AcrIF11 is a potent CRISPR-specific ADP-ribosyltransferase encoded by phage and plasmid

Daphne F. Chen[1] *, Leah T. Roe[2] *, Yuping Li[3] , Adair L. Borges[4], Jenny Y. Zhang[3], Palak Babbar[5], Sourobh Maji[5], Maisie G.V. Stevens[6], Galen J. Correy[1], Morgan E. Diolaiti[6], Dominique H. Smith[6], Alan Ashworth[6], Robert M. Stroud[5], Mark J.S. Kelly[7], Joseph Bondy-Denomy[3,8,9,#], James S. Fraser[1,8,#]

[1] Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, San Francisco, CA
[2] Department of Chemistry, University of California, Berkeley, CA
[3] Department of Microbiology and Immunology, University of California San Francisco, San Francisco, CA
[4] Arcadia Science, Berkeley, CA, USA
[5] Department of Biochemistry and Biophysics, University of California San Francisco, San Francisco, CA
[6] Helen Diller Family Comprehensive Cancer Center, University of California San Francisco, San Francisco, CA
[7] Department of Pharmaceutical Chemistry, University of California San Francisco, San Francisco, CA
[8] Quantitative Biosciences Institute, University of California San Francisco, San Francisco, CA
[9] Innovative Genomics Institute, Berkeley, CA, USA
[#] Corresponding authors (jfraser@fraserlab.com, joseph.bondy-denomy@ucsf.edu)

Abstract

Phage-encoded anti-CRISPR (Acr) proteins inhibit CRISPR-Cas systems to allow phage replication and lysogeny maintenance. Most of the Acrs characterized to date are stable stoichiometric inhibitors, and while enzymatic Acrs have been characterized biochemically, little is known about their potency, specificity, and reversibility. Here, we examine AcrIF11, a widespread phage and plasmid-encoded ADP-ribosyltransferase (ART) that inhibits the Type I-F CRISPR-Cas system. We present an NMR structure of an AcrIF11 homolog that reveals chemical shift perturbations consistent with NAD (cofactor) binding. In experiments that model both lytic phage replication and MGE/lysogen stability under high targeting pressure, AcrIF11 is a highly potent CRISPR-Cas inhibitor and more robust to Cas protein level fluctuations than stoichiometric inhibitors. Furthermore, we demonstrate that AcrIF11 is remarkably specific, predominantly ADP-ribosylating Csy1 when expressed in *P. aeruginosa*. Given the reversible nature of ADP-ribosylation, we hypothesized that ADPr eraser enzymes (macrodomains) could remove ADPr from Csy1, a potential limitation of PTM-based CRISPR inhibition. We demonstrate that diverse macrodomains can indeed remove the modification from Csy1 in *P. aeruginosa* lysate. Together, these experiments connect the *in vitro* observations of AcrIF11's enzymatic activity to its potent and specific effects *in vivo*, clarifying the advantages and drawbacks of enzymatic Acrs in the evolutionary arms race between phages and bacteria.

Introduction

The evolutionary arms race between bacteriophages and their hosts has resulted in a plethora of bacteria-encoded immune systems and phage-encoded anti-immune factors. Amidst prolific characterization of systems in this arms race, CRISPR-Cas still distinguishes itself from the rest due to its unique adaptive nature. Phages can evade CRISPR-Cas by encoding anti-CRISPR proteins (Acrs), which can fall under one of two generalized mechanisms for inhibition. The first

and most common mechanism is characterized by the stable, stoichiometric association of Acr to the CRISPR-Cas complex. Depending on the Acr, this stable binding can inhibit different functions of the complex: target DNA binding, nuclease cleavage, nuclease recruitment, and so forth[1]. More recently, a second mechanistic class has emerged: enzymatic, substoichiometric inhibition of CRISPR-Cas. Several distinct enzymatic modifications have been observed, including crRNA cleavage[2 3 4], degradation of Type III CRISPR-Cas signaling molecules[5], and post-translational modifications such as acetylation[6 7] and ADP-ribosylation[8].

In contrast to the dozens of stoichiometric Acrs discovered, there are only five biochemically confirmed enzymatic Acrs. They were discovered using methods such as genome fragment screening or guilt-by-association bioinformatics – methods not specifically targeted towards enzymes. For example, AcrIF11 was previously discovered as a neighbor of transcriptional regulator *aca1* and described as a widespread Type I-F anti-CRISPR protein, which enabled the discovery of Cas12 Acrs via guilt-by-association[9]. Although the sequence was too diverged to permit functional assignment upon discovery, a crystal structure revealed similarity to diphtheria toxin, an ADP-ribosyltransferase[8]. Further *in vitro* biochemical assays demonstrated that AcrIF11 modified N250 of Csy1 in the Type I-F Csy complex, which inhibited target DNA binding[8].

The emergence of enzymatic Acrs, with more likely being annotated using structural conservation aided by AlphaFold and related methods, raises a question: under what conditions is it favorable for a mobile genetic element (MGE) to encode a catalytic Acr as opposed to a stable, stoichiometric Acr? We hypothesize that the substoichiometric activity of an enzymatic Acr would make it more potent at inhibiting CRISPR-Cas function compared to a stable stoichiometric Acr. Furthermore, we hypothesize that enzymatic Acrs must be specific to their target to avoid producing off-target effects that might damage the host at inopportune times or interfere with host-encoded enzymes with the same chemical activity. Hyper-potent Acr proteins could also be well suited to stabilizing MGE/prophage-host symbiosis, to prevent CRISPR-Cas self-targeting[10].

Here, we find that ADP-ribosyl transferase AcrIF11 surpasses stable stoichiometric Acrs in protecting phage from Csy complex upregulation and multi-spacer phage targeting. We also show that AcrIF11 effectively rescues lysogens from prophage-induced autoimmunity. Additionally, we establish that AcrIF11 is highly specific to endogenous Csy1/Cas8 in the *P. aeruginosa* intracellular environment using ADP-ribose-specific immunoblotting. The observed potency and specificity likely explain the observed wide distribution across diverse mobile genetic elements. Furthermore, we propose that the reversible nature of some post-translational modifications would allow for removal by the host and demonstrate this possibility using diverse macrodomain proteins. Our characterization of AcrIF11 illustrates the versatility, specificity, and potency of ADP-ribosylation in the evolutionary arms race between phages and bacteria.

Results

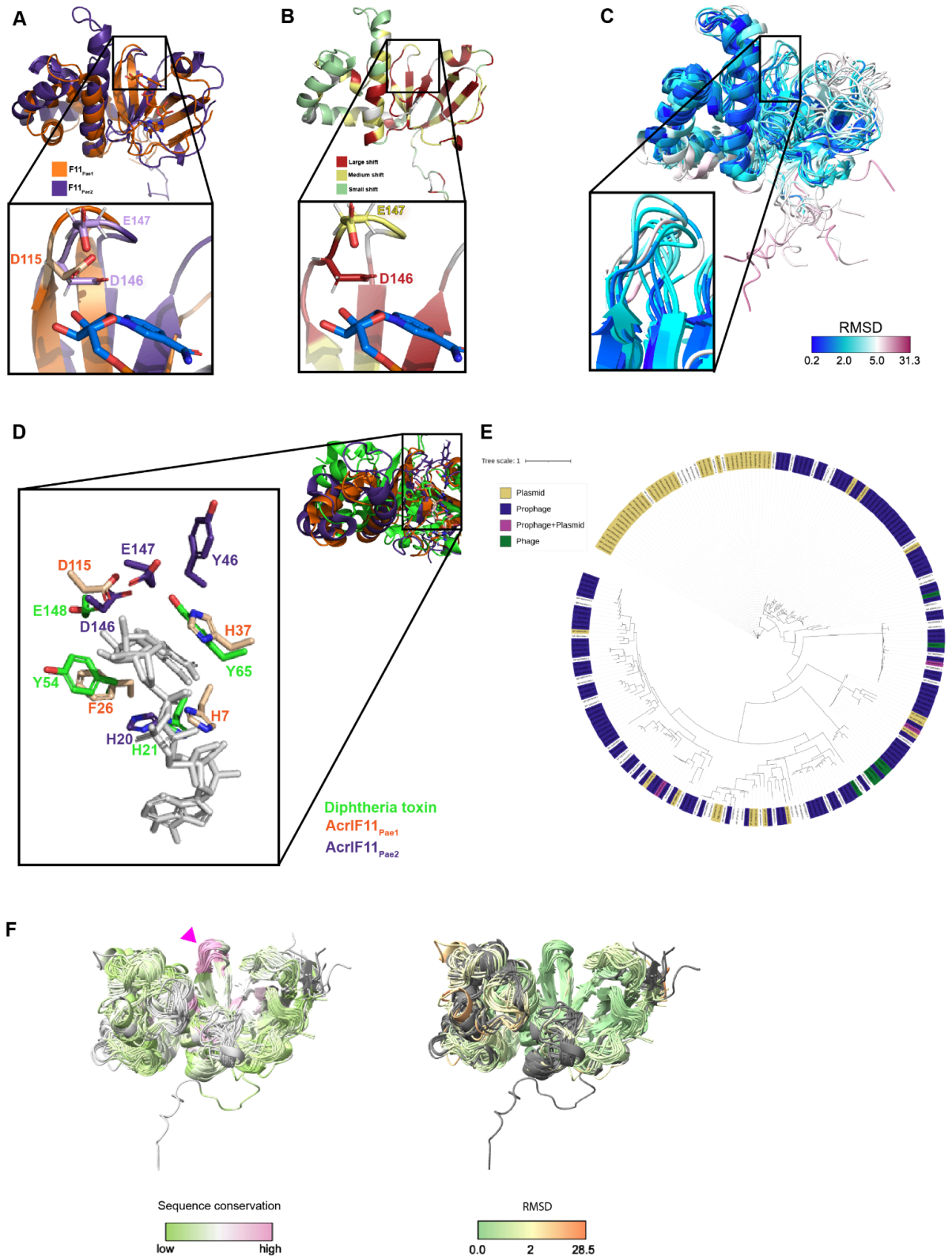
A
F11Pae1
F11Pae2
E147
D115
D146
B
Large shift
Medium shift
Small shift
E147
D146
C
RMSD
0.2
2.0
5.0
31.3
D
E147
Y46
D115
E148
D146
H37
Y65
Y54
F26
H7
H20
H21
Diphtheria toxin
AcrIF11Pae1
AcrIF11Pae2
E
Tree scale: 1
Plasmid
Prophage
Prophage+Plasmid
Phage
F
Sequence conservation
low
high
RMSD
0.0
2
28.5

Figure 1 **NMR structure of AcrIF11$_{Pae2}$**
A) Alignment of AcrIF11$_{Pae1}$ (orange) crystal structure (PDB: 6KYF), and AcrIF11$_{Pae2}$ (purple) NMR structure (PDB: 8DWQ). The inset shows D115 of AcrIF11$_{Pae1}$, a residue previously shown to be important for ART activity[8], as well as D146 and E147, negatively charged residues of AcrIF11$_{Pae2}$ in a similar position to D115.
B) AcrIF11$_{Pae2}$ NMR structure colored according to the magnitude of chemical shift perturbations observed upon NAD (cofactor) titration. The inset shows the same region as the panel A inset.
C) AcrIF11$_{Pae2}$ NMR ensemble colored by RMSD. The inset shows the same region as the inset in panel A and B.
D) Alignment of AcrIF11$_{Pae1}$, AcrIF11$_{Pae2}$, and catalytic domain of monomer of diphtheria toxin (PDB: 1TOX) based on the NAD (light gray) molecules of AcrIF11$_{Pae1}$ and diphtheria toxin.
E) Phylogeny of *acrIF11* homologs found after three iterations of PSI-BLAST. Nodes labeled according to the mobile genetic element they are found on: plasmid (gold), prophage (navy), prophage and plasmid (magenta), or phage (green).
F)Alphafold2 predictions of AcrIF11$_{Pae1}$ homologs, colored by sequence conservation and RMSD. Pink arrow indicates the same region as panel A, B, and C insets.

The NMR structure of AcrIF11$_{Pae2}$ reveals conserved ADP-ribosylation machinery.

Our original motivation for determining the solution structure of AcrIF11 was to leverage the structural information to assign its molecular function. During the time we worked toward that goal, the X-ray structure of a distinct homolog was determined[8], permitting functional annotation as an ADP-ribosyltransferase and validation of that activity *in vitro*. In addition, the advent of Alphafold and related methods allows for high confidence structure prediction of AcrIF11 homologs. These two advances position us to compare the experimental structures and conduct broader structural analyses based on predictions for the family.

