Title

Bacterial WYL domain transcriptional repressors sense single-stranded DNA to control gene expression.

Authors

Blankenchip, Chelsea

Corbett, Kevin

Publication Date

2024-11-26

DOI

10.1093/nar/gkae1101

Peer reviewed

Nucleic Acids Research, 2024, **52**, 13723–13732
https://doi.org/10.1093/nar/gkae1101
Advance access publication date: 26 November 2024
Gene regulation, Chromatin and Epigenetics

Bacterial WYL domain transcriptional repressors sense single-stranded DNA to control gene expression

Chelsea L. Blankenchip[1,2] and Kevin D. Corbett[2,3,*]

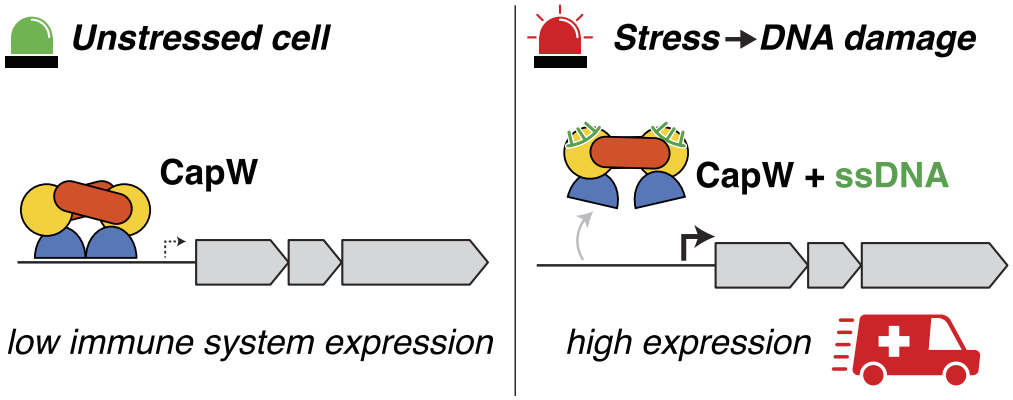
[1]Biomedical Sciences Graduate Program, University of California, San Diego, 9500 Gilman Dr. La Jolla, CA 92093, USA
[2]Department of Cellular and Molecular Medicine, University of California, San Diego, 9500 Gilman Dr. La Jolla, CA 92093, USA
[3]Department of Molecular Biology, University of California, San Diego, 9500 Gilman Dr. La Jolla, CA, 92093, USA

[*]To whom correspondence should be addressed. Tel: +1 858 534 7267; Fax: +1 858 534 7750; Email: kcorbett@ucsd.edu

Abstract

Bacteria encode a wide array of immune systems to protect themselves against ubiquitous bacteriophages and foreign DNA elements. While these systems' molecular mechanisms are becoming increasingly well known, their regulation remains poorly understood. Here, we show that an immune system-associated transcriptional repressor of the wHTH–WYL–WCX family, CapW, directly binds single-stranded DNA to sense DNA damage and activate expression of its associated immune system. We show that CapW mediates increased expression of a reporter gene in response to DNA damage in a host cell. CapW directly binds single-stranded DNA by-products of DNA repair through its WYL domain, causing a conformational change that releases the protein from double-stranded DNA. In an *Escherichia coli* CBASS system with an integrated *capW* gene, we find that CapW-mediated transcriptional activation is important for this system's ability to prevent induction of a λ prophage. Overall, our data reveal the molecular mechanisms of WYL-domain transcriptional repressors, and provide an example of how bacteria can balance the protective benefits of carrying anti-phage immune systems against the inherent risk of these systems' aberrant activation.

Graphical abstract



Introduction

Bacteria are constantly under threat from bacteriophages (phages), foreign DNA elements and environmental stresses, and have evolved a wide array of immune and stress-response systems to guard against these threats. In recent years, over 100 distinct immune systems have been identified that protect bacteria from infection by phages or transformation by plasmids (1–4). A typical bacterial genome encodes at least 5–10 distinct immune systems, often clustered in 'defense islands' (3,5,6). In general, these immune systems function either by directly targeting and destroying foreign DNA, slowing cell growth by inhibiting metabolism or depleting key metabolites, or by triggering cell death to prevent phage propagation (termed abortive infection) (5). The molecular mechanisms of many anti-phage and anti-plasmid immune systems are now known, but whether and how their expression is regulated in response to infection or other signals remains poorly understood.

Recently, two distinct sets of proteins have been identified that control the expression of diverse bacterial immune systems. One set comprises the DNA-binding protein CapH and the metalloprotease CapP. In unperturbed cells, CapH binds to the promoter of its associated immune system and represses transcription (7). In response to DNA damage, CapP

Received: June 26, 2024. **Revised:** October 4, 2024. **Editorial Decision:** October 21, 2024. **Accepted:** October 26, 2024

becomes active and cleaves CapH, releasing it from DNA to allow transcription (7). A second group comprises the related transcription factors CapW and BrxR, which function as transcriptional repressors and are associated with a broad set of anti-phage immune systems like BREX, CBASS, restriction-modification systems and others (8–10). While a CBASS-associated CapW was shown to mediate increased immune system expression upon phage infection, the activating signal and the molecular mechanism of de-repression in this protein family are unknown (8).

CapW and BrxR are part of a broad family of homodimeric transcriptional regulators with a common domain architecture comprising an N-terminal wHTH (winged helix–turn–helix) domain, a central WYL domain (named after a highly–conserved tryptophan–tyrosine–leucine motif) and a C-terminal WCX (WYL C-terminal extension) domain (11,12). Within this family, there are two distinct groups that function either as transcriptional activators or as repressors. WYL-domain transcriptional activators like PafBC and DriD bind promoters and recruit RNA polymerase to activate transcription when bound to single-stranded DNA (ssDNA) through their WYL domain (13–15). In contrast, transcriptional repressors like CapW and BrxR bind to promoters in the absence of ligand to repress transcription (8–10). Biochemical and genetic data suggest that these proteins also bind a ligand through their WYL domain, then undergo a conformational change to release DNA and enable transcription (8–12). However, these proteins' ligands are not known.

Here, we show that a CBASS-associated CapW transcription factor binds ssDNA through its WYL domain in order to respond to DNA damage. Using biochemical assays and X-ray crystallography, we show that ssDNA binding to CapW's WYL domain causes a conformational change in the CapW dimer that compromises double-stranded DNA (dsDNA) binding. Finally, we find that CapW-mediated expression of CBASS genes contributes to this system's ability to protect against induction of a λ prophage. Overall, our data reveal the molecular mechanisms of WYL domain transcriptional repressors, and reveal one specific biological role for these proteins in protection against lysogens.

Materials and methods

Strain verification

Keio collection strains, including the parent strain (BW25113, Horizon Discovery Biosciences product # OEC5042), Δ*recB* (clone JW2788, product # OEC4987-200828046) or Δ*recA* (clone JW2669, product # OEC4987-200827940) (28), were verified by polymerase chain reaction (PCR) with primer pairs designed to amplify 500 bp regions of either the *recA* (forward: CTTGGGGCAGGTGGTCTGC, reverse: GAGGCGTAGAATTTCAGCGC) or *recB* (forward: GGCAGTTCTGGATCTTAATGCTG; reverse: TGATGTGGTGTTAACGTCGGTTC) genes. For diagnostic PCR, a small amount of cells were lysed by boiling in 0.25% SDS for 5 min, then the resulting mixture was used as template for PCR. Reactions were performed with Phusion DNA Polymerase in Buffer HF (New England Biolabs). The PCR program included 30 cycles of (i) denaturation at 95°C for 5 s; (ii) annealing at 70°C for 15 s; and (iii) extension at 72°C for 15 s. Products were separated by agarose gel electrophoresis and visualized by ethidium bromide staining.

