg183, Phe185, and Arg189 (**Fig 4a,c, S2b, S3**). Comparing these residues to a sequence logo constructed from PFAM13280, which represents over 18,000 WYL-WCX domain proteins, we found that *Sm* CapW Tyr145 aligns with a highly-conserved tyrosine in this family, and Arg183/Phe185/Arg189 are situated in a region that shows high conservation in the broader WYL domain family (**Fig. S2b**). Based on this conservation and on the previously-identified ligand binding role for WYL domains (Andis *et al.*, 2018; Makarova *et al.*, 2014; Zhang *et al.*, 2019), we propose that this conserved surface on the CapW WYL domain may bind a nucleic acid or small-molecule ligand. Notably, the two putative ligand-binding sites on CapW are situated near one another on the top face of the dimer at either end of a groove defined by the two WCX domains (**Fig. 4a**). The positioning of these sites suggest that they may cooperate to bind an extended nucleic acid ligand.

A major question for the function of CapW and related transcription factors is how these proteins are regulated in order to sense and respond to bacteriophage infection. Based on the putative ligand-binding role for the WYL domain, a compelling model is that ligand binding induces a conformational change that alters the ability of CapW to bind DNA. We determined a crystal structure of a second CapW protein, *Pa* CapW, that sheds light on this potential mode of regulation. We crystallized and determined a 2.3 Å resolution crystal structure of *Pa* CapW in a low-pH condition (pH 5.0). The structure reveals that *Pa* CapW shares the same overall architecture as *Sm* CapW, forming a homodimer of protomers with wHTH, WYL, and WCX domains (**Fig. 4b**). The symmetric arrangement of WYL domains linked by WCX domains is similar to *Sm* CapW, but in *Pa* CapW the WYL domains are each rotated ~10° downward and inward toward the wHTH domains. This motion is accompanied by a slight widening of the groove between the WYL domains' ligand-binding sites and bordered by the two WCX domains (**Fig. 4b**), and also by a rearrangement of the C-termini of each WCX domain. In *Sm* CapW, the WCX domain C-terminus forms two short α-helices with an intervening loop that folds along the top of the domain (**Fig. 4a**). In *Pa* CapW, this region undergoes a domain swap to fold against the opposite protomer's WCX domain (**Fig. 4b**). Finally, we observe that a salt bridge between a conserved arginine in the WYL domain and an aspartate on the WCX domain is broken in the *Pa* CapW structure, enabling the WCX domain to move out and away from the dimer-related WYL domain (**Fig. 4d**).

The highly-conserved tryptophan residue of the WYL domain family is situated on the opposite face of the WYL domain compared to the putative ligand binding site. Why, then, is this residue so highly conserved? In our structure of *Sm* CapW, Trp170 is pointed down from the WYL domain and packs against a hydrophobic face of an α-helix in the extended wHTH-WYL domain linker (α4; **Fig. S4a**). In our low-pH structure of *Pa* CapW, the WYL domains pinch inward by ~6 Å (**Fig. 4b**), positioning the equivalent tryptophan residues (Trp181) and the nearby α5 helices too close to one another to accommodate the α4 helices in their original configuration. As a result, the α4 helix of each CapW protomer rotates downward away from the WYL domains (**Fig. S4b**). This change, and the motion of the WYL domains in general, in turn cause a striking ~70° rotation of each wHTH domain compared to its position in *Sm* CapW (**Fig. 4b**). Whereas the wHTH domains are aligned for cooperative DNA binding in our structure of *Sm* CapW, in *Pa* CapW they are completely misaligned and would be unable to bind a contiguous DNA sequence.

Our biochemical data shows that *Pa* CapW binds DNA with an affinity equivalent to that of *Sm* CapW or *Ec* CapW, yet our structure of this protein reveals a conformation that is clearly unable to bind DNA. We propose that the low-pH crystallization condition for *Pa* CapW induces a conformational change to a state that mimics the structure of ligand-bound CapW. Comparing the DNA-binding and non-DNA binding states of CapW reveals a key role for the WYL domain's conserved tryptophan residue in driving conformational changes between these two states.

CapW is a transcriptional repressor for CBASS

Our identification of CapW binding sites overlapping the -10 site of its cognate CBASS promoter suggested that CapW binding may repress CBASS transcription by blocking RNA polymerase binding to this site. To test this idea, we generated an expression reporter system for *Ec* CapW. We first generated

DNA-binding mutants of *Ec* CapW equivalent to *Sm* CapW^{R32A} (R43A) and CapW$^{SQS\text{-}AAA}$ (S53A/Q56A/S58A), and found that both mutants eliminated detectable binding of *Ec* CapW to its palindromic site by fluorescence polarization (**Fig. 5a**). We next constructed an expression reporter system with the *E. coli* upec-117 *capW* gene and CBASS promoter linked to a gene encoding sfGFP (**Fig. 5b**). To track CapW expression directly in this system, we also fused a C-terminal FLAG tag to the *capW* gene. We measured both GFP and CapW expression using Western blots, in the presence and absence of *capW* or a DNA-binding mutant. In the presence of wild-type *capW*, the levels of both GFP and CapW were too low to detect by Western blotting (**Fig. 5b**). In contrast, disrupting *capW* or eliminating DNA binding through the R43A or SQS-AAA mutants resulted in strong expression of GFP (**Fig. 5b**). We observed a similar increase in expression of CapW itself in both DNA-binding mutants, suggesting that CapW regulates its own transcription in addition to that of the core CBASS genes. We could identify a promoter in the *E. coli* upec-117 CBASS operon that is oriented in the reverse direction compared to the promoter driving core CBASS expression (**Fig. S5**). This promoter likely drives expression of *capW* and *cap17*, and our data suggests that CapW binding to the CBASS promoter region affects transcription in both directions. Overall, these data show that CapW is a strong transcriptional repressor, capable of repressing transcription bidirectionally from the CBASS promoter even when the levels of CapW itself are undetectable by Western blotting.

We next sought to understand CapW's role in bacteriophage defense by the *E. coli* upec-117 CBASS system. We found that a plasmid encoding the entire system confers strong resistance to bacteriophage λ cI-, and that disrupting second messenger synthesis by the system's cGAS-like enzyme eliminates this resistance (**Fig. 5c**). Despite repeated attempts, we were unable to generate mutants that either disrupted the *capW* gene or eliminated the protein's DNA binding activity in the plasmid encoding this system. After mutagenesis, clones containing these mutations consistently also showed large deletions of critical regions of the CBASS operon (not shown). Our inability to isolate these mutants in the context of a full CBASS system, when the same mutants are readily obtainable in our GFP reporter system, suggests that the CBASS system may be toxic without a functional *capW* gene. Combined with our expression reporter data, this observation implies that *capW* disruption results in overexpression of CBASS genes and toxicity in host cells. This model agrees with contemporary studies on the related BrxR transcription factor, in which deletion of BrxR leads to increased expression of BREX proteins and toxicity to host cells (Luyten *et al.*, 2021; Picton *et al.*, 2021).