We used NMR spectroscopy to determine a structure of a homolog of AcrIF11, hereafter noted as AcrIF11$_{Pae2}$ to avoid confusion with the homolog used in a previous study[8] (PDB: 6KYF), that we will call AcrIF11$_{Pae1.}$ Despite the low sequence similarity (approximately 27% amino acid sequence similarity) between AcrIF11$_{Pae1}$ and AcrIF11$_{Pae2}$, their beta sheet domains align well, and there are conserved negatively charged residues in positions that have been previously noted as crucial for catalysis [11] [12] (Fig 1A). Furthermore, upon titration of cofactor NAD, the largest chemical shift perturbations occurred in the beta sheet region, reinforcing the notion that this region, which is also the most structurally conserved region, is functionally important for NAD binding (Fig 1B). Additionally, the negatively charged AcrIF11$_{Pae2}$ residues D146 and E147, which are conserved across representative AcrIF11 homologs (Supplementary Figure 1), experience large and medium chemical shift perturbations (Fig 1B), highlighting their importance in NAD binding. The NMR structural ensemble of AcrIF11$_{Pae2}$ shows a broader range of RMSD values for the beta sheet region, including the loop where D146 and E147 are located, than the alpha helical region (Fig 1C). This observation of possible loop flexibility is consistent with the need for structural changes to occur upon NAD binding, as observed in previous bacterial ARTs [13]. Further studies are required to characterize dynamic changes in AcrF11$_{Pae2}$ on NAD-binding. The position of AcrIF11$_{Pae2}$ residues in the NAD-binding/catalytic region appear to be conserved

compared to $AcrIF11_{Pae1}$ and diphtheria toxin (Figure 1D), although the exact residue identity may not be conserved.

acrIF11 is widely dispersed among mobile genetic elements

When *acrIF11* was first discovered, it was difficult to annotate due to its low sequence similarity with functionally annotated proteins. However, structural comparisons with the solved crystal structure permitted identification of ART function[8]. With this annotation in hand, we investigated the extent of *acrIF11* spread. We found *acrIF11* homologs via PSI-BLAST and queried these representative homologs' presence on phages and plasmids via a database of bacteriophage and plasmid proteins collated from NCBI. In parallel, we also queried these homologs against a broader provirus and plasmid detector[14] to account for any unannotated prophage and plasmid regions not present in our homemade database. Combining the results of these two methods, we found that the vast majority of representative *acrIF11* homologs in this phylogeny are present on plasmids and prophages/phages (Fig 1E).

Having observed the structural similarity between $AcrIF11_{Pae1}$ and $AcrIF11_{Pae2}$ despite their low sequence similarity, we were curious if other AcrIF11 homologs displayed a similar structure. Using homologs in the phylogeny in Fig 1E, we clustered homologs with sequence similarity higher than 95% to prevent redundancy, and predicted structures of representative sequences from those clusters using Alphafold2[15]. We used a template date set to earlier than the $AcrIF11_{Pae1}$ deposition date to avoid template bias. The predicted $AcrIF11_{Pae1}$ and $AcrIF11_{Pae2}$ structures aligned well with the experimental structures (Supplementary Figure 2), and most of the structures were predicted with high confidence (Supplementary Figure 3). Alignment of all predicted structures to the predicted $AcrIF11_{Pae1}$ structure revealed that the most variable structural feature is the alpha helical region, while the most structurally conserved region is the loop where $AcrIF11_{Pae2}$ D146/E147 and $AcrIF11_{Pae1}$ D115/E116 are located (pink arrow, Fig 1E). These results show that Alphafold2 is capable of detecting conserved catalytic regions across structures, even when the overall sequence similarity is low.

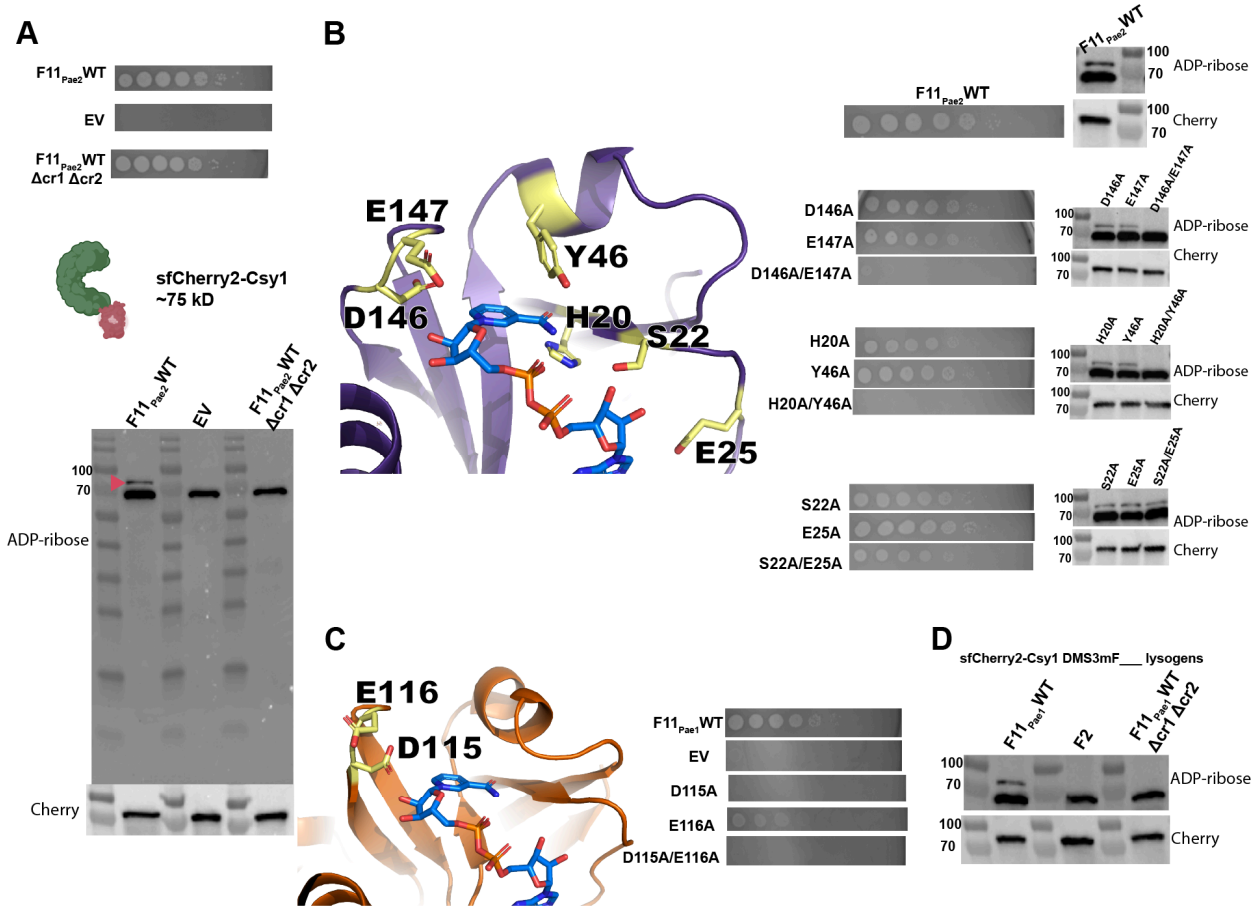

Figure 2 **ADP-ribosylation activity of AcrIF11**
A) Plaque assays of PA14 chromosomally expressing sfCherry2-Csy1 and also containing a plasmid expressing wildtype AcrIF11$_{Pae2}$ (F11$_{Pae2}$WT) or empty vector (EV). Δcr1Δcr2 indicates PA14 with both CRISPR arrays deleted. These strains were also lysed and probed for ADPr and Cherry. Pink arrow indicates sfCherry2-Csy1 protein modified with ADPr.
B) Plaque assays and Western blots as in panel A, but with wildtype and mutant versions of AcrIF11$_{Pae2}$. Mutants are colored yellow on the AcrIF11$_{Pae2}$ NMR structure (PDB: 8DWQ).
C) Plaque assays and Western blots as in panel A, but with wildtype and mutant versions of AcrIF11$_{Pae1}$. Mutants are colored yellow on the AcrIF11$_{Pae1}$ crystal structure (PDB: 6KYF).
D) Western blots of lysate from PA14 sfCherry2-Csy1 DMS3m lysogens encoding the indicated Acr. AcrIF2 is a stable stoichiometric binder with no enzymatic activity.

ADP-ribosylation from AcrIF11 is specific to Csy1 and required for protecting phage against CRISPR-Cas.

Although AcrIF11's ADP-ribosyltransferase activity has been validated *in vitro*[8], the specificity of AcrIF11 in the cellular environment is still unknown. To detect the ADP-ribose modification imparted by AcrIF11 on Csy1, we expressed AcrIF11$_{Pae2}$ on a plasmid in PA14, and then blotted the lysate for ADPr-modified protein with an ADPr-specific antibody (Fig 2A). A previously engineered PA14 strain was used, which has an sfCherry2-fusion to Csy at the endogenous location in the genome[16] to enable us to blot for Cherry for both a loading control and to detect any changes in Csy1/Cas8 length/stability upon modification. Upon expression of WT

AcrIF11$_{Pae2}$, we observed an ADPr band at the expected mass of sfCherry2-Csy1 (Fig 2A, F11$_{Pae2}$WT lane, red arrow). The strong band slightly below 70 kDa is presumably an ADP-ribosylated protein in the PA14 lysate, as shown by its presence in the absence of AcrIF11$_{Pae2}$ (Fig 2A, EV lane). An anti-Cherry immunoblot indeed confirmed equal loading and protein levels of the different samples, confirming that modification does not induce proteolysis. Furthermore, in the absence of the PA14 CRISPR arrays CRISPR1 and CRISPR2, AcrIF11$_{Pae2}$ was unable to ADP-ribosylate Csy1 (Fig 2A, F11$_{Pae2}$WT Δcr1Δcr2 lane), suggesting that the fully assembled Csy complex is necessary for AcrIF11 to recognize and/or bind to Csy1. This observation is in line with previous *in vitro* studies showing that residues on a neighboring protein in the Csy complex, Csy3/Cas7f, are necessary for AcrIF11 activity. Regarding specificity in the intracellular environment, the only detectable covalent modification imparted in an AcrIF11-dependent manner was on Csy1/Cas8, suggesting that this is a very specific anti-CRISPR mechanism.

In order to directly correlate ADP-ribosylation to successful CRISPR-Cas inhibition and phage replication, we tested AcrIF11$_{Pae2}$ mutants' ability to protect CRISPR-targeted DMS3m, and also detected ADP ribosylation in the lysate (Fig 2B). The D146A/E147A pair was chosen for its conservation across homologs as well as the importance of their negative charge in stabilizing the positively charged transition state of NAD[11 12]. The H20A/Y46A pair was chosen because the structure suggests they are important residues for stabilizing the nicotinamide leaving group. Lastly, we hypothesized that the S22A/E25A pair were important for NAD binding. For D146A/E147A and H20A/Y46A, disappearance of ADPr in lysate is matched by the absence of phage replication, showing that without AcrIF11's ADP-ribosylation of Csy1, CRISPR-Cas activity has targeted the phage. We also mutated the equivalent catalytic residues based on the AcrIF11$_{Pae1}$ crystal structure 6KYF (Fig 2C) and observed no phage replication in D115A, consistent with the absence of ADPr from *in vitro* Western blot in a previous study[8]. Lastly, the S22A/E25A mutant in AcrIF11$_{Pae2}$ still inhibited CRISPR-Cas function and appended an ADPr-modification, demonstrating that the residues are redundant for Acr catalytic activity.