GFP reporter assays

Green fluorescent protein (GFP) reporter plasmids were generated as previously described (8). Briefly, a synthetic DNA sequence encoding *Escherichia coli* upec-117 CapW, its associated bidirectional promoter, and superfolder GFP was cloned into plasmid pBR322. CapW point mutants were generated by mutagenic PCR, and verified by full-plasmid sequencing. Vectors were transformed into *E. coli* strain JP313 (16) or Keio collection strains: parent strain BW25113, Δ*recB* or Δ*recA*. Overnight cultures were diluted into fresh LB broth plus ampicillin and grown at 37°C and shaking to an OD_{600} of 0.25–0.35. Cultures were treated with either 100 μg/ml zeocin or 1 μg/ml mitomycin C. After treatment, cultures were grown at 30°C with shaking. For each time point, 500 μl of culture was pelleted by centrifugation and resuspended in 50 μl of 2× SDS–PAGE loading buffer (125 mM Tris–HCl, pH 6.8, 20% glycerol, 4% SDS, 200 mM DTT, 0.02% bromophenol blue). Sample volumes were adjusted based on the culture density at each time point. Samples were boiled for 5 min, then 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 1:3000 dilution; Mouse anti-RNA polymerase primary antibody (clone NT63; BioLegend #10019-878) at 1:3000 dilution; and Goat anti-mouse HRP-linked secondary antibody (Millipore Sigma) at 1:30000 dilution.

GFP reporter plasmids were also transformed into the *E. coli* temperature sensitive λ lysogen strain JP858 (17). Overnight cultures were diluted into fresh LB broth plus ampicillin and kanamycin and grown at 30°C and shaking to an OD_{600} of 0.45–0.55. Subcultures were then heat shocked in a 42°C water bath for 15 min. After heat shock, cultures were incubated at 37°C with shaking. At indicated time points, samples were collected and western blots performed as described above.

Protein expression and purification

A codon-optimized gene encoding the full-length CapW from *E. coli* upec-117 (NCBI #WP 001534693.1), was synthesized (Invitrogen/GeneArt) and cloned into UC Berkeley Macrolab vector 2BT (Addgene #29666) to generate a construct with an N-terminal TEV protease-cleavable His_6-tag. The wild-type (WT) gene encoding the full-length CapW from *Rhizobium leguminosarum* (NCBI #WP 033184384.1) was also synthesized (IDT) and cloned into UC Berkeley Macrolab vector 2BT (Addgene #29666). 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 OD_{600} = 0.6, then induced with 0.25 mM IPTG and shifted to 20°C for 16 h. Cells were harvested by centrifugation and resuspended in buffer A (25 mM Tris–HCl, pH 8.5, 10% glycerol, 1 mM NaN_3) plus 300 mM NaCl, 5 mM imidazole and 5 mM β-mercaptoethanol. Resuspended cells were lysed by sonication (Branson Sonifier, 40 s at output control 5, 50–60% duty cycle), then cell debris was removed by centrifugation (Beckman Coulter JA-17 rotor and 17 000 RPM, 30 min, at 4°C). Proteins were purified by Ni^{2+}-affinity (Ni-NTA agarose, Qiagen) then passed over an anion-exchange column (Hitrap

Q HP, Cytiva) in Buffer A plus 100 mM NaCl, 5 mM imidazole and 5 mM β-mercaptoethanol, collecting flow-through fractions.

For structural studies, His_6-tags were cleaved with TEV protease (18). Briefly, 0.4 mg of His_6-TEV protease S219V mutant [recombinantly expressed and purified in-house; see (18)] was added to 22.5 ml of purified CapW in buffer A plus 100 mM NaCl, 5 mM imidazole and 5 mM β-mercaptoethanol, and the mixture was incubated at 4C for 4 days. 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 then passed over a size exclusion column (Superdex 200, Cytiva) in buffer GF [buffer A plus 300 mM NaCl and 1 mM dithiothreitol (DTT)], then concentrated by ultrafiltration (Amicon Ultra, EMD Millipore) to ∼14 mg/ml and stored at 4°C. All point mutants showed equivalent migration on a size exclusion column compared to wild type. For biochemical studies, TEV cleavage was not performed and proteins were directly passed over a size exclusion column following the anion-exchange chromatography step, then concentrated and stored −80°C.

Crystallization and structure determination

For crystallization of *Rl* CapW bound to ssDNA, protein in a buffer containing 20 mM Tris, pH 8.5, 75 mM NaCl and 1 mM DTT [13 mg/ml protein with 1.5× molar excess of a 7mer poly-T oligonucleotide (IDT)] was mixed in a ratio of 2 protein to 1 well solution with a buffer containing 0.2 M ammonium sulfate, 0.1 M sodium acetate and 22% PEG 4000 in sitting drop format. Crystals were transferred to a cryoprotectant solution containing an additional 15% glycerol and flash-frozen in liquid nitrogen. Diffraction data were collected at the Stanford Synchrotron Radiation Lightsource beamline 9-2 and processed with the *autoxds* script, which uses XDS (19) for data indexing and reduction, AIMLESS (20) for scaling and TRUNCATE (21) for conversion to structure factors (Supplementary Table S1). We determined the structure by molecular replacement in PHASER (22), using separate models of the protein's wHTH, WYL and WCX domains generated by Alphafold 2 (entry A0A6M5ZVM6 at the AlphaFold Protein Structure Database: https://alphafold.ebi.ac.uk/) (23). One copy of the wHTH domain, two copies of the WYL domain and two copies of the WCX domain were successfully placed and used to calculate initial electron density maps. The second wHTH domain was manually placed using symmetry and difference electron density. Both linker domains were manually built into difference density, using the AlphaFold 2 model as a reference. Two copies of a 3-base ssDNA were placed into difference density by overlaying the WYL domain with the ssDNA-bound WYL domain of *Caulobacter crescentus* DriD [PDB ID 8TP8; (15)] and each DNA base was mutated to thymine. The model was manually rebuilt in COOT (24), then refined in phenix.refine (25) using positional, B-factor and TLS refinement (separate TLS groups for each chain's wHTH + linker, WYL and WCX domains) with reference-model restraints supplied by the AlphaFold 2-predicted structure (Supplementary Table S1).

Fluorescence polarization assays

For dsDNA binding fluorescence polarization (FP) assays, a 30 bp dsDNA representing the preferred binding site for *E. coli* upec-117 CapW (top strand sequence: 5′-TAAAAACTGATCGTATAATAATGCTACGCT-3′) (8) was produced by annealing complementary oligos, one of which was 5′-6-FAM labeled. FP reactions (30 µl) contained 25 mM Tris–HCl, pH 8.5, 5 mM $MgCl_2$, 1 mM DTT, 5% glycerol, 50 mM monopotassium glutamate, 0.01% nonidet p40 substitute, 50 nM DNA and the indicated amounts of protein. For competition assays, an unlabeled 7-nucleotide poly-T ssDNA was added at the indicated concentrations. After a 10 min incubation at room temperature, FP was read using a Tecan Spark plate reader, and binding data were analyzed with Graphpad Prism v.10.1.1 using a single-site binding model. For ssDNA binding FP assays, a 5′-6-FAM labeled 7-nucleotide poly-T ssDNA was used.

Prophage induction assays

To measure phage escape after prophage induction, *E. coli* strain JP858 (17) was transformed with a pLOCO2 plasmid encoding the entire CBASS system from *E. coli* strain upec-117 [bases 69 873–76 528 of NCBI accession # CP054230.1, encoding proteins QKY44554.1 (Cap17; MTA/SAH nucleoside phosphorylase), QKY44555.1 (CapW), QKY44556.1 (CdnC), QKY44557.1 (Cap7; HORMA), QKY445558.1 (Cap6; TRIP13), QKY44559.1 (Cap18; 3′-5′ exonuclease) and QKY44560.1 (Cap19; 3-transmembrane protein)] (gift from A. Whiteley) or the indicated mutants. As a negative control, we used the same pLOCO2 backbone encoding LacI and sfGFP. A transformant was picked from a fresh LB agar plate and grown at 30°C in LB broth plus ampicillin and kanamycin overnight. One hundred microliter of overnight culture was added to 5 ml fresh LB broth plus ampicillin and kanamycin and grown at 30°C to an OD_{600} of 0.45–0.55. We heat shocked the cultures at 42°C for 15 min then incubated the cultures for 1.5 h at 37°C. After incubation, phages were harvested by passing 1 ml of culture over a 0.22 µM filter. In parallel, top agar plates were prepared by growing *E. coli* strain JP313 (16) overnight at 37°C in LB broth, inoculating 50 µl of overnight culture into 5 ml fresh LB and growing at 37°C to an OD_{600} of 0.15–0.25. 500 µl of this culture was mixed with 4.5 ml of 0.35% LB top agar, then poured onto LB plates. Plates were spotted with 3 µl of harvested bacteriophage solution diluted in phage buffer plus 1 mM $CaCl_2$ at six 10-fold dilutions. Plates were incubated at 37°C for 18 h, imaged and scored for plaque formation.