Discussion

Bacterial CBASS immune systems are highly diverse, and an emerging theme in these systems is that they encode multiple redundant regulatory mechanisms in order to prevent anti-phage signaling, and associated cell killing, outside the context of infection. All CBASS systems encode a cGAS-like oligonucleotide cyclase that likely possesses an inherent phage-dependent activation mechanism (Burroughs *et al.*, 2015; Millman *et al.*, 2020). Type II, III, and IV CBASS systems additionally encode regulators - many of which represent ancestral forms of important eukaryotic signaling proteins - that provide a second level of control over the activity of their cognate cGAS-like enzymes (Millman *et al.*,

2020; Ye *et al.*, 2020). Here, we identify a third mode of regulation in some CBASS systems: a WYL domain-containing transcription factor, CapW. We propose that CapW strongly represses CBASS expression in uninfected cells, and that this repression is released upon infection by the production of a particular nucleic acid species or small-molecule ligand by the infected cell. Our two structures suggest that ligand binding causes a conformational change in CapW's WYL and WCX domains that triggers a large-scale rotation of the protein's wHTH domains to release the protein from DNA (**Fig. 5d**). A key next step will be to identify the CapW-binding ligand, and to dissect how phage infection leads to its production.

CapW shares a similar overall architecture to another recently-identified transcription factor, BrxR, which controls the expression of BREX immune systems (Luyten *et al.*, 2021; Picton *et al.*, 2021). Like CapW, BrxR binds BREX promoter sequences to repress expression in uninfected cells. Upon infection, BrxR-mediated repression is released to enable BREX gene expression and activate the anti-phage response (Luyten *et al.*, 2021). We propose that the allosteric mechanism we identify for ligand-induced conformational changes in CapW also applies to BrxR, and to the larger family of bacterial immune system-associated WYL domain transcription factors. Sequence alignments of BrxR and similar transcription factors associated with CRISPR-Cas and restriction-modification systems (Picton *et al.*, 2021) shows that all of these proteins share a conserved ligand-binding surface (**Fig. S2b-c**), suggesting that they may bind the same ligand. Moreover, they also share the conserved tryptophan residue after which the WYL domain is named, which we implicate in ligand-induced conformational changes in CapW. Thus, CapW and BrxR likely represent the first examples of a universal anti-phage signaling factor in bacteria, which can sense phage infection and activate expression of a broad array of immune systems.

Outside bacterial immune systems, the WYL domain and WYL-WCX domain pair likely play a range of signaling roles. For example, the transcription factor PafBC possesses a tandem array of wHTH, WYL, and WCX domains, and plays a role in regulating DNA damage in mycobacteria (Müller *et al.*, 2019). The Cas13d regulator WYL1 shares a similar domain structure with an N-terminal ribbon-helix-helix domain, a central WYL domain, and a C-terminal dimerization domain (Yan *et al.*, 2018; Zhang *et al.*, 2019). Examining sequence conservation in PFAM13280, which includes over 18,000 bacterial proteins that possess the WYL-WCX domain pair embedded in a variety of protein scaffolds, reveals that the ligand-binding site is much more variable across this family than within the smaller group of proteins associated with immune systems (**Fig. S2b**). Thus, while immune system-associated WYL proteins likely bind a common ligand, other family members have likely evolved to bind a variety of ligands to control diverse signaling pathways. Key directions for future work will be to determine the range of ligands bound by diverse WYL proteins, how ligand binding is coupled to conformational changes within different protein scaffolds, and how these conformational changes are coupled to signaling.

Materials and Methods

Bioinformatics

To comprehensively search CBASS systems for homologs of *E. coli* upec-117 CapW (NCBI #WP_001534693.1), we exported the genomic DNA sequence +/-10 kb of 6233 previously-reported CD-NTases (Cohen *et al.*, 2019) using the Integrated Microbial Genomes (IMG) database at the DOE Joint Genome Institute (https://img.jgi.doe.gov). We used NCBI Genome Workbench (https://www.ncbi.nlm.nih.gov/tools/gbench/) to perform custom TBLASTN searches for proteins related to *E. coli* upec-117 CapW. CBASS system type and effector assignments for each hit were taken from Cohen et al. (Cohen *et al.*, 2019) and manually updated. Each hit was manually inspected for the presence of CapW specifically associated with CBASS rather than a neighboring operon.

Protein Expression and Purification

Codon-optimized sequences encoding full-length CapW from *E. coli* upec-117 (NCBI #WP_001534693.1), *P. aeruginosa* PA17 (NCBI #WP_023098969.1), and *S. maltophilia* C11 (IMG #2657474953) were synthesized (Invitrogen/GeneArt) and cloned into UC Berkeley Macrolab vector 2BT (Addgene #29666) to generate constructs with N-terminal TEV protease-cleavable His_6-tags. Point mutations were generated by PCR-based mutagenesis.

Proteins were expressed in *E. coli* strain Rosetta 2 (DE3) pLysS (EMD Millipore). Cultures were grown at 37°C to A_{600}=0.7, then induced with 0.25 mM IPTG and shifted to 20°C for 16 hours. Cells were harvested by centrifugation and resuspended in buffer A (25 mM Tris pH 7.5, 10% glycerol, and 1 mM NaN_3) plus 300 mM NaCl, 5 mM imidazole, 5 mM β-mercaptoethanol. Proteins were purified by Ni^{2+}-affinity (Ni-NTA agarose, Qiagen) then passed over an anion-exchange column (Hitrap Q HP, GE Life Sciences) in Buffer A plus 100 mM NaCl and 5 mM β-mercaptoethanol, collecting flow-through fractions. Tags were cleaved with TEV protease (Tropea *et al*, 2009), and cleaved protein was passed over another Ni^{2+} column (collecting flow-through fractions) to remove uncleaved protein, cleaved tags, and tagged TEV protease. The protein was passed over a size exclusion column (Superdex 200, GE Life Sciences) in buffer GF (buffer A plus 300 mM NaCl and 1 mM dithiothreitol (DTT)), then concentrated by ultrafiltration (Amicon Ultra, EMD Millipore) to 10 mg/ml and stored at 4°C. All point mutants showed equivalent migration on size exclusion column compared to wild type. For selenomethionine derivatization, protein expression was carried out in M9 minimal media supplemented with amino acids plus selenomethionine prior to IPTG induction (Van Duyne *et al*, 1993).