Having observed successful ADP-ribosylation of Csy1 using plasmid-expressed AcrIF11$_{Pae2}$, we were curious if native protein expression levels would also result in robust ADP-ribosylation. To investigate this possibility, we made a PA14 sfCherry2-Csy1 DMS3mAcrIF11$_{Pae1}$ lysogen. We first engineered the DMS3m phage to express *acrIF11* from the native anti-CRISPR locus, replacing the endogenous *acrIE3* gene via recombination. This approach was similarly used previously to compare distinct Acr proteins in an otherwise isogenic phage background[17]. Using the PA14 sfCherry2-Csy1 DMS3mAcrIF11$_{Pae1}$ lysogen, we observed an ADP-ribose band at the mass of the sfCherry2-Csy1 fusion (Fig 2D, lane F11$_{Pae1}$WT). As a negative control, we used a previously constructed DMS3m phage with AcrIF2 and made a lysogen; AcrIF2 is a stable stoichiometric Acr that targets the same region of the Csy complex but has no ADP-ribosyltransferase activity. As expected, we saw no ADP-ribose band at the sfCherry2-Csy1 mass (Fig 2D, lane F2). To test if assembly of the Csy complex is important as it was previously (Fig 2A F11$_{Pae2}$WT Δcr1Δcr2 lane), we made a PA14 Δcr1Δcr2 sfCherry2-Csy1 DMS3mAcrIF11$_{Pae1}$ lysogen. In line with our previous observations, we did not see an ADP-ribose band at sfCherry2-Csy1 mass. These results show that even under a native

promoter where protein expression is more than an order of magnitude less than plasmid expression[18], AcrIF11's substoichiometric nature allows it to ADP-ribosylate and inactivate the Csy complex.

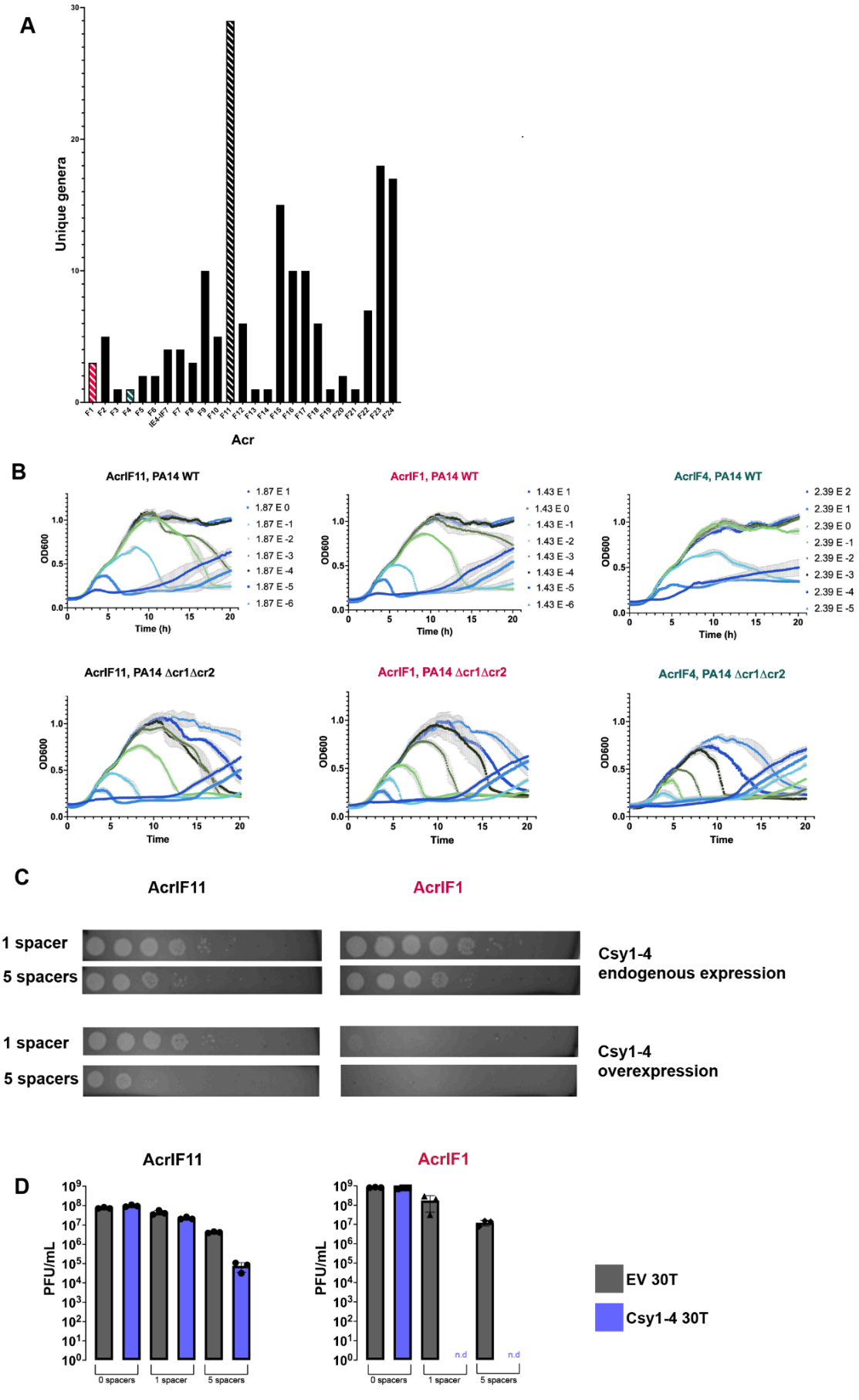
A
Unique genera
Acr
B
AcrIF11, PA14 WT
AcrIF1, PA14 WT
AcrIF4, PA14 WT
AcrIF11, PA14 Δcr1Δcr2
AcrIF1, PA14 Δcr1Δcr2
AcrIF4, PA14 Δcr1Δcr2
OD600
Time (h)
Time
C
AcrIF11
AcrIF1
1 spacer
5 spacers
Csy1-4
endogenous expression
Csy1-4
overexpression
D
PFU/mL
0 spacers
1 spacer
5 spacers
n.d
EV 30T
Csy1-4 30T

Figure 3 ***acrIF11* distribution and protection of lytic phage.**
A) Plot of the number of bacterial genera that each Type I-F Acr can be found in, via BLASTp. Acrs of interest have been marked with striped bars: AcrIF11 (black), AcrIF1 (pink), AcrIF4 (teal).
B) Liquid growth curves of PA14 WT (1 spacer targeting DMS3mvir) infected by DMS3mvir encoding the indicated Acr, across a range of multiplicities of infection (MOI). Plots are the average of three biological replicates. The second row of plots labeled "Δcr1Δcr2" are liquid growth curves of PA14 Δcr1Δcr2 (no spacers targeting DMS3mvir) infected by the same phage as above.
C) Plaque assays of PA14 spotted with 10-fold serial dilutions of DMS3m phage encoding either $AcrIF11_{Pae1}$ (left column) or AcrIF1 (right column). PA14 strains encoded either 1 spacer (wildtype targeting) or 5 spacers (high targeting) in their CRISPR arrays targeting DMS3m; each spacer condition was also combined with endogenous chromosomal Csy1-4 expression or overexpressed Csy1-4 on pHERD30T induced with 0.1% arabinose. Plaque assays are representative of three biological replicates.
D) Quantification of full plate infections using the same strains and phage as panel D.

The substoichiometric activity of AcrIF11 protects phages in situations where stoichiometric Acr activity fails

In addition to our observation of *acrIF11* homologs in diverse MGEs (Fig 1E) , we also observed that *acrIF11* is far more widespread than other type I-F Acrs (Fig 3A). *acrIF11* is encoded by MGEs present in over 25 bacterial genera, a full list of which can be found in Supplementary Table 2. These findings suggest that *acrIF11* is a versatile Acr that confers a fitness advantage in many environments and/or niches. We hypothesized that post-translational modifications are a potent mechanism of rapid and complete CRISPR-Cas inhibition, perhaps explaining the widespread nature of *acrIF11*. Borges et. al. have previously shown that Acr proteins vary in potency during phage infection. For "strong" Acr proteins such as AcrIF1, a low concentration is needed for both successful lytic infection and establishment of lysogeny. In contrast to AcrIF1, "weak" Acrs such as AcrIF4 must be in relatively higher concentration for successful infection[17]. When expressed from a phage infecting PA14, $AcrIF11_{Pae1}$ allowed for robust phage replication across a wide range of input concentrations similar to that of AcrIF1, inducing culture lysis at MOIs as low as 1.87E-5 (Fig 3B). This shows that $AcrIF11_{Pae1}$ exhibits strong inhibition of type I-F CRISPR-Cas during lytic infection, although this experiment does not distinguish between stoichiometric vs substoichiometric activity. Interestingly, during infection with exceedingly low MOIs, the strain lacking CRISPR arrays (Δcr1Δcr2) succumbed to lysis at phage concentrations that failed to lyse when CRISPR was intact. Therefore, it remains clear that successful inhibition of CRISPR-Cas activity, whether by an enzyme or a stoichiometric binder, still requires a critical threshold of anti-CRISPR protein, which is determined by the phage population size[17 19].

Having established that AcrIF11 and AcrIF1 are both strong Acrs, we hypothesized that AcrIF11's substoichiometric enzymatic activity might allow it to overcome "high pressure" scenarios such as during CRISPR-Cas upregulation or when multiple spacers target the same phage. Recent work has revealed mechanisms for CRISPR-Cas systems to increase Cas

protein expression levels after ‘detecting’ an Acr protein [20] [21]. To test this idea, we overexpressed Csy1-4 in PA14 backgrounds with 1 (wildtype) or 5 (high) CRISPR spacers targeting the phage carrying the Acr. $AcrIF11_{Pae1}$ and AcrIF1 allowed phage to replicate at similar levels in both targeting conditions under endogenous levels of Csy1-4 expression. However, with additional Csy1-4 overexpression, phages encoding AcrIF1 were completely unable to replicate in either the 1 or 5 spacer targeting condition, while phages encoding $AcrIF11_{Pae1}$ were still able to replicate in both the 1 and 5 spacer targeting conditions (Fig 3C, 3D). This shows that despite both Acr proteins binding to the Csy complex to exert their inhibition, the transient binding and catalytic activity of AcrIF11 allows phage to overcome increases in target concentration through a substoichiometric mechanism.

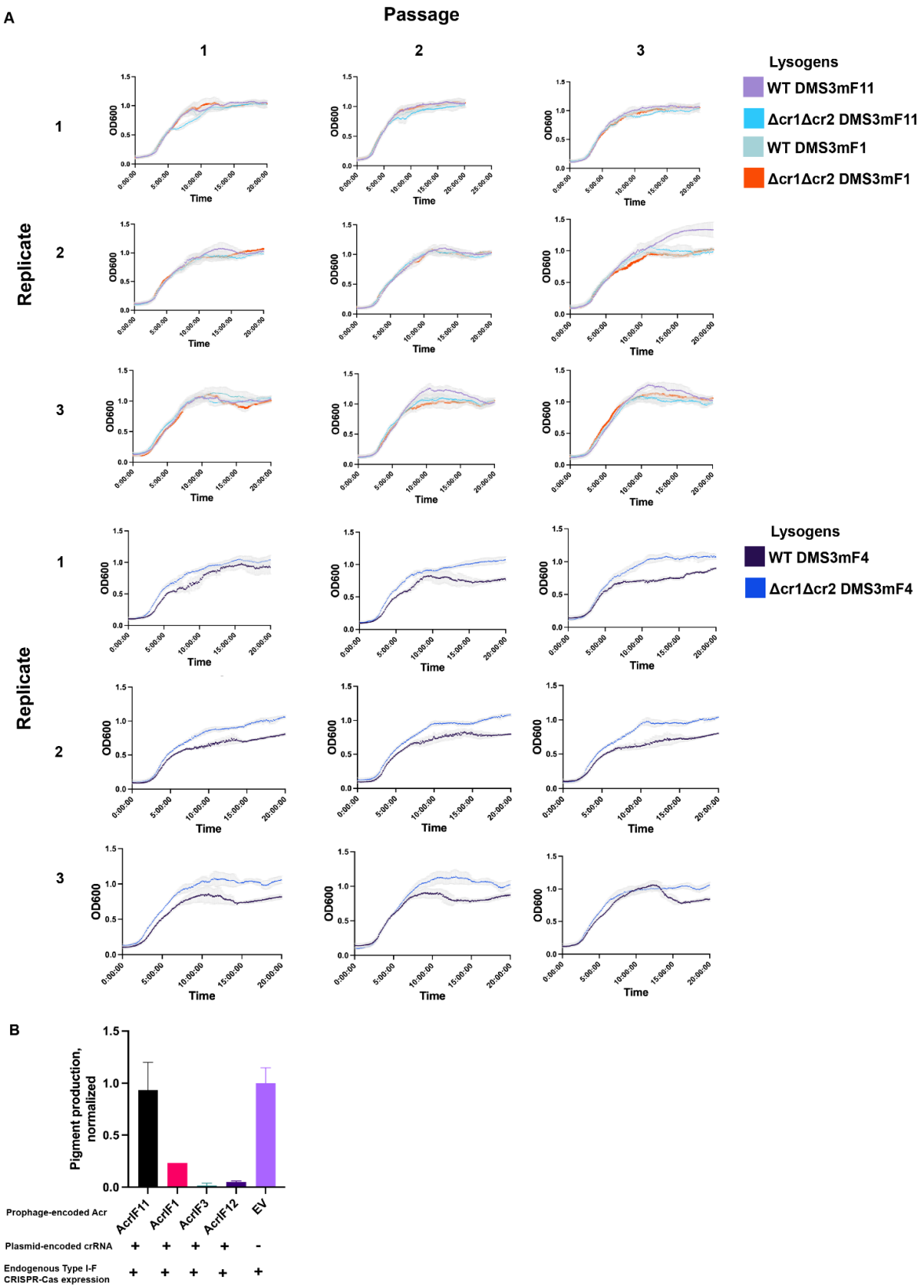

Figure 4 **Prophage-encoded *acrIF11* prevents self-targeting**

A) Liquid growth curves of PA14 lysogens with a DMS3m prophage encoding either AcrIF11$_{Pae1}$, AcrIF1, or AcrIF4, as indicated in the legend. “WT” refers to wild type PA14, and “Δcr1Δcr2” refers to PA14 with both CRISPR arrays deleted.
B) CRISPRi experiment assessing self-targeting via pyocyanin production. Pyocyanin production was measured after growth of PA14 DMS3m lysogens containing a plasmid encoding *phzM*-targeted (pyocyanin synthesis gene-targeted) crRNA, with DMS3m encoding the Acr indicated on the x axis.