Results

CapW responds to DNA damage

We previously used a reporter plasmid with *E. coli* upec-117 CapW (*Ec* CapW) controlling the expression of GFP to show that CapW is a transcriptional repressor, and that repression is relieved upon bacteriophage infection (Figure 1A) (8). To determine whether CapW responds to DNA damage, we treated cells carrying this reporter plasmid with mitomycin C or zeocin, both of which cause DNA double-strand breaks (26,27). We observed strong GFP expression within 30–60 min of treatment, indicating that CapW responds directly to DNA damage (Figure 1A). To further investigate CapW's response to DNA damage, we tested CapW-mediated GFP expression in *E. coli* strains lacking the key DNA repair proteins RecB and RecA (28). In homologous recombination, the RecBCD complex first resects DNA double-strand breaks to

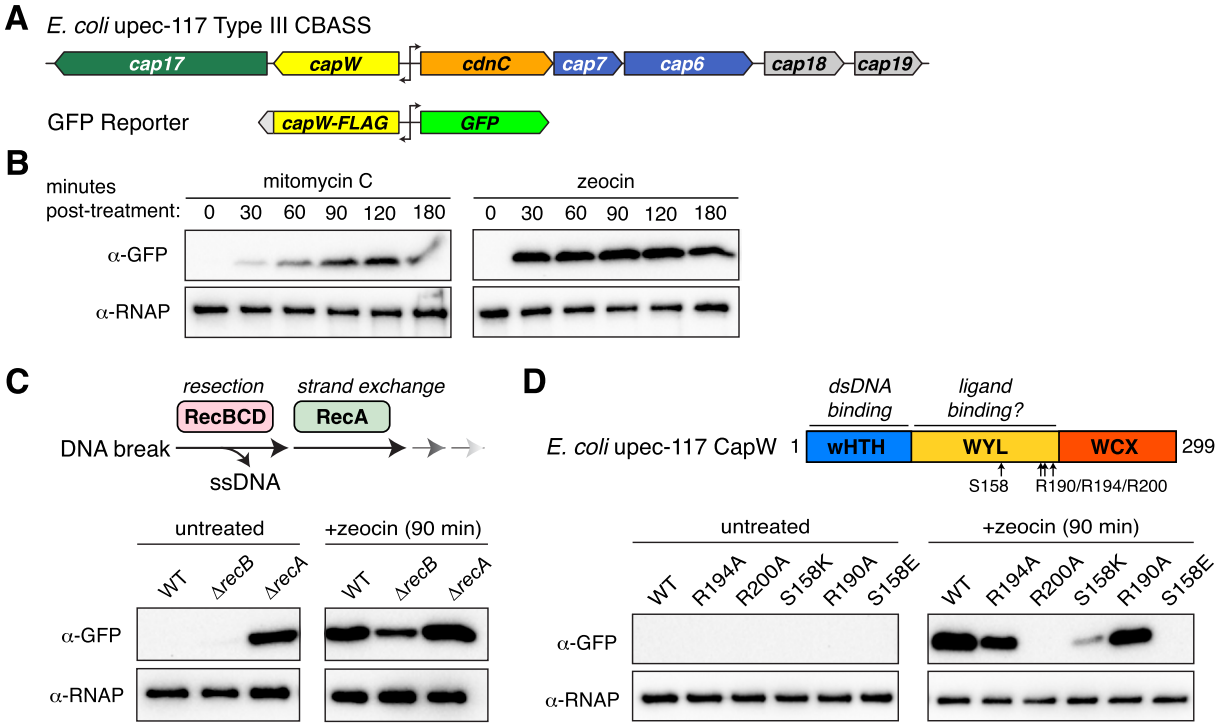


Figure 1. CapW activates transcription in response to DNA damage. (**A**) *Top:* Schematic of the *E. coli* upec-117 CBASS operon. The *capW* gene is shown in yellow, *cdcC* (CD-NTase) in orange, *cap17* (TPALS-PNP effector) in green and *cap7*/*cap8* (HORMA and Trip13 regulators) in blue. Putative regulators *cap18* and *cap19* are shown in gray. The bidirectional promoter is indicated by two arrows. *Bottom:* Schematic of the capW GFP reporter construct. The intergenic region containing the bidirectional promoter is the same as in the native *E. coli* upec-117 CBASS operon. (**B**) Western blot showing activation of GFP reporter expression upon treatment of log-phase *E. coli* cultures with DNA damaging agents mitomycin C (1 μg/ml, left) or zeocin (100 μg/ml, right). α-RNAP: anti-RNA polymerase loading control. (**C**) *Top:* Schematic of homologous recombination in *E. coli*. RecBCD resects DNA double-strand breaks to enable strand exchange by RecA. In Δ*recA* cells, the ssDNA products of resection are expected to accumulate, whereas ssDNA production is limited in Δ*recB* cells. *Bottom:* Western blot showing GFP reporter expression in *E. coli* strains lacking functional RecBCD (Δ*recB*) or RecA (Δ*recA*), in the absence of DNA damage (left) or 90 min after the addition of zeocin (right). (**D**) *Top:* Domain schematic of *E. coli* upec-117 CapW, with positions of WYL domain point mutations highlighted. *Bottom:* Western blot showing GFP reporter expression driven by CapW WYL domain mutants, in the absence of DNA damage (left) or 90 min after the addition of zeocin (right).

expose single-stranded ends, which are then bound by RecA to mediate DNA strand exchange (29). In an *E. coli* strain lacking *recB*, we observed reduced expression of the GFP reporter compared to a WT control when cells were treated with zeocin (Figure 1C). In contrast, a strain lacking *recA* showed high expression of GFP even in the absence of DNA damage and showed a further increase in GFP expression upon zeocin treatment (Figure 1C). These data suggest that reaction products of RecBCD, which include fragments of ssDNA, activate CapW. In the absence of RecA, RecBCD may show higher overall activity due to a block in DNA strand exchange and resulting persistence of DNA breaks. The low level of GFP expression observed upon zeocin treatment in *recB*-deleted cells can likely be attributed to the resection activity of other helicases and exonucleases in the cell.

We previously showed that mutating conserved residues on the surface of the CapW WYL domain eliminates its ability to respond to bacteriophage infection. In particular, mutation of *Ec* CapW S158 and R200 resulted in a complete loss of detectable GFP expression upon infection with bacteriophage λ (8). We tested the same mutants for their ability to respond to DNA damage. We found that *Ec* CapW S158E and R200A mutants were unable to drive GFP expression after treatment with zeocin (Figure 1D). Overall, these data suggest that a byproduct of DNA repair, potentially ssDNA generated by resection of DNA double-strand breaks, binds to the CapW WYL domain to mediate increased expression of its associated operon.

CapW binds ssDNA to control dsDNA binding

To directly test for ssDNA binding by CapW, we used a FP assay with purified *Ec* CapW and fluorescently labeled 7mer DNA oligonucleotides. We found that poly-thymidine, poly-cytosine and poly-guanosine bound CapW with K_d's ranging from 7 to 13 μM, while poly-adenosine did not detectably bind CapW (Figure 2A, Supplementary Figure S1A). We also tested a 7mer poly-uracil single-stranded RNA, and did not detect binding to CapW (Supplementary Figure S1B). The two CapW point mutants that failed to respond to DNA damage in our reporter assays—S158E and R200A—did not detectably bind to ssDNA, suggesting that ssDNA binds to the CapW WYL domain (Figure 2A).