For characterization of oligomeric state by size exclusion chromatography coupled to multi-angle light scattering (SEC-MALS), 100 µL of purified *P. aeruginosa* PA17 or *S. maltophilia* C11 CapW at 5 mg/mL was injected onto a Superdex 200 Increase 10/300 GL column (GE Life Sciences) in buffer GF. Light scattering and refractive index profiles were collected by miniDAWN TREOS and Optilab T-rEX detectors (Wyatt Technology), respectively, and molecular weight was calculated using ASTRA v. 6 software (Wyatt Technology).

Crystallization and Structure Determination

For crystallization of *P. aeruginosa* PA17 CapW, selenomethionine-derivatized protein in buffer GF (9 mg/mL) was mixed 1:1 with well solution containing 0.27 M $LiSO_4$, 1% PEG 400, and 0.1 M sodium acetate (pH 5.0) in hanging-drop format. Crystals were cryoprotected by the addition of 30% glycerol, and flash-frozen in liquid nitrogen. Diffraction data were collected at the Advanced Photon Source NE-CAT beamline 24ID-E (see support statement below) and processed with the RAPD data-processing pipeline (https://github.com/RAPD/RAPD), which uses XDS (Kabsch, 2010) for data indexing and reduction, AIMLESS (Evans & Murshudov, 2013) for scaling, and TRUNCATE (Winn *et al*, 2011) for conversion to structure factors. We determined the structure by single-wavelength anomalous diffraction methods in the PHENIX Autosol wizard (Terwilliger *et al*, 2009). We manually rebuilt the initial model in COOT (Emsley *et al*, 2010), and refined in phenix.refine (Afonine *et al*, 2012) using positional and individual B-factor refinement (**Table S3**).

For crystallization of *S. maltophilia* C11 CapW, protein in a buffer containing 20mM Tris pH 8.5, 1 mM DTT, and 100 mM NaCl (14mg/mL) was mixed 2:1 with well solution containing 0.1 M Tris pH 8.5, and 1.5 M Lithium sulfate in sitting drop format. Crystals were cryoprotected by the addition of 24% glycerol and flash-frozen in liquid nitrogen. Diffraction data were collected at the Advanced Light Source BCSB beamline 5.0.2 (see support statement below) and processed with the DIALS data-processing pipeline (https://dials.github.io) (Beilsten-Edmands *et al*, 2020). We determined the structure by molecular replacement in PHASER (McCoy *et al*, 2007), using individual wHTH and WYL domain structures from *P. aeruginosa* PA17 CapW as search models. We manually rebuilt the initial model in COOT (Emsley *et al.*, 2010), and refined in phenix.refine (Afonine *et al.*, 2012) using positional and individual B-factor refinement (**Table S3**).

DNA binding

For electromobility shift DNA-binding assays (EMSA), the *S. maltophilia* C11 CBASS promoter region was amplified via PCR with one primer 5'-labeled with 6-carboxyfluorescein (5'-6-FAM), followed by gel-purification (Machery-Nagel Nucleospin). For the EMSA control, a random sequence from within the *S. maltophilia* C11 CapW gene that was similar in length and GC content to the *S. maltophilia* C11 promoter region was amplified via PCR with one primer 5'-labeled with 6-carboxyfluorescein (5'-6-FAM), followed by gel-purification (Machery-Nagel Nucleospin). 10 µL reactions with 100 nM DNA and indicated concentrations of protein were prepared in a buffer containing 50 mM Tris-HCl pH 8.5, 50 mM NaCl, 5 mM $MgCl_2$, 5% glycerol, and 1 mM DTT. After a 1 hour incubation at room temperature, reactions were loaded onto 2% TBE-agarose gels in running buffer with 0.5X TBE pH 8.5 running buffer, run for 2 hours at 60 V at 4°C, and imaged using a Bio-Rad ChemiDoc system (Cy2 filter settings).

For DNA binding fluorescence polarization (FP) assays, a 30 bp double-stranded DNA was produced by annealing complementary oligos, one of which was 5'-6-FAM labeled. Binding reactions (30 µL) contained 25 mM Tris pH 8.5, 50 mM NaCl, 5% glycerol, 5 mM $MgCl_2$, 1 mM DTT, 0.01% nonidet p40 substitute, 50 nM DNA, and the indicated amounts of protein. After a 10 minute incubation at room temperature, fluorescence polarization was read using a Tecan Infinite M1000 PRO fluorescence plate reader, and binding data were analyzed with Graphpad Prism v.9.2.0 using a single-site binding model.

GFP Reporter Assays

To generate a GFP reporter plasmid, a DNA sequence encoding full-length *E. coli* upec-117 CapW and its promoter, adjacent to a gene encoding super-folder GFP was synthesized (IDT) and cloned via isothermal assembly into pBR322. Point mutations were generated by PCR-based mutagenesis. Vectors were transformed into *E. coli* strain JP313 (Economou *et al*, 1995). 100 µL of saturated overnight culture was added to 5 mL LB broth plus ampicillin and grown at 37°C to an OD_{600} of 0.4-0.5. 500 µL of culture was pelleted by centrifugation and resuspended in 100 µL of 2x SDS-PAGE loading buffer (125 mM Tris pH 6.8, 20% Glycerol, 4% SDS, 200 mM DTT, 180 µM bromophenol blue). Samples were boiled for 5 minutes, then 1-10 µL was loaded onto an SDS-PAGE gel. Proteins were transferred to a PVDF membrane (Bio-Rad Trans-Blot Turbo), then the membranes were blocked with 5% nonfat dry milk and blotted for appropriate proteins. Blots were imaged using a Bio-Rad ChemiDoc system using filters to image horseradish peroxidase activity. Antibodies used: Mouse anti-GFP primary antibody (Roche) at 1:3,000 dilution; Mouse anti-FLAG primary antibody (Sigma-Aldrich) at 1:3,000 dilution; Goat anti-mouse HRP-linked secondary antibody (Millipore Sigma) at 1:30,000 dilution.

Bacteriophage infection assays

To measure anti-phage activity of *E. coli* upec-117 CBASS, we transformed *E. coli* strain JP313 (Economou *et al.*, 1995) with a pLOCO2 plasmid encoding the entire CBASS system from this species (bases 69873 to 76528 of NCBI accession # CP054230.1) (gift from A. Whiteley). As a negative control, we used the same pLOCO2 backbone encoding lacI and sfGFP. For bacteriophage infection plaque assays, a single bacterial colony or equivalent mass was picked from a fresh LB agar plate and grown at 37°C in LB broth plus ampicillin to an OD_{600} of 0.2-0.3. 500 µL of cells were mixed with 4.5 mL of 0.35% LB top agar, then poured onto LB plates containing carbenicillin. Plates were spotted with 3 µL of bacteriophage λcI- (cI gene deleted to inhibit lysogeny and promote clear plaque formation) diluted in phage buffer (150 mM NaCl, 40 mM Tris pH 7.5, 10 mM $MgSO_4$) plus 1 mM $CaCl_2$ at 6 dilutions: ~$4x10^7$ PFU/mL and 5 10-fold dilutions thereof. Plates were incubated at 37°C for 18 hours, then imaged.