AcrIF11 is a strong Acr that prevents self-targeting

In addition to protecting phage during lytic infection, Acr proteins can be vital for stabilization of the lysogenic state. Rollie et. al. observed autoimmunity between PA14 type I-F CRISPR-Cas and DMS3 prophages that manifested as a growth defect, demonstrating the necessity of factors that mediate phage-host symbiosis by alleviating autoimmunity[10]. In addition to spacers present in the CRISPR array *a priori*, priming spacer acquisition can also lead to new spacers that would perfectly target a self prophage. To test AcrIF11’s ability to prevent prophage-targeted autoimmunity, we conducted growth experiments with PA14 DMS3m lysogens encoding Acrs. The DMS3m phage is targeted by PA14’s endogenous CRISPR array, such that integration into the genome during lysogeny results in self-targeting by CRISPR-Cas. DMS3m phage expressing AcrIF4 was able to establish lysogeny, but a growth defect emerged demonstrating self-targeting over time. Lysogens expressing AcrIF1 and AcrIF11$_{Pae1}$, however, displayed growth comparable to a ∆cr1∆cr2 control (Fig 4A), demonstrating full protection. Moreover, after three rounds of passaging these lysogens, we observed no growth defect. These results show that both AcrIF11 and AcrIF1 are strong inhibitors of CRISPR-Cas in the lysogenic cycle, stabilizing a self-targeted MGE.

To quantitatively query the *in vivo* self DNA binding ability of the Csy complex as a proxy for self-targeting risk, we assessed the degree of CRISPR interference (CRISPRi) enacted by a crRNA targeting *phzM*, a gene involved in pyocyanin synthesis. Repression of *phzM* silences pyocyanin production, a colored pigment, while de-repression by lysogen-encoded Acrs restores pyocyanin production (Fig 4B). While AcrIF1 has been previously shown to disable CRISPRi, overexpression of the *phzM*-targeting crRNA overwhelmed this Acr, leading to repression of pigment production. However, AcrIF11 fully disabled the Csy complex under overexpression of the *phzM*-targeting crRNA, allowing full pyocyanin production compared to a strain with an non-targeting crRNA. This shows that, while AcrIF11 and AcrIF1 are capable of stabilizing lysogeny in wildtype targeting conditions (Fig 4A), only AcrIF11 can prevent Csy from stably binding to its own genome under “high” targeting conditions. Thus, AcrIF11 activity is fully capable of inhibiting Csy complex from stably binding to its own genome, likely an important aspect of the biology of this MGE-encoded Acr.

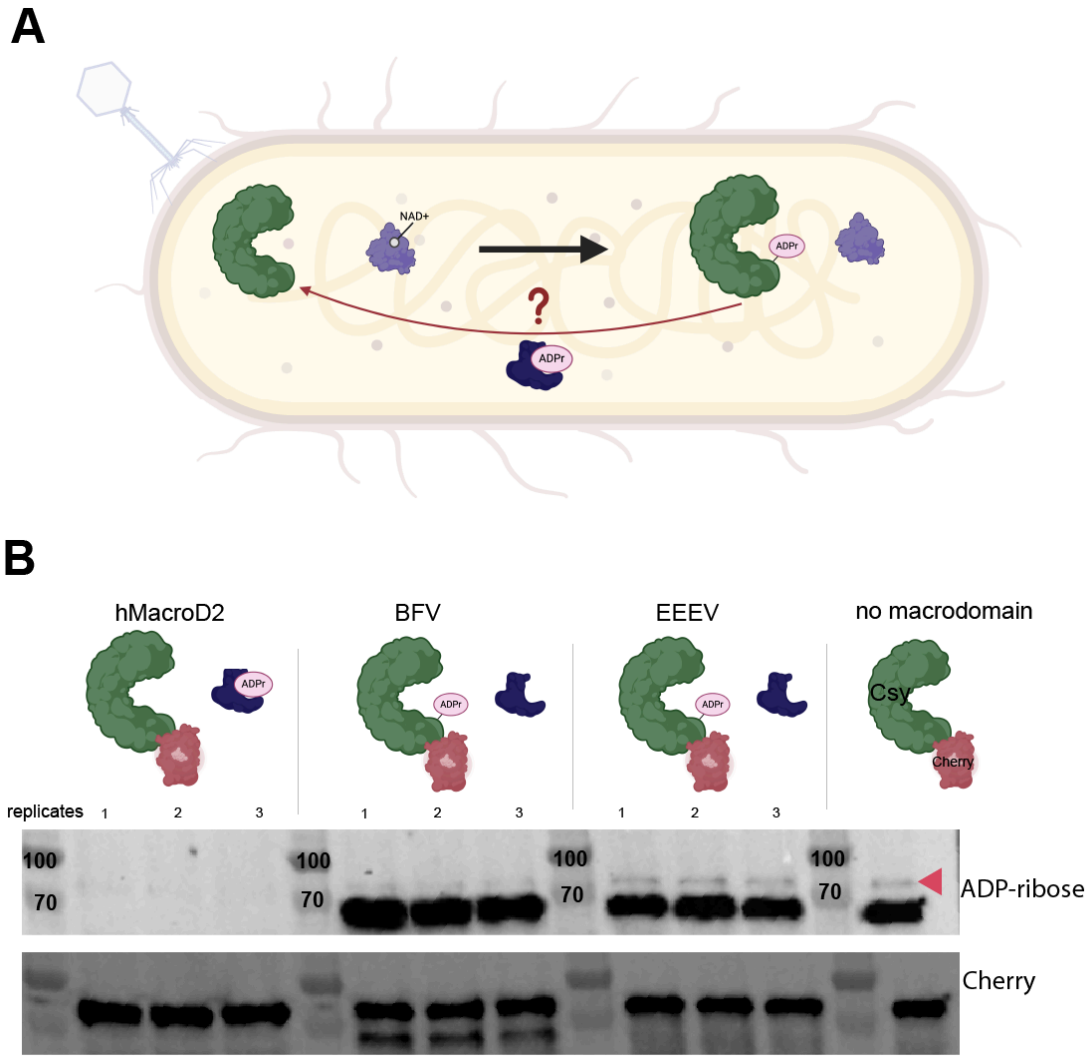

Figure 5 **Macrodomain removal of ADPr**
A) Schematic of Csy complex (green) modified by AcrIF11 (purple) using NAD as a cofactor. We hypothesize that a macrodomain (dark blue) could remove ADPr from the Csy complex.
B) Western blot of PA14 sfCherry2-Csy1 lysate incubated with the indicated macrodomain, and then blotted for ADP-ribose and Cherry as in Figure 2. hMacroD2 final concentration was 9 uM, BFV final concentration was 16 uM, and EEEV final concentration was 43 uM. The sfCherry2-Csy1 ADPr band is indicated by the pink arrow. The same background band around 70 kD from ADPr blots in Figure 2 is also observed in this blot. In the Cherry blot there is also an additional unknown/likely nonspecific band (~70 kD) below the Csy-Cherry band (~75 kD) in the BFV lanes.

ADP-ribosylation of Csy1 is reversible

Lastly, we hypothesized that enzymatic modifications such as ADPr could be removed by host-encoded eraser enzymes such as macrodomains, thereby restoring CRISPR-Cas activity (Fig 5A) as a continuation of the phage-bacteria arms race. To investigate this possibility, we

introduced purified macrodomains into lysate from our sfCherry2-Csy1 strain expressing plasmid-encoded $AcrIF11_{Pae2}$. Due to ongoing interest in one of our labs (J.S.F.) in macrodomain activity in the context of human viral infections, the macrodomains we chose to introduce were from Eastern Equine Encephalitis Virus (EEEV), Barmah Forest Virus (BFV), and humans (hMacroD2). Compared to our control lysate that did not receive any macrodomain (Fig 5B, pink arrow), we observed disappearance of the ADPr signal from sfCherry2-Csy1, and from the background ADP-ribosylated protein (~70 kD), with hMacroD2 added in. Although addition of BFV did not result in the same complete removal of ADPr that we observed with hMacroD2, we still observed a fainter ADPr signal compared to our control, indicating partial ADPr removal. Lastly, the ADPr signal from incubation with EEEV appeared to be the same as our control, indicating no removal of ADPr.

To further test the activity of these macrodomains, we used a phage encoding AcrIF11 to infect PA14 overexpressing the macrodomains in a liquid infection setup. However, we were unable to observe increased bacterial growth that would indicate antagonism of the anti-CRISPR mechanism compared to our negative control (Supplementary Figure 4). This could be due to potential low/no expression of these evolutionarily distant macrodomains in the cellular environment of PA14. Despite this, our observation of purified macrodomain removing ADPr in lysate shows that the ADPr on Csy1 is accessible and able to be removed *in vitro*. Macrodomain specificity, activity levels, and native substrates are an active field of exploration, so it is difficult to reach a conclusion about why hMacroD2 was able to remove ADPr while BFV and EEEV were not. Although these results are not definitive, they open the possibility of another aspect of the phage-bacteria arms race: host-encoded erasers of enzymatic Acr modifications. We hypothesize that the single-residue specificity of AcrIF11's modification, as well as its transient binding to the Csy complex, could be a disadvantage in certain strains compared to a stoichiometric binder such as AcrIF1, which is stably bound to a larger interface of the Csy complex.