We next tested whether ssDNA inhibits dsDNA binding by CapW. We found that WT *Ec* CapW binds its palindromic dsDNA binding site with ∼3 μM affinity in the absence of ssDNA, and that this binding is progressively disrupted by increasing concentrations of 7mer poly-T ssDNA (Supplementary Figure S1C). dsDNA binding was nearly undetectable in the presence of 0.5 mM ssDNA (Figure 2B). The CapW WYL domain mutants S158E and R200A both bound dsDNA equivalently to WT CapW, but their binding to ds-

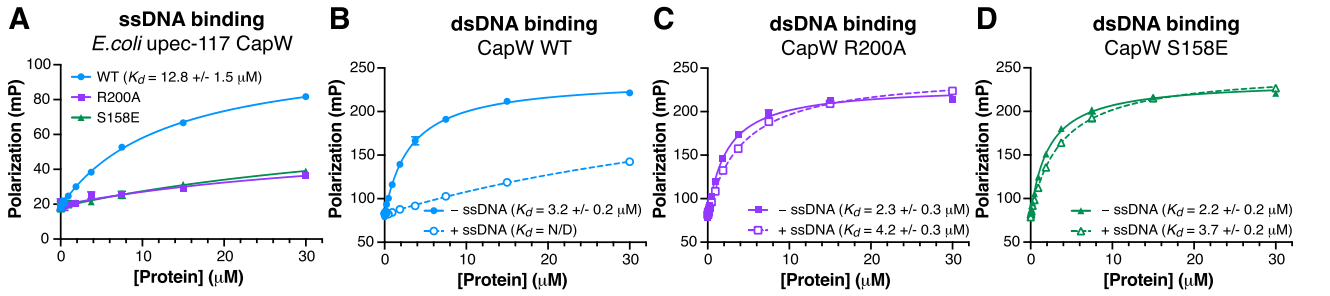


Figure 2. CapW–ssDNA binding inhibits dsDNA binding. (**A**) FP assay showing the binding of a 7mer poly-T ssDNA to *E. coli* upec-117 CapW (WT, circles; R200A, squares; and S158E, triangles). Data represent the mean ± standard deviation of triplicate samples, fit with a one-site binding model. K_d values could not be determined for R200A and S158E mutants. See Supplementary Figure S1A for CapW binding different 7mer ssDNAs, and Supplementary Figure S1B for ssDNA versus ssRNA binding. (**B**) FP assay showing the binding of WT *E. coli* upec-117 CapW to its preferred palindromic binding site (30mer dsDNA), in the absence of ssDNA (solid circles and solid line) or in the presence of 0.5 mM 7mer poly-T ssDNA (open circles and dashed line). See Supplementary Figure S1C for dsDNA binding in different concentrations of ssDNA. (**C**) FP assay showing the binding of CapW R200A to dsDNA, in the absence of ssDNA (solid squares and solid line) or in the presence of 0.5 mM 7mer poly-T ssDNA (open squares and dashed line). (**D**) FP assay showing the binding of CapW S158E to dsDNA, in the absence of ssDNA (solid triangles and solid line) or in the presence of 0.5 mM 7mer poly-T ssDNA (open triangles and dashed line).

DNA was only marginally affected by the presence of 0.5 mM ssDNA (Figure 2C and D). Together with our prior data showing that dsDNA binding is mediated by the CapW wHTH domain (8), these data support the idea that the CapW WYL domain binds ssDNA, and that ssDNA binding allosterically inhibits dsDNA binding by the protein's wHTH domain.

Structure of a CapW–ssDNA complex

To determine the structural basis for CapW ssDNA binding, we performed crystallization trials with three CapW proteins (from *E. coli* upec-117, *S. maltophilia* and *R. leguminosarum*) in the presence of 7mer poly-T ssDNA. We crystallized and determined a 2.77 Å resolution structure of (*Rl*) CapW using molecular replacement (Supplementary Table S1). *Rl* CapW forms a homodimer with an overall structure similar to prior structures of CapW and BrxR (8–10), with an N-terminal wHTH domain (residues 1–70), a linker domain (residues 71–136), a WYL domain (residues 137–193) and a WCX domain (residues 193–297). The linker domain contains two α-helices that pack between the wHTH and WYL domains, and the dimer-related linker domains cross one another to generate a domain-swapped dimer with each protomer's wHTH domain packed against the opposite protomer's WYL domain (Figure 3A).

Based on *Fo–Fc* difference electron density, we identified three ordered DNA bases bound to the WYL domain of each protomer (Figure 3A, Supplementary Figure S2A). ssDNA interacts with the CapW WYL domain in an essentially identical manner as that previously observed with the WYL domain transcriptional activator DriD (14,15), despite that protein's WCX domain adopting a different orientation relative to its WYL domain than in CapW (Supplementary Figure S2B and C). In CapW, ssDNA binds the top surface of the WYL domain in a deep, positively charged crevice created by the WCX domains of both protomers (Figure 3B and C). The 3′ end of the ssDNA is docked against the WCX domain, while the 5′ end extends outward away from the WYL domain. We observe additional density past the third ordered DNA base, suggesting that the additional bases of the 7mer DNA used for co-crystallization are present but disordered in our structure. Several highly conserved residues line the ssDNA-binding pocket, including S144 (equivalent to *Ec* CapW S158), R164 (equivalent to *Ec* CapW R180), R168 (equivalent to *Ec* CapW R184) and R184 (equivalent to *Ec* CapW R200). These residues are conserved across CapW and BrxR homologs (Figure 3D), suggesting that all WYL domain transcriptional repressors bind ligands in a similar manner.

Our biochemical data suggest that ssDNA binding to CapW allosterically inhibits dsDNA binding. To determine the structural basis for this effect, we compared the structure of ssDNA-bound *Rl* CapW to two related structures: a DNA-free structure of *S. maltophilia* (*Sm*) CapW (PDB ID 7TB6; 35.7% sequence identity to *Rl* CapW) (8) and a dsDNA-bound structure of *Acinetobacter* BrxR (PDB ID 7T8K; 29.4% sequence identity to *Rl* CapW) (9). In the BrxR–dsDNA structure, the dimer-related wHTH domains of BrxR bind the major groove of a palindromic DNA sequence ∼20 bp in length, and induce a gentle bend in the bound DNA (Supplementary Figure S2D). The DNA-free structure of *Sm* CapW shows a near-identical conformation of wHTH domains, with a Cα r.m.s.d. of 1.7 Å over both domains (136 Cα atoms aligned) (Supplementary Figure S2E and F). Thus, the DNA-free structure of *Sm* CapW likely represents the dsDNA-binding conformation of the protein.