APS NE-CAT Support Statement

This work is based upon research conducted at the Northeastern Collaborative Access Team beamlines, which are funded by the National Institute of General Medical Sciences from the National Institutes of Health (P30 GM124165). The Eiger 16M detector on the 24-ID-E beam line is funded by a NIH-ORIP HEI grant (S10OD021527). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357.

ALS BCSB Support Statement

The Berkeley Center for Structural Biology is supported in part by the Howard Hughes Medical Institute. The Advanced Light Source is a Department of Energy Office of Science User Facility under Contract No. DE-AC02-05CH11231. The Pilatus detector on 5.0.1. was funded under NIH grant S10OD021832. The

ALS-ENABLE beamlines are supported in part by the National Institutes of Health, National Institute of General Medical Sciences, grant P30 GM124169.

Acknowledgements

We thank A. Whiteley for the gift of a plasmid encoding the *E. coli* upec-117 CBASS system, and members of the Corbett lab for helpful discussions and comments on the manuscript. K.D.C. acknowledges support from UC San Diego and NIH/NIAID grant R21 AI148814. C.L.B. was supported by the UCSD Molecular Biophysics Training Grant, NIH Grant T32 GM008326.

Author Contributions

CLB planned the study with RKL and KDC, determined the structure of *Sm* CapW and performed most biochemical and genetic assays, generated figures, and wrote the paper with KDC. **JVN** performed SEC-MALS analysis and determined the structure of *Pa* CapW with assistance from QY. **RKL** planned the study with CLB and KDC, performed bioinformatics analysis and assisted CLB with protein purification, biochemical and genetic assays, and structure determination of *Sm* CapW. **QY** assisted JVN with purification, SEC-MALS analysis, and structure determination of *Pa* CapW. **YG** assisted CLB with biochemical assays including SEC-MALS measurements. **KDC** planned the study with CLB and RKL, assisted with structure determination, interpreted data, generated figures, and wrote the paper with CLB.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability

The datasets produced in this study are available in the following databases:

Protein structure data: Protein Data Bank IDs 7TB5 (http://dx.doi.org/10.2210/pdb7tb5/pdb) and 7TB6 (http://dx.doi.org/10.2210/pdb7tb6/pdb).

X-ray diffraction data: SBGrid Data Bank IDs 869 (https://data.sbgrid.org/dataset/869/) and 870 (https://data.sbgrid.org/dataset/870/).

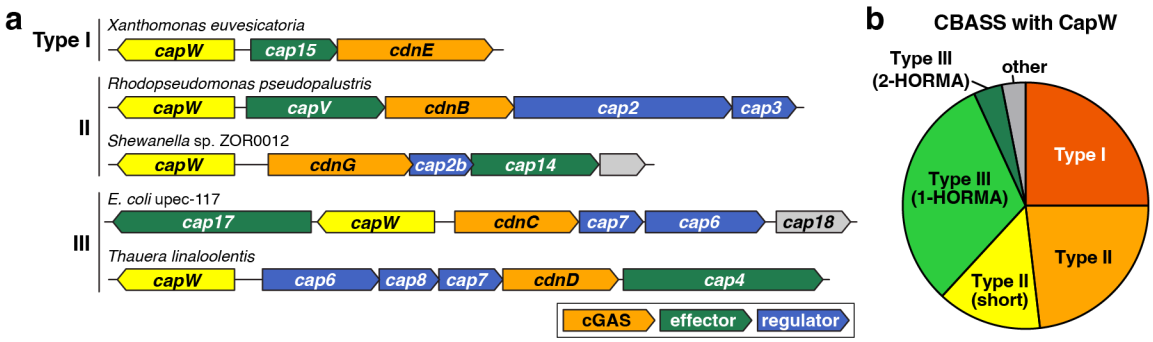

Figure 1. Identification of CapW in CBASS systems

(a) Operon schematics of five CBASS systems encoding *capW* genes (yellow). Each system's CBASS type is noted at left, and genes are color-coded with cGAS-like enzymes orange, effectors green, regulators blue, and unknown proteins gray. cGAS-like enzymes are labeled according to their clade, as previously assigned (Whiteley *et al.*, 2019). See **Table S1** for a comprehensive list of CapW-associated CBASS systems, and **Table S2** for gene names and descriptions. (b) Pie chart showing the number of CapW-associated CBASS systems, with Type I red, Type II orange, Type II (short) yellow, Type III (1-HORMA) light green, Type III (2-HORMA) dark green, and other/unknown gray (see **Table S1**).