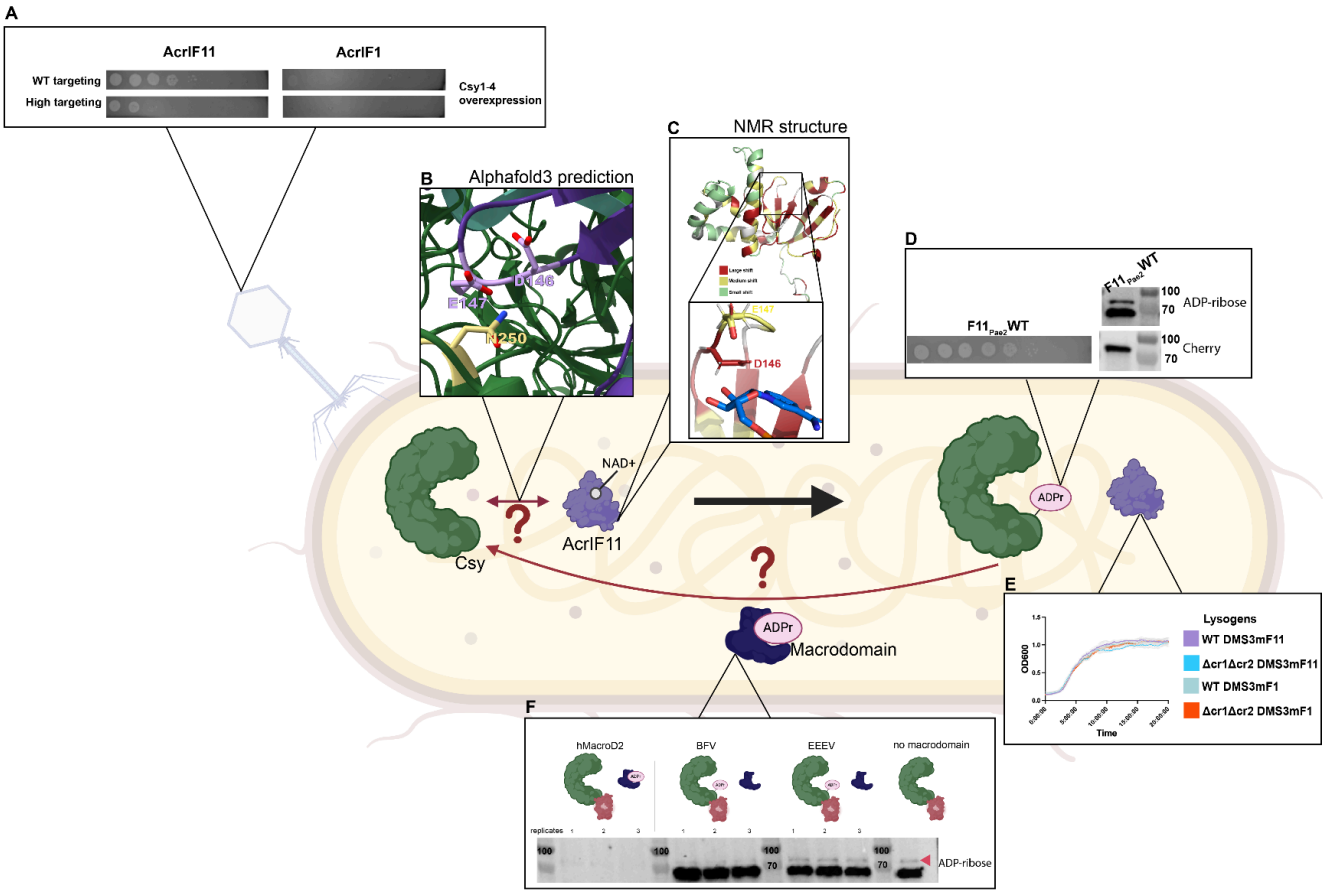

Figure 6 **Summary of AcrIF11 function in PA14**

A) AcrIF11's versatility in the face of multi-spacer CRISPR targeting and Csy overexpression allow it to protect phage under "high pressure" scenarios where stable stoichiometric Acrs such as AcrIF1 fail.

B) Inside the host, AcrIF11 recognizes its target Csy complex through a currently unknown mechanism. Alphafold3 predictions position the catalytic residues above the target Csy1 residue, and suggest interactions in the alpha helical region could contribute to a potential recognition interface (Supplementary Figure 5).

C) AcrIF11$_{Pae2}$, upon binding to cofactor NAD, experiences large chemical shift perturbations in its catalytic beta sheet region.

D) ADP-ribosylated Csy1 can be detected in PA14 lysate and directly correlates to success of phage infection. Our lysate blots also indicate it is highly specific to Csy1 in the intracellular environment of PA14.

E) AcrIF11's specific, potent activity allows it to stabilize lysogens by preventing autoimmunity from self-targeted prophages.

F) It is currently unknown whether there are host-encoded enzymes that can remove ADPr from Csy1 (and therefore re-activate CRISPR-Cas immunity), but we have shown that ADPr on Csy1 is accessible and able to be removed by hMacroD2.

Discussion

In this study we investigated the cellular behavior of the ADP-ribosyltransferase AcrIF11 and found that it confers several benefits to a phage's ability to infect its host. We showed that enzymatic Acrs outperform stoichiometric Acrs when the phage faces a "high pressure" scenario, such as increased targeting from multiple spacers and Csy upregulation. Additionally, we observed AcrIF11's potency in lysogeny by measuring the extent to which it prevents host Csy from self-targeting a prophage in the host genome. We also demonstrated that AcrIF11 can rescue lysogens from growth defects due to autoimmunity resulting from prophage targeting. Furthermore, we established that AcrIF11 is highly specific in its host's intracellular environment – presumably so it does not interfere with activity from host enzymes (Figure 6). AcrIF11's potency and specificity illustrates its versatility as an Acr, which is corroborated by our phylogenetic analysis showing it is the most widespread Type I-F Acr and can be found in many distinct mobile genetic elements.

AcrIF11's enzymatic activity as an ADP-ribosyltransferase was discovered via crystal structure determination and further validated with *in vitro* biochemistry in 2020. Had structure prediction been as widely available as it is now, AcrIF11's enzymatic activity could have been hypothesized after running a predicted structure through DALI and noticing structural similarity to diphtheria toxin. Our Alphafold predictions of AcrIF11 homologs showed that catalytic sites could be predicted, presumably because catalytic sites are structurally conserved across many enzymes in the PDB. We envision structure prediction and comparison to play a larger role in determining enzymatic Acr function in the future; the functional hypotheses that structure prediction provides could narrow the search space for a protein of unknown function, leading to faster experimental validation and iteration.

We found that numerous mobile genetic elements encode AcrIF11 homologs. Our structure of one particular homolog $AcrIF11_{Pae2}$, in conjunction with Alphafold2 predicted structures of the homologs from our phylogeny, demonstrates that dissimilar amino acid sequences from across the sequence space of AcrIF11 homologs can still converge on a common catalytic ART fold. That being said, the details of binding within the ART fold are currently difficult to classify for the AcrIF11 family due to small sample size. Although $AcrIF11_{Pae1}$ appears to be most structurally similar to diphtheria-like ARTs, its key residues appear to be H-F-H-D[8] instead of the canonical H-Y-Y-E[22]. Furthermore, the exact classification of $AcrIF11_{Pae2}$ cannot be determined from our NMR structure alone because this structure was not solved in complex with NAD. The details of NAD-$AcrIF11_{Pae2}$ binding remain unknown, but our mutagenesis of $AcrIF11_{Pae2}$ suggests it is similar to diphtheria-like ARTs, given the importance of H20 and Y46. The chemical shifts of $AcrIF11_{Pae2}$ measured upon NAD titration illustrate the importance of loops analogous to the A- and D- loops near the catalytic site, in line with previous observations of these loops' importance[23].

An unanswered question about AcrIF11, and the ART family more broadly, is how target recognition occurs. Because the catalytic loops are known to be important for both catalysis and target recognition[23], more *in vitro* studies will be needed to determine if catalysis and Csy1

recognition residues can be disentangled for AcrIF11, or if the exact same set of residues plays both roles. Mutagenesis has shown that residues in Csy3/Cas7f are required for $AcrIF11_{Pae1}$ ART activity on N250 of Csy1[8]; whether these Csy3 residues interact with an AcrIF11 loop residue in the beta sheet region, or a residue further away in the alpha helical region, remains to be seen. We observed the alpha helical region of AcrIF11 to be mostly poorly conserved and experience little to no chemical shift perturbations upon NAD titration. However, whether that is because this region solely plays a role in Csy1 recognition and not NAD binding, or whether this region only exists for structural integrity and plays no catalytic or target recognition role at all, cannot be answered without a structure of the Csy complex bound to an AcrIF11 homolog. Remarkably, an Alphafold3 prediction of the Csy complex and $AcrIF11_{Pae2}$ together (Figure 6) places D146 and E147, the catalytic residues predicted to stabilize the transition state of NAD, in close proximity to N250 of Csy1, the experimentally determined[8] target residue. Furthermore, the Alphafold3 prediction suggests that the helical region of $AcrIF11_{Pae2}$ is important for target recognition: residues K58 and K60 of Csy3, residues that have been experimentally confirmed to be important for $AcrIF11_{Pae1}$ activity[8], are in close proximity to negatively charged residues in the alpha helical region of $AcrIF11_{Pae2}$, revealing a possible charged interaction interface (Supplementary Figure 5). Because the helical region is poorly conserved across AcrIF11 homologs, we hypothesize that this region confers specificity of an AcrIF11 homolog to a particular target Csy ortholog, or more broadly facilitates these homologs' ability to distinguish between different Csy orthologs. The structural basis for the high specificity of $AcrIF11_{Pae2}$ that we observed in our lysate blots remains to be explored.

The potency and specificity we observed with AcrIF11 seems at odds with the current scarce prevalence of enzymatic Acrs. Although this problem may be solved in the future by integrating structure prediction for *in silico* screening[24] of common enzymatic folds, it is possible that enzymatic Acrs are rarer because there is a drawback to this evolutionary strategy: the chemical modifications these Acrs impart could be undone by host enzymes. For example, a macrodomain could remove the ADP-ribose placed on Csy1 by AcrIF11. Although PA14 has an endogenous predicted macrodomain, this is likely the wrong host to search in, as we have shown that AcrIF11 is potent in PA14. However, in this study we have established it is at least possible to remove ADP-ribose from Csy1 using hMacroD2. Furthermore, it would make sense for bacteria to have evolved a way to undo enzymatic modifications from phages, especially if the modifications are on an essential residue, such as N250 of Csy1, which functions in PAM recognition. Stable stoichiometric Acrs, on the other hand, would not be so easily removed by the host because they have a larger binding interface with the CRISPR-Cas complex[25]. We envision host-encoded proteins that undo Acr-induced PTMs or PTMs from other phage modification enzymes (e.g. the ART Mod from phage T4)[26] as a new area to explore in host-phage biology.

Methods

AcrIF11$_{Pae2}$ protein purification

pET28b-AcrIF11$_{Pae2}$ vector:
AcrIF11$_{Pae2}$ coding sequence
MGSSHHHHHHSSGLVPRGSHMASMTGGQQMGRMEIFHTSPVEITTINTQGRFGEFLCFAADE
YVMTAGDHVTYRIKVDESDIIMAGSIFYHERAADLSGLVERVMQLTGCDEDTAEELISQRIDVFNL
DDIDASDAAELSWEIQAITAKAAKTLGFRGVSMQDEQGTCYMIDMLGHDAELVRVK*

AcrIF11$_{Pae2}$ sequence after Thrombin cleavage
GSHMASMTGGQQMGRMEIFHTSPVEITTINTQGRFGEFLCFAADEYVMTAGDHVTYRIKVDES
DIIMAGSIFYHERAADLSGLVERVMQLTGCDEDTAEELISQRIDVFNLDDIDASDAAELSWEIQAIT
AKAAKTLGFRGVSMQDEQGTCYMIDMLGHDAELVRVK*
6x His-tag
Thrombin cleavage site

AcrIF11$_{Pae2}$ accession: WP_033936089.1

Transformation protocol: BL21(DE3) E. coli were transformed as follows. Frozen stocks were thawed on ice. Upon thawing, 100 ng of the relevant plasmid was added. After a 30 min incubation on ice, cells were heat-shocked for 30 sec at 42 °C and allowed to recover on ice for 2 min. Following, 350 µL of Super Optimal broth with Catabolite repression (S.O.C.) was immediately added and cells were recovered at 37 °C shaking for 1 hour before plating 50 µL on LB agar plates containing appropriate antibiotics. Plates were incubated overnight at 37 °C.

M9 Minimal Media Expression for labeled protein expression: Starter cultures of 20 mL of Miller's LB Broth supplemented with 50 µg/mL kanamycin were inoculated with a single colony of BL21(DE3) E. coli cells harboring pET28b-AcrIF11$_{Pae2}$ and grown overnight for 16 h at 37 °C with shaking at 220 rpm until the culture was saturated. 8 mL of the starter culture was harvested by centrifugation at 4000 x g at 4 °C for 15 minutes. The supernatant was discarded and the pellet was gently resuspended in 5 mL 1X M9 salts. The 5 mL inoculum was then transferred into 1L of M9 minimal media containing appropriate labeling components (see M9 minimal media assembly) and 50 µg/mL kanamycin. Cells were then grown at 37 °C with shaking at 220 rpm to an OD of 1.0 at which point the culture was induced with 1mL of 1 mM IPTG and grown for 16 hours at 16 °C. The expression culture was harvested by centrifugation at 5000 x g at 4 °C for 20 minutes. Pellets were either stored at -80 °C or immediately used for purification.