To understand global conformational changes induced by ssDNA binding to CapW, we next overlaid the WYL domains of ssDNA-bound *Rl* CapW with those of DNA-free *S. maltophilia* CapW (Figure 4A and B). While the individual domains of *Rl* CapW and *Sm* CapW overlay closely [Cα r.m.s.d. of 0.8 Å for the wHTH domain (53 Cα atoms aligned); 2.1 Å for the WYL domain (106 Cα atoms aligned); and 1.5 Å for the WCX domain (78 Cα atoms aligned)], this analysis revealed significant conformational differences between the two proteins that are likely caused by ssDNA binding to the WYL domain. The WYL domain is named after a three-residue tryptophan-tyrosine-leucine (WYL) motif, which is positioned on a β-strand adjacent to the ssDNA binding site (11,12). In CapW/BrxR-like transcriptional repressors, the equivalent residues are embedded in a highly-conserved RWHVR motif, with WHV equivalent to the canonical WYL motif and the surrounding arginine residues (*Rl* CapW R164 and R168) on the top face of the WYL domain and contacting ssDNA. The conserved tryptophan residue of the RWHVR motif is positioned on the bottom face of the WYL domain, and packs against the first α-helix of the linker domain (LH1; Figure

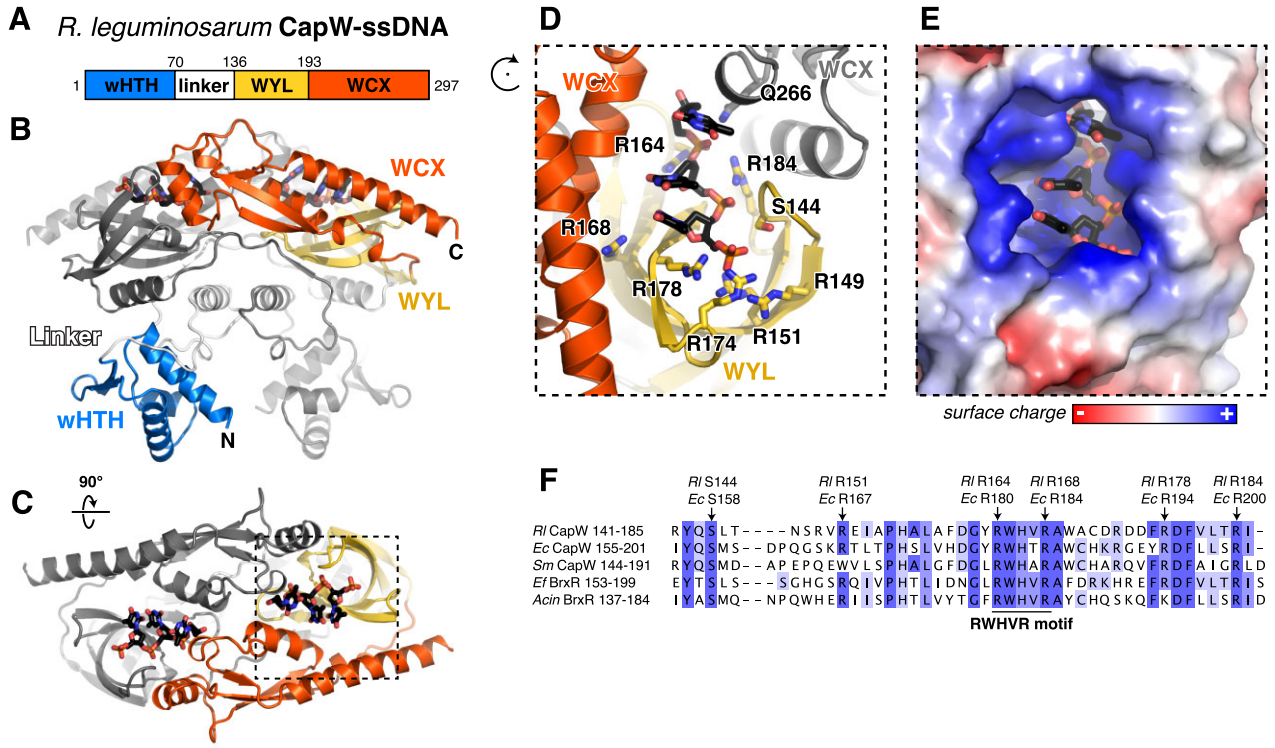


Figure 3. Structure of a CapW–ssDNA complex. (**A**) Domain schematic of *Rl* CapW. (**B**) Side view of the ssDNA-bound *Rl* CapW dimer, with one protomer colored as in panel (A) with individual domains and N- and C-termini labeled; and the second protomer colored gray. (**C**) Top view of the ssDNA-bound *Rl* CapW dimer. Bound ssDNA is colored black and shown as sticks. Dotted box indicates the area shown in panels (D) and (E). (**D**) Closeup view of ssDNA binding to *Rl* CapW. Conserved amino acids that contact DNA are shown as sticks. See Supplementary Figure S2A for *Fo–Fc* electron density at 3.0 σ from an omit map calculated without modeled DNA. (**E**) View as in panel (D), showing the electrostatic surface of CapW (red indicates negative charge and blue indicates positive charge). (**F**) Sequence alignment of the WYL domains of *Rl* CapW (IMG # 2631366338), *Ec* CapW (NCBI #WP_001534693.1), *Sm* CapW (IMG #2657474953), *E. fergusonii* BrxR (UniProt ID B7L3Y3) and *Acinetobacter* sp. NEB 394 BrxR (UniProt ID A0A7H8SL41). Conserved residues that contact ssDNA are marked, and the conserved RWHVR motif is indicated.

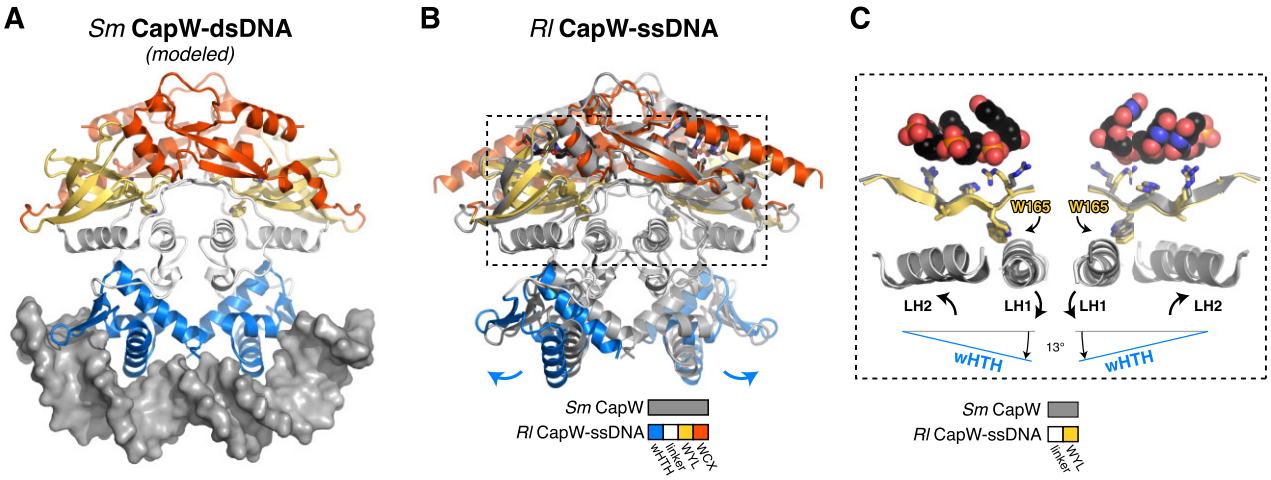


Figure 4. ssDNA binding induces conformational changes in CapW. (**A**) Structure of the *Sm* CapW dimer, with domains colored as in Figure 3. Bound DNA is modeled from a structure of *Acinetobacter* BrxR bound to DNA (see Supplementary Figure S2D–F) (9). (**B**) Overlay of *Rl* CapW bound to ssDNA (domains colored as in Figure 3) and *Sm* CapW (gray). (**C**) Closeup view of the *Rl* CapW RWHVR motif (yellow) contacting ssDNA (black), and conformational changes in the WYL (yellow), linker (white) and wHTH domains (blue) that result from ssDNA binding. Equivalent structural elements of *Sm* CapW are shown in gray.