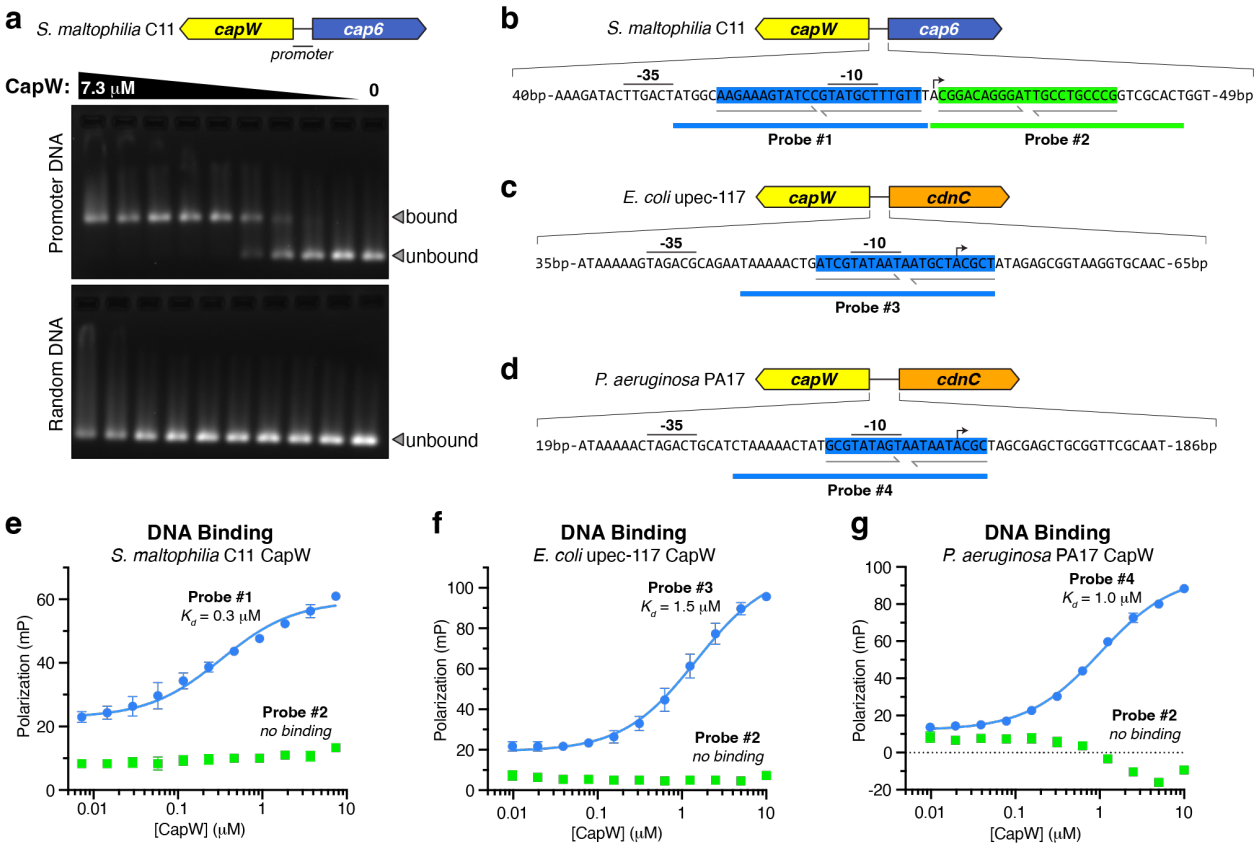

Figure 2. CapW binds the CBASS promoter region

(a) Electrophoretic mobility shift assay (EMSA) for *Sm* CapW binding the full promoter region of its cognate CBASS system (upper panel), or an equivalent-length random DNA (lower panel). The leftmost lane contains 7.3 μM CapW (monomer), and lanes 2-9 contain progressive 2-fold dilutions. Estimated *Kd* = 0.2-0.3 μM. (b) Diagram of the *S. maltophilia* C11 CBASS promoter region, with two palindromic sequences highlighted in blue and green. Two 30-bp DNA probes used for fluorescence polarization binding assays are also shown. Promoter sequences (-35, -10, and TSS sites) were predicted with the BPROM server (Solovyev & Salamov, 2011). (c) Diagram of the *E. coli* upec-117 CBASS promoter region, with palindromic sequence highlighted in blue. (d) Diagram of the *P. aeruginosa* PA17 CBASS promoter region, with palindromic sequence highlighted in blue. (e) Fluorescence polarization assay showing robust binding of *Sm* CapW to Probe #1 (blue, *Kd* = 0.3 +/- 0.1 μM), but not Probe #2 (green). For panels (e), (f), and (g), datapoints represent the mean +/- standard deviation of three measurements. (f) Fluorescence polarization assay showing binding of *Ec* CapW to Probe #3 (blue: *Kd* = 1.5 +/- 0.3 μM), but not Probe #2 from *S. maltophilia* C11 CBASS (green). (g) Fluorescence polarization assay showing binding of *Pa* CapW to Probe #4 (blue: *Kd* = 1.0 +/- 0.1 μM), but not Probe #2 from *S. maltophilia* C11 CBASS (green).