Protein purification: The resulting cell pellet was suspended in 20 mL of Buffer A (30 mM Imidazole, 250 mM NaCl, 20 mM HEPES pH 7.0, 5% Glycerol, and 0.5 mM TCEP) containing a tablet of cOmplete, mini EDTA-free ULTRA protease inhibitor cocktail. The cell suspension was disrupted by sonication on ice (1 second pulse at 50% duty cycle followed by 1 second pause

for a total time of 5 min). The cell lysate was cleared by centrifugation at 30,000xg at 4 °C for 30 minutes. Nickel affinity purification was conducted with a Cytiva 5 mL HisTrap HP column. The column was first equilibrated with Buffer A over 5 column volumes (CV), then the lysed sample was applied to the column at a flow rate of 3 mL/min. After protein was bound to the column, a wash was performed with 95% Buffer A and 5% Buffer B (500 mM Imidazole, 250 mM NaCl, 20 mM HEPES pH 7.0, 5% Glycerol, and 0.5 mM TCEP) over 10 CV. The protein was eluted over a linear gradient from 95% Buffer A and 5% Buffer B to 5% Buffer A and 95% Buffer B over 5 CV and fractionated. Fractions containing protein were concentrated using a Amicon® Ultra Centrifugal Filter 10 kDa MWCO to 10 mL. 10 units/mg thrombin was added to the protein, and the mixture was then dialyzed into Buffer C (100 mM NaCl, 20 mM HEPES pH 7.0, and 0.5 mM TCEP) at 4 °C overnight. The following day, a reverse nickel affinity purification was run using the same method as nickel affinity purification, but with replacing Buffer A with Buffer C. Fractions were collected as the sample was applied to the column to capture cleaved $AcrIF11_{Pae2}$ protein. $AcrIF11_{Pae2}$ protein fractions were concentrated to 2 mL and immediately applied to a HiLoad® 16/600 Superdex® 75 pg SEC column using Buffer C over 1.3 CV and fractionated. Fractions containing the protein were concentrated to 3 mL and then desalted using a Bio-Gel p-6DG gel Desalting column in accordance with manufacturer instructions. The purified H3 protein was quantified using absorbance at 280 nm and concentrated to a final protein concentration of 1 mM for NMR analysis.

M9 minimal media assembly: M9 media was prepared accordingly for either ^{15}N-labeled $AcrIF11_{Pae2}$ protein expression or ^{15}N-/^{13}C-labeled H3 protein expression (Supplementary Table 3 and 4). First, 10x M9 salts, ammonium sulfate, and MilliQ H2O were autoclaved together. The remaining components were sterile filtered individually into the autoclaved solution to create 1 L of M9 media for labeled protein expression.

$AcrIF11_{Pae2}$ NMR structure determination

The final protein concentration for the [$U^{13}C$, $U^{15}N$]-labeled protein for the data collection for the structure determination[27] was 1mM $AcrIF11_{Pae2}$ in 50 mM KPi pH 7.4 with 5% (v/v) D2O. NMR spectra were all measured at 298.0 K[28]. Two dimensional (2D) 1H,15N-HSQC (pulse program: fhsqcf3gpph), 2D multiplicity-edited CT 1H,13C-HSQC (pulse program: hsqcctetgpsp), 36ms 3D HCcH-TOCSY (pulse program hcchdigp3d) and 120ms 3D simultaneous 13C-/15N-NOESY-HSQC (pulse program: noesyhsqcgpsismsp3d) spectra were measured using a Bruker Avance NEO 800 MHz spectrometer with a 5mm TCI H&F-C/N-D-05 Z-gradient CP2.1 CryoProbe. 3D CACBcoNH (pulse program: cbcaconhgp3d), 3D HNCACB (pulse program: hncacbgp3d), 3D HcccoNH (pulse program: hccconhgp3d2), 3D hCcccoNH (pulse program: hccconhgp3d3) experiments were collected on a Bruker Avance NEO 600 MHz spectrometer with an 5mm TCI H&F-C/N-D-05 Z-gradient CP2.1 CryoProbe. Spectra were processed in TopSpin version 4.0.6 and referenced indirectly to an external DSS standard[29].

Resonances were assigned using the program CCPN Analysis version 2.4.2[30]. Backbone resonances were assigned using the 2D 15N-HSQC, 3D CACBcoNH and 3D HNCACB spectra. Sidechain resonances were assigned using the 2D constant-time 13C-HSQC and 3D HCcH-TOCSY spectra. Distance restraints were generated using CCPN Analysis. Dihedral

restraints were generated using the program DANGLE[31]. The programs ARIA version 2.3.2[32] and CNS version 1.2.1[33] were used to calculate the NMR structures. To solve the structures, 9 iterations of simulated annealing were performed using CNS. For the first 8 rounds of simulated annealing, the n_structures parameter was set to 50 and the n_best_structures parameter was set to 15. For the 8th round, the n_structures parameter was set to 200 and the n_best_structures parameter was set to 25. Finally, a refinement in water was performed on the lowest energy structures from the 8th iteration. Otherwise, the default values were used for the remaining ARIA parameters. Initial structure calculations were conducted without hydrogen bond restraints. Hydrogen bond donors were then identified, and the corresponding hydrogen bond restraints were included in later calculations. Hydrogen bond restraints included in the structure calculations were based on measurements of amide chemical exchange with solvent detected by 2D ^{15}N-CLEANEX-HSQC experiments[34]. Structures were validated using the Protein Structure Validation Software (PSVS) suite 1.5. The chemical shifts, restraints, and structural coordinates have been deposited with the BMRB (31035) and PDB (8DWQ).

AcrIF11 phylogenetic analysis

To find AcrIF11 homologs, AcrIF11$_{Pae1}$ (WP_038819808.1) was blasted against the nr database for 3 iterations of psiBLASTp. Hits with greater than 70% coverage and expected value less than 0.0005 were used for generating the PSSM in each iteration, and the maximum hitlist size was set to 500. Sequences were aligned via MAFFT[35], and the multiple sequence alignment was trimmed manually to prevent large gaps in the alignment. The phylogenetic tree was created from the final multiple sequence alignment using FastTree[36], and visualization was done using iTOL[37].

For determining the distribution of Type I-F Acrs, the Type I-F Acr sequences were first acquired from the Acr database[38]. The sequences were blasted against the nr database using BLASTp, with an expectation value of 0.0005 and a maximum hitlist size of 500. Unique genera were manually counted from the hitlists.

Alphafold structure prediction

The hits used to make the phylogeny in “AcrIF11 phylogenetic analysis” were clustered using MMseqs2[39], using minimum sequence identity of 95%, cluster mode 2, and coverage mode 1. Representative sequences of each cluster were used for structure prediction. Sequences were run through the SBGrid installation of Alphafold2, with a maximum template date set to 2020-09-20. The highest confidence results (structures named ranked_0.pdb) were used for sequence conservation and RMSD analysis via MatchMaker in ChimeraX[40].

The Alphafold3 web server (https://alphafoldserver.com/) was used to predict AcrIF11 with the Csy complex. Sequences used for Alphafold3 prediction were from Guo et. al[41] PDB 6B45:
AcrIF11$_{Pae2}$:
MEIFHTSPVEITTINTQGRFGEFLCFAADEYVMTAGDHVTYRIKVDESDIIMAGSIFYHE
RAADLSGLVERVMQLTGCDEDTAEELISQRIDVFNLDDIDASDAAELSWEIQAITAKAAK
TLGFRGVSMQDEQGTCYMIDMLGHDAELVRVK
Csy1:

MTSPLPTPTWQELRQFIESFIQERLQGKLDKLQPDEDDKRQTLLATHRREAWLADAARRV
GQLQLVTHTLKPIHPDARGSNLHSLPQAPGQPGLAGSHELGDRLVSDVVGNAAALDVFKF
LSLQYQGKNLLNWLTEDSAEALQALSDNAEQAREWRQAFIGITTVKGAPASHSLAKQLYF
PLPGSGYHLLAPLFPTSLVHHVHALLREARFGDAAKAAREARSRQESWPHGFSEYPNLAI
QKFGGTKPQNISQLNNERRGENWLLPSLPPNWQRQNVNAPMRHSSVFEHDFGRTPEVSRL
TRTLQRFLAKTVHNNLAIRQRRAQLVAQICDEALQYAARLRELEPGWSATPGCQLHDAEQ
LWLDPLRAQTDETFLQRRLRGDWPAEVGNRFANWLNRAVSSDSQILGSPEAAQWSQELSK
ELTMFKEILEDERD

Csy2:

MSVTDPEALLLLPRLSIQNANAISSPLTWGFPSPGAFTGFVHALQRRVGISLDIELDGVG
IVCHRFEAQISQPAGKRTKVFNLTRNPLNRDGSTAAIVEEGRAHLEVSLLLGVHGDGLDD
HPAQEIARQVQEQAGAMRLAGGSILPWCNERFPAPNAELLMLGGSDEQRRKNQRRLTRRL
LPGFALVSREALLQQHLETLRTTLPEATTLDALLDLCRINFEPPATSSEEEASPPDAAWQ
VRDKPGWLVPIPAGYNALSPLYLPGEVRNARDRETPLRFVENLFGLGEWLSPHRVAALSD
LLWYHHAEPDKGLYRWSTPRFVEHAIA

Csy3:

MSKPILSTASVLAFERKLDPSDALMSAGAWAQRDASQEWPAVTVREKSVRGTISNRLKTK
DRDPAKLDASIQSPNLQTVDVANLPSDADTLKVRFTLRVLGGAGTPSACNDAAYRDKLLQ
TVATYVNDQGFAELARRYAHNLANARFLWRNRVGAEAVEVRINHIRQGEVARAWRFDALA
IGLRDFKADAELDALAELIASGLSGSGHVLLEVVAFARIGDGQEVFPSQELILDKGDKKG
QKSKTLYSVRDAAAIHSQKIGNALRTIDTWYPDEDGLGPIAVEPYGSVTSQGKAYRQPKQ
KLDFYTLLDNWVLRDEAPAVEQQHYVIANLIRGGVFGEAEEK

Csy4:

MDHYLDIRLRPDPEFPPAQLMSVLFGKLHQALVAQGGDRIGVSFPDLDESRSRLGERLRI
HASADDLRALLARPWLEGLRDHLQFGEPAVVPHPTPYRQVSRVQAKSNPERLRRRLMRRH
DLSEEEARKRIPDTVARALDLPFVTLRSQSTGQHFRLFIRHGPLQVTAEEGGFTCYGLSK
GGFVPWF

crRNA:

CUAAGAAAUUCACGGCGGGCUUGAUGUCCGCGUCUACCUGGUUCACUGCCGUGUAGGC
AG

Determining the presence of AcrIF11 on mobile genetic elements

The first mode of analysis consisted of blasting AcrIF11 homologs from the phylogenetic analysis above (only the representative sequences of each cluster) against a homemade database of plasmid and phage proteins. Plasmid proteins were downloaded from the NCBI RefSeq Plasmid database (https://ftp.ncbi.nlm.nih.gov/genomes/refseq/plasmid/), and phage proteins were downloaded from the NCBI Virus web database (Find Data > Search by virus > Bacteriophages). The BLAST database combining the plasmid and phage dataset was created using makeblastdb in the BLAST command line application. AcrIF11 homologs were blasted against this database using BLASTp with an expectation value of 0.0005. Results were marked as “phage” if the species indicated in the hitlist (or species associated with the query accession) was a phage, and “plasmid” if the species indicated in the hitlist was a bacterium. Hits were only considered valid if the query and target length matched exactly, and if the sequence identity was 100% with zero gaps.

The second mode of analysis was using geNomad[14]. For every AcrIF11 homolog, the associated genome accessions were identified using eLink from NCBI eUtilities, or manually in cases where eLink failed. geNomad was run on each genome, with score calibration enabled and post-classification filtering set to conservative. The AcrIF11 homologs associated with each genome were then blasted against said genome to identify its genomic location. The genomic location of the AcrIF11 homolog was then compared to the prophage location identified by geNomad, or to the plasmid genes location list identified by geNomad, to ensure that each homolog fell within the bounds of the mobile genetic element. In the case of plasmid hits, which were identified by coding region accession numbers instead of translated protein accession numbers, the hits were also confirmed to be the homolog of interest via NCBI Identical Protein Groups.

Results appearing as “prophage + plasmid” either had both hits in geNomad, or a hit in geNomad (prophage) and a hit in the homemade database (plasmid).

CRISPRi pyocyanin assay

This assay was performed as previously described[42]. Briefly, a plasmid encoding a Type I-F crRNA targeting the *phzM* (pyocyanin synthesis gene) promoter was used to transform the desired lysogen. An empty vector was also transformed into the lysogen as a control. Lysogens were grown as overnight cultures with gentamicin for plasmid maintenance and 0.1% arabinose for induction of crRNA expression. Pyocyanin was extracted with an equal volume of chloroform, mixed with a half volume of 0.2M HCl, and quantified by measuring absorbance at 520 nm. For the plot in Fig 2C, pyocyanin levels were normalized to the empty vector control.