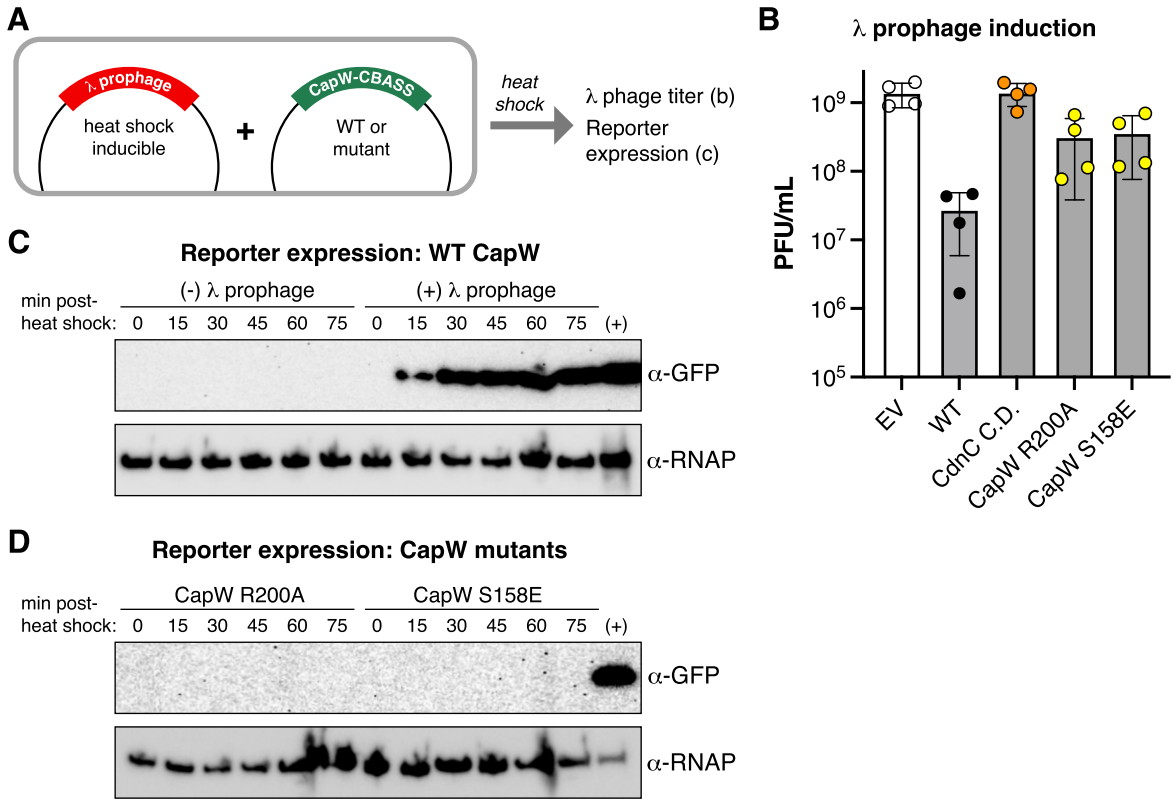


Figure 5. CapW aids CBASS-mediated protection against lysogenic phage induction. (**A**) Schematic of experiment, showing *E. coli* cells carrying a heat shock-inducible λ prophage and a plasmid carrying either the full *E. coli* upec-117 CBASS system (for panel B) or a GFP reporter plasmid (for panel C). Cells were heat-shocked to induce the lytic cycle, then the resulting phages were titered (panel B) or GFP reporter expression was measured by western blot (panel C). (**B**) λ phage titer produced by lytic induction in *E. coli* cells carrying an empty vector (EV), WT *E. coli* upec-117 CBASS, or CBASS with a catalytically dead CdnC (CdnC C.D.: D73N/D75N) or with CapW R200A or S158E mutations. Significance values of paired *t*-tests: EV versus WT ($P = 0.0153$); EV versus CdnC C.D. (0.9734); EV versus CapW R200A (0.0045); and EV versus CapW S158E (0.0042). See Supplementary Figure S3A for plaque data. (**C**) Western blot showing activation of GFP reporter expression (α-GFP) controlled by WT CapW after heat shock, in cells either lacking a λ prophage (left) or carrying a λ prophage (right). The lane marked (+) is a positive control for cells expressing GFP. α-RNAP: anti-RNA polymerase loading control. (**D**) Western blot showing activation of GFP reporter expression (α-GFP) controlled by CapW mutants (R200A or S158E) after heat shock, in cells carrying a λ prophage. The lane marked (+) is a positive control for cells expressing GFP. α-RNAP: anti-RNA polymerase loading control.

4B and C). When comparing the structures of DNA-free *Sm* CapW and ssDNA-bound *Rl* CapW, these tryptophan residues shift subtly downward upon ssDNA binding, causing a rotation of the LH1 helices and a coupled movement of the second α-helix of the linker domain (LH2; Figure 4B and C). The movements of LH1 and LH2, in turn, cause each wHTH domain to rotate ∼13° outward, resulting in the two wHTH domains shifting ∼13 Å away from one another in the *Rl* CapW–ssDNA structure relative to DNA-free *Sm* CapW (Figure 4B and C). The two wHTH domains in ssDNA-bound *Rl* CapW are therefore too far apart to bind the major groove of their palindromic dsDNA binding site. Thus, our structure shows that subtle motions in the WYL and WCX domains of CapW induced by ssDNA binding are amplified through the linker domain α-helices to rotate the wHTH domains into a conformation unable to bind dsDNA.

CapW facilitates defense against prophage induction

CapW/BrxR-family transcriptional regulators are associated with a variety of anti-phage immune systems, including CBASS, BREX, CRISPR/Cas and restriction/modification (10). In the *E. coli* upec-117 CBASS system, mutation of the CapW WYL domain to prevent transcriptional activation does not compromise the system's ability to protect against a lytic bacteriophage λ infection (8). We theorized that one role of CapW-mediated transcriptional activation could be to protect against induction of prophages integrated into the host genome, which are often induced to enter the lytic cycle when the host's DNA is damaged (30). To test this idea, we transformed an *E. coli* strain carrying a heat shock-inducible λ prophage (17) with plasmids carrying the full *E. coli* upec-117 CBASS system, or mutant versions encoding the CapW WYL domain mutants S158E or R200A (Figure 5A). We grew liquid cultures, induced prophage induction by heat shock and measured the amount of λ phage produced by each strain. We found that the WT CBASS system strongly protected against prophage induction, with cells carrying CBASS producing ∼100-fold fewer λ virions than cells without CBASS (Figure 5B, Supplementary Figure S3). This protection was completely eliminated by mutation of the system's CD-NTase, and was strongly compromised by both CapW WYL domain mutations (Figure 5B). Thus, CapW contributes to protection against λ prophage induction.

We next used our reporter plasmid to directly measure whether CapW drives increased expression of its associated operon in response to prophage induction. We found that

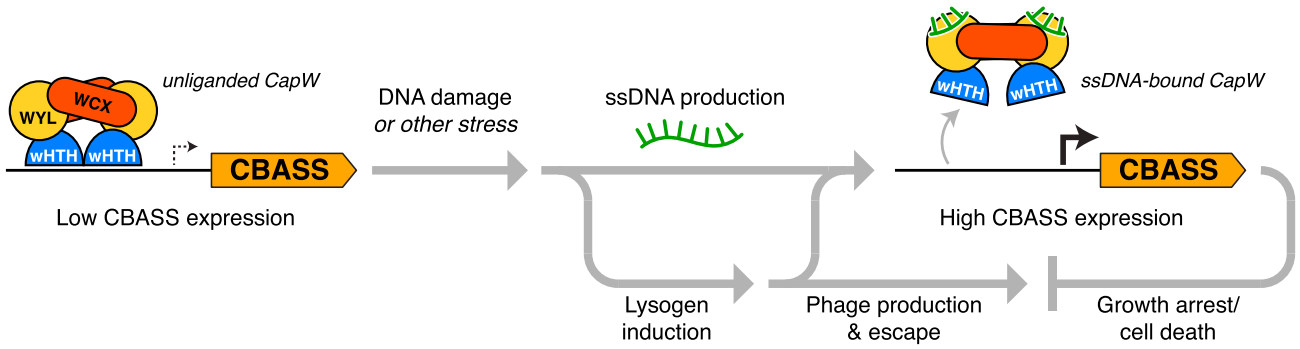


Figure 6. Model for CapW-mediated CBASS expression. In unperturbed cells, CapW binds the CBASS promoter to repress transcription. Upon DNA damage or another stress signal, lysogenic phages are induced (lower path), which contributes to the production of ssDNA fragments. ssDNA binds to the CapW WYL domains to induce a conformational change that leads to dissociation from the CBASS promoter and increased CBASS expression. In *E. coli* upec-117 CBASS, the Cap17 effector depletes ATP to arrest growth and potentially cause cell death (36), preventing phages from escaping the cell.