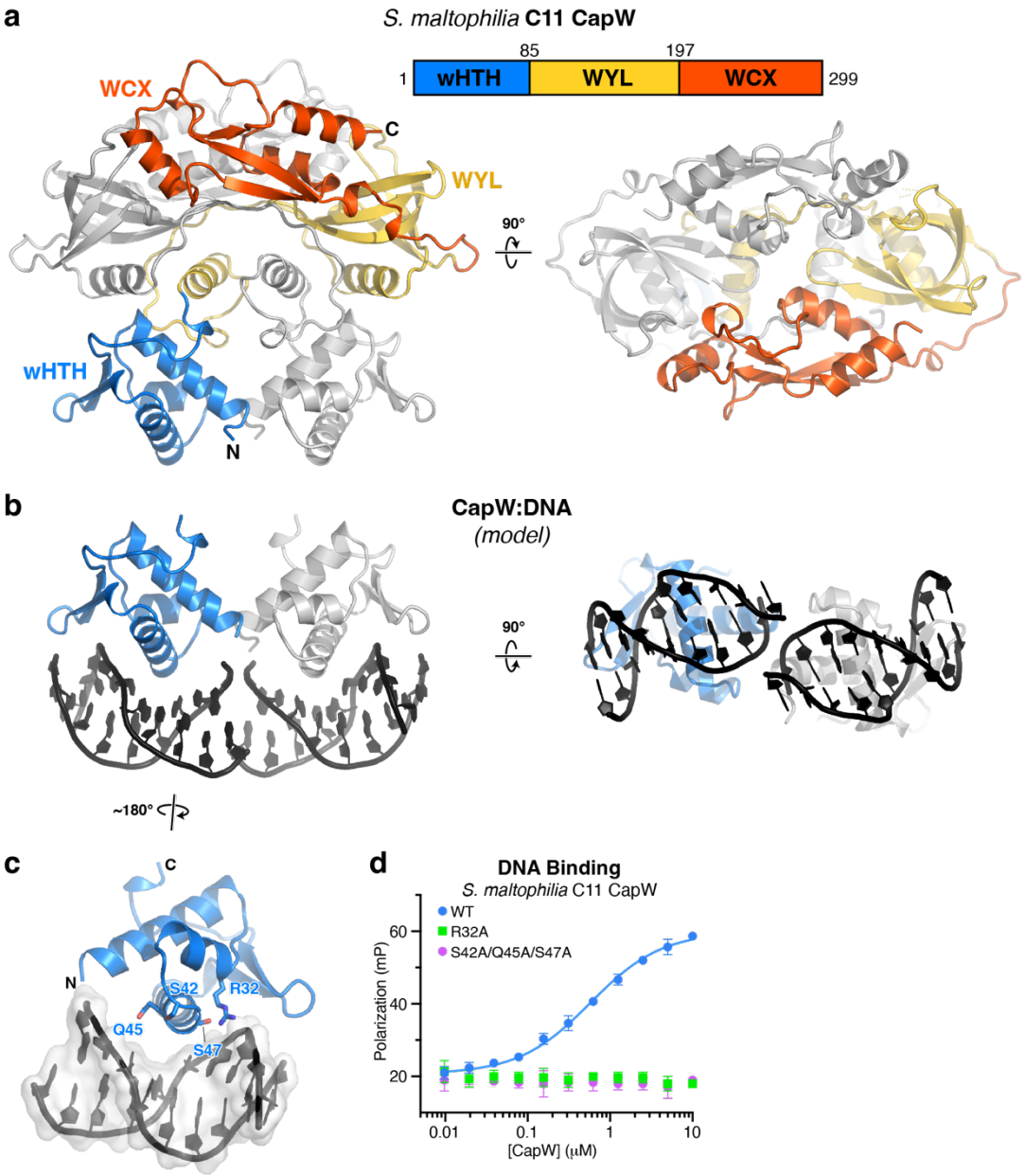

Figure 3. Structure of *Sm* CapW

(a) Structure of the *Sm* CapW dimer, with wHTH domain colored blue, WYL domain yellow, and WYL C-terminal extension (WCX) domain orange. (b) Model of the CapW wHTH domains binding DNA, generated by docking a DNA-bound wHTH domain (PDB ID 1CF7) (Zheng *et al*, 1999) onto each CapW wHTH domain. The bottom view (right) shows that the two wHTH domains are nearly aligned for binding a contiguous 16-20 bp palindromic DNA sequence. (c) Closeup view of the CapW:DNA model, with residues chosen for mutagenesis labeled. (d) Fluorescence polarization assay showing binding of *Sm* CapW (WT and indicated mutants) to Probe #1 from **Fig. 2b**.

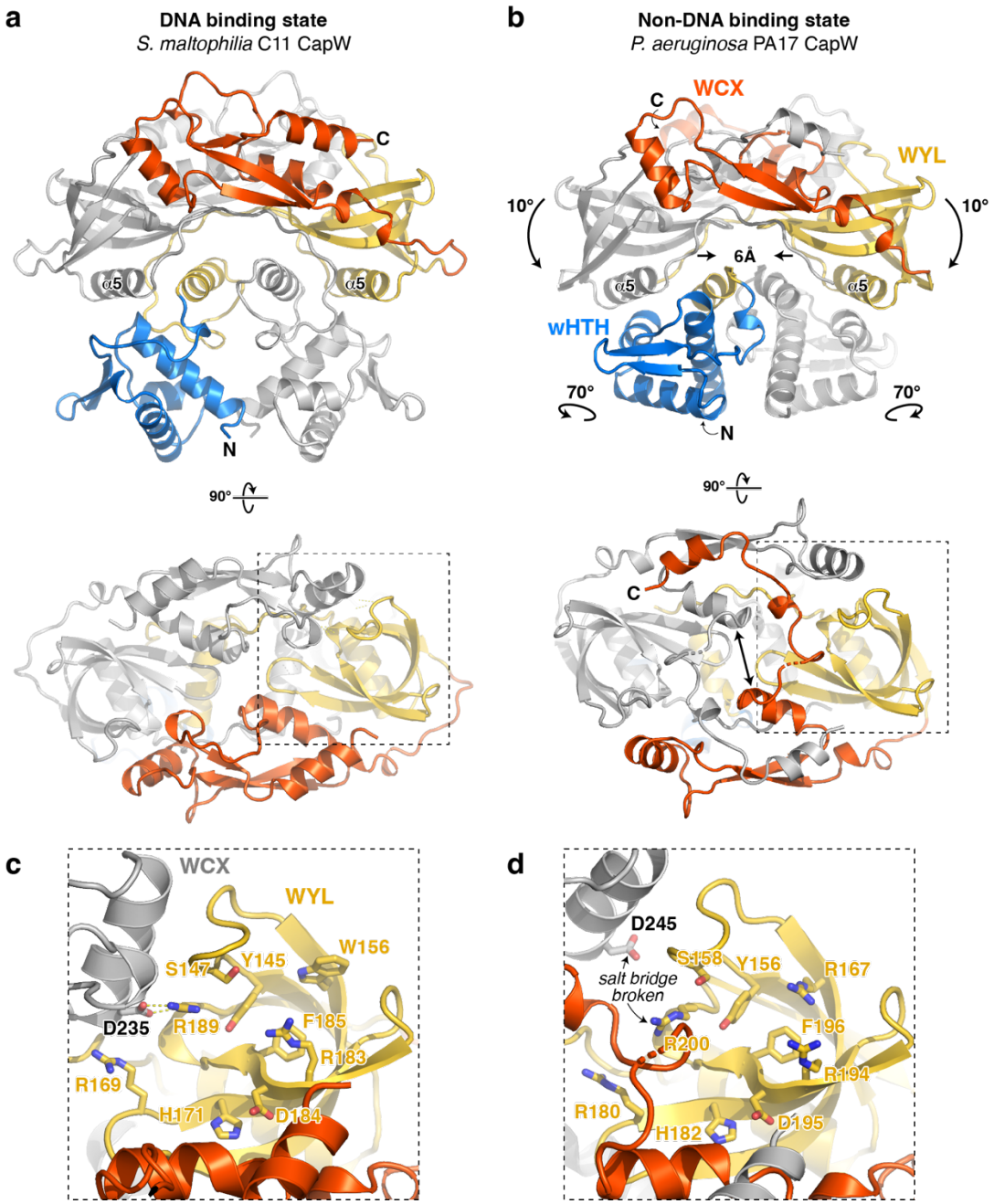

Figure 4 - Structure of *Pa* CapW

(a) Structure of *Sm* CapW, with wHTH domain colored blue, WYL domain yellow, and WYL C-terminal extension (WCX) domain orange. (b) Low-pH structure of *Pa* CapW, with domains colored as in panel (a). (c) Closeup of the putative ligand-binding site on the *Sm* CapW WYL domain. WYL domain residues are labeled in yellow, and residues from the dimer-related WCX domain (gray) are labeled in black. See **Figure S3** for surface conservation. (d) Closeup of the putative ligand-binding site on the *Pa* CapW WYL domain. WYL domain residues are labeled in yellow, and residues from the dimer-related WCX domain (gray) are labeled in black. The arginine-asparate salt bridge observed between WYL and the dimer-related WCX domain in *Sm* CapW is broken in this structure (arrow).