Molecular cloning & PA14 conjugation

Genes of interest were synthesized or PCRed from template DNA, and then inserted into pHERD20T or pHERD30T backbone using the NEB-recommended HiFi protocol. pHERD vectors were digested with NcoI-HF and Sbf-HF (for 20T) or NcoI-HF and HindIII-HF (for 30T) for the HiFi reaction. XL1-B cells were transformed with the HiFi reactions, and transformants were verified via Sanger sequencing. Miniprepped plasmids were verified again via whole-plasmid sequencing from Primordium.

To introduce plasmids into PA14, *E. coli* SM10 cells were used as donors. Plasmids were electroporated into SM10 cells and plated onto LB + antibiotic plates (carbenicillin for 20T, gentamicin for 30T) after recovery. SM10 transformed colonies were cross-streaked with PA14 for conjugation. After incubating conjugation plates overnight at 37°C, PA14 colonies that received the plasmid were selected via VBMM + antibiotic plates. Selected colonies were verified via PCR and Sanger sequencing of PCR products.

Liquid phage infections

PA14 strains were grown in LB or LB + the appropriate antibiotic overnight at 37°C and diluted 1:100 in the same LB formulation. 10-fold serial dilutions of the desired phage were made using SM buffer. In each well of a 96-well plate, 10 µL of each phage dilution was added to 140 µL of 1:100 diluted overnight culture, and grown at 37°C on a plate reader that monitored OD600.

Plaque assays

PA14 strains were grown in LB or LB + the appropriate antibiotic overnight at 37°C. 150 µL of the overnight culture was mixed with 3 mL of LB top agar supplemented with 10 mM Mg. The top agar + overnight culture mixture was plated onto LB + 10 mM Mg + appropriate antibiotic plates. 10-fold serial dilutions of the desired phage were made using SM buffer, and 2.5 µL of each dilution was spotted onto the plate. Plates were incubated at 30°C overnight.

Full plate phage infections

PA14 strains were grown in LB or LB + the appropriate antibiotic overnight at 37°C. 150 µL of overnight culture was added to 10 µL of phage, and mixed via shaking at room temperature for 15 minutes. This mixture was then added to 3 mL of LB top agar supplemented with 10 mM Mg, and spread out over an LB + 10 mM Mg + appropriate antibiotic plate. Plates were incubated at 30°C overnight. Plaques were manually counted. Phage titers were calculated via full plate infections with a PA14 Δcr1Δcr2 lawn.

Lysogen construction

Plaque assays with spot titrations of the desired phage were done following the protocol outlined above. Clearings from plaque assays were streaked out onto LB + 10 mM Mg plates and incubated overnight at 37°C. The resulting colonies were grown up as overnight cultures in LB + 10 mM Mg, and lysogeny was verified via two methods: plating the putative lysogen culture and spotting with the same phage to check for superinfection exclusion, and chloroform extraction of the putative lysogen culture followed by spotting the extract onto a PA14 lawn to check for the presence of virions.

Lysogen growth experiments

Lysogens were streaked out onto LB + 10 mM Mg plates from glycerol stocks. LB + 10 mM Mg cultures were inoculated with a lysogen colony and grown overnight at 37°C. The overnight culture was passaged 1:100 into LB+Mg three times, with the OD600 of each passage monitored via plate reader for 20 hours. Three replicates were performed, with each replicate consisting of three passages.

Western blot of lysates

Three replicates of the $AcrIF11_{Pae2}$ mutant blots were done. For each replicate, the desired plasmid was conjugated into PA14 sfCherry2-Csy1 following the conjugation protocol outlined above. Conjugated colonies were verified via PCR of the multiple cloning site and Sanger sequencing, then grown overnight in LB + Carb250. Overnight cultures were diluted 1:100 in fresh LB + Carb250 + 0.1% arabinose, and grown for 13 hours at 37°C. After 13 hours, cultures were pelleted, washed, resuspended in lysis buffer, incubated in lysis buffer for 15 minutes, and then lysed via sonication with the Bioruptor Pico. The wash buffer used was 50 mM Tris pH 7.4, 150 mM NaCl, 1 mM EDTA, 1 mM $MgCl_2$. The lysis buffer contained the same formulation as the wash buffer, but with the addition of 1 mg/mL lysozyme, 1 protease inhibitor tablet, 0.5 mM TCEP, and 125 U/mL Pierce Universal Nuclease. After lysis, lysates were clarified with a 20

minute spin in a benchtop centrifuge at 4°C and approximately 21,000 x g.Clarified lysates were then flash frozen in liquid nitrogen for future blotting.
For lysogen lysate blots, three replicates were also grown and lysed in the same manner as above, but grown with LB + 10 mM Mg instead of LB + Carb250 + 0.1% arabinose.

Total protein concentration in lysate was quantified via Bradford assay. Lysates were prepared for SDS-PAGE by resuspending in SDS loading buffer and heating at 95°C for 5 minutes. Gel samples were loaded onto an SDS-PAGE gel, with loading volumes adjusted to normalize total protein concentration across all wells. Depending on the particular concentrations from lysis in each replicate, 20-30 µg/well was loaded. Gels were transferred onto a nitrocellulose membrane using the BioRad TurboBlot, and normalization was verified via Ponceau staining. After washing off Ponceau stain, blots were blocked for 1 hour at room temperature in 5% casein, and then incubated in 1:1000 of primary antibody (ADPr: Cell Signaling D9P7Z, Cherry: Cell Signaling E5D8F) overnight at 4°C. The next day, the blots were washed in TBS-T for 10 minutes per round, for 3 rounds. Blots were blocked in 1:10,000 HRP secondary antibody (Cell Signaling 7074) at room temperature for 1 hour, then washed in TBS-T for 10 minutes per round, for 3 rounds. BioRad Clarity Max ECL substrate was added to the blots in accordance with the manufacturer protocol, and blots were imaged using the BioRad Chemidoc. Blots for ADPr and Cherry were run in parallel with the same lysates and same loading volumes.

hMacroD2 macrodomain protein purification

hMacroD2 sequence:
MHHHHHHSSGVDLGTENLYFQSYPSNKKKKVWREEKERLLKMTLEERRKEYLRDYIPLNSILS
WKEEMKGKGQNDEENTQETSQVKKSLTEKVSLYRGDITLLEVDAIVNAANASLLGGGGVDGCI
HRAAGPCLLAECRNLNGCDTGHAKITCGYDLPAKYVIHTVGPIARGHINGSHKEDLANCYKSSL
KLVKENNIRSVAFPCISTGIYGFPNEPAAVIALNTIKEWLAKNHHEVDRIIFCVFLEVDFKIYKKKMN
EFFSVDDNNEEEEDVEMKEDSDENGPEEKQSVEEMEEQSQDADGVNTVTVPGPASEEAVED
CKDEDFAKDENITKGGEVTDHSVRDQDHPDGQENDSTKNEIKIETESQSSYMETEELSSNQED
AVIVEQPEVIPLTEDQEEKEGEKAPGEDTPRMPGKSEGSSDLENTPGPDAGAQDEAKEQRNGT
KGLNDIFEAQKIEWHE*
6xHIS / AVI / TEV
hMacroD2 was expressed and purified as described previously for SARS-CoV-2 Mac1[43].

BFV macrodomain protein purification

BFV sequence:

MSYYHHHHHHLESTSLYKKAGFLEVLFQGPEVNSFSGYLKLAPAYRVKRGDISNAPEDAVVNAA
NQQGVKGAGVCGAIYRKWPDAFGDVATPTGTAVSKSVQDKLVIHAVGPNFSKCSEEEGDRDLA
SAYRAAAEIVMDKKITTVAVPLLSTGIYAGGKNRVEQSLNHLFTAFDNTDADVTIYCMDKTWEKKI
KEAIDHRT

Cloning and expression

The macrodomain of BFV was synthesized and cloned into the pET28 vector, incorporating an N-terminal 6xHis tag and a long linker before the macrodomain, resulting in the expression

construct 6xHis-LINKER-Macrodomain. The plasmid was then transformed into LOBSTR-BL21(DE3) *E. coli* for protein expression. Following a standard expression protocol, cells were grown in LB medium at 37 °C until the OD at 600 nm reached 0.8. Protein expression was then induced by the addition of 500 μM IPTG, and the cells were allowed to grow overnight at 18 °C. After incubation, the cells were harvested by centrifugation. The cell pellets were recovered and stored at -80 °C until purification.

Protein purification

The cell pellet was suspended in lysis buffer (50 mM Tris, pH 7.0, 250 mM NaCl, 10 mM imidazole, 5% glycerol, and 2 mM β-mercaptoethanol) containing one tablet of Complete, Mini EDTA-free ULTRA protease inhibitor cocktail and DNase I (10 μg/mL). The cell suspension was disrupted by passaging three times through a chilled Emulsiflex at 15,000 psi. The cell lysate was clarified by centrifugation at 30,000 × g for 30 minutes at 4 °C.

Nickel affinity purification was conducted using a gravity flow column packed with 5 mL of resin. The column was first equilibrated with water and then with lysis buffer. Following equilibration, the lysate was added to the resin and rocked at 4 °C for 30 minutes. After the protein bound to the column, the column was washed with wash buffer (50 mM Tris, pH 7.0, 250 mM NaCl, 10 mM imidazole, 5% glycerol, and 2 mM β-mercaptoethanol) over 2 × 10 column volumes (CV). The protein was eluted with elution buffer (50 mM Tris, pH 7.0, 250 mM NaCl, 500 mM imidazole, 5% glycerol, and 2 mM β-mercaptoethanol) in 6 × 5 mL fractions (30 mL total). Fractions containing the protein were pooled and dialyzed overnight at 4 °C to remove imidazole using a dialysis buffer (50 mM Tris, 150 mM NaCl, 1 mM DTT, 5% glycerol).

The protein was concentrated to 1 mL and immediately applied to a Superdex® 75 10/300 SEC column using size exclusion buffer (50 mM Tris, pH 7.0, 150 mM NaCl, 5% glycerol). Fractions containing the protein were concentrated to 0.7 mg/mL, flash-frozen, and stored at -80 °C until required for assays.

EEEV macrodomain protein purification

EEEV sequence:
MGHHHHHHHHHHENLYFQSGAPAYRVVRGDITKSNDEVIVNAANNKGQPGGGVCGALYRKW
PGAFDKQPVATGKAHLVKHSPNVIHAVGPNFSRLSENEGDQKLSEVYMDIARIINNERFTKVSIP
LLSTGIYAGGKDRVMQSLNHLFTAMDTTDADITIYCLDKQWESRIKEAI

Cloning and expression
The macrodomain of EEEV was synthesized and cloned into the pET28 vector, incorporating an N-terminal 6X-HIS tag, TEV protease site and linker before macrodomain, generating the expression as 6XHis-TEV-SG-Macrodomain. The plasmid was then transformed into LOBSTR-BL21(DE3) E. coli for protein expression. Following a standard expression protocol, cells were grown in LB media at 37°C until OD 600 nm reached 0.8, then protein expression was induced by the addition of 500 μM IPTG to the media, and cells were allowed to grow overnight at 18°C, after which they were harvested by centrifugation. Cell pellets were recovered and stored at -80°C until purification.

Protein purification

Purifications were carried out under standard NTA purification conditions. Briefly, cells pellets were resuspended in lysis buffer, 50 mM Tris pH 8, 500 mM NaCl, 10mM imidazole, 5% glycerol, 2 mM β-mercaptoethanol. Cells were then lysed by passing through chilled Emulsiflex at ~10,000psi for three cycles. The lysate was clarified by centrifugation at 35,000 g for 1 hour. The clarified lysate was applied to 5 ml of pre-equilibrated Ni-NTA resin and incubated for 1 hour. The resin was then washed with 5 times 10 CV of 50 mM Tris pH 8, 500 mM NaCl, 30mM imidazole, 5% glycerol, 2 mM β-mercaptoethanol. Elution of the protein was done with 5 CV buffer 50 mM Tris pH 8, 500 mM NaCl, 500 mM imidazole, 5% glycerol, 2mM BME. The eluted protein was dialyzed in 50 mM Tris, pH 8.0, 150 mM NaCl, 5% glycerol, 1 mM DTT. The His-tag was cleaved with purified His-TEV protease (1:20 mass ratio with EEEV eluted protein). His-TEV was further removed by reverse NTA purification. Further EEEV macrodomain was purified using SEC S75 10/300 column in 20 mM Tris, pH 7, 150 mM NaCl and fractions with the purest protein were mixed, concentrated, flash frozen and stored at -80°C until required for assays.