WT CapW mediates strong induction of GFP expression within 15–30 min of heat shock in the strain carrying the λ prophage, but no expression upon heat shock in cells lacking the prophage (Figure 5C and D). The CapW WYL domain mutants S158E and R200A were unable to drive GFP expression in response to prophage induction (Figure 5D). Combined with our data on λ virion production in the presence of CBASS systems with WT versus mutant CapW, these data suggest that *E. coli* upec-117 CapW mediates protection against prophage induction by boosting the expression of its associated CBASS system genes.

Discussion

Here we show that CapW, a wHTH–WYL–WCX transcriptional repressor, senses DNA damage by binding ssDNA. In unperturbed cells, CapW binds the promoter region of its associated operon to repress transcription (Figure 6) (8). Upon DNA damage, resection of DNA double-strand breaks results in the production of ssDNA that CapW binds through its WYL domain. As a result of ssDNA binding, CapW undergoes a conformational change that disrupts dsDNA binding by its wHTH domain, resulting in increased expression of its associated immune system. Finally, we find that the CapW-associated CBASS immune system of *E. coli* upec-117 protects against λ prophage induction, and that CapW-mediated induction of CBASS expression contributes to this protection.

Within the broad family of wHTH–WYL–WCX transcriptional regulators, two distinct groups mediate either transcription activation or repression. In the transcriptional activator family, ssDNA binding by the WYL domain promotes dsDNA binding by the wHTH domains, enabling promoter binding and recruitment of RNA polymerase (13–15). In contrast, CapW and other repressors including BrxR bind their associated promoters in the absence of ligand to repress transcription (8–10). Here, we find that ssDNA binding to the CapW WYL domains triggers a conformational change that is transmitted through the linker domains to cause rotation of the wHTH domains and loss of dsDNA binding. Based on the high sequence and structural conservation within the WYL-domain repressor family, including the conservation of the RWHVR motif, we hypothesize that all members of this family bind ssDNA similarly to CapW.

In addition to the ssDNA-bound *Rl* CapW discussed here, we previously determined two DNA-free structures of CapW (8). As noted above, the *Sm* CapW structure likely represents the dsDNA-binding conformation of the protein, with its wHTH domains perfectly positioned to bind a palindromic dsDNA (8). The second structure, of *P. aeruginosa* CapW, showed a dramatic ∼70° rotation of the wHTH domains compared to *Sm* CapW (8). Based on these data, we previously hypothesized that the *P. aeruginosa* CapW structure represented the ligand bound, non-dsDNA binding conformation of the protein (8). The structure of the ssDNA-bound *Rl* CapW suggests that ligand binding instead induces a relatively subtle rotation of the wHTH domains that nonetheless disrupts dsDNA binding. Given these data, we propose that our prior *P. aeruginosa* CapW structure likely represents a crystallographic artifact caused by the low-pH conditions used for crystallization of this protein (8).

Our prior work on the CapW-associated *E. coli* upec-117 CBASS system showed that this system robustly protects host cells against a lytic λ phage infection (8). Curiously, this protection does not depend on CapW-mediated induction of high-level expression, since the CapW S158E and R200A mutants do not compromise CBASS-mediated protection (8). This finding suggested that the low-level expression of CBASS proteins present in unperturbed cells is sufficient to mediate protection against lytic λ phage. Consistent with this idea, our GFP reporter assay showed relatively slow and weak CapW-mediated gene expression in response to lytic λ phage (at a multiplicity of infection of 10), with peak expression 90–120 min after lytic infection, likely too slow to mount a strong defensive response (8). In contrast, we now find that CapW-mediated CBASS expression is important to protect host cells against induction of λ prophage, and our GFP reporter assay shows strong expression by 30 min after heat shock (Figure 5). While the reasons for these differences are unclear, it is possible that induction of a λ prophage results in faster and more extensive DNA damage than lytic infection, resulting in stronger CapW-mediated CBASS expression. Overall, our data show that cellular stress—likely in the form of DNA damage—can serve as a signal to boost immune system expression when needed.

While our data show that CapW can contribute to defense against prophage induction, CapW/BrxR is often associated with systems that are unlikely to protect against prophage induction, such as restriction-modification systems

and the anti-plasmid Wadjet system (10). This correlation suggests that stress-induced immune system expression is likely useful in a variety of contexts. We suggest that tight transcriptional control over immune system expression can help bacteria balance the protective benefits of immune systems with the potential burden of carrying multiple such systems, especially those that could cause cell death if aberrantly activated.

Despite the burgeoning knowledge of bacterial immune systems and their molecular mechanisms, our understanding of how these systems are regulated is still in its infancy. Bacterial immune systems that immediately counteract phage infection or plasmid transformation, such as the Shedu nuclease family (31,32) or the Wadjet anti-plasmid system (33–35), should be constitutively expressed to be maximally responsive against external threats. At the same time, constitutive expression of abortive infection systems like CBASS is potentially dangerous for the cell, given the risk that these systems might activate and cause cell death outside the context of an infection. We find that many bacterial immune systems are associated with DNA damage-responsive transcription factors, including CapH/CapP and CapW, that likely enable cells to express these immune systems only in specific circumstances. We anticipate that CapW/BrxR is only one of an array of regulators that control immune system expression in response to DNA damage and other stress signals, enabling bacteria to activate specific immune systems when circumstances warrant. The benefits of this tight control must in some fashion outweigh the obvious drawbacks, including the time delay arising from activating transcription and translation upon stress, and the risk that phage-mediated hijacking of transcription and/or translation may reduce expression of immune system proteins. Future work will be needed to measure the relative short- and long-term benefits and drawbacks to individual bacteria and bacterial communities of constitutive versus stress-inducible immune system expression.

Data availability

Final refined coordinates and structure factors for the *Rl* CapW–ssDNA structure have been deposited in the Protein Data Bank (https://rcsb.org) under accession code 9C5G. Raw diffraction images for the *Rl* CapW–ssDNA structure have been deposited in the SBGrid Data Bank (https://data.sbgrid.org) under dataset ID 1112.

Supplementary data

Supplementary Data are available at NAR Online.

Acknowledgements

The authors acknowledge J. Pogliano, M. Raffatellu, J. Meyer, A. Whiteley and members of the Corbett lab for helpful discussions.

Funding

National Institute of General Medical Sciences [P30 GM133894, R35 GM144121 to K.D.C., T32 GM139795 and F31 GM150210 to C.L.B.).

Conflict of interest statement

None declared.

References

1. Georjon,H. and Bernheim,A. (2023) The highly diverse antiphage defence systems of bacteria. *Nat. Rev. Microbiol.*, **21**, 686–700.
2. Doron,S., Melamed,S., Ofir,G., Leavitt,A., Lopatina,A., Keren,M., Amitai,G. and Sorek,R. (2018) Systematic discovery of antiphage defense systems in the microbial pangenome. *Science*, **359**, eaar4120.
3. Tesson,F., Hervé,A., Mordret,E., Touchon,M., d'Humières,C., Cury,J. and Bernheim,A. (2022) Systematic and quantitative view of the antiviral arsenal of prokaryotes. *Nat. Commun.*, **13**, 2561.
4. Payne,L.J., Todeschini,T.C., Wu,Y., Perry,B.J., Ronson,C.W., Fineran,P.C., Nobrega,F.L. and Jackson,S.A. (2021) Identification and classification of antiviral defence systems in bacteria and archaea with PADLOC reveals new system types. *Nucleic Acids Res.*, **49**, 10868–10878.
5. Bernheim,A. and Sorek,R. (2020) The pan-immune system of bacteria: antiviral defence as a community resource. *Nat. Rev. Microbiol.*, **18**, 113–119.
6. Makarova,K.S., Wolf,Y.I., Snir,S. and Koonin,E.V. (2011) Defense islands in bacterial and archaeal genomes and prediction of novel defense systems. *J. Bacteriol.*, **193**, 6039–6056.
7. Lau,R.K., Enustun,E., Gu,Y., Nguyen,J.V. and Corbett,K.D. (2022) A conserved signaling pathway activates bacterial CBASS immune signaling in response to DNA damage. *EMBO J.*, **41**, e111540.
8. Blankenchip,C.L., Nguyen,J.V., Lau,R.K., Ye,Q., Gu,Y. and Corbett,K.D. (2022) Control of bacterial immune signaling by a WYL domain transcription factor. *Nucleic Acids Res.*, **50**, gkac343.
9. Luyten,Y.A., Hausman,D.E., Young,J.C., Doyle,L.A., Higashi,K.M., Ubilla-Rodriguez,N.C., Lambert,A.R., Arroyo,C.S., Forsberg,K.J., Morgan,R.D., *et al.* (2022) Identification and characterization of the WYL BrxR protein and its gene as separable regulatory elements of a BREX phage restriction system. *Nucleic Acids Res.*, **50**, 5171–5190.
10. Picton,D.M., Harling-Lee,J.D., Duffner,S.J., Went,S.C., Morgan,R.D., Hinton,J.C.D. and Blower,T.R. (2022) A widespread family of WYL-domain transcriptional regulators co-localizes with diverse phage defence systems and islands. *Nucleic Acids Res.*, **50**, 5191–5207.
11. Makarova,K.S., Anantharaman,V., Grishin,N.V., Koonin,E.V. and Aravind,L. (2014) CARF and WYL domains: ligand-binding regulators of prokaryotic defense systems. *Front. Genet.*, **5**, 102.
12. Müller,A.U., Leibundgut,M., Ban,N. and Weber-Ban,E. (2019) Structure and functional implications of WYL domain-containing bacterial DNA damage response regulator PafBC. *Nat. Commun.*, **10**, 4653.
13. Müller,A.U., Kummer,E., Schilling,C.M., Ban,N. and Weber-Ban,E. (2021) Transcriptional control of mycobacterial DNA damage response by sigma adaptation. *Sci. Adv.*, **7**, eabl4064.
14. Gozzi,K., Salinas,R., Nguyen,V.D., Laub,M.T. and Schumacher,M.A. (2022) ssDNA is an allosteric regulator of the *C. crescentus* SOS-independent DNA damage response transcription activator, DriD. *Genes Dev.*, **36**, 618–633.
15. Schumacher,M.A., Cannistraci,E., Salinas,R., Lloyd,D., Messner,E. and Gozzi,K. (2024) Structure of the WYL-domain containing transcription activator, DriD, in complex with ssDNA effector and DNA target site. *Nucleic Acids Res.*, **52**, 1435–1449.
16. Economou,A., Pogliano,J.A., Beckwith,J., Oliver,D.B. and 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.
17. Sussman,R. and Jacob,F. (1962) On a thermosensitive repression system in the *Escherichia coli* lambda bacteriophage. *C. R. Hebd. Seances Acad. Sci.*, **254**, 1517–1519.

18. Tropea,J.E., Cherry,S. and Waugh,D.S. (2009) Expression and purification of soluble His(6)-tagged TEV protease. *Methods Mol. Biol.*, **498**, 297–307.
19. Kabsch,W. (2010) XDS. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 125–132.
20. Evans,P.R. and Murshudov,G.N. (2013) How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.*, **69**, 1204–1214.
21. Winn,M.D., Ballard,C.C., Cowtan,K.D., Dodson,E.J., Emsley,P., Evans,P.R., Keegan,R.M., Krissinel,E.B., Leslie,A.G.W., McCoy,A., *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.*, **67**, 235–242.
22. McCoy,A.J., Grosse-Kunstleve,R.W., Adams,P.D., Winn,M.D., Storoni,L.C. and Read,R.J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.*, **40**, 658–674.
23. Jumper,J., Evans,R., Pritzel,A., Green,T., Figurnov,M., Ronneberger,O., Tunyasuvunakool,K., Bates,R., Žídek,A., Potapenko,A., *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**, 583–589.
24. Emsley,P., Lohkamp,B., Scott,W.G. and Cowtan,K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 486–501.
25. Afonine,P.V., Grosse-Kunstleve,R.W., Echols,N., Headd,J.J., Moriarty,N.W., Mustyakimov,M., Terwilliger,T.C., Urzhumtsev,A., Zwart,P.H. and Adams,P.D. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.*, **68**, 352–367.
26. Kidane,D., Sanchez,H., Alonso,J.C. and Graumann,P.L. (2004) Visualization of DNA double-strand break repair in live bacteria reveals dynamic recruitment of Bacillus subtilis RecF, RecO and RecN proteins to distinct sites on the nucleoids. *Mol. Microbiol.*, **52**, 1627–1639.
27. Chankova,S.G., Dimova,E., Dimitrova,M. and Bryant,P.E. (2007) Induction of DNA double-strand breaks by zeocin in *Chlamydomonas reinhardtii* and the role of increased DNA double-strand breaks rejoining in the formation of an adaptive response. *Radiat. Environ. Biophys.*, **46**, 409–416.
28. Baba,T., Ara,T., Hasegawa,M., Takai,Y., Okumura,Y., Baba,M., Datsenko,K.A., Tomita,M., Wanner,B.L. and Mori,H. (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.*, **2**, 2006.0008.
29. Dillingham,M.S. and Kowalczykowski,S.C. (2008) RecBCD enzyme and the repair of double-stranded DNA breaks. *Microbiol. Mol. Biol. Rev.*, **72**, 642–671.
30. Little,J.W. (2005) Lysogeny, prophage induction, and lysogenic conversion. In: Waldor,M.K., Friedman,D.I. and Adhya,S.L. (eds.) *Phages: Their Role in BacterialPathogenesis and Biotechnology*, John Wiley & Sons, Ltd, Hoboken, NJ, pp. 37–54.
31. Gu,Y., Li,H., Deep,A., Enustun,E., Zhang,D. and Corbett,K.D. (2023) Bacterial Shedu immune nucleases share a common enzymatic core regulated by diverse sensor domains. bioRxiv doi: https://doi.org/10.1101/2023.08.10.552793, 10 August 2023, preprint: not peer reviewed.
32. Loeff,L., Walter,A., Rosalen,G.T. and Jinek,M. (2023) DNA end sensing and cleavage by the Shedu anti-phage defense system. bioRxiv doi: https://doi.org/10.1101/2023.08.10.552762, 12 August 2023, preprint: not peer reviewed.
33. Deep,A., Gu,Y., Gao,Y.-Q., Ego,K.M., Herzik,M.A., Zhou,H. and Corbett,K.D. (2022) The SMC-family Wadjet complex protects bacteria from plasmid transformation by recognition and cleavage of closed-circular DNA. *Mol. Cell*, **82**, 4145–4159.
34. Liu,H.W., Roisné-Hamelin,F., Beckert,B., Li,Y., Myasnikov,A. and Gruber,S. (2022) DNA-measuring Wadjet SMC ATPases restrict smaller circular plasmids by DNA cleavage. *Mol. Cell*, **82**, 4727–4740.
35. Roisné-Hamelin,F., Liu,H.W., Taschner,M., Li,Y. and Gruber,S. (2024) Structural basis for plasmid restriction by SMC JET nuclease. *Mol. Cell*, **84**, 883–896.
36. Rousset,F., Yirmiya,E., Nesher,S., Brandis,A., Mehlman,T., Itkin,M., Malitsky,S., Millman,A., Melamed,S. and Sorek,R. (2023) A conserved family of immune effectors cleaves cellular ATP upon viral infection. *Cell*, **186**, 3619–3631.

Received: June 26, 2024. **Revised:** October 4, 2024. **Editorial Decision:** October 21, 2024. **Accepted:** October 26, 2024