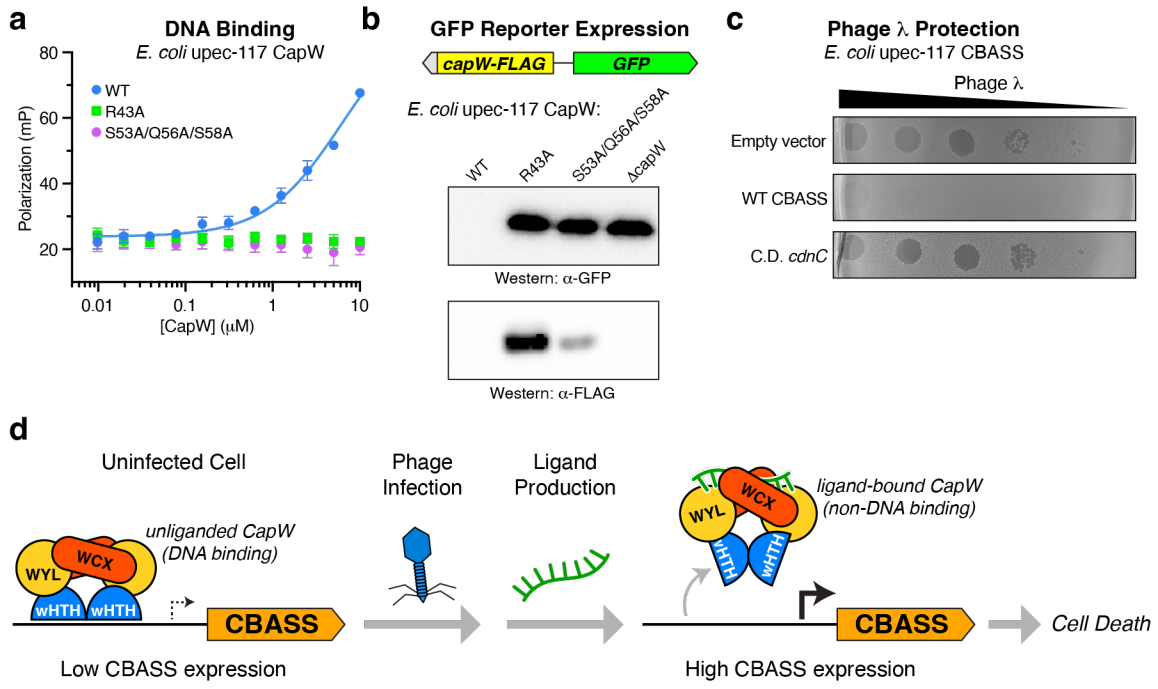

Figure 5. CapW is a transcriptional repressor

(a) Fluorescence polarization assay showing binding of *Ec* CapW (WT and indicated mutants) to Probe #3 from **Fig. 2c**. (b) Reporter assay with *E. coli* upec-117 CapW and promoter region, with GFP in place of the CBASS core genes. Western blots show expression of GFP (top) and CapW-FLAG (bottom) in log-phase cells in the presence of wild-type or mutant CapW. *ΔcapW*: stop codon inserted after codon 10. See **Fig. S5** for a diagram of the *E. coli* upec-117 CBASS promoter with both forward and reverse promoter sequences. (c) Plaque assays with the *E. coli* upec-117 CBASS system and bacteriophage λ cI- (10-fold dilutions). C.D. *cdcC*: D73N/D75N catalytic-dead mutant. (d) Model for CapW and related wHTH-WYL-WCX domain transcription factors in anti-phage defense. In an uninfected cell, the dimeric transcription factor binds the promoter of its cognate defense system and suppresses its expression. Upon infection, a small-molecule or nucleic acid ligand is produced that binds to the transcription factor and induces a conformational change, releasing it from DNA to enable expression and trigger the defensive response.

References

Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, Terwilliger TC, Urzhumtsev A, Zwart PH, Adams PD *et al* (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta crystallographica Section D, Biological crystallography* 68: 352-367

Andis NM, Sausen CW, Alladin A, Bochman ML (2018) The WYL Domain of the PIF1 Helicase from the Thermophilic Bacterium Thermotoga elfii is an Accessory Single-Stranded DNA Binding Module. *Biochemistry* 57: 1108-1118

Beilsten-Edmands J, Winter G, Gildea R, Parkhurst J, Waterman D, Evans G (2020) Scaling diffraction data in the DIALS software package: algorithms and new approaches for multi-crystal scaling. *Acta Crystallogr D Struct Biol* 76: 385-399

Burroughs AM, Zhang D, Schäffer DE, Iyer LM, Aravind L (2015) Comparative genomic analyses reveal a vast, novel network of nucleotide-centric systems in biological conflicts, immunity and signaling. *Nucleic Acids Research* 43: 10633-10654

Cohen D, Melamed S, Millman A, Shulman G, Oppenheimer-Shaanan Y, Kacen A, Doron S, Amitai G, Sorek R (2019) Cyclic GMP–AMP signalling protects bacteria against viral infection. *Nature* 574: 691-695

Doron S, Melamed S, Ofir G, Leavitt A, Lopatina A, Keren M, Amitai G, Sorek R (2018) Systematic discovery of antiphage defense systems in the microbial pangenome. *Science* 359: eaar4120

Duncan-Lowey B, McNamara-Bordewick NK, Tal N, Sorek R, Kranzusch PJ (2021) Effector-mediated membrane disruption controls cell death in CBASS antiphage defense. *Mol Cell* 81: 5039-5051

Economou A, Pogliano JA, Beckwith J, Oliver DB, Wickner W (1995) SecA membrane cycling at SecYEG is driven by distinct ATP binding and hydrolysis events and is regulated by SecD and SecF. *Cell* 83: 1171-1181

Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta crystallographica Section D, Biological crystallography* 66: 486-501

Evans PR, Murshudov GN (2013) How good are my data and what is the resolution? *Acta crystallographica Section D, Biological crystallography* 69: 1204-1214

Fatma S, Chakravarti A, Zeng X, Huang RH (2021) Molecular mechanisms of the CdnG-Cap5 antiphage defense system employing 3',2'-cGAMP as the second messenger. *Nat Commun* 12: 6381

Fudrini Olivencia B, Muller AU, Roschitzki B, Burger S, Weber-Ban E, Imkamp F (2017) Mycobacterium smegmatis PafBC is involved in regulation of DNA damage response. *Sci Rep* 7: 13987

Gao L, Altae-Tran H, Böhning F, Makarova KS, Segel M, Schmid-Burgk JL, Koob J, Wolf YI, Koonin EV, Zhang F (2020) Diverse enzymatic activities mediate antiviral immunity in prokaryotes. *Science* 369: 1077-1084

Goldfarb T, Sberro H, Weinstock E, Cohen O, Doron S, Charpak-Amikam Y, Afik S, Ofir G, Sorek R (2015) BREX is a novel phage resistance system widespread in microbial genomes. *EMBO J* 34: 169-183

Gordeeva J, Morozova N, Sierro N, Isaev A, Sinkunas T, Tsvetkova K, Matlashov M, Truncaite L, Morgan RD, Ivanov NV *et al* (2019) BREX system of Escherichia coli distinguishes self from non-self by methylation of a specific DNA site. *Nucleic Acids Res* 47: 253-265