Macrodomain reactions

$AcrIF11_{Pae2}$ WT was conjugated into PA14 sfCherry2-Csy1, grown, and lysed following the same protocol as outlined above in “Western blot of lysates”. Before loading onto the SDS-PAGE gel, purified macrodomain was added and incubated with lysates for 1 hour at room temperature. SDS loading buffer was added to the reactions, and heated for 5 minutes at 95°C before running the SDS-PAGE gel. The blotting procedure was the same as outlined above, but after imaging with the ADPr primary antibody, the blot was stripped using Restore Plus Stripping Buffer for 10 minutes at 37°C, blocked using the same protocol as above, and incubated overnight in Cherry primary antibody solution. The next day, the same protocol as above was done for washing, secondary incubation, and imaging.

Phage genetic editing

Acrs of interest were introduced into DMS3m as previously described[17]. Briefly, the *acr* gene was introduced into the pHERD plasmid with homology regions flanking DMS3 gene 30. Cells containing the plasmid were infected with DMS3m, and phages isolated from this infection were then plated onto PA14, which contains a Type I-F CRISPR system to select for recombinant DMS3m. Individual plaques from this selection were purified and the presence of the *acr* was verified via Sanger sequencing.

References

1. Wiegand, T., Karambelkar, S., Bondy-Denomy, J. & Wiedenheft, B. Structures and Strategies of Anti-CRISPR-Mediated Immune Suppression. *Annu. Rev. Microbiol.* **74**, 21–37 (2020).
2. Zhang, H. *et al.* Structural Basis for the Inhibition of CRISPR-Cas12a by Anti-CRISPR Proteins. *Cell Host Microbe* **25**, 815-826.e4 (2019).
3. Mahendra, C. *et al.* Broad-spectrum anti-CRISPR proteins facilitate horizontal gene transfer. *Nat. Microbiol.* **5**, 620–629 (2020).
4. Wang, X. *et al.* Inhibition mechanisms of CRISPR-Cas9 by AcrIIA17 and AcrIIA18. *Nucleic Acids Res.* **50**, 512–521 (2022).
5. Athukoralage, J. S. *et al.* An anti-CRISPR viral ring nuclease subverts type III CRISPR immunity. *Nature* **577**, 572–575 (2020).
6. An anti-CRISPR protein disables type V Cas12a by acetylation | Nature Structural & Molecular Biology. https://www.nature.com/articles/s41594-019-0206-1.
7. Kang, X. *et al.* Reversible regulation of Cas12a activities by AcrVA5-mediated acetylation and CobB-mediated deacetylation. *Cell Discov.* **8**, 1–4 (2022).
8. Niu, Y. *et al.* A Type I-F Anti-CRISPR Protein Inhibits the CRISPR-Cas Surveillance Complex by ADP-Ribosylation. *Mol. Cell* **80**, 512-524.e5 (2020).
9. Discovery of widespread type I and type V CRISPR-Cas inhibitors | Science. https://www.science.org/doi/10.1126/science.aau5174?url_ver=Z39.88-2003&rfr_id=ori:rid:crossref.org&rfr_dat=cr_pub%20%200pubmed.
10. Rollie, C. *et al.* Targeting of temperate phages drives loss of type I CRISPR–Cas systems. *Nature* **578**, 149–153 (2020).
11. Parikh, S. L. & Schramm, V. L. Transition State Structure for ADP-Ribosylation of Eukaryotic Elongation Factor 2 Catalyzed by Diphtheria Toxin. *Biochemistry* **43**, 1204–1212 (2004).
12. Jørgensen, R., Wang, Y., Visschedyk, D. & Merrill, A. R. The nature and character of the transition state for the ADP-ribosyltransferase reaction. *EMBO Rep.* **9**, 802–809 (2008).
13. Ménétrey, J. *et al.* NAD Binding Induces Conformational Changes in Rho ADP-ribosylating Clostridium botulinum C3 Exoenzyme *. *J. Biol. Chem.* **277**, 30950–30957 (2002).
14. Camargo, A. P. *et al.* Identification of mobile genetic elements with geNomad. *Nat. Biotechnol.* 1–10 (2023) doi:10.1038/s41587-023-01953-y.
15. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
16. Govindarajan, S., Borges, A., Karambelkar, S. & Bondy-Denomy, J. Distinct Subcellular Localization of a Type I CRISPR Complex and the Cas3 Nuclease in Bacteria. *J. Bacteriol.* **204**, e00105-22 (2022).
17. Borges, A. L. *et al.* Bacteriophage Cooperation Suppresses CRISPR-Cas3 and Cas9 Immunity. *Cell* **174**, 917-925.e10 (2018).
18. Bondy-Denomy, J., Pawluk, A., Maxwell, K. L. & Davidson, A. R. Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature* **493**, 429–432 (2013).
19. Landsberger, M. *et al.* Anti-CRISPR Phages Cooperate to Overcome CRISPR-Cas Immunity. *Cell* **174**, 908-916.e12 (2018).
20. Osuna, B. A. *et al.* Critical Anti-CRISPR Locus Repression by a Bi-functional Cas9 Inhibitor. *Cell Host Microbe* **28**, 23-30.e5 (2020).
21. Workman, R. E. *et al.* Anti-CRISPR proteins trigger a burst of CRISPR-Cas9 expression that enhances phage defense. *Cell Rep.* **43**, (2024).

22. Pinto, A. F. & Schüler, H. Comparative Structural Analysis of the Putative Mono-ADP-Ribosyltransferases of the ARTD/PARP Family. in *Endogenous ADP-Ribosylation* (ed. Koch-Nolte, F.) 153–166 (Springer International Publishing, Cham, 2015). doi:10.1007/82_2014_417.
23. Cohen, M. S. & Chang, P. Insights into the biogenesis, function, and regulation of ADP-ribosylation. *Nat. Chem. Biol.* **14**, 236–243 (2018).
24. Duan, N., Hand, E., Pheko, M., Sharma, S. & Emiola, A. Structure-guided discovery of anti-CRISPR and anti-phage defense proteins. *Nat. Commun.* **15**, 649 (2024).
25. Chowdhury, S. *et al.* Structure Reveals Mechanisms of Viral Suppressors that Intercept a CRISPR RNA-Guided Surveillance Complex. *Cell* **169**, 47-57.e11 (2017).
26. ModA and ModB, Two ADP-Ribosyltransferases Encoded by Bacteriophage T4: Catalytic Properties and Mutation Analysis | Journal of Bacteriology. https://journals.asm.org/doi/10.1128/jb.186.21.7262-7272.2004.
27. *Protein NMR Spectroscopy*. (2006).
28. Findeisen, M., Brand, T. & Berger, S. A 1H-NMR thermometer suitable for cryoprobes. *Magn. Reson. Chem.* **45**, 175–178 (2007).
29. Wishart, D. S. *et al.* 1H, 13C and 15N chemical shift referencing in biomolecular NMR. *J. Biomol. NMR* **6**, 135–140 (1995).
30. Vranken, W. F. *et al.* The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins Struct. Funct. Bioinforma.* **59**, 687–696 (2005).
31. Cheung, M.-S., Maguire, M. L., Stevens, T. J. & Broadhurst, R. W. DANGLE: A Bayesian inferential method for predicting protein backbone dihedral angles and secondary structure. *J. Magn. Reson.* **202**, 223–233 (2010).
32. Rieping, W. *et al.* ARIA2: Automated NOE assignment and data integration in NMR structure calculation. *Bioinformatics* **23**, 381–382 (2007).
33. Brünger, A. T. *et al.* Crystallography & NMR System: A New Software Suite for Macromolecular Structure Determination. *Acta Crystallogr. D Biol. Crystallogr.* **54**, 905–921 (1998).
34. Hwang, T.-L., van Zijl, P. C. M. & Mori, S. Accurate Quantitation of Water-amide Proton Exchange Rates Using the Phase-Modulated CLEAN Chemical EXchange (CLEANEX-PM) Approach with a Fast-HSQC (FHSQC) Detection Scheme. *J. Biomol. NMR* **11**, 221–226 (1998).
35. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).
36. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix. *Mol. Biol. Evol.* **26**, 1641–1650 (2009).
37. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**, 127–128 (2007).
38. Bondy-Denomy, J. *et al.* A Unified Resource for Tracking Anti-CRISPR Names. *CRISPR J.* **1**, 304–305 (2018).
39. Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, 1026–1028 (2017).
40. Meng, E. C. *et al.* UCSF ChimeraX: Tools for structure building and analysis. *Protein Sci.* **32**, e4792 (2023).
41. Guo, T. W. *et al.* Cryo-EM Structures Reveal Mechanism and Inhibition of DNA Targeting by a CRISPR-Cas Surveillance Complex. *Cell* **171**, 414-426.e12 (2017).
42. Borges, A. L. *et al.* Bacterial alginate regulators and phage homologs repress CRISPR–Cas immunity. *Nat. Microbiol.* **5**, 679–687 (2020).
43. Schuller, M. *et al.* Fragment binding to the Nsp3 macrodomain of SARS-CoV-2 identified through crystallographic screening and computational docking. *Sci. Adv.* **7**, eabf8711 (2021).

44. Bhattacharya, A., Tejero, R. & Montelione, G. T. Evaluating protein structures determined by structural genomics consortia. *Proteins Struct. Funct. Bioinforma.* **66**, 778–795 (2007).
45. Brunger, A. T. Version 1.2 of the Crystallography and NMR system. *Nat. Protoc.* **2**, 2728–2733 (2007).
46. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291 (1993).
47. Lovell, S. C. *et al.* Structure validation by Cα geometry: ϕ,ψ and Cβ deviation. *Proteins Struct. Funct. Bioinforma.* **50**, 437–450 (2003).
48. Lüthy, R., Bowie, J. U. & Eisenberg, D. Assessment of protein models with three-dimensional profiles. *Nature* **356**, 83–85 (1992).
49. Sippl, M. J. Recognition of errors in three-dimensional structures of proteins. *Proteins Struct. Funct. Bioinforma.* **17**, 355–362 (1993).
50. Zhang, Y. & Skolnick, J. Scoring function for automated assessment of protein structure template quality. *Proteins Struct. Funct. Bioinforma.* **57**, 702–710 (2004).
51. Xu, J. & Zhang, Y. How significant is a protein structure similarity with TM-score = 0.5? *Bioinformatics* **26**, 889–895 (2010).

Author contributions

A.L.B., J.B.D, and J.S.F. conceived the project. A.L.B., J.Y.Z, Y.L. and D.F.C. executed phage & biochemistry experiments. L.T.R. and M.J.S.K. determined the NMR structure. P.B., S.M., M.G.V.S., G.J.C., M.E.D., D.H.S., A.A., R.M.S. provided macrodomains. J.B.D. and J.S.F. supervised experiments and acquired funding. D.F.C., J.B.D, and J.S.F. wrote the manuscript with input from coauthors.

Acknowledgements

Experiments were funded by awards to J.B.D by the Vallee, Searle, and Kleberg Foundations and to J.S.F. by NIH GM145238 and U19AI171110. D.F.C is supported by the UCSF Discovery Fellowship. D.F.C. and J.S.F would like to thank Dr. Allison Williams for guidance on biochemistry experiments, and Dr. Amy Diallo and Dr. Kliment Verba for BFV and EEEV constructs.

Competing interest statement

J.B.D. is a scientific advisory board member of SNIPR Biome, Excision Biotherapeutics, and LeapFrog Bio, consults for BiomX, and is a scientific advisory board member and co-founder of Acrigen Biosciences and ePhective Therapeutics. The Bondy-Denomy lab received research support from Felix Biotechnology. A.A. is a co-founder of Tango Therapeutics, Azkarra Therapeutics and Kytarro; a member of the board of Cytomx, Ovibio Corporation, Cambridge Science Corporation; a member of the scientific advisory board of Genentech, GLAdiator, Circle, Bluestar/Clearnote Health, Earli, Ambagon, Phoenix Molecular Designs, Yingli/280Bio, Trial Library, ORIC and HAP10; a consultant for ProLynx, Next RNA and Novartis; receives research support from SPARC; and holds patents on the use of PARP inhibitors held jointly with AstraZeneca from which he has benefited financially (and may do so in the future). J.S.F. is a consultant to, shareholder of, and receives sponsored research support from Relay Therapeutics.