Title

Cellular, biochemical, and structural characterization of anti-immune factors in the phage-bacteria arms race

ISBN

9798270227098

Author

Chen, Daphne Flora

Publication Date

2025-12-17

Cellular, biochemical, and structural characterization of anti-immune factors in the phage-bacteria arms race

by
Daphne Flora Chen

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Biophysics

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:
James Fraser
430BBB9A04D24A3...

James Fraser
Chair

DocuSigned by:
Joseph Bondy-Denomy
[illegible]47C...

Joseph Bondy-Denomy

DocuSigned by:
Allison Williams
53CA70BD89CD49F...

Allison Williams

Committee Members

Acknowledgements

First, thank you to my mentors, Dr. James Fraser and Dr. Joseph Bondy-Denomy. I didn't anticipate falling down the microbiology rabbit hole in graduate school, and I'm grateful to both of you for letting me pursue my interests and carve my own path between microbiology and structural biology. I've enjoyed juggling both of these fields – and I've only been able to do so because you have worked hard to keep your labs steady through uncertain times. Because of the two of you, I got the best of both worlds. Thank you for your eternal patience and encouragement.

Thank you to Dr. Allison Williams, for sitting through the good and the bad thesis committee meetings. Both your technical advice and career advice have been invaluable.

I wouldn't have survived graduate school without the support of my labmates. Eric Greene, your positivity and perseverance continue to inspire me to this day. Wearn Xin Yee, thank you for enduring my constant chaos and panic with patience, snacks, and stationery. Yuping Li, thank you for teaching me phage techniques and answering every single question I had with unrelenting kindness. Claire Kokontis, Lexie Villani, and Alex Hong: thank you for commiserating with me and taking me in as your own, even though I am nowhere near cool enough to be hanging out with you all. On the 4th floor: thank you to Sofia Bali, Galen Correy, Kara Zielinski, CJ San Felipe, Karson Chrispens, Gabby Estevam, Jenna Pellegrino, John Lee, Joey Olmos, Stephanie Wankowicz, Robbie Diaz, Ashraya Ravikumar, Donovan Trinidad, Patrick Rockefeller Grimes, Sonya Lee, and Jessica Flowers. On the 3rd floor: thank you to Alex Chan, Iana Fedorova, Tim Klein, Erin Huiting, Nicole Marino, Heloise Carion, Matthew Johnson, and Michelle Grunberg. Thank you all for being part of my scientific journey (and for keeping me sane!).

Thank you to Nicole Flowers, who somehow manages to keep the Biophysics program afloat despite all the turbulence over the past few years. Nicole, you are our hidden gem!

Thank you to Dr. Giovanni Zocchi and Yilin Wong at UCLA for introducing me to the world of scientific research. Thank you also to Francoise Queval, Liya Oster, and Liz Mills

for encouraging me to pursue research. Erika Hoffman, Rafi Hessami, and Sonia Chung: thanks for keeping in touch all these years!

To Dr. Kliment Verba, my first homie at UCSF: thank you for taking a chance on me back when I was a clumsy undergraduate. Spending the summer doing research at UCSF gave me the confidence and motivation to pursue graduate school.

Thank you to all the friends who believed in me, even when I didn't believe in myself. Caroline Viola, you have always grounded me and reminded me of the bigger picture - thank you for being there for me, from middle school onwards! Brian Sullivan, Naomi Genuth, and John Schwarz, I've thoroughly enjoyed your chaos - thanks for making me laugh. Gabby Estevam, thank you for all the dramatic adventures and schemes. Lieza Chan, thank you for laughing and crying with me through everything. Estelle Ronayne, thank you for nurturing my creativity. Duncan Muir, thanks for a shuckin' good time. Adamo Mancino, thank you for your unwavering support. Dominic Grisingher, thanks for cheering me on since day one of graduate school; team work makes the dream work!

Last but not least, thank you to my family, for giving me a life filled with opportunities you never had. I can't imagine the scale of sacrifice it took for you to raise me and help me get to where I am today. Thank you for instilling curiosity and determination into me; every success I've had is because of the foundation you've set for me.

Contributions

Chapter 1

Chapter 1 contains work that has been previously published:

Chen DF*, Roe LT*, Yuping L, Borges AL, Zhang JY, Babbar P, Maji S, Stevens MGV, Correy GJ, Diolaiti ME, Smith DH, Ashworth A, Stroud RM, Kelly MJS, Bondy-Denomy J, Fraser JS. 2025.

AcrIF11 is a potent CRISPR-specific ADP-ribosyltransferase encoded by phage and plasmid. mBio 16:e01698-25. https://doi.org/10.1128/mbio.01698-25

Chapter 2

Kokontis C, Chen DF, Zheng I, Klein TA, Bondy-Denomy J, Fraser JS. 2025.

Biochemical and structural characterization of jumbophage nuclear import factors.

*denotes equal contribution

"Form is certainty. All nature knows this, and we have no greater adviser...Each form sets a tone, enables a destiny, strikes a note in the universe unlike any other. How can we ever stop looking? How can we ever turn away?"

-Mary Oliver, *Upstream*

Cellular, biochemical, and structural characterization of anti-immune factors in the phage-bacteria arms race

Daphne Chen

Abstract

Bacteria are subject to predation by bacteriophages, leading to an evolutionary arms race where bacteria-encoded immune systems and phage-encoded anti-immune factors evolve to counteract each other. A classic example of this evolutionary arms race dynamic is bacterial CRISPR-Cas immune systems, which are inhibited by phage-encoded anti-CRISPR proteins (Acrs). While CRISPR-Cas and Acr interactions have been characterized *in vitro*, their impact on fitness, and the evolutionary benefits and drawbacks of the mechanistic diversity in the Acr family, are best investigated in the native context of phage infection, which I explore in Chapter 1. On the other hand, for anti-immune factors that are less well-characterized than Acrs such as the jumbophage nucleus, *in vitro* biochemistry can aid in establishing molecular mechanisms of protein import, which I present preliminary work for in Chapter 2. Altogether, my thesis characterizes anti-immune factors at different scales, from molecular mechanism to cell-level fitness.

In Chapter 1, I investigate the advantages and disadvantages of enzymatic Acr activity using phage-based experiments. Acrs can inhibit CRISPR-Cas by a variety of mechanisms, most of which involve stable direct binding of Acr to CRISPR-Cas complex. An Acr's mechanism of action is typically confirmed via structure determination and *in vitro* biochemical assays using purified protein. In contrast to these techniques, and in order to understand how Acrs behave during phage infection, I characterized the anti-CRISPR enzyme AcrIF11 using phage-based techniques. Based on the NMR structure of an AcrIF11 homolog, I made mutations to inhibit AcrIF11's ADP-ribosyltransferase activity, and directly correlated this activity to phage infection outcomes in *P. aeruginosa*. I also showed that AcrIF11 is effective under high CRISPR targeting and CRISPR-Cas overexpression conditions, allowing phage

survival in conditions where stably bound, stoichiometric Acrs are overwhelmed. Furthermore, I showed that AcrIF11 is capable of protecting lysogens from deleterious self-targeting, stabilizing their hosts to the same extent as a strong stable binder Acr. To investigate the fitness advantage that enzymatic Acrs confer, I performed a phage competition experiment and observed that AcrIF11 can outcompete a stable binding Acr. Finally, I demonstrated a potential drawback to enzymatic Acrs: removal of AcrIF11's ADP-ribose modification via ADP-ribose erasers (macrodomains).

In Chapter 2, I turn my attention from Acrs protecting phage against one DNA-targeting immune system to a mechanism capable of protecting phage against multiple DNA-targeting immune systems. Jumbophages such as ϕKZ form a proteinaceous nuclear shell inside their host during infection, protecting the phage genome from cytoplasmic bacterial immune systems such as CRISPR-Cas and restriction-modification systems. However, this phage nucleus also forms a barrier against proteins that must be imported into the nucleus, such as genome maintenance proteins and non-virion RNA-polymerase (nvRNAP). By toxifying imported cargo and identifying escape mutants, Kokontis et. al. identified *imp1* and other import-related genes in ϕKZ. Interestingly, Kokontis et. al. observed a pattern of escape mutations suggesting that different interfaces of Imp1 are responsible for importing different cargo. Now we aim to answer the question of specificity, stoichiometry, and physical mechanism of the central import complex at a structural level by solving a cryoEM structure of Imp1 and its partners. I purified Imp1 with its partner Imp6 using a co-expression construct. However, attempts to separate the Imp1-Imp6 complex from abundant Imp6 homotetramers were unsuccessful. Aside from the central import complex, I also investigated candidates in the nvRNAP import pathway, performing co-immunoprecipitation experiments to test for stable binding. Preliminary experiments show no stable binding interaction between nvRNAP and individual hits from the genetic selection, providing the beginnings of a mechanistic hypothesis for import of this crucial protein complex.

Contents

1 AcrIF11 is a potent CRISPR-specific ADP-ribosyltransferase encoded by phage and plasmid. 1

Introduction 1

Results 3

Discussion 18

Methods 22

Supplementary Material 40

References 51

2 Biochemical and structural characterization of jumbophage nuclear import factors. 57

Introduction 57

Results 59

Discussion 64

Methods 65

References 71

List of Figures

1.1 NMR structure of AcrIF11$_{Pae2}$ 4

1.2 ADP-ribosylation activity of AcrIF11 7

1.3 *acrIF11* distribution and protection of lytic phage. 11

1.4 Prophage-encoded *acrIF11* prevents self-targeting 14

1.5 Macrodomain removal of ADPr 16

1.6 Alphafold3 prediction of AcrIF11-Csy recognition interface. 19

1.7 Sequence logo of AcrIF11 homologs. 43

1.8 Alignment of experimental and predicted structures of AcrIF11$_{Pae1}$ and AcrIF11$_{Pae2}$ 44

1.9 Confidence of Alphafold2 predictions for AcrIF11 homologs. 44

1.10 Enlarged version of the AcrIF11 phylogeny in Figure 1F. 44

1.11 AcrIF11 phylogeny sequence alignment 45

1.12 Replicates of lysogen growth experiment 46

1.13 Quantification of macrodomain lysate blot 47

1.14 Verification of macrodomain lysate blot loading via Ponceau 47

1.15 Liquid growth curves of PA14 WT overexpressing non-endogenous macrodomains 48

1.16 Alphafold3 prediction of Csy complex + NAD + crRNA + AcrIF11$_{Pae2}$ colored by pLDDT 49

2.1 Imp6 strongly associates with Imp1 during protein purification 60

2.2 nvRNAP does not stably associate with import and adapter proteins . . . 62

List of Tables

1.1 NMR assignment, structure calculation and validation statistics 40

1.2 List of bacterial genera containing AcrIF11 homologs 41

1.3 1 L M9 Minimal Media Recipe for ^{15}N labeled proteins 42

1.4 1 L M9 Minimal Media Recipe for ^{15}N ^{13}C labeled proteins 42

Chapter 1. AcrIF11 is a potent CRISPR-specific ADP-ribosyltransferase encoded by phage and plasmid.

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 three 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 activity, nuclease recruitment, and so forth [1]. The second mechanism is CRISPR-Cas complex dissociation or degradation [2, 3]. The third mechanism is enzymatic, substoichiometric modification of CRISPR-Cas. Several distinct enzymatic modifications have been observed, including crRNA cleavage [4, 5, 6], degradation of Type III CRISPR-Cas signaling molecules [7], and post-translational modifications such as acetylation [8, 9] and ADP-ribosylation [10].

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 toward enzymes. For example, AcrIF11 was previously discovered as a neighbor of a gene encoding an anti-CRISPR repressor protein *aca1* and described as a widespread Type I-F anti-CRISPR protein, which enabled the discovery of Cas12 Acrs via guilt-by-association [11]. Although the sequence was too diverged to permit functional assignment upon discovery, a crystal structure revealed similarity to diphtheria toxin, an ADP-ribosyltransferase [10]. Furthermore, *in vitro* biochemical assays demonstrated that AcrIF11 modified N250 of

Csy1 in the Type I-F Csy complex, which inhibited target DNA binding [10].

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 interfere with the host, especially when encoded by symbiotic prophages or plasmids. Moreover, hyper-potent and specific Acr proteins could be well suited to stabilizing MGE-host symbiosis, to prevent CRISPR-Cas self-targeting [12].

Here, we find that ADP-ribosyltransferase 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 CRISPR-Cas autoimmunity. In addition, 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. However, as a potential cost to enzymatic mechanisms, we propose that the reversible nature of post-translational modifications would allow for removal by the host. Indeed, we provide proof of principle for 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

The NMR structure of AcrIF11$_{\mathbf{Pae2}}$ 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 [10], 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 the structure of a homolog of AcrIF11, hereafter noted as AcrIF11$_{\mathrm{Pae2}}$ to avoid confusion with the homolog used in a previous study [10] (PDB: 6KYF), that we will call AcrIF11$_{\mathrm{Pae1}}$. Despite the low sequence similarity (approximately 27% amino acid sequence similarity) between AcrIF11$_{\mathrm{Pae1}}$ and AcrIF11$_{\mathrm{Pae2}}$, their beta sheet domains align well, and there are conserved negatively charged residues in positions that have been previously noted as crucial for catalysis [13, 14] (Fig 1.1A). 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. 1.1B). In addition, the negatively charged AcrIF11$_{\mathrm{Pae2}}$ residues D146 and E147, which are conserved across representative AcrIF11 homologs (Fig 1.7), experience large and medium chemical shift perturbations (Fig. 1.1B), respectively, highlighting their importance in NAD binding. The NMR structural ensemble of AcrIF11$_{\mathrm{Pae2}}$ 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. 1.1C). 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 [15].

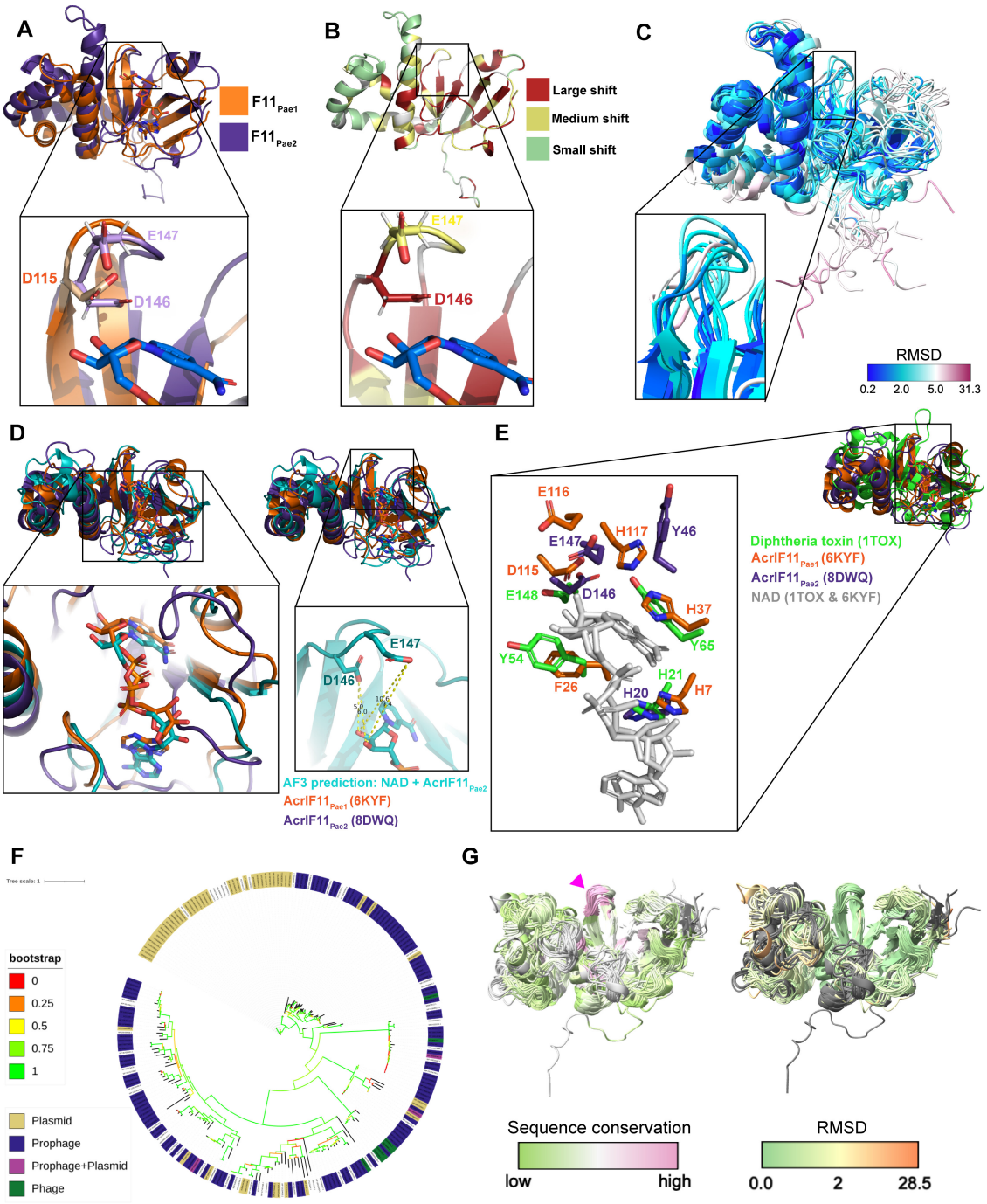

Figure 1.1: NMR structure of AcrIF11$_{Pae2}$. (A) Alignment of AcrIF11$_{Pae1}$ (orange) crystal structure (PDB: 6KYF), and AcrIF11$_{Pae2}$ (purple) NMR structure (PDB: 8DWQ). The inset shows D115 of AcrIF11$_{Pae1}$, a residue previously shown to be important for ART activity [10], as well as D146 and E147, negatively charged residues of AcrIF11$_{Pae2}$ in a similar position to D115. (Figure caption continued on the next page)

(Figure caption continued from the previous page) (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$_{Pae1}$ (orange) crystal structure (PDB: 6KYF), AcrIF11$_{Pae2}$ (purple) NMR structure (PDB: 8DWQ), and AcrIF11$_{Pae2}$ (teal) Alphafold3 prediction with NAD. The left inset shows the position of NAD in the AcrIF11$_{Pae1}$ crystal structure in comparison to its position in the Alphafold3 prediction of AcrIF11$_{Pae2}$. The right inset shows the positioning of NAD in the Alphafold3 prediction of AcrIF11$_{Pae2}$ relative to the catalytic residues discussed in the panels above. (E) Alignment of AcrIF11$_{Pae1}$, AcrIF11$_{Pae2}$, and the catalytic domain of the monomer of diphtheria toxin (PDB: 1TOX) based on the NAD (light gray) molecules of AcrIF11$_{Pae1}$ and diphtheria toxin. The loop where D146 and E147 are located on AcrIF11$_{Pae2}$ is more extended in the diphtheria toxin, such that there are no diphtheria toxin residues in the equivalent position of Y46. (F) 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). An enlarged version of this phylogeny can be found in Fig. 1.10, and a sample of the alignment used to build the phylogeny is provided in Fig. 1.11. (G) Alphafold2 predictions of AcrIF11$_{Pae1}$ homologs, colored by sequence conservation and RMSD. The pink arrow indicates the same region as panel A, B, and C insets.

An Alphafold3 prediction of AcrIF11$_{Pae2}$ in complex with NAD also predicted that NAD binds in a similar position to what appears in the crystal structure of AcrIF11$_{Pae1}$ (Fig. 1.1D). Although the prediction places the nicotinamide ribose too far away to interact with D146 and E147 (Fig. 1.1D, right inset), the overall configuration of the NAD molecule is similar to what is observed in the crystal structure (Fig. 1.1D, left inset). Further experimental studies would be required to detail the dynamic changes in AcrIF11$_{Pae2}$ upon NAD binding and how they differ from AcrIF11$_{Pae1}$. The position of AcrIF11$_{Pae2}$ residues in the NAD-binding/catalytic region appears to be conserved compared to AcrIF11$_{Pae1}$ and diphtheria toxin (Fig. 1.1E), 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 [10]. With this annotation

in hand, we investigated the extent of *acrIF11* spread (an enlarged version of this phylogeny can be found in Fig. 1.10). 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 [16] 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. 1.1F).

Having observed the structural similarity between AcrIF11$_{\text{Pae1}}$ and AcrIF11$_{\text{Pae2}}$ despite their low sequence similarity, we were curious whether other AcrIF11 homologs displayed a similar structure. Using homologs in the phylogeny in Fig. 1.1F, we clustered homologs with a sequence identity cutoff of 95% to prevent redundancy, trimmed the alignment of representative sequences to prevent large gaps in their alignment, and predicted structures of those representative sequences from the alignment using Alphafold2 [17]. We used a template date set earlier than the AcrIF11$_{\text{Pae1}}$ deposition date to avoid template bias. The predicted AcrIF11$_{\text{Pae1}}$ and AcrIF11$_{\text{Pae2}}$ structures aligned well with the experimental structures (Fig. 1.8), and most of the structures were predicted with high confidence (Fig. 1.9). Alignment of all predicted structures to the predicted AcrIF11$_{\text{Pae1}}$ structure revealed that the most variable structural feature is the alpha helical region (approximately residues 51–107 of AcrIF11$_{\text{Pae1}}$ PDB 6KYF), while the most structurally conserved region is the loop where AcrIF11$_{\text{Pae2}}$ D146/E147 and AcrIF11$_{\text{Pae1}}$ D115/E116 are located (pink arrow, Fig. 1G). These results show that Alphafold2 is capable of detecting conserved catalytic regions across structures, even when the overall sequence similarity is low.

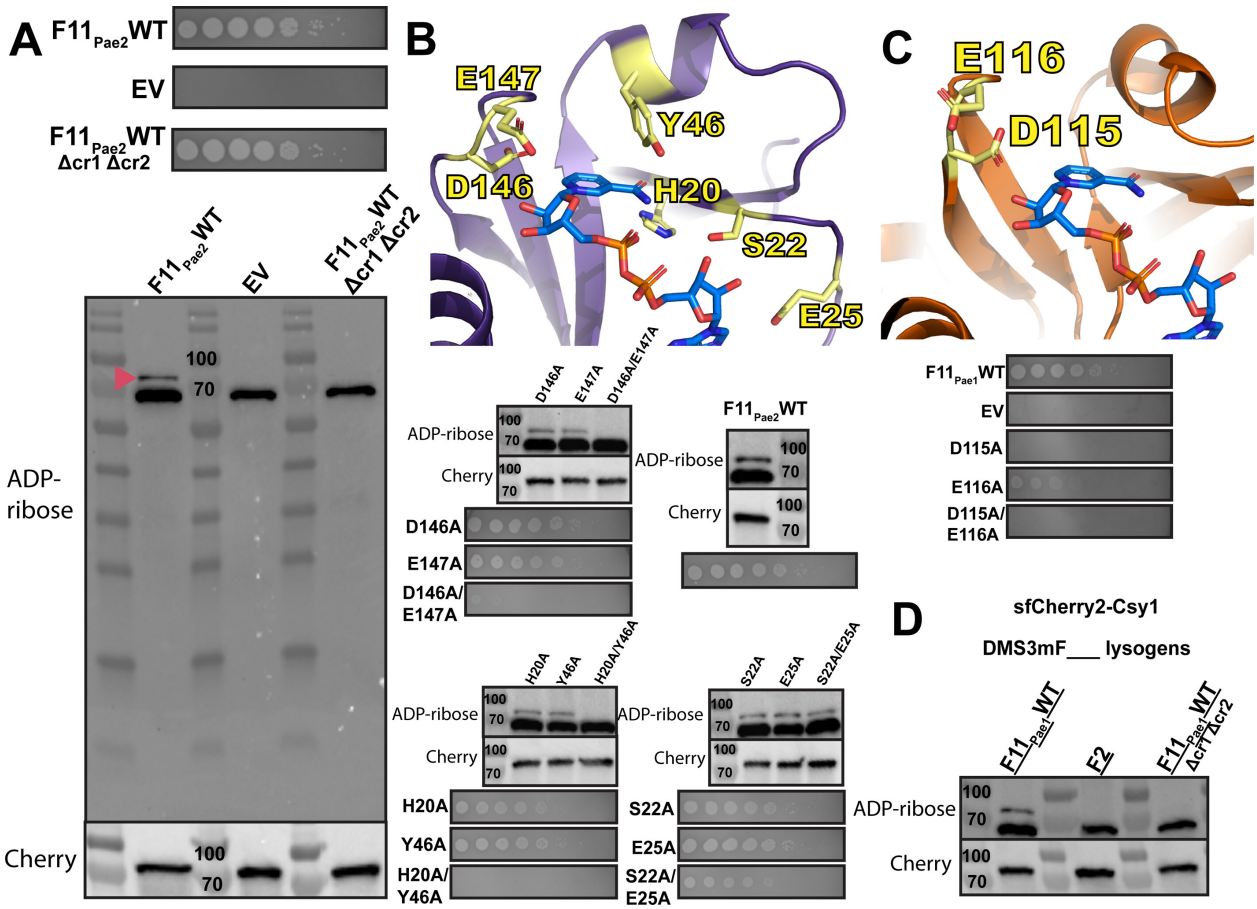

Figure 1.2: ADP-ribosylation activity of AcrIF11. (A) Plaque assays of PA14 chromosomally expressing sfCherry2-Csy1 and also containing a plasmid expressing wild-type AcrIF11$_{Pae2}$ (F11$_{Pae2}$WT) or empty vector (EV). Δcr1Δcr2 indicates PA14 with both CRISPR arrays deleted. These strains were also lysed and probed for ADPr and Cherry. The pink arrow indicates the sfCherry2-Csy1 protein modified with ADPr. (B) Plaque assays and Western blots as in panel A, but with wild-type and mutant versions of AcrIF11$_{Pae2}$. Mutants are colored yellow on the AcrIF11$_{Pae2}$ NMR structure (PDB: 8DWQ). The NAD molecule was solved as part of the AcrIF11$_{Pae1}$ crystal structure (PDB: 6KYF). (C) Plaque assays and Western blots as in panel A, but with wild-type 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* [10], the specificity of AcrIF11 in the cellular environment is still unknown. To detect the ADP-ribose modification imparted by AcrIF11 on Csy1 (Cas8 homolog), we expressed AcrIF11$_{\text{Pae2}}$ on a plasmid in PA14, and then blotted the lysate for ADPr-modified protein with an ADPr-specific antibody (Fig. 1.2A). A previously engineered PA14 strain was used, which has an sfCherry2-fusion to Csy1 at the endogenous location in the genome [18] to enable us to blot for Cherry for both a loading control and to detect any changes in Csy1 length/stability upon modification. Upon expression of WT AcrIF11$_{\text{Pae2}}$, we observed an ADPr band at the expected mass of sfCherry2-Csy1 (Fig. 1.2A, F11$_{\text{Pae2}}$2WT lane, red arrow). The strong band slightly below 70 kDa is presumably an ADP-ribosylated protein in the PA14 lysate, as shown by its presence in the absence of AcrIF11$_{\text{Pae2}}$ (Fig. 1.2A, 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 a strain where CRISPR RNAs will not be expressed and processed, AcrIF11$_{\text{Pae2}}$ was unable to ADP-ribosylate Csy1 (Fig. 1.2A, F11$_{\text{Pae2}}$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 [10]. Regarding specificity in the intracellular environment, the only detectable covalent ADP-ribosylation imparted in an AcrIF11-dependent manner was on Csy1, suggesting that this is a highly specific anti-CRISPR mechanism.

To directly correlate ADP-ribosylation to successful CRISPR-Cas inhibition and phage replication, we tested AcrIF11$_{\text{Pae2}}$ mutants' ability to protect CRISPR-targeted DMS3m, and also blotted for ADP ribosylation (Fig. 1.2B). The D146/E147 pair was chosen for its conservation across homologs as well as the potential importance of their negative charge in

stabilizing the positively charged transition state of NAD [13, 14]. The H20/Y46 pair was chosen because the structure suggests they are important residues for stabilizing the nicotinamide leaving group. Lastly, we hypothesized that the S22/E25 pair was 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$_{\text{Pae1}}$ crystal structure 6KYF (Fig. 1.2C) and observed no CRISPR inhibition in D115A, consistent with the absence of ADPr from *in vitro* Western blotting in a previous study [10]. Lastly, the S22A/E25A mutant in AcrIF11$_{\text{Pae2}}$ 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$_{\text{Pae2}}$, 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$_{\text{Pae1}}$ 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 [19]. Using the PA14 sfCherry2-Csy1 strain with a DMS3mAcrIF11$_{\text{Pae1}}$ prophage, we observed an ADP-ribose band at the mass of the sfCherry2-Csy1 fusion (Fig. 1.2D, lane F11$_{\text{Pae1}}$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. 1.2D, lane F2). To test whether assembly of the Csy complex is important as it was previously (Fig. 1.2A F11$_{\text{Pae2}}$WT Δcr1Δcr2 lane), we made a PA14 Δcr1Δcr2 sfCherry2-Csy1 DMS3mAcrIF11$_{\text{Pae1}}$ 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 [20], AcrIF11's substoichiometric nature allows it to ADP-ribosylate and inactivate the Csy complex.

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. 1.1F), we also observed that *acrIF11* is far more widespread than other type I-F Acrs (Fig. 1.3A). *acrIF11* is encoded by MGEs present in over 35 bacterial genera, a full list of which is shown in Table 1.2. 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 [19]. When expressed from a phage infecting PA14, $\text{AcrIF11}_{\text{Pae1}}$ 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. 1.3B). This shows that $\text{AcrIF11}_{\text{Pae1}}$ 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 [19, 21].

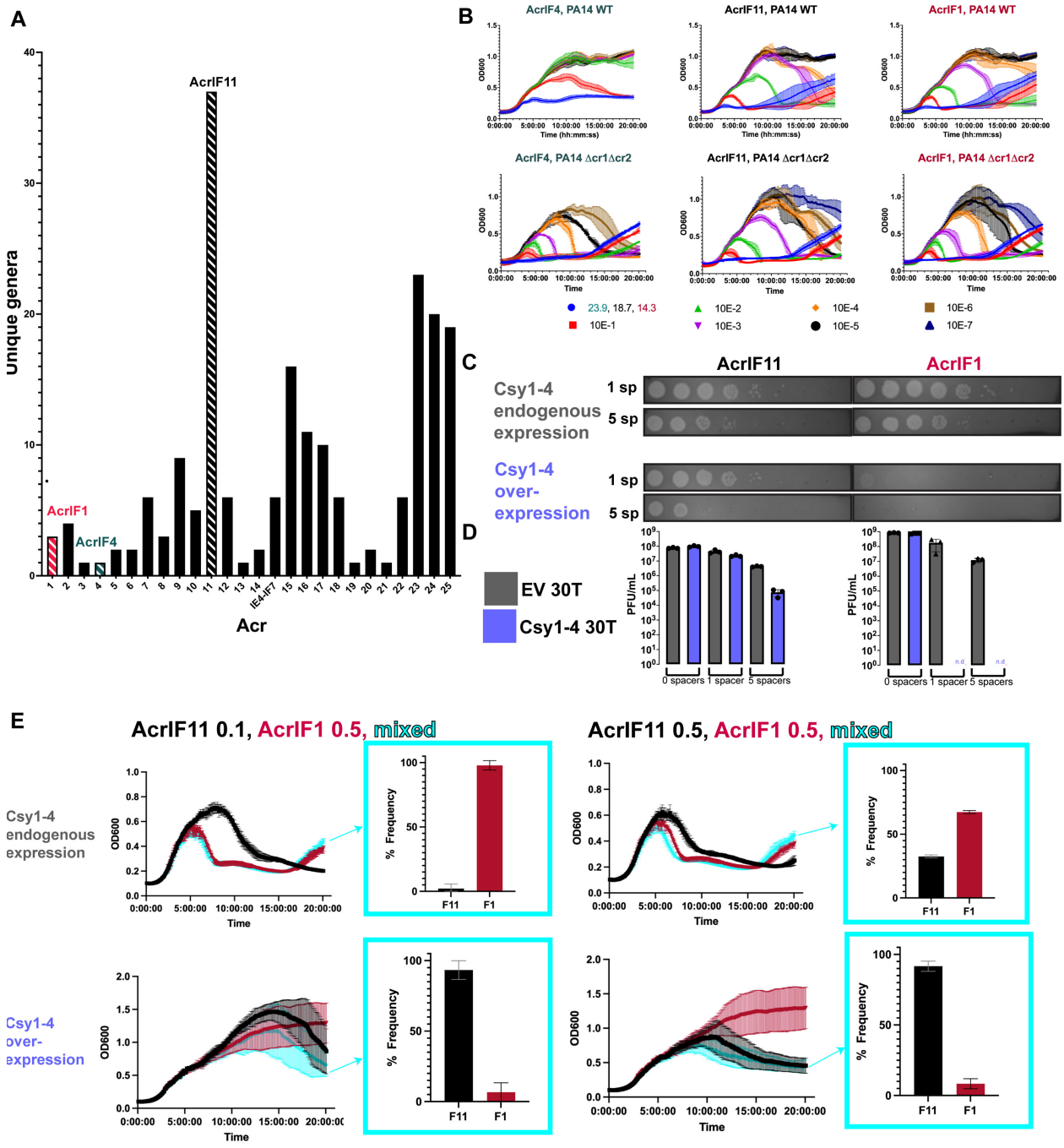

Figure 1.3: *acrIF11* distribution and protection of lytic phage. (A) Plot of the number of bacterial genera that each Type I-F Acr can be found in, via BLASTp. Acrs of interest have been marked with striped bars: AcrIF11 (black), AcrIF1 (pink), AcrIF4 (teal). B) Liquid growth curves of PA14 WT (1 spacer targeting DMS3mvir) infected by DMS3mvir encoding the indicated Acr, across a range of multiplicities of infection (MOI). (Figure caption continued on the next page)

(Figure caption continued from the previous page) 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 (wild-type 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 C. (E) Phage competition growth curves, using the same wild-type and high targeting backgrounds of PA14 as in panel C. PA14 strains were infected with DMS3mvir phages encoding the indicated Acr. Cyan curves show PA14 strains infected with both DMS3mF11vir and DMS3mF1vir at the indicated MOI. The resulting frequencies of observing AcrIF1 or AcrIF11 by genotyping at the end of the time course are plotted. Growth curves and frequencies are the average of three biological replicates.

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 [22, 23]. To test this idea, we overexpressed Csy1-4 in PA14 backgrounds with 1 (wild type) 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. 1.3C and D). This suggests 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.

Next, to directly examine the fitness advantage that AcrIF11 might confer, we conducted a phage competition experiment where phages encoding either $AcrIF11_{Pae1}$ or AcrIF1 were mixed together in equal abundance (Fig. 1.3E, right half) or with AcrIF1 in fivefold excess

of AcrIF11$_{Pae1}$ (Fig. 1.3E, left half). Under wild-type targeting (1sp) and endogenous Csy expression conditions, AcrIF1 was the most frequently occurring Acr in the population at the end of the competition period (Fig. 1.3E, upper half), in both the equal abundance and fivefold excess conditions. However, under Csy overexpression conditions, AcrIF11$_{Pae1}$ became the most frequently occurring Acr in the population (Fig. 1.3E, lower half) after inputs of both equal abundance and fivefold AcrIF1 excess conditions. Remarkably, even when AcrIF1 is in fivefold excess of AcrIF11$_{Pae1}$, it is AcrIF11$_{Pae1}$ that becomes the dominant Acr in the phage population under Csy overexpression (Fig. 1.3E, lower left quadrant). Furthermore, the difference between AcrIF1 and AcrIF11$_{Pae1}$ frequency under Csy overexpression is much larger than the difference between AcrIF1 and AcrIF11$_{Pae1}$frequency under Csy endogenous expression (Fig. 1.3E lower right quadrant vs Fig. 1.3E upper right quadrant). This indicates that under endogenous levels of Csy expression, phages can use either AcrIF11$_{Pae1}$ or AcrIF1 to inhibit the Csy complex, but when Csy expression levels are increased, it is only AcrIF11$_{Pae1}$ that is robust enough to inhibit Csy and allow for phage replication. Overall, our experiments show that environmental pressures such as increased Cas protein overexpression and multi-spacer targeting can select for the substoichiometric mechanism of AcrIF11$_{Pae1}$ over the more conventional stable binding mechanism of AcrIF1.

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 [12]. 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.

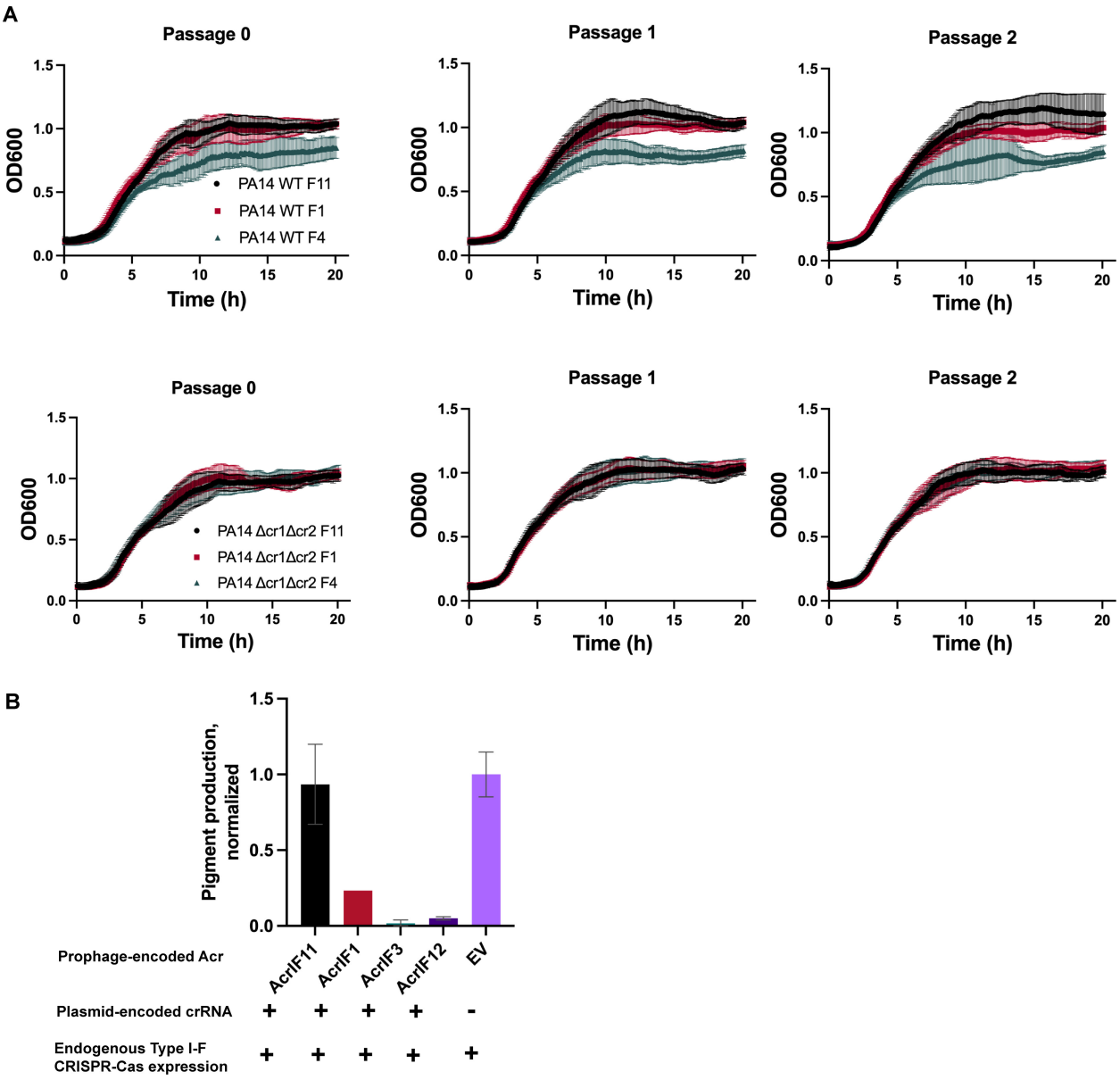

Figure 1.4: Prophage-encoded *acrIF11* prevents self-targeting (A) Liquid growth curves of PA14 lysogens with a DMS3m prophage encoding either AcrIF11$_{\text{Pae1}}$, AcrIF1, or AcrIF4, as indicated in the legend. "WT" refers to wild type PA14, and "Δcr1Δcr2" refers to PA14 with both CRISPR arrays deleted. Plots are the average of three biological replicates. (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.

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$_{\text{Pae1}}$, however, displayed growth comparable to a Δcr1Δcr2 control (Fig. 1.4A), demonstrating full protection. To address the potential for priming spacer acquisition over time, we passaged the lysogens for 3 days but observed no growth defect in the presence of AcrIF11 and AcrIF1. 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. 1.4B). While AcrIF1 has been previously shown to disable CRISPRi, overexpression of the *phzM*-targeting crRNA overwhelmed this Acr, leading to repression of pigment production. In a previous study, AcrIF4 was also shown to partially inhibit CRISPR-Cas [24] in the same CRISPRi setup, as AcrIF1 did in our experiment. Unlike AcrIF1 and AcrIF4, AcrIF11 fully disabled the Csy complex under overexpression of the *phzM*-targeting crRNA, allowing full pyocyanin production compared to a strain with a non-targeting crRNA. This shows that, while AcrIF11 and AcrIF1 are capable of stabilizing lysogeny in wild-type targeting conditions (Fig. 1.4A), 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.

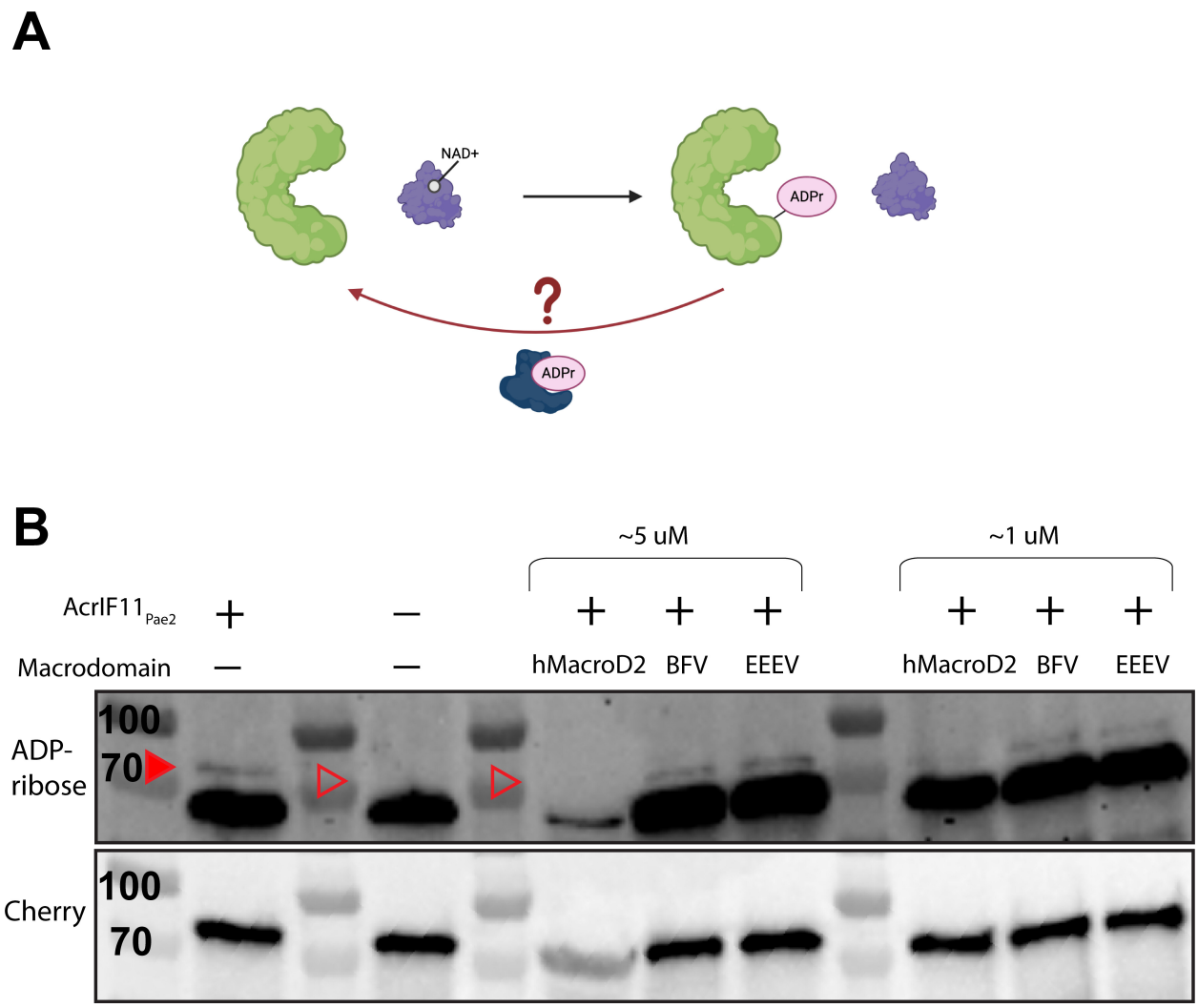

Figure 1.5: Macrodomain removal of ADPr. (A) Schematic of Csy complex (green) modified by AcrIF11 (purple) using NAD as a cofactor. We hypothesize that a macrodomain (dark blue) could remove ADPr from the Csy complex. (B) Western blot of purified macrodomain incubated with lysate from PA14 sfCherry2-Csy1 overexpressing $\text{AcrIF11}_{\text{Pae2}}$ on a plasmid. The mixture was blotted for ADP-ribose and Cherry as in Fig. 1.2. Quantification of the band intensities is available in Fig. 1.13. The sfCherry2-Csy1 ADPr band is indicated by the solid red arrow, and its absence is indicated by the hollow red arrows. We believe the decrease in Cherry intensity in the hMacroD2 5 μM lane is not due to decreased loading, but rather the similar molecular weight of our hMacroD2 construct and sfCherry2-Csy1 leading to overcrowding in that molecular weight range as observed by Ponceau stain in Fig. 1.14, and thus decreased antibody binding.

ADP-ribosylation of Csy1 is reversible

Lastly, we hypothesized that enzymatic modifications such as ADPr could be removed by enzymes known as macrodomains, or host-encoded eraser enzymes (Fig. 1.5A) as a continuation of the phage-bacteria arms race. Macrodomains have previously been found in defense and anti-defense contexts; for example, DarG of the DarTG toxin/anti-toxin system can remove ADPr from ADP-ribosylated DNA [25, 26], and ThsA of the Thoeris system can cleave NAD into nicotinamide and ADPr [27]. However, neither DarG nor ThsA has been shown to remove ADPr from a protein, and little is known about *Pseudomonas* macrodomain proteins. To investigate the possibility of ADPr removal from Csy1 via macrodomain, we introduced purified macrodomains into lysate from our sfCherry2-Csy1 strain expressing plasmid-encoded AcrIF11$_{\mathrm{Pae2}}$. 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. 1.5B, solid red arrow), we observed disappearance of the ADPr signal from sfCherry2-Csy1 (Fig. 1.5B, hollow red arrow), and from the background ADP-ribosylated protein (~70 kD), with hMacroD2 added in. The ADPr signals from incubation with BFV and EEEV appeared to be the same as our control, indicating no removal of ADPr.

To 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 (Fig. 1.15). 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 i*n 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 and its transient binding to the Csy complex could be a disadvantage in strains that encode a modification eraser, compared to a stoichiometric binder such as AcrIF1, which is stably bound to a larger interface of the Csy complex.

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. In addition, 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. AcrIF11's potency and specificity illustrate 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 previously via crystal structure determination and further validated with *in vitro* biochemistry. 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.

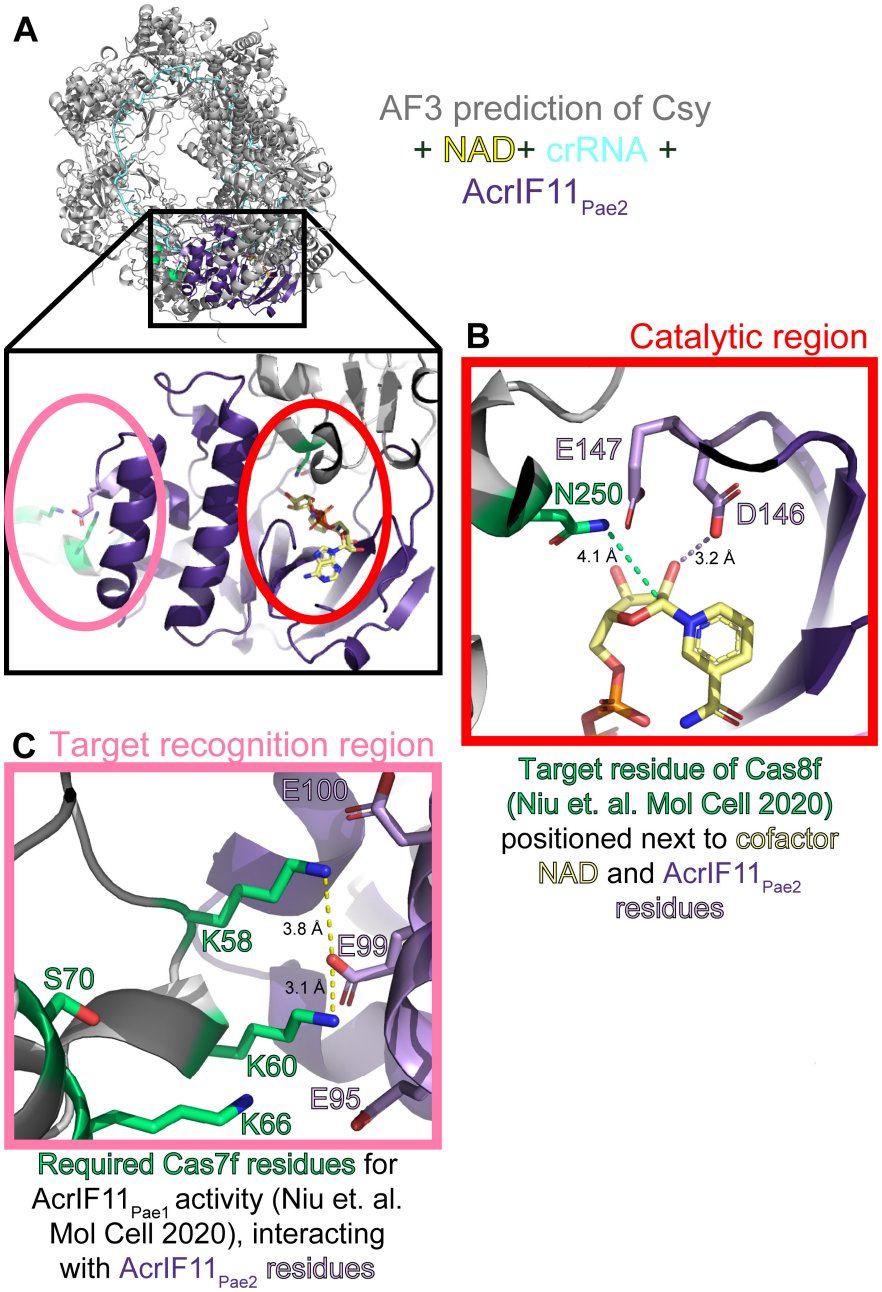

Figure 1.6: Alphafold3 prediction of AcrIF11-Csy recognition interface. (A) Alphafold3 prediction of Csy complex (gray) with crRNA (blue, sequence from PDB 6B45), NAD (yellow), and AcrIF11$_{Pae2}$ (purple). Scores of iPTM = 0.69 and pTM = 0.72 indicate this is a medium confidence prediction [28, 29] The inset depicts the overall position of AcrIF11$_{Pae2}$ on the Csy complex, as well as the two main interfaces of interest: the catalytic region where NAD is bound, and the target recognition region. (B) Catalytic region of AcrIF11$_{Pae2}$ with NAD-interacting residues D146 and E147 highlighted in light purple. The dashed purple line indicates a potential hydrogen bond in this static structure, although the dynamics of catalysis could place D146 and E147 closer. N250, the target residue identified by a previous study [10], is highlighted in green. The green dashed line is the distance between atoms that would be covalently linked after the ADP-ribosylation reaction. (C) Target recognition region between AcrIF11$_{Pae2}$ and the Csy complex. Csy3/Cas7f residues that have been previously identified as important for AcrIF11$_{Pae1}$ activity are highlighted in green. The dashed yellow lines indicate potential charged interactions between Csy3/Cas7f and AcrIF11$_{Pae2}$.

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$_{\text{Pae2}}$, 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$_{\text{Pae1}}$ appears to be most structurally similar to diphtheria-like ARTs, its key residues appear to be H-F-H-D [10] instead of the canonical H-Y-Y-E [30]. Furthermore, the exact classification of AcrIF11$_{\text{Pae2}}$ cannot be determined from our NMR structure alone because this structure was not solved in complex with NAD. The details of NAD-AcrIF11$_{\text{Pae2}}$ binding remain unknown, but our mutagenesis of AcrIF11$_{\text{Pae2}}$ suggests it is similar to diphtheria-like ARTs. The chemical shifts of AcrIF11$_{\text{Pae2}}$ 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 [31].

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 [31], more *in vitro* studies will be needed to determine if catalysis and Csy1 recognition residues can be disentangled for AcrIF11, or if the exact set of residues plays both roles. Mutagenesis has shown that residues in Csy3/Cas7f are required for AcrIF11$_{\text{Pae1}}$ ART activity on N250 of Csy1 [10]; 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$_{\text{Pae2}}$ together (Fig. 1.6A) 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 [10] target residue (Fig 1.6B). Furthermore, the Alphafold3 prediction suggests that the helical region of AcrIF11$_{\text{Pae2}}$ is important for target recognition: residues K58 and K60 of Csy3, residues that have been experimentally confirmed to be important for AcrIF11$_{\text{Pae1}}$ activity [10], are in close proximity to negatively charged residues in the alpha helical region of AcrIF11$_{\text{Pae2}}$, revealing a possible charged interaction interface. Likewise, Csy3 residues that were previously determined to have a weaker effect on AcrIF11$_{\text{Pae1}}$ activity upon mutation, K66 and S70, are placed further away from AcrIF11$_{\text{Pae2}}$ residues in the Alphafold3 prediction (Fig. 1.6C). 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$_{\text{Pae2}}$ that we observed in our lysate blots remains to be explored.

While the enzymatic mechanism of AcrIF11 appears highly potent and specific, enzymatic Acr proteins remain scarce in the literature. This likely represents a discovery challenge that may be solved in the future by integrating structure prediction for *in silico* screening [32] of common enzymatic folds. It is additionally possible that enzymatic Acrs are rarer for one of two potential reasons: first, the evolution of Acr activity from an enzyme capable of a specific chemistry is likely to happen less often than Acr activity arising from any given non-enzyme.

Second, there could be a mechanistic drawback to covalent modifications, as the modification could be undone by host enzymes. The reversal of ADP-ribosylation is an emerging theme in toxin/antitoxin systems important for bacterial and phage ecology [33]. While we could not identify an endogenous bacterial macrodomain that reverses the modification of Csy1, in this study, we have established it is at least possible to remove ADP-ribose from Csy1 using human MacroD2. Furthermore, it would make sense for bacteria to have evolved a way to undo enzymatic modifications from phages, especially if the modifications are on a critical residue, such as N250 of Csy1, which functions in PAM recognition. In this context, an enzymatic Acr's effectiveness would be limited by its high specificity. Stable stoichiometric Acrs, on the other hand, would not be so easily counteracted by the host because they have a larger binding interface with the CRISPR-Cas complex [34]. 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) [35] as a new area to explore in host-phage biology.

Methods

AcrIF11$_{Pae2}$ protein purification

pET28b-AcrIF11$_{Pae2}$ vector

The AcrIF11$_{Pae2}$ coding sequence was as follows:

MGSSHHHHHHSSGLVPRGSHMASMTGGQQMGRMEIFHTSPVEITTINTQGR-FGEFLCFAADEYVMTAGDHVTYRIKVDESDIIMAGSIFYHERAADLSGLVERVMQLT-GCDEDTAEELISQRIDVFNLDDIDASDAAELSWEIQAITAKAAKTLGFRGVSMQDE-QGTCYMIDMLGHDAELVRVK*

The AcrIF11$_{Pae2}$ sequence after thrombin cleavage was as follows:

GSHMASMTGGQQMGRMEIFHTSPVEITTINTQGRFGEFLCFAADEYVM-TAGDHVTYRIKVDESDIIMAGSIFYHERAADLSGLVERVMQLTGCDEDTAEEL-ISQRIDVFNLDDIDASDAAELSWEIQAITAKAAKTLGFRGVSMQDEQGTCYMIDML-GHDAELVRVK*

The AcrIF11$_{Pae2}$ accession is 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 s 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 medium 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. Eight milliliters of the starter culture was harvested by centrifugation at 4,000 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 1 L of M9 minimal media containing appropriate labeling components (see "M9 minimal medium 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 1 mL of 1 mM IPTG and grown for 16 hours at 16°C. The expression culture was harvested by centrifugation at 5,000 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 30,000 g 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 five column volumes (CV), and 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 an 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 buffer A was replaced 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 protein was quantified using absorbance at 280 nm and concentrated to a final protein concentration of 1 mM for NMR analysis.

M9 minimal medium assembly

M9 medium was prepared accordingly for either ^{15}N-labeled AcrIF11$_{Pae2}$ protein expression or ^{15}N-/^{13}C-labeled AcrIF11$_{Pae2}$ protein expression (Tables 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 medium for labeled protein expression.

$\mathbf{AcrIF11_{Pae2}}$ 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 [36] was 1 mM $AcrIF11_{Pae2}$ in 50 mM KPi (pH 7.4) with 5% (vol/vol) D_2O. NMR spectra were all measured at 298.0 K [37]. Two-dimensional (2D) 1H,15N-HSQC (pulse program: fhsqcf3gpph), 2D multiplicity-edited CT 1H,13C-HSQC (pulse program: hsqcctetgpsp), 36 ms 3D HCcH-TOCSY (pulse program hcchdigp3d), and 120 ms 3D simultaneous 13C-/15N-NOESY-HSQC (pulse program: noesyhsqcgpsismsp3d) spectra were measured using a Bruker Avance NEO 800 MHz spectrometer with a 5 mm 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), and 3D hCccoNH (pulse program: hccconhgp3d3) experiments were collected on a Bruker Avance NEO 600 MHz spectrometer with a 5 mm 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 [38].

Resonances were assigned using the program CCPN Analysis version 2.4.2 [39]. 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 [40]. The programs ARIA version 2.3.2 [41] and CNS version 1.2.1 [42] were used to calculate the NMR structures. To solve the structures, nine iterations of simulated annealing were performed using CNS. For the first eight 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 iden-

tified, 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 [43]. 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$_{\text{Pae1}}$ (WP_038819808.1) was used in PSI-BLAST with the nr database for three iterations. 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 [44], and the multiple sequence alignment was trimmed manually to prevent large gaps in the alignment. Sequences larger than 200 residues were also discarded to prevent large gaps in the alignment. Iterative rounds of trimming and alignment were performed to refine the alignment until gaps were smaller than 25 residues. The phylogenetic tree was created from the final multiple sequence alignment using FastTree [45], and visualization was done using iTOL [46]. For determining the distribution of Type I-F Acrs, the Type I-F Acr sequences were first acquired from the Acr database [47]. The sequences were blasted against the clustered_nr database using BLASTp, with an expectation value of 0.0005 and a maximum hitlist size of 5,000. 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 [48], using minimum sequence identity of 95%, cluster mode 2, and coverage mode 1. Sequences were also trimmed for length and to prevent large gaps in alignment. Then the representative sequences of each cluster were used for structure prediction. Se-

quences 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 [49]. 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. [50] PDB 6B45.

AcrIF11$_{\text{Pae2}}$ was as follows:

MEIFHTSPVEITTINTQGRFGEFLCFAADEYVMTAGDHVTYRIKVDESDI-
IMAGSIFYHERAADLSGLVERVMQLTGCDEDTAEELISQRIDVFNLDDIDASDAAEL-
SWEIQAITAKAAKTLGFRGVSMQDEQGTCYMIDMLGHDAELVRVK

Csy1 was as follows:

MTSPLPTPTWQELRQFIESFIQERLQGKLDKLQPDEDDKRQTL-
LATHRREAWLADAARRVGQLQLVTHTLKPIHPDARGSNLH-
SLPQAPGQPGLAGSHELGDRLVSDVVGNAAALDVFKFLSLQYQGKNLLNWLT-
EDSAEALQALSDNAEQAREWRQAFIGITTVKGAPASHSLAKQLYFPLPGSGYH-
LLAPLFPTSLVHHVHALLREARFGDAAKAAREARSRQESWPHGFSEYPNLAIQKFG-
GTKPQNISQLNNERRGENWLLPSLPPNWQRQNVNAPMRHSSVFEHDFGRTPEVS-
RLTRTLQRFLAKTVHNNLAIRQRRAQLVAQICDEALQYAARLRELEPGWSATPGC-
QLHDAEQLWLDPLRAQTDETFLQRRLRGDWPAEVGNRFANWLNRAVSSDSQILGS-
PEAAQWSQELSKELTMFKEILEDERD

Csy2 was as follows:

MSVTDPEALLLLPRLSIQNANAISSPLTWGFPSPGAFTGFVHALQRRVG-
ISLDIELDGVGIVCHRFEAQISQPAGKRTKVFNLTRNPLNRDGSTAAIVEE-
GRAHLEVSLLLGVHGDGLDDHPAQEIARQVQEQAGAMRLAGGSILPWCNERF-
PAPNAELLMLGGSDEQRRKNQRRLTRRLLPGFALVSREALLQQHLETLRTTL-
PEATTLDALLDLCRINFEPPATSSEEEASPPDAAWQVRDKPGWLVPIPAGYNAL-
SPLYLPGEVRNARDRETPLRFVENLFGLGEWLSPHRVAALSDLLWYHHAEPDKG-
LYRWSTPRFVEHAIA

Csy3 was as follows:

MSKPILSTASVLAFERKLDPSDALMSAGAWAQRDASQEWPAVTVREKSVRGTIS-
NRLKTKDRDPAKLDASIQSPNLQTVDVANLPSDADTLKVRFTLRVLGGAGTPSACN-
DAAYRDKLLQTVATYVNDQGFAELARRYAHNLANARFLWRNRVGAEAVEVRIN-
HIRQGEVARAWRFDALAIGLRDFKADAELDALAELIASGLSGSGHVLLEV-
VAFARIGDGQEVFPSQELILDKGDKKGQKSKTLYSVRDAAAIHSQKIGNALR-
TIDTWYPDEDGLGPIAVEPYGSVTSQGKAYRQPKQKLDFYTLLDNWVLRDEA-
PAVEQQHYVIANLIRGGVFGEAEEK

Csy4 was as follows:

MDHYLDIRLRPDPEFPPAQLMSVLFGKLHQALVAQGGDRIGVSFPDLDESRSRL-
GERLRIHASADDLRALLARPWLEGLRDHLQFGEPAVVPHPTPYRQVSRVQAK-
SNPERLRRRLMRRHDLSEEEARKRIPDTVARALDLPFVTLRSQSTGQHFRLFIRHG-
PLQVTAEEGGFTCYGLSKGGFVPWF

crRNA was as follows:

CUAAGAAAUUCACGGCGGGCUUGAUGUCCGCGUCUACCUGGUUCACUGC-
CGUGUAGGCAG

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 data set 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 [16]. 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 [24]. 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. 1.4B, pyocyanin levels were normalized to the empty vector control.

Molecular cloning and 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. The pHERD vector was 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 recovering in LB at 37°C for 2–3 hours. 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, with 0.1% arabinose if induction of a pHERD vector was necessary. 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 140μL 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 (and arabinose at 0.1% when induction of a pHERD vector was necessary). 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. Overnight culture (150μL) was added to 10μL 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 arabinose at 0.1% when induction of a pHERD vectory was necessary) 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.

Phage competition

PA14 WT, containing either Csy1-4 on pHERD30T or empty vector pHERD30T, was grown in LB + 10 mM Mg + Gent50 overnight at 37°C and diluted 1:100 in LB + 10 mM

Mg + Gent50 + 0.1% arabinose. DMS3mF11$_{\text{Pae1}}$vir and DMS3mF1vir were diluted to the appropriate concentration in SM buffer. In each well of a 96-well plate, 140μL of the 1:100 diluted overnight culture was added, and either 10μL of one phage was added for single-phage growths, or 5μL of each phage for mixed phage growths. Plates were grown at 37°C on a plate reader that monitored OD600. After the growth period was completed, phages were extracted from plate reader cultures using 6%–8% chloroform. These extracted phages were used in full plate infections with PA14 Δcr1Δcr2, with each MOI or MOI combination indicated in Fig. 1.3E on a separate plate. 16 plaques were picked for each MOI or MOI combination and subjected to PCR using primers surrounding DMS3 gene 30 (where the appropriate acr gene was inserted). Based on the size of the PCR product viewed on an agarose gel, the identity of the Acr was determined (AcrIF1 is approximately 300 bp, while AcrIF11$_{\text{Pae1}}$ is approximately 500 bp).

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, and 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 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, 15–30 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 1:1,000 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 three 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 three rounds. Bio-Rad Clarity Max ECL substrate was added to the blots in accordance with the manufacturer's protocol, and blots were imaged using the Bio-Rad Chemidoc. Blots for ADPr and Cherry were run in parallel with the same lysates and same loading volumes. During initial trial runs of the ADPr blots, we noticed significant differences in signal detection between ADPr antibodies from different manufacturers, as well as sample preparation and storage conditions. This is in line with previous observations of ADPr reagent variability [51], and for consistency, we decided to use only the Cell Signaling primary mentioned above.

Western blot quantification

Image Lab 6.1 from Bio-Rad was used for quantification of band intensity in Fig. 1.5B. In the Analysis Toolbox, lanes and the appropriate bands were selected using the tools in the Lanes and Bands menu. Background subtraction and accurate band selection were manually verified using the Lane Profile tool. Adjusted volume of each band was normalized to the adjusted volume of the +AcrIF11$_{\mathrm{Pae2}}$, -macrodomain band (leftmost lane in Fig. 1.5B).

hMacroD2 macrodomain protein purification

The hMacroD2 sequence is as follows:

MHHHHHHSSGVDLGTENLYFQSYPSNKKKKVWREEKERLLKMTLEERRKEYL-
RDYIPLNSILSWKEEMKGKGQNDEENTQETSQVKKSLTEKVSLYRGDITLLEV-
DAIVNAANASLLGGGGVDGCIHRAAGPCLLAECRNLNGCDTGHAKITCGY-
DLPAKYVIHTVGPIARGHINGSHKEDLANCYKSSLKLVKENNIRSVAFPCIST-
GIYGFPNEPAAVIALNTIKEWLAKNHHEVDRIIFCVFLEVDFKIYKKKMN-
EFFSVDDNNEEEEDVEMKEDSDENGPEEKQSVEEMEEQSQDADGVNTVTVPG-
PASEEAVEDCKDEDFAKDENITKGGEVTDHSVRDQDHPDGQENDSTKNEIKI-
ETESQSSYMETEELSSNQEDAVIVEQPEVIPLTEDQEEKEGEKAPGEDTPRMPGK-
SEGSSDLENTPGPDAGAQDEAKEQRNGTKGLNDIFEAQKIEWHE*.

hMacroD2 was expressed and purified as described previously for SARS-CoV-2 Mac1 [52].

BFV macrodomain protein purification

The BFV sequence is as follows:

MSYYHHHHHHLESTSLYKKAGFLEVLFQGPEVNSFSGYLKLAPAYRVKRGDIS-
NAPEDAVVNAANQQGVKGAGVCGAIYRKWPDAFGDVATPTGTAVSKSVQDKLVI-
HAVGPNFSKCSEEEGDRDLASAYRAAAEIVMDKKITTVAVPLLSTGIYAGGKNRVE-
QSLNHLFTAFDNTDADVTIYCMDKTWEKKIKEAIDHRT.

Cloning and expression

The macrodomain of BFV was synthesized and cloned into the pET28 vector, incorporating an N-terminal 6×His tag and a long linker before the macrodomain, resulting in the expression construct 6×His-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 -mercaptoethanol) containing one tablet of Complete, Mini EDTA-free ULTRA protease inhibitor cocktail and DNase I (10μg/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 x 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 $\times$ 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

The EEEV sequence is as follows:

MGHHHHHHHHHHENLYFQSGAPAYRVVRGDITKSNDEVIV-
NAANNKGQPGGGVCGALYRKWPGAFDKQPVATGKAHLVKHSPNVIHAVGPNF-
SRLSENEGDQKLSEVYMDIARIINNERFTKVSIPLLSTGIYAGGKDRVMQSLNHLF-
TAMDTTDADITIYCLDKQWESRIKEAI.

Cloning and expression

The macrodomain of EEEV was synthesized and cloned into the pET28 vector, incorporating an N-terminal 6×-His tag, TEV protease site, and linker before macrodomain, generating the expression as 6×His-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 OD600 reached 0.8, then protein expression was induced by the addition of 500μM 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, cell pellets were resuspended in lysis buffer, 50 mM Tris (pH 8), 500 mM NaCl, 10 mM imidazole, 5% glycerol, and 2 mM β-mercaptoethanol. Cells were then lysed by passing through chilled Emulsiflex at 10,000 psi for three cycles. The lysate was clarified by centrifugation at 35,000 x 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, 30 mM imidazole, 5% glycerol, and 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, and 2 mM βME. The eluted protein was dialyzed in 50 mM Tris, pH 8.0, 150 mM NaCl, 5% glycerol, and 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$_{\text{Pae2}}$ 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

acrIF1 naturally overlaps with a downstream gene *aca1*, a transcriptional regulator of Acr expression. The native architecture of this operon was maintained when introducing *acrIF1* into DMS3m/DMS3mvir phages. *acrIF11*$_{Pae1}$ does not co-occur with *aca1* and thus was engineered in with *aca1* downstream, but no overlap. These genes were engineered into DMS3m/DMS3mvir phages as previously described [19]. Briefly, each gene was introduced into the pHERD plasmid with homology regions flanking DMS3 gene 30. Cells containing the plasmid were infected with DMS3m/DMS3mvir, and phages isolated from this infection were then plated onto PA14, which contains a Type I-F CRISPR system to select for recombinant DMS3m/DMS3mvir. Individual plaques from this selection were purified, and the presence of the acr was verified via Sanger sequencing.

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.

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., and 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.

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; is a member of the board of Cytomx, Ovibio Corporation, and Cambridge Science Corporation; is 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; is 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, is a shareholder of, and receives sponsored research support from Relay Therapeutics.

Supplementary Material

Table 1.1: NMR assignment, structure calculation and validation statistics

Degree of assignment[a]	
Backbone (N and H^N) (%)	89.8
Side-chain H (%)	72.4
Side-chain non-H (%)	56.9
Number of restraints[b]	
NOE restraints	
Intra-residue (|i-j| = 0)	591
Sequential (|i-j| = 1)	545
Medium range (1 < |i-j| < 5)	423
Long range (|i-j| 5)	537
Ambiguous[a]	640
Total	2069
H-bond restraints	54
Long range (|i - j| 5)	40
Dihedral angle restraints (/)	150/150
Restraint statistics[c]	
r.m.s. of NOE violations (Å)	0.525 ± 0.498
r.m.s. of dihedral violations (°)	2.60 ± 2.488
r.m.s. from idealised covalent geometry[d]	
Bonds (Å)	0.0042 ± 0.0002
Angles (°)	0.63 ± 0.041
Impropers (°)	2.28 ± 0.24
Structural quality	
Ramachandran statistics[e/f]	
Most favoured regions (%)	84.4 / 90.2
Allowed regions (%)	14.9 / 8.5
Generously allowed regions (%)	0.7 / NA
Disallowed regions (%)	0.0/1.3
Verify3D Z-score[g]	-4.98
Prosa II Z-score[h]	-0.99
Procheck Z-score (/)[e]	-1.46
Procheck Z-score (all)[e]	-3.02
MolProbity Z-score[f]	-5.86
No. of close contacts[i]	11
Coordinates precision (rmsd)[b]	
All backbone atoms (Å)	2.8 / 2.3
All heavy atoms (Å)	3.3 / 2.8

Values reported by: [a]CCPNMR 2.5.2[1];[b]Protein Structure Validation Software suite 1.5 & [c]PDBStat 5.12[2]; [d]Crystallography and NMR system (CNS) 1.2 [3]; [e]Procheck[4] & [f]MolProbity[5]. The structural validation programs used were as follows: [g]Verify3D[6], [h]Prosa II[7], [e]Procheck[4], [f]MolProbity[5] and [i]PDB validation software.

Table 1.2: List of bacterial genera containing AcrIF11 homologs

Pseudomonas
Pseudoxanthomonas
Halopseudomonas
Xanthomonas
Pigmentiphaga
Marinobacterium
Enterobacter
Citrobacter
Klebsiella
Raoultella
Delftia
Paramixta
Dickeya
Cedecea
Yersinia
Pectobacterium
Brenneria
Serratia
Buttiauxella
Moraxella
Lelliottia
Escherichia
Erwinia
Actinobacillus
Haemophilus
Frischella
Rouxiella
Megasphaera
Billgrantia
Mitsuokella
Halomonas
Chromohalobacter
Sphaerochaeta
Desulfobulbus
Alcanivorax
Enterococcus
Photorhabdus

Table 1.3: 1 L M9 Minimal Media Recipe for ^{15}N labeled proteins

Component	**Amount**
10x M9 salts	100 mL
1M Magnesium sulphate	1 mL
1M Calcium chloride	100 µL
20% Thiamine	100 µL
0.003 g/mL Iron II sulphate heptahydrate	1 mL
10x MEM Vitamin mix	10 mL
20% Glucose	40 mL
15N Ammonium sulfate (15N source)	1 g
MilliQ H2O	860 mL

All solutions are dissolved in MilliQ H2O.

Table 1.4: 1 L M9 Minimal Media Recipe for ^{15}N ^{13}C labeled proteins

Component	**Amount**
10x M9 salts	100 mL
1M Magnesium sulphate	1 mL
1M Calcium chloride	100 µL
20% Thiamine	100 µL
Iron II sulphate heptahydrate	3 mg
10x MEM Vitamin mix	10 mL
10% 13C Glucose	40 mL
15N Ammonium sulfate	1 g
MilliQ H2O	860 mL

All solutions are dissolved in MilliQ H2O.

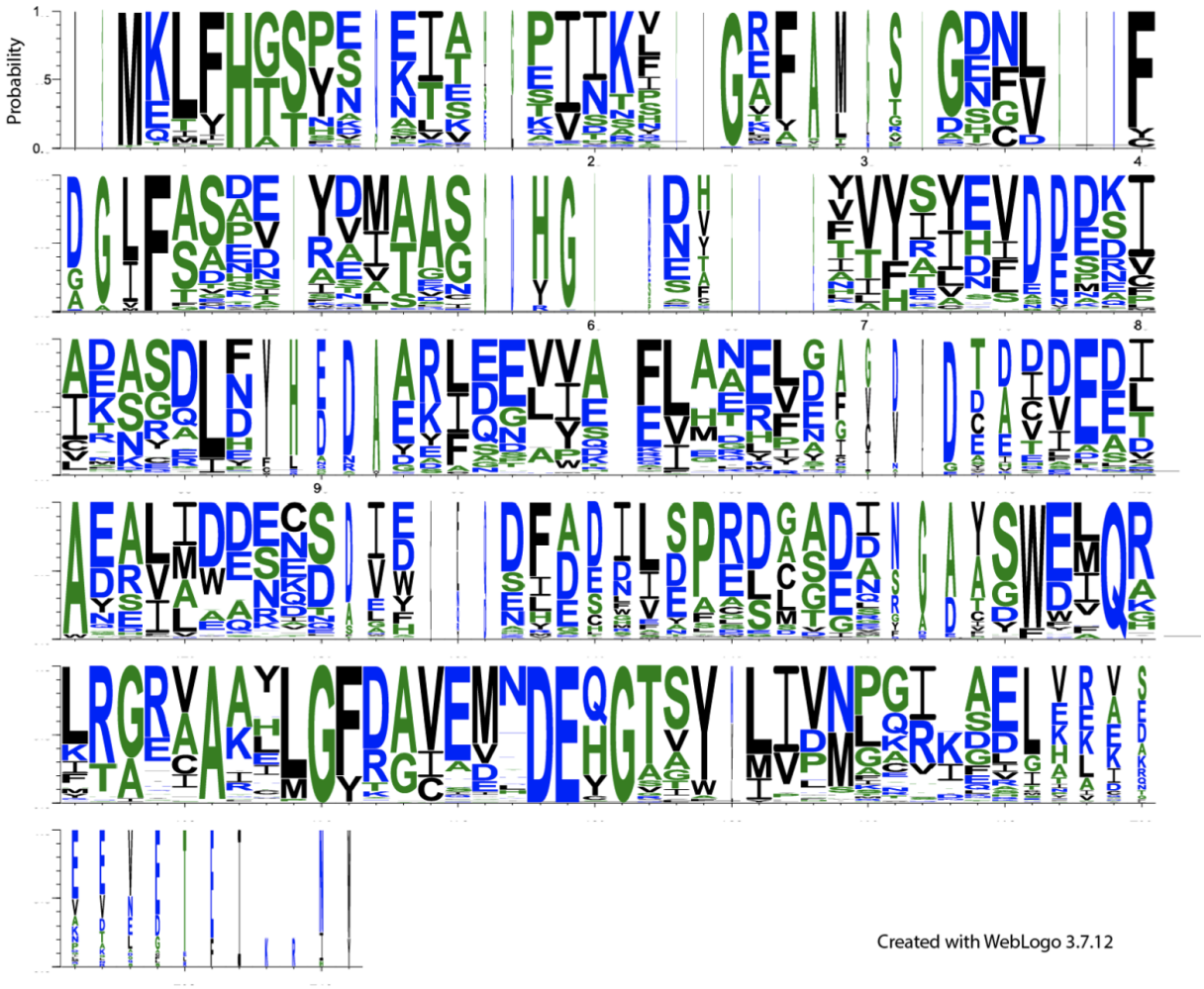

Figure 1.7: Sequence logo of AcrIF11 homologs.

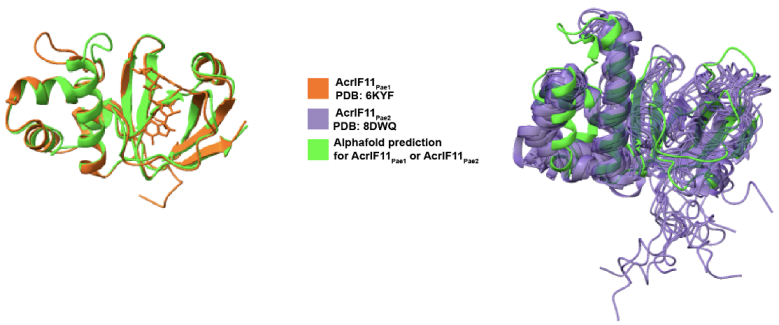

Figure 1.8: Alignment of experimental and predicted structures of AcrIF11$_{\text{Pae1}}$ and AcrIF11$_{\text{Pae2}}$

Figure 1.9: Confidence of Alphafold2 predictions for AcrIF11 homologs.

Figure 1.10: See attached file.

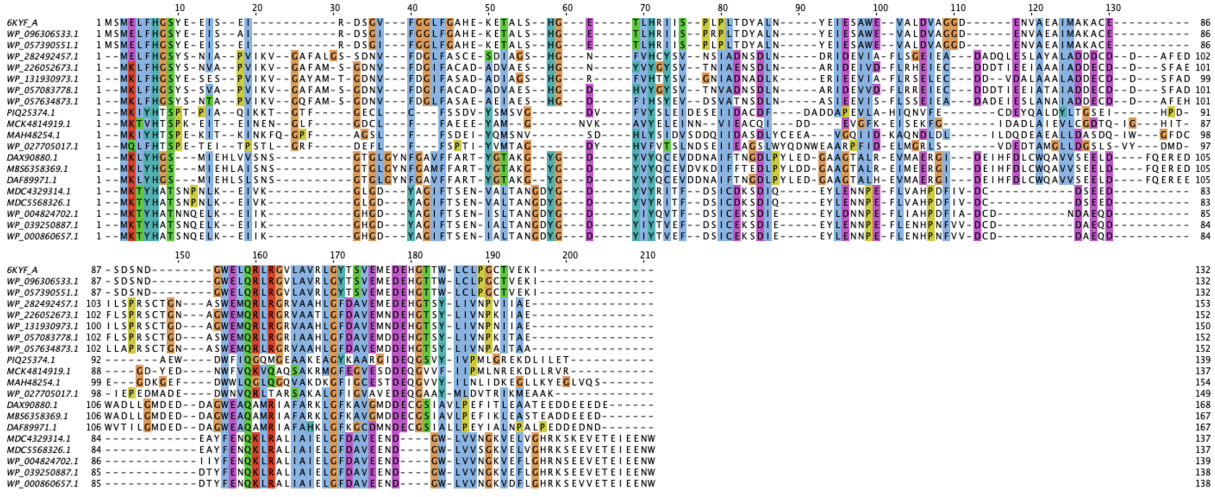

Figure 1.11: AcrIF11 phylogeny sequence alignment

Above is a diverse sampling of the sequence alignment used to build the AcrIF11 phylogeny. Phylogeny construction is discussed in the Methods section.

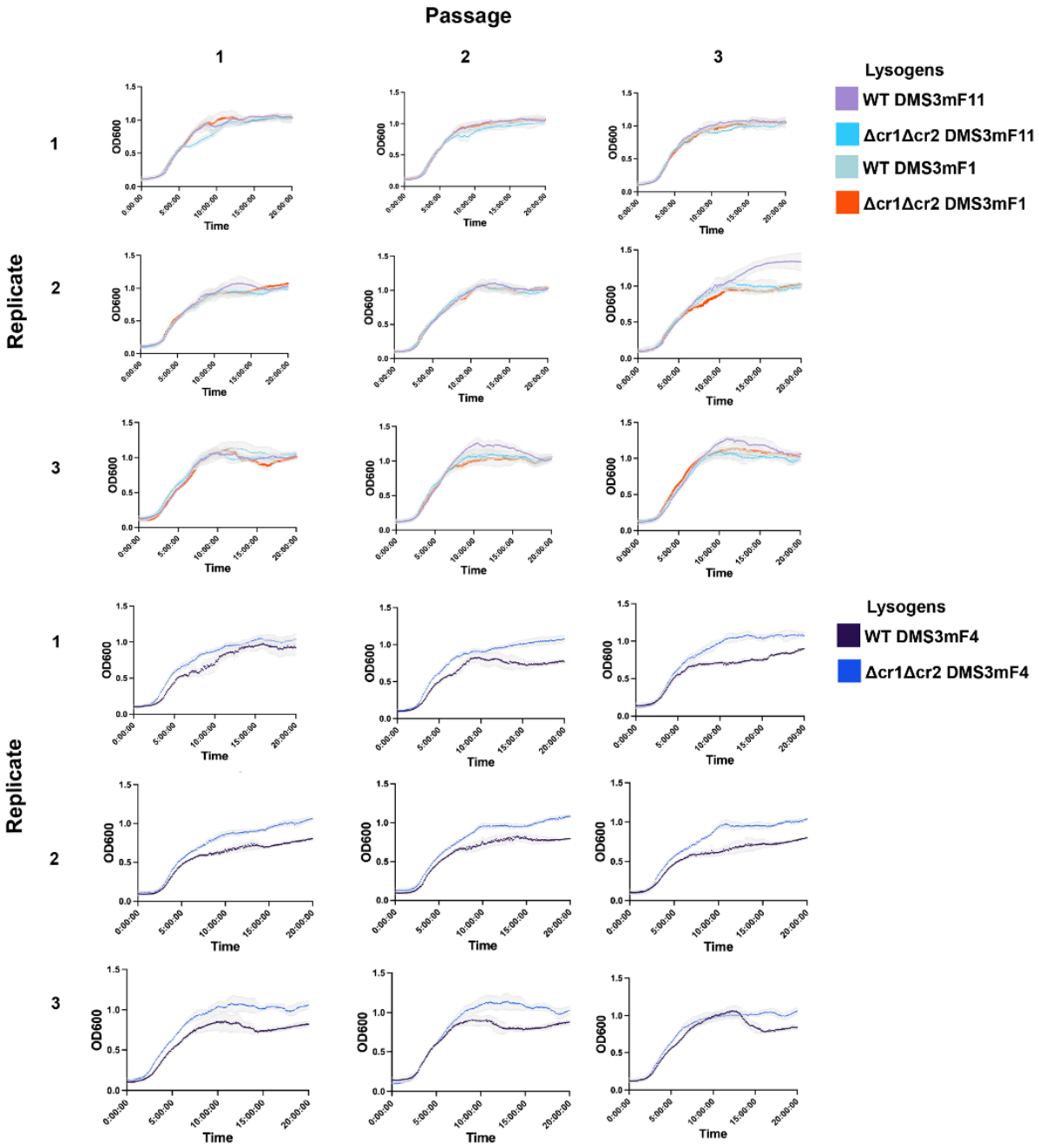

Figure 1.12: Replicates of lysogen growth experiment

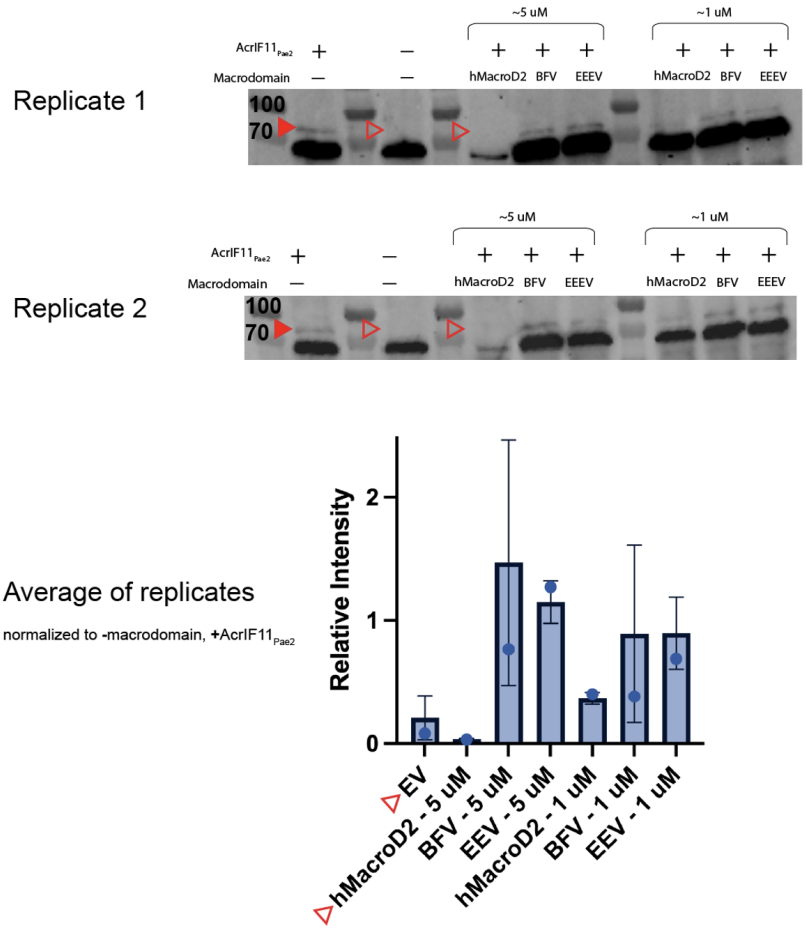

Figure 1.13: Quantification of macrodomain lysate blot

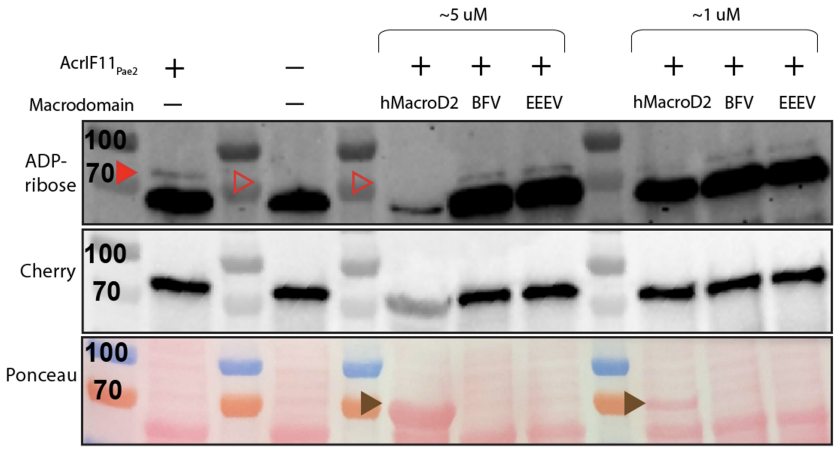

Figure 1.14: Verification of macrodomain lysate blot loading via Ponceau

All labels are the same as in Fig 5B. Brown arrow indicates hMacroD2.

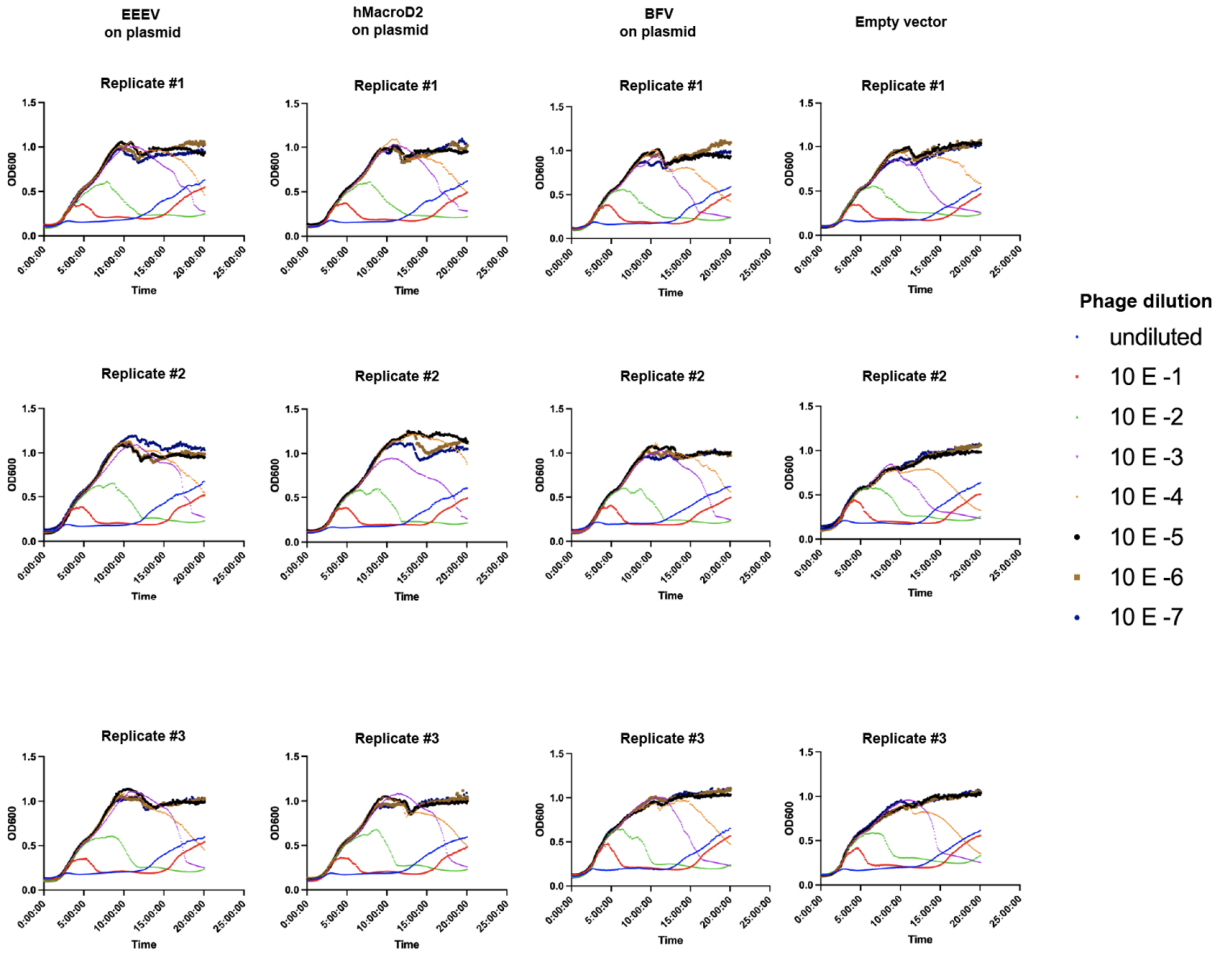

Figure 1.15: Liquid growth curves of PA14 WT overexpressing non-endogenous macrodomains

PA14 WT was infected with DMS3mF11$_{\text{Pae1}}$vir, following the lytic infection protocol listed in the Methods section above.

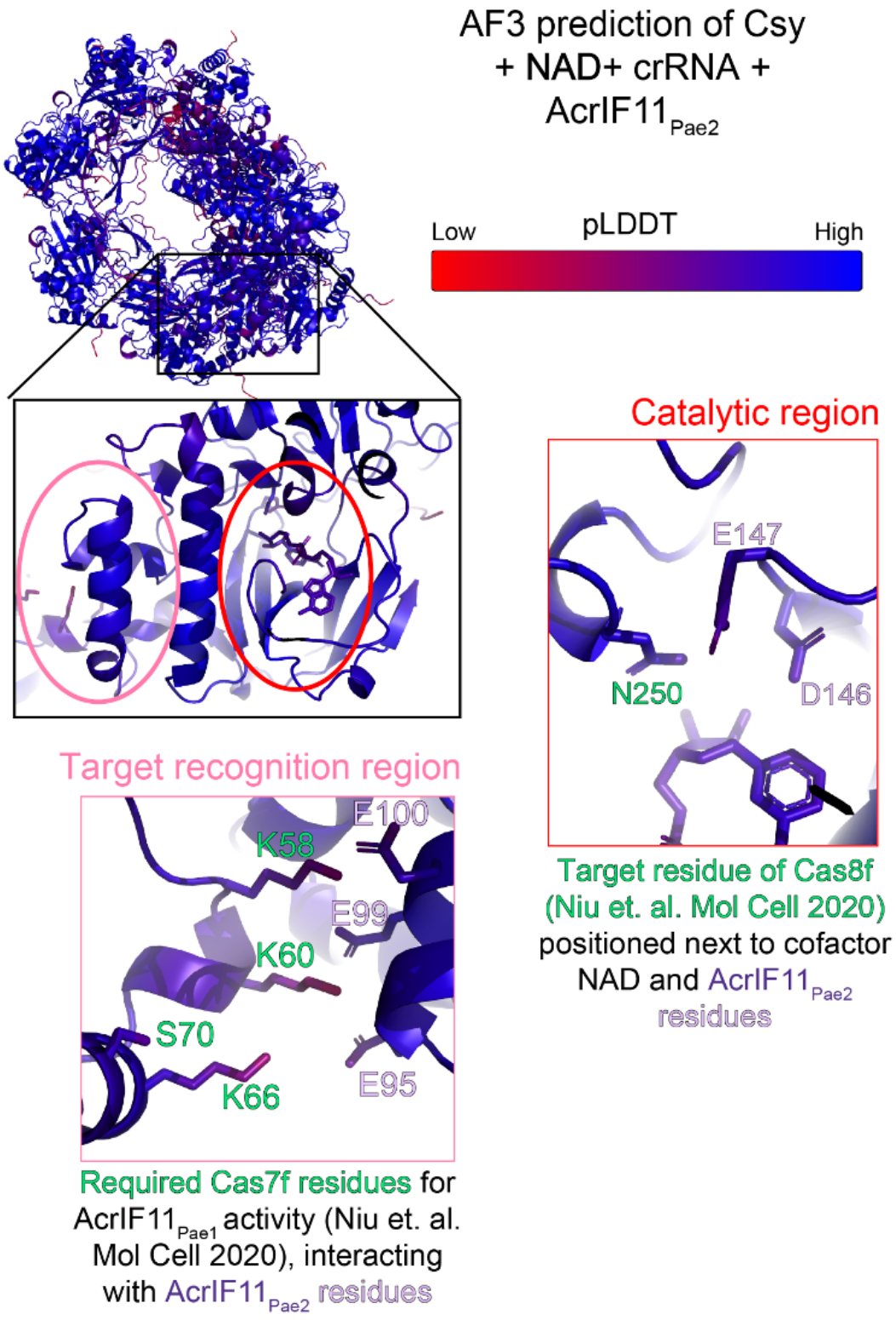

Figure 1.16: Alphafold3 prediction of Csy complex + NAD + crRNA + AcrIF11Pae2 colored by pLDDT

Supplementary Information References

[1] Vranken, Wim F. et al. "The CCPN data model for NMR spectroscopy: Development of a software pipeline". *PROTEINS: Structure, Function, and Bioinformatics* 59.687-696 (2005). DOI: https://doi.org/10.1002/prot.20449.

[2] Bhattacharya, Aneerban, Tejero, Roberto, and Montelione, Gaetano T. "Evaluating protein structures determined by structural genomics consortia". *PROTEINS: Structure, Function, and Bioinformatics* 66.778-795 (2007). DOI: https://doi.org/10.1002/prot.21165.

[3] Brunger, Axel T. "Version 1.2 of the Crystallography and NMR system". *Nature Protocols* 2.2728-2733 (2007). DOI: https://doi.org/10.1038/nprot.2007.406.

[4] Laskowski, Roman A. et al. "PROCHECK: a program to check the stereochemical quality of protein structures". *Journal of Applied Crystallography* 26.283-291 (1993). DOI: https://doi.org/10.1107/S0021889892009944.

[5] Lovell, Simon C. et al. "Structure validation by C geometry: , and C deviation". *PROTEINS: Structure, Function, and Bioinformatics* 50.437-450 (2003). DOI: https://doi.org/10.1002/prot.10286.

[6] Lüthy, Roland, Bowie, James U., and Eisenberg, David. "Assessment of protein models with three-dimensional profiles". *Nature* 356.83-85 (1992). DOI: https://doi.org/10.1038/356083a0.

[7] Sippl, Manfred J. "Recognition of errors in three-dimensional structures of proteins". *PROTEINS: Structure, Function, and Bioinformatics* 17.355-362 (1993). DOI: https://doi.org/10.1002/prot.340170404.

References

[1] Wiegand, Tanner et al. "Structures and strategies of anti-CRISPR-mediated immune suppression". *Annu Rev Microbiol* 74.21-37 (2020). DOI: https://doi.org/10.1146/annurev-micro-020518-120107.

[2] Thavalingam, Annoj et al. "Inhibition of CRISPR-Cas9 ribonucleoprotein complex assembly by anti-CRISPR AcrIIC2". *Nat Commun* 10.2806 (2019). DOI: https://doi.org/10.1038/s41467-019-10577-3.

[3] Trost, Chantel N. et al. "An anti-CRISPR that pulls apart a CRISPR–Cas complex". *Nature* 632.375–382 (2024). DOI: https://doi.org/10.1038/s41586-024-07642-3.

[4] Zhang, Heng et al. "Structural Basis for the Inhibition of CRISPR-Cas12a by Anti-CRISPR Proteins". *Cell Host & Microbe* 25.6 (2019). DOI: https://doi.org/10.1016/j.chom.2019.05.004.

[5] Mahendra, Caroline et al. "Broad-spectrum anti-CRISPR proteins facilitate horizontal gene transfer". *Nature Microbiology* 5.620-629 (2020). DOI: https://doi.org/10.1038/s41564-020-0692-2.

[6] Wang, Xiaoshen et al. "Inhibition mechanisms of CRISPR-Cas9 by AcrIIA17 and AcrIIA18". *Nucleic Acids Research* 50.512-521 (2021). DOI: https://doi.org/10.1093/nar/gkab1197.

[7] Athukoralage, Januka S. et al. "An anti-CRISPR viral ring nuclease subverts type III CRISPR immunity". *Nature* 577.572-575 (2020). DOI: https://doi.org/10.1038/s41586-019-1909-5.

[8] Dong, Liyong et al. "An anti-CRISPR protein disables type V Cas12a by acetylation". *Nature Structural & Molecular Biology* 26.308-314 (2019). DOI: https://doi.org/10.1038/s41594-019-0206-1.

[9] Kang, Xiaoman et al. “Reversible regulation of Cas12a activities by AcrVA5-mediated acetylation and CobB-mediated deacetylation”. *Cell Discovery* 8.45 (2022). DOI: https://doi.org/10.1038/s41421-022-00396-0.

[10] Niu, Yiying et al. “A type I-F anti-CRISPR protein inhibits the CRISPR-Cas surveillance complex by ADP-ribosylation.” *Molecular Cell* 80.512-524 (2020). DOI: https://doi.org/10.1016/j.molcel.2020.09.015.

[11] Marino, Nicole D. et al. “Discovery of widespread type I and type V CRISPR-Cas inhibitors”. *Science* 362.240-242 (2018). DOI: https://doi.org/10.1126/science.aau5174.

[12] Rollie, Clare et al. “Targeting of temperate phages drives loss of type I CRISPR-Cas systems”. *Nature* 578.149-153 (2020). DOI: https://doi.org/10.1038/s41586-020-1936-2.

[13] Parikh, Sapan L. and Schramm, Vern L. “Transition state structure for ADP-ribosylation of eukaryotic elongation factor 2 catalyzed by diphtheria toxin”. *Biochemistry* 43.1204-1212 (2004). DOI: https://doi.org/10.1021/bi035907z.

[14] Jørgensen, René et al. “The nature and character of the transition state for the ADP-ribosyltransferase reaction”. *EMBO Reports* 9.802-809 (2008). DOI: https://doi.org/10.1038/embor.2008.90.

[15] Ménétrey, Julie et al. “NAD binding induces conformational changes in Rho ADP-ribosylating clostridium botulinum C3 exoenzyme.” *Journal of Biological Chemistry* 277.30950-30957 (2002). DOI: https://doi.org/10.1074/jbc.M201844200.

[16] Camargo, Antonio Pedro et al. “Identification of mobile genetic elements with geNomad”. *Nature Biotechnology* 42.1303-1312 (2024). DOI: https://doi.org/10.1038/s41587-023-01953-y.

[17] Jumper, John et al. “Highly accurate protein structure prediction with AlphaFold”. *Nature* 596.583-589 (2021). DOI: https://doi.org/10.1038/s41586-021-03819-2.

[18] Govindarajan, Sutharsan et al. “Distinct subcellular localization of a type I CRISPR complex and the Cas3 nuclease in bacteria.” *Journal of Bacteriology* 204.e00105-22 (2022). DOI: https://doi.org/10.1128/jb.00105-22.

[19] Borges, Adair L. et al. “Bacteriophage cooperation suppresses CRISPR-Cas3 and Cas9 immunity”. *Cell* 174.917-925 (2018). DOI: https://doi.org/10.1016/j.cell.2018.06.013.

[20] Bondy-Denomy, Joesph et al. “Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system”. *Nature* 493.429-432 (2013). DOI: https://doi.org/10.1038/nature11723.

[21] Landsberger, Mariann et al. “Anti-CRISPR phages cooperate to overcome CRISPR-Cas immunity”. *Cell* 174.908-916 (2018). DOI: https://doi.org/10.1016/j.cell.2018.05.058.

[22] Osuna, Beatriz A. et al. “Critical anti-CRISPR locus repression by a bi-functional Cas9 inhibitor”. *Cell Host & Microbe* 28.23-30 (2020). DOI: https://doi.org/10.1016/j.chom.2020.04.002.

[23] Workman, Rachael E. et al. “Anti-CRISPR proteins trigger a burst of CRISPR-Cas9 expression that enhances phage defense”. *Cell Reports* 43.113849 (2024). DOI: https://doi.org/10.1016/j.celrep.2024.113849.

[24] Borges, Adair L. et al. “Bacterial alginate regulators and phage homologs repress CRISPR-Cas immunity”. *Nature Microbiology* 5.679-687 (2020). DOI: https://doi.org/10.1038/s41564-020-0691-3.

[25] Jankevicius, Gytis et al. “The toxin-antitoxin system DarTG catalyzes reversible ADP-ribosylation of DNA”. *Molecular Cell* 64.1109–1116 (2016). DOI: https://doi.org/10.1016/j.molcel.2016.11.014.

[26] LeRoux, Michele et al. “The DarTG toxin-antitoxin system provides phage defence by ADP-ribosylating viral DNA”. *Nature Microbiology* 7.1028–1040 (2022). DOI: https://doi.org/10.1038/s41564-022-01153-5.

[27] Ka, Donghyun et al. “Structural and functional evidence of bacterial antiphage protection by Thoeris defense system via NAD+ degradation”. *Nature Communications* 11.2816 (2020). DOI: https://doi.org/10.1038/s41467-020-16703-w.

[28] Zhang, Yang and Skolnick, Jeffrey. “Scoring function for automated assessment of protein structure template quality”. *Proteins* 57.702-710 (2004). DOI: https://doi.org/10.1002/prot.20264.

[29] Xu, Jinrui and Zhang, Yang. “How significant is a protein structure similarity with TM-score = 0.5?” *Bioinformatics* 26.889-895 (2010). DOI: https://doi.org/10.1093/bioinformatics/btq066.

[30] Pinto, Ana Filipa and Schüler, Herwig. *Endogenous ADP-Ribosylation*. Current Topics in Microbiology and Immunology. Springer Cham, 2015. Chap. Comparative Structural Analysis of the Putative Mono-ADP-Ribosyltransferases of the ARTD/PARP Family.

[31] Cohen, Michael S and Chang, Paul. “Insights into the biogenesis, function, and regulation of ADP-ribosylation”. *Nature Chemical Biology* 14.236-243 (2018). DOI: https://doi.org/10.1038/nchembio.2568.

[32] Duan, Ning et al. “Structure-guided discovery of anti-CRISPR and anti-phage defense proteins”. *Nature Communications* 15.649 (2024). DOI: https://doi.org/10.1038/s41467-024-45068-7.

[33] Cihlova, Bara et al. “Specificity of DNA ADP-ribosylation reversal by NADARs”. *Toxins* 16.208 (2024). DOI: https://doi.org/10.3390/toxins16050208.

[34] Chowdhury, Saikat et al. “Structure reveals mechanisms of viral suppressors that intercept a CRISPR RNA-guided surveillance complex”. *Cell* 169.47-57 (2017). DOI: https://doi.org/10.1016/j.cell.2017.03.012.

[35] Tiemann, Bernd et al. “ModA and ModB, two ADP-ribosyltransferases encoded by bacteriophage T4: catalytic properties and mutation analysis”. *Journal of Bacteriology* 186.7262-7272 (2004). DOI: https://doi.org/10.1128/jb.186.21.7262-7272.2004.

[36] Cavanagh, John et al. *Protein NMR spectroscopy principles and practice*. 2nd ed. Elsevier Academic Press, 2006.

[37] Findeisen, M., Brand, T., and Berger, S. “A 1H-NMR thermometer suitable for cryoprobes”. *Magnetic Resonance in Chemistry* 45.175-178 (2007). DOI: https://doi.org/10.1002/mrc.1941.

[38] Wishart, David S. et al. “1H, 13C and 15N chemical shift referencing in biomolecular NMR”. *Journal of Biomolecular NMR* 6.135-140 (1995).

[39] Vranken, Wim F. et al. “The CCPN data model for NMR spectroscopy: development of a software pipeline”. *Proteins* 59.687-696 (2005). DOI: https://doi.org/10.1002/prot.20449.

[40] Cheung, Ming-Sin et al. “DANGLE: a Bayesian inferential method for predicting protein backbone dihedral angles and secondary structure”. *Journal of Magnetic Resonance* 202.223-233 (2010). DOI: https://doi.org/10.1016/j.jmr.2009.11.008.

[41] Rieping, Wolfgang et al. “ARIA2: automated NOE assignment and data integration in NMR structure calculation”. *Bioinformatics* 23.381-382 (2007). DOI: https://doi.org/10.1093/bioinformatics/btl589.

[42] Brünger, Axel T. et al. “Crystallography & NMR system: a new software suite for macromolecular structure determination”. *Acta Crystallographica Section D* 54.905-921 (1998). DOI: https://doi.org/10.1107/S0907444998003254.

[43] Hwang, Tsang-Lin, Zijl, Peter C.M. van, and Mori, Susumu. “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”. *Journal of Biomolecular NMR* 11.221-226 (1998).

[44] Katoh, Kazutaka et al. “MAFFT: a novel method for rapid multiple sequence alignment based on fast fourier transform”. *Nucleic Acids Research* 30.3059-3066 (2002). DOI: https://doi.org/10.1093/nar/gkf436.

[45] Price, Morgan N., Dehal, Paramvir S., and Arkin, Adam P. “FastTree: computing large minimum evolution trees with profiles instead of a distance matrix”. *Molecular Biology and Evolution* 26.1641-1650 (2009). DOI: https://doi.org/10.1093/molbev/msp077.

[46] Letunic, Ivica and Bork, Peer. “Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation”. *Bioinformatics* 23.127-128 (2007). DOI: https://doi.org/10.1093/bioinformatics/btl529.

[47] Bondy-Denomy, Joseph et al. “A unified resource for tracking Anti-CRISPR names”. *The CRISPR Journal* 1.304-305 (2018). DOI: https://doi.org/10.1089/crispr.2018.0043.

[48] Steinegger, Martin and Söding, Johannes. “MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets”. *Nature Biotechnology* 35.1026-1028 (2017). DOI: https://doi.org/10.1038/nbt.3988.

[49] Meng, Elaine C. et al. “UCSF ChimeraX: tools for structure building and analysis”. *Protein Science* 32.e4792 (2023). DOI: https://doi.org/10.1002/pro.4792.

[50] Guo, Tai Wei et al. “Cryo-EM structures reveal mechanism and inhibition of DNA targeting by a CRISPR-Cas surveillance complex”. *Cell* 171.414-426 (2017). DOI: https://doi.org/10.1016/j.cell.2017.09.006.

[51] Weixler, Lisa et al. “Protein and RNA ADP-ribosylation detection is influenced by sample preparation and reagents used”. *Life Science Alliance* 6.e202201455 (2023). DOI: http://doi.org/10.26508/lsa.202201455.

[52] Schuller, Marion et al. “Fragment binding to the Nsp3 macrodomain of SARS-CoV-2 identified through crystallographic screening and computational docking”. *Science Advances* 7.eabf8711 (2021). DOI: https://doi.org/10.1126/sciadv.abf8711.

Chapter 2. Biochemical and structural characterization of jumbophage nuclear import factors.

Introduction

Bacteriophages and their hosts are under high selective pressure to evade and outsmart each other's defenses. Over time, this evolutionary arms race has spawned enormous mechanistic diversity in immune systems and anti-immune factors. Of the bacterial immune systems discovered so far, the most abundant ones degrade nucleic acid, with prominent examples being CRISPR-Cas and restriction-modification [1]. This has led to the evolution of phage-encoded anti-immune factors that disable nucleic acid-targeting systems, such as anti-CRISPR (Acr) and overcoming classical restriction (Ocr) proteins [2]. However, these anti-immune factors are generally highly specific to a particular bacterial immune system.

Phages in the ϕKZ-like jumbophage family can create a proteinaceous, nucleus-like compartment inside their host during infection, sequestering their genome from the rest of the host's intracellular environment. This allows the phage genome to be broadly protected from DNA-targeting immune systems such as CRISPR-Cas and restriction-modification [3], instead of encoding several anti-immune proteins that individually confer protection against a particular system. The nucleus is crucial to the infection cycle, allowing for protected genome replication and a central location for virion packaging [4]. However, in order for the nucleus to function properly, it must selectively import proteins important for the phage lifecycle while also excluding host immune nucleases. Nuclear protein import has been unexplored until recently when Kokontis et. al. identified import factors in ϕKZ using an unbiased genetic selection. Escaper phages capable of reducing import of the toxified cargo had mutations in a central import gene *imp1*, with certain cargo requiring other import factors such *imp2-5* [5]. Imp1 also formed a direct physical interaction with a non-specific RNA binding protein called ChmC or Imp6. [5]

From the genetic selection, Kokontis et. al. noted that the pattern of escape mutations in *imp1* suggested unique interfaces for each cargo [5]. However, the physical mechanism of transport remains uncharacterized. As a step towards answering the question of how cargo is transported into the nucleus, we aim to understand how Imp1 interacts with other import factors and cargo. Solving a structure of an Imp1-containing import complex would allow us to understand the stoichiometry of cargo transport, and how such interactions allow for the specificity observed from escape mutations. Here, we present preliminary attempts to purify an Imp1-Imp6 complex and structurally characterize it via cryoEM.

During middle and late infection, ϕKZ transcripts are produced by the non-virion RNA polymerase (nvRNAP), which must be imported into the nucleus to perform its function. The presence of nvRNAP, in addition to its crucial role in producing middle and late transcripts, allows ϕKZ to continue infection even when rifampicin inhibits host RNAP [6], highlighting an unusual aspect of jumbophage biology. A toxified cargo selection, similar to the one outlined in Kokontis et. al., identified *imp7*, a previously uncharacterized gene that is required for nvRNAP import into the nucleus. Selection for factors required for Imp7 to localize into the nucleus revealed mutations in phage gene *imp1*. Escape mutations were also present in the nvRNAP genes themselves, suggesting that nvRNAP could be imported as a complete complex instead of individual subunits. Here, we present preliminary data showing that nvRNAP does not interact directly with Imp7 and other potential interactors in a heterologous expression system.

Altogether, the nucleus of ϕKZ is a potent anti-immune barrier in the phage-bacteria evolutionary arms race. However, compartmentalization of the phage genome presents a challenge for localizing genome-associated proteins. Kokontis et. al. identified key import factors responsible for import of proteins into the phage nucleus, but the physical mechanism of transport remains unknown. Our attempts at biochemical and structural characterization of Imp1-Imp6 and nvRNAP interactors via cryoEM could elucidate interactions to help us understand the specificity and mechanism of import, as well as rationalize the escaper

mutations observed from the genetic selection.

Results

Imp6 strongly associates with Imp1 during protein purification

Due to previous work from Kokontis et. al. showing a stable binding interaction between Imp1 and Imp6 [5], we co-expressed Imp1 and Imp6 together for structural characterization. We expected Imp1 and Imp6 to form a heterotrimer of approximately 120 kD, according to previous mass photometry experiments [5]. Our initial purification scheme consisted of nickel affinity purification followed by size exclusion chromatography (SEC), out of which Imp1 and Imp6 were able to be expressed and purified to an acceptable degree for cryoEM screening. CryoEM screening of this sample showed distinct and defined particles, but after processing a larger dataset, the 2D classes were strongly reminiscent of predicted Imp6 homotetramer structures (Fig 2.1A). Imp6, also known as ChmC, has been previously characterized as a non-specific RNA binding protein and putative mRNA exporter and confirmed to be a tetramer by SEC-MALS [7]. Alphafold3 [8] also predicts a high confidence homotetramer structure (Fig 2.1B). Our purification of predominantly Imp6 homotetramer instead of Imp1-Imp6 heterotrimer was unexpected because the nickel affinity purification should have selected for only His-tagged Imp1. We reasoned that the Imp1-Imp6 complex was less stable than the Imp6 homotetramer, allowing for Imp6 to dissociate from Imp1 and homooligomerize. Furthermore, because the Imp6 ($\sim$30 kDa) homotetramer and Imp6 ($\sim$30 kDa):Imp1 ($\sim$60 kDa) heterotrimer are similar sizes, they could not be separated via SEC. We hypothesized that the overall charges of the complexes were different enough to separate them via ion exchange chromatography (IEX). Although IEX yielded fractions with slightly different stoichiometries of Imp1:Imp6, as seen by a difference of intensity on a Coomassie gel (Fig 2.1C), the resulting 2D classes from data collection were still highly reminiscent of Imp6 homotetramer (Fig 2.1D).

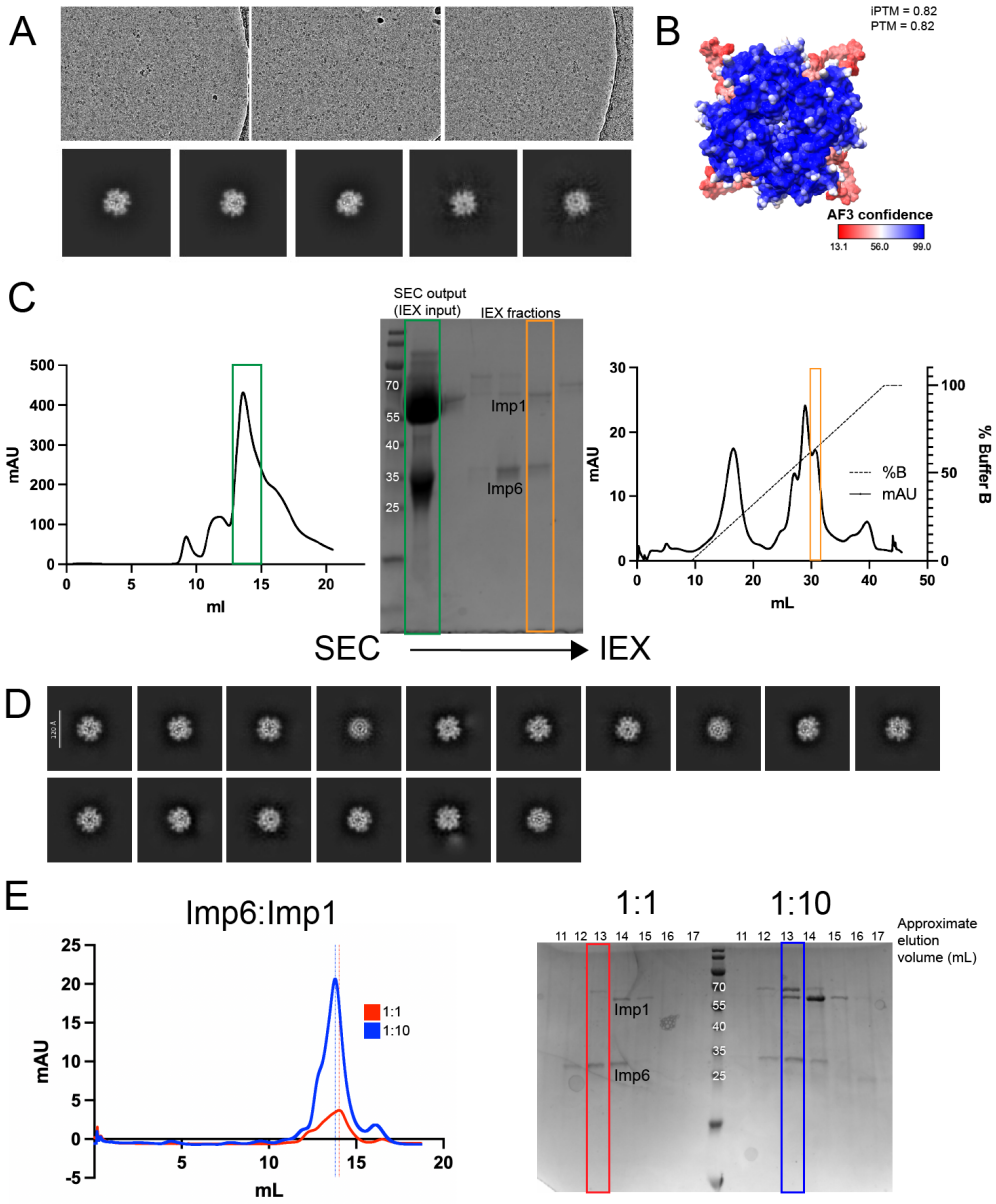

Figure 2.1: Imp6 strongly associates with Imp1 during protein purification. (A) Micrographs and 2D classes from nickel affinity purification followed by size exclusion chromatography of the Imp1-Imp6 coexpression strain. (B) AF3 structure prediction of Imp6 homotetramer, colored by confidence. (C) Chromatograms and Coomassie-stained SDS-PAGE gel of an attempted Imp1-Imp6 copurification optimization using nickel affinity purification, then size exclusion chromatography, and followed by ion exchange chromatography as the final step. (D) 2D classes from a cryoEM data collection on the sample outlined in orange in panel C. (E) Titration of Imp1 into Imp6 sample, and investigation via size exclusion chromatography and SDS-PAGE gel of size exclusion chromatography fractions (n=1). The red and blue boxes indicate the fraction where Imp1 is not present in 1:1 but present in the 1:10 titration.

Attempts to generate a 3D volume from the 2D classes, as well as 3D classification, resulted in volumes that were too low resolution to distinguish parts of the complex. *In vitro* complex formation by individually purifying Imp1 and Imp6, incubating together, and then using SEC to detect complex formation resulted in a slight shift of Imp1 to an earlier fraction, but stoichiometry of the fraction as seen on SDS-PAGE (Fig 2.1E, boxes) was similar to what had been observed previously from coexpression and purification leading to Imp6 homotetramerization (Fig 2.1C). Based on these attempts, future work could focus on different methods of complex stabilization such as chemical crosslinking, alternative methods of purification, or coexpression with a cargo protein or other import factors.

nvRNAP does not stably associate with import and adapter proteins

In addition to investigating interactions in the central import pathway, we also investigated the import of one cargo in particular: nvRNAP. Toxified genetic selections similar to the screen in Kokontis et. al., revealed an nvRNAP-specific import factor, Imp7. We tested for stable binding interactions between nvRNAP and Imp7 via co-immunoprecipitation from a heterologous *E. coli* expression system, but were unable to detect an interaction (Fig 2.2A). We also tested for stable binding to the central import factor Imp1, but were unable to detect an interaction (Fig 2.2B). One hypothesis for nvRNAP function is that it is not imported like other cargo, but rather plays a role in genome transfer from the EPI vesicle [9] to the nascent/premature nucleus. Because of this hypothesis, we also tested for an interaction with the major nuclear shell protein ChmA, but were unable to detect an interaction (Fig 2.2C). The apparent lack of stable interactions with individual potential interactors led us to wonder if multiple potential interactors should be combined as opposed to investigating interactions with nvRNAP one-by-one. Mixing purified nvRNAP (either the full 5-subunit complex or 4-subunit complex without the gp68 sigma factor), Imp7, and Imp1 together and investigating coelution of these proteins yielded a potential interaction between 4s nvRNAP and Imp1.

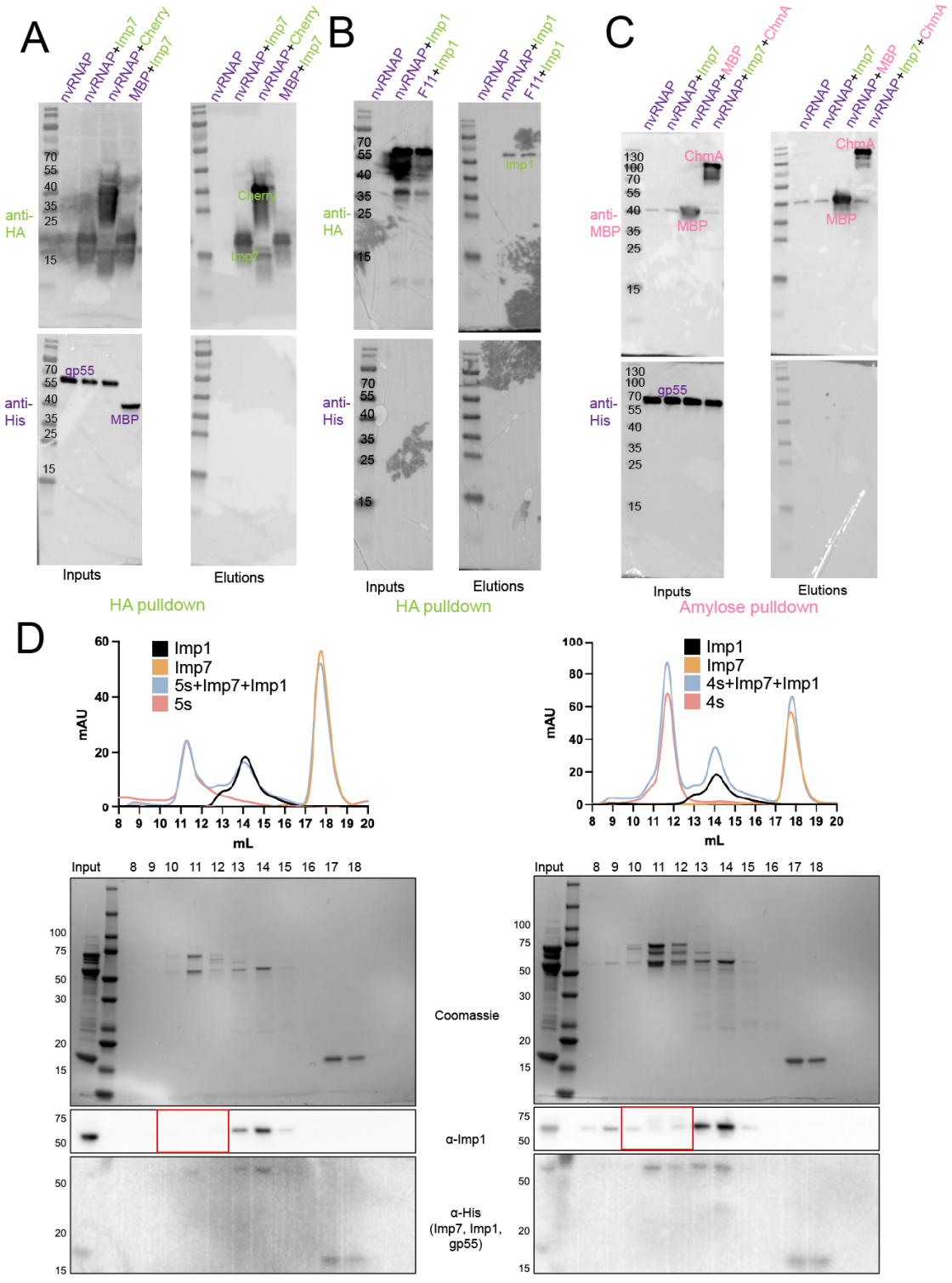

Figure 2.2: nvRNAP does not stably associate with import and adapter proteins (A) Co-immunoprecipitation via pulldown of HA-tagged Imp7. (Figure caption continued on the next page)

(Figure caption continued from the previous page) HA-Cherry (third lane) served as a nonspecific HA binding control, while His-MBP (fourth lane) served as a nonspecific His binding control. Gp55, a subunit of nvRNAP, is His-tagged. His-tagged proteins are labeled in purple, and HA-tagged proteins are labeled in green. (B) Co-immunoprecipitation via pulldown of HA-tagged Imp1 (n=1). His-AcrIF11 (third lane) served as a nonspecific His binding control. (C) Co-immunoprecipitation via amylose pulldown of MBP-tagged ChmA (n=1). The MBP-ChmA construct was coexpressed on a pDuet plasmid with HA-Imp7, so HA-Imp7 alone (second lane) served as a negative binding control. MBP (third lane) served as a positive binding control. His-tagged proteins are labeled in purple, and MBP-tagged proteins are labeled in pink. (D) Comparison of 5-subunit/4-subunit nvRNAP + Imp7 + Imp1 mixing and investigation via size exclusion chromatography and SDS-PAGE gel of SEC fractions (n=1). 5s/4s:Imp7:Imp1 was mixed at a ratio of 1:10:5. SEC curves were plotted after subtracting baseline UV background from each curve, which varied with each run. Experiment done by I.Z. under supervision of D.F.C and C.K.

This potential interaction was not immediately obvious by visual inspection of SEC curves, but could be visualized by Western blotting for Imp1 in SEC fractions: in the 4s + Imp7 + Imp1 mixture, Imp1 can be detected in the same fractions as nvRNAP, whereas it cannot for the 5s + Imp7 + Imp1 mixture (Fig 2.2D, red boxes). These SEC experiments also recapitulate earlier observations in panels A,B,and C that 5s nvRNAP does not stably interact with Imp1, nor Imp7. Reproducing this 4s-Imp1 interaction, and focusing on multi-partner nvRNAP interactions more broadly, should be the focus of future investigations.

Discussion

Initial attempts at structural characterization of ϕKZ import-related proteins have been met with technical obstacles. Although Imp1 and Imp6 were previously shown to interact by co-immunoprecipitation [5], the strong tendency of Imp6 to homooligomerize presents a challenge when attempting to isolate the Imp1-Imp6 complex. Separation by affinity purification, SEC, and ion exchange chromatography was unsuccessful. Other methods could be incorporated into future purification workflows, such as a reverse purification where tagged Imp6 homotetramer is captured on resin while leaving Imp1-Imp6 heterotrimer in flowthrough. Imp1-Imp6 chemical crosslinking could also be attempted to stabilize the complex. Alternatively, our inability to stabilize Imp1-Imp6 *in vitro* could suggest that another binding partner (i.e. Imp3, ChmA) is needed to stabilize the complex. A biologically relevant protein to introduce into our coexpression setup could be an imported cargo protein, which presumably must interact with Imp1 and Imp6 for import.

nvRNAP is a particular imported protein complex of interest, due to its crucial function of producing middle and late infection transcripts. Using a genetic selection with toxified cargo, the import of nvRNAP has been found to require Imp7, and import of Imp7 requires Imp1. However, our experiments show that nvRNAP and Imp7 do not stably interact; nvRNAP also does not stably interact with Imp1 or ChmA. The absence of stable interaction between nvRNAP and single interactors could imply that multiple interactors need to be present for stable association with nvRNAP. Indeed, proteomics studies after phage infection have shown that nvRNAP of ϕPA3 interacts with ChmA [10], suggesting that there could be something missing from our heterologous expression setup but present in the native cellular environment of *Pseudomonas* that is facilitating nvRNAP interaction. Additionally, although the 5-subunit version of nvRNAP is the active form, there also exists a 4-subunit complex that has been previously observed to be stable in vitro, which lacks the gp68 sigma factor[11, 12]. Interactions between the 4-subunit nvRNAP and potential interactors should be investigated

more systematically because it is currently unknown which version of nvRNAP gets imported, or if both versions can be imported. Further investigation into nvRNAP's nucleic acid interaction could be helpful as well; perhaps import depends on nucleic acid binding state.

Methods

Imp1-Imp6 large scale protein expression & purification

Small scale starter cultures were diluted 1:100 into large scale LB + antibiotic cultures. Large scale cultures were grown at 37°C until OD 0.4 - 0.6, and induced with 1 mM IPTG at 18°C for 16-18 hours. Cells were pelleted at 3,500g for 20 minutes, flash frozen in liquid nitrogen, and stored at -80°C for future use. To lyse, cell pellets were thawed and resuspended in 20 mM Tris pH 8, 300 mM NaCl, 20 mM imidazole supplemented with 0.5 mM TCEP, 1 protease inhibitor tablet, 25 U/mL Pierce Universal Nuclease. Cell resuspensions were sonicated for 30s on 30s off per round, with a total sonication on time for 5 minutes. Lysates were clarified at 30,000g for 30 minutes at 4°C. Clarified lysates were incubated with Ni-NTA resin (1 mL bed volume per 1 liter of culture lysed) at 4°C for 1 hr, shaking. The Ni resin was washed with approximately 35 bed volumes of 20 mM Tris pH 8, 300 mM NaCl, 40 mM imidazole, 0.5 mM TCEP. Approximately 6 bed volumes were eluted with 20 mM Tris pH 8, 300 mM NaCl, 400 mM imidazole, 0.5 mM TCEP, concentrated with an Amicon 3K MWCO centrifugal concentrator, and loaded onto the Superdex 200 Increase 10/300 GL. The size exclusion buffer used was 20 mM Tris pH 8, 0.5 mM TCEP, 100 mM NaCl. Sizing fractions were examined via SDS-PAGE, and the appropriate fractions were further concentrated with an Amicon 3K MWCO centrifugal concentrator before loading onto the PE Source 15Q for ion exchange chromatography. Buffer A was the previously mentioned size exclusion buffer, and Buffer B was 20 mM Tris pH 8, 0.5 mM TCEP, 300 mM NaCl. Ion exchange fractions were examined by SDS-PAGE for a relatively higher amount of Imp1 compared to Imp6, and the relevant fractions were concentrated in an Amicon 3K MWCO centrifugal concentrator,

then flash frozen in liquid nitrogen for future use. Protein concentration was determined via Bradford assay. Purification of Imp1 and Imp6 for size exclusion chromatography titrations followed a similar protocol, but without the ion exchange step.

Imp1-Imp6 and nvRNAP-Imp1-Imp7 titrations

Purified Imp1 and Imp6 were mixed at either 2μM:2μM or 2μM:20μM ratios of Imp6:Imp1. Purified nvRNAP, Imp7, and Imp1 were mixed at 1:10:5 ratios of 5s/4s:Imp7:Imp1. The appropriate size exclusion buffers for each protein, listed under "Imp1-Imp6 large scale protein expression & purification" and "nvRNAP large scale protein expression & purification" was used for diluting reactions to a fixed volume. Reactions were incubated on ice for 30 minutes before loading onto the Superdex 200 Increase 10/300 GL.

Imp1-Imp6 cryo-EM sample preparation

Graphene oxide (GO) grids were prepared from Quantifoil Au400 1.2/1.3 grids. Grids were glow discharged using the PELCO easiGlow at 15 mA, 30s hold, 30s glow. Home-made GO flakes were diluted using a 5:1 v/v methanol:H2O solution, with 8-10 uL of GO flakes added to 292-290 μL of the methanol:H_2O solution. Glow-discharged grids were submerged in milliQ H_2O, and 200 μL of the GO flake dilution was added to evenly coat the surface of the milliQ water. GO flakes were deposited on the grids by pumping water out, and grids were left at room temperature overnight to dry. Grids were then stored at 4°C for future use.

The appropriate fractions from ion exchange chromatography were diluted using Buffer B to approximately 0.1 mg/mL, and 3μL of this diluted sample was added to GO grids. Grids were frozen using a FEI Vitrobot Mark IV. During freezing, the Vitrobot was at 4-7°C with 94-100% humidity. Grids were blotted with blot force 0, wait time 30s, blot time 6s.

Imp1-Imp6 cryo-EM data collection

His purification-IEX-SEC sample: data was collected at the UCSF EM Core Facility on a Thermo Scientific Glacios 200kV transmission electron microscope at a nominal magnification of 54,000x, pixel size 0.73Å, dose rate 30 $e^-/Å^2/s$, total exposure time 2s, and total dose 60 $e^-/Å^2$.

His purification-SEC sample: data was collected at the UCSF EM Core Facility on a Thermo Scientific Arctica 200kV transmission electron microscope at a nominal magnification of 45,000x, pixel size 0.865Å, dose rate 21 $e^-/Å^2/s$, total exposure time 3s, and total dose 64 $e^-/Å^2$.

Both datasets: Multiple images were collected from each hole through the use of fringe-free illumination and image shift. Automated data collection was conducted using SerialEM software [13]. Movies were aligned, motion-corrected, and dose-weighted in MotionCor [14] as part of on-the-fly data processing in the UCSF EM Core Facility.

Imp1-Imp6 cryo-EM data processing

Both datasets: Dose-weighted micrographs from on-the-fly processing were imported into Cryosparc [15]. CTF estimation was done using CTFFIND4 [16] through Cryosparc.

His purification-IEX-SEC dataset : Micrographs with resolution worse than 6Å(low quality image and/or crystalline ice) and a CTF fit cross-correlation higher than 0.4 (off-hole micrographs) were discarded, leaving 3,945 micrographs. Particles were blob picked and extracted from micrographs with box size 384 px. Multiple rounds of 2D classification were done with 40 online-EM iterations, and batch size of 400 to optimize for small particles.

His purification-SEC dataset: Micrographs with a CTF fit cross-correlation higher than 0.5 (off-hole micrographs) were discarded and the rest manually curated to discard crystalline ice and overall low-resolution micrographs, leaving 2,888 micrographs. Particles were blob picked and extracted from micrographs with box size 384 px. Multiple rounds of 2D classification were done with 40 online-EM iterations, and batch size of 200 to optimize for

small particles.

nvRNAP molecular cloning

nvRNAP: A plasmid containing 4-subunit nvRNAP, including the spliced version of gp55, was obtained from the Yakunina Lab [12]. To construct a plasmid containing the full 5-subunit nvRNAP, gp68 was cloned into the 4-subunit plasmid by joining restriction-digested plasmid and gp68 together via HiFi reaction, following manufacturer-recommend protocol. A T7 promoter and RBS were added in front of gp68 using the NEB Q5 Site-Directed Mutagenesis Kit, following the manufacturer-recommended protocol.

nvRNAP potential interactors: The NEB Q5 Site-Directed Mutagenesis Kit was used to add affinity tags to potential interactors of interest. Tagged genes were then cloned into pDuet vectors via HiFi reaction, following manufacturer-recommended protocol.

XL1B cells were used for all cloning steps. Cloning reactions were incubated with 50 μL XL1B cells on ice for 20 minutes, heat shocked at 42°C for 45s, recovered for 1 hour in LB, and plated onto LB agar plates containing the appropriate antibiotic. Constructs were confirmed via Sanger sequencing and whole-plasmid sequencing.

nvRNAP co-immunoprecipitation

Plasmids containing nvRNAP and the potential interactor were co-transformed into BL21(DE3) cells following the manufacturer's protocol, using approximately equal concentrations of each plasmid. Transformations were plated on LB agar plates containing antibiotics for both plasmids to ensure successful co-transformation. Colonies were grown up in LB + the appropriate antibiotics and preserved as glycerol stocks.

From glycerol stocks, strains were grown up in small-scale starter cultures with LB + the appropriate antibiotics, then diluted 1:100 into 50 mL LB + antibiotic cultures. The 50 mL cultures were grown at 37°C to OD 0.4 - 0.7 (variability across constructs and replicates). Cultures were induced with 1 mM IPTG and grown for 16-17 hours at 18°C. Cells were

spun down at 3,000g for 20 minutes. Cell pellets were resuspended in 1 mL lysis buffer (25 mM Tris pH 8, 150 mM NaCl supplemented with protease inhibitor tablets, 25 units/mL Pierce Universal Nuclease, 0.5mM TCEP), and sonicated for 3 rounds at 20% amplitude, with 1s on 1s off for 10s being 1 round. Lysates were clarified at 21,000g for 30 minutes in a 4°C benchtop centrifuge. Clarified lysates were added to 100 μL resin (bed volume) and incubated at 4°C for 1 hour, then washed with at least 100 bed volumes of 25 mM Tris-HCl pH 8, 300 mM NaCl, 0.5 mM TCEP. The resin was incubated with 5-10 volumes of elution buffer for 30 minutes at 4°C, and elutions were collected for Western blot. Wash buffer supplemented with 500 μg/mL HA peptide was used to elute HA-tagged proteins, and wash buffer supplemented with 100 mM maltose was used to elute MBP-tagged proteins. These pulldowns targeted the potential interactor, with detection of coeluting nvRNAP via Western blot. We initially attempted using NiNTA resin to pull down on His-tagged nvRNAP, but results were uninterpretable due to negative controls displaying signal.

Western blotting

Samples were prepared for SDS-PAGE by mixing with 5X SDS loading buffer and heating at 95°C for 5 minutes. After running, SDS-PAGE gels were transferred onto PVDF or nitrocellulose membranes using the BioRad Turboblot. Quality of transfer was confirmed by Ponceau. After washing off Ponceau with 1X TBS-T, blots were blocked in 5% casein at room temperature for 1 hour. Blots were incubated in 1:1,000 (His, HA, MBP) or 1:5,000 (Imp1) of primary antibody overnight at 4°C, then washed in TBS-T for 10 minutes per round, for 3 rounds total. Blots were then incubated in 1:10,000 HRP secondary (His, HA, MBP) or 1:5,000 HRP secondary (Imp1) for 1 hour at room temperature. Blots were washed in TBS-T for 10 minutes per round, for 3 rounds total to remove excess secondary, then imaged with ClarityMax ECL following the manufacturer-recommended protocol. All primary and secondary antibodies were from Cell Signaling Technology, with the exception of Imp1 primary antibody, which was custom-made.

nvRNAP large scale protein expression & purification

Small scale starter cultures were diluted 1:100 into large scale LB + antibiotic cultures. Large scale cultures were grown at 37°C until approximately OD 0.5, and induced with 1 mM IPTG at 18°C for 16-18 hours. After growth, cultures were pelleted and cell pellets resuspended in 40 mM Tris pH 8, 500 mM NaCl, 5 mM imidazole supplemented with 1 mM DTT, 1 protease inhibitor tablet, and 25 U/mL Pierce Universal Nuclease. This mixture was lysed via sonication for 20 rounds at 50% amplitude, with each round being 15s on, 45s off. Lysate was clarified at 30,000g for 30 minutes at 4°C. Clarified lysate was incubated with Ni-NTA resin in a gravity column for 1 hour at 4°C. 500 μL - 1 mL of bed volume was used per liter of large scale culture lysed. After incubation of clarified lysate and Ni-NTA resin, the column was washed with at least 100 bed volumes of 40 mM Tris pH 8, 500 mM NaCl, 20 mM imidazole, 1 mM DTT. The resin was incubated with 40 mM Tris pH 8, 500 mM NaCl, 400 mM imidazole, 1 mM DTT for 10 minutes at 4°C before elution over 