Govande AA, Duncan-Lowey B, Eaglesham JB, Whiteley AT, Kranzusch PJ (2021) Molecular basis of CD-NTase nucleotide selection in CBASS anti-phage defense. *Cell Rep* 35: 109206

Iyer LM, Burroughs AM, Aravind L (2006) The prokaryotic antecedents of the ubiquitin-signaling system and the early evolution of ubiquitin-like β-grasp domains. *Genome Biology* 7: R60

Kabsch W (2010) XDS. *Acta crystallographica Section D, Biological crystallography* 66: 125-132

Karplus PA, Diederichs K (2012) Linking crystallographic model and data quality. *Science (New York, NY)* 336: 1030-1033

Lau RK, Ye Q, Birkholz EA, Berg KR, Patel L, Mathews IT, Watrous JD, Ego KM, Whiteley AT, Lowey B *et al* (2020) Structure and mechanism of a cyclic trinucleotide-activated bacterial endonuclease mediating bacteriophage immunity. *Molecular Cell* 77: 723-733

Link TM, Valentin-Hansen P, Brennana RG (2009) Structure of Escherichia coli Hfq bound to polyriboadenylate RNA. *Proceedings of the National Academy of Sciences of the United States of America* 106: 19292-19297

Lowey B, Whiteley AT, Keszei AFA, Morehouse BR, Mathews IT, Antine SP, Cabrera VJ, Kashin D, Niemann P, Jain M *et al* (2020) CBASS Immunity Uses CARF-Related Effectors to Sense 3'-5'- and 2'-5'-Linked Cyclic Oligonucleotide Signals and Protect Bacteria from Phage Infection. *Cell* 182: 38-49 e17

Luyten Y, Hausman D, Young J, Doyle LA, Ubilla-Rodriguez NC, Lambert AR, Arroyo C, Forsberg K, Morgan RM, Stoddard BL *et al* (2021) Identification and characterization of BrxR as a regulatory gene in the BREX phage restriction system. *bioRxiv*: 10.1101/2021.1112.1119.473356

Makarova KS, Anantharaman V, Grishin NV, Koonin EV, Aravind L (2014) CARF and WYL domains: ligand-binding regulators of prokaryotic defense systems. *Frontiers in Genetics* 5: 102

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. *Journal of applied crystallography* 40: 658-674

Millman A, Melamed S, Amitai G, Sorek R (2020) Diversity and classification of cyclic-oligonucleotide-based anti-phage signalling systems. *Nature Microbiology* 5: 1608-1615

Morehouse BR, Govande AA, Millman A, Keszei AFA, Lowey B, Ofir G, Shao S, Sorek R, Kranzusch PJ (2020) STING cyclic dinucleotide sensing originated in bacteria. *Nature* 586: 429-433

Muller AU, Imkamp F, Weber-Ban E (2018) The Mycobacterial LexA/RecA-Independent DNA Damage Response Is Controlled by PafBC and the Pup-Proteasome System. *Cell Rep* 23: 3551-3564

Müller AU, Leibundgut M, Ban N, Weber-Ban E (2019) Structure and functional implications of WYL domain-containing bacterial DNA damage response regulator PafBC. *Nature Communications* 10: 4653

Picton DM, Harling-Lee JD, Duffner SJ, Went SC, Morgan RM, Hinton JCD, Blower TR (2021) A widespread family of WYL-domain transcriptional regulators co-localises with diverse phage defence systems and islands. *bioRxiv*: 10.1101/2021.1112.1119.473342

Schumacher MA, Pearson RF, Møller T, Valentin-Hansen P, Brennan RG (2002) Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: A bacterial Sm-like protein. *EMBO Journal* 21: 3546-3556

Severin GB, Ramliden MS, Hawver LA, Wang K, Pell ME, Kieninger AK, Khataokar A, O'Hara BJ, Behrmann LV, Neiditch MB *et al* (2018) Direct activation of a phospholipase by cyclic GMP-AMP in El Tor Vibrio cholerae. *Proceedings of the National Academy of Sciences of the United States of America* 115: E6048-E6055

Solovyev V, Salamov A (2011) Automatic Annotation of Microbial Genomes and Metagenomic Sequences. *Metagenomics and its Applications in Agriculture, Biomedicine and Environmental Studies*: 61-78

Terwilliger TC, Adams PD, Read RJ, McCoy AJ, Moriarty NW, Grosse-Kunstleve RW, Afonine PV, Zwart PH, Hung L-W (2009) Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta crystallographica Section D, Biological crystallography* 65: 582-601

Tromer EC, van Hooff JJE, Kops GJPL, Snel B (2019) Mosaic origin of the eukaryotic kinetochore. *Proceedings of the National Academy of Sciences of the United States of America* 116: 12873-12882

Tropea JE, Cherry S, Waugh DS (2009) Expression and purification of soluble His(6)-tagged TEV protease. *Methods in molecular biology (Clifton, NJ)* 498: 297-307

Van Duyne GD, Standaert RF, Karplus PA, Schreiber SL, Clardy J (1993) Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *Journal of molecular biology* 229: 105-124

Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ (2009) Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics (Oxford, England)* 25: 1189-1191

Whiteley AT, Eaglesham JB, de Oliveira Mann CC, Morehouse BR, Lowey B, Nieminen EA, Danilchanka O, King DS, Lee ASY, Mekalanos JJ *et al* (2019) Bacterial cGAS-like enzymes synthesize diverse nucleotide signals. *Nature* 567: 194-199

Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AGW, McCoy A *et al* (2011) Overview of the CCP4 suite and current developments. *Acta crystallographica Section D, Biological crystallography* 67: 235-242

Yan WX, Chong S, Zhang H, Makarova KS, Koonin EV, Cheng DR, Scott DA (2018) Cas13d Is a Compact RNA-Targeting Type VI CRISPR Effector Positively Modulated by a WYL-Domain-Containing Accessory Protein. *Mol Cell* 70: 327-339 e325

Ye Q, Lau RK, Mathews IT, Birkholz EA, Watrous JD, Azimi CS, Pogliano J, Jain M, Corbett KD (2020) HORMA Domain Proteins and a Trip13-like ATPase Regulate Bacterial cGAS-like Enzymes to Mediate Bacteriophage Immunity. *Molecular Cell* 77: 709-722

Zhang H, Dong C, Li L, Wasney GA, Min J (2019) Structural insights into the modulatory role of the accessory protein WYL1 in the Type VI-D CRISPR-Cas system. *Nucleic Acids Res* 47: 5420-5428

Zheng N, Fraenkel E, Pabo CO, Pavletich NP (1999) Structural basis of DNA recognition by the heterodimeric cell cycle transcription factor E2F-DP. *Genes Dev* 13: 666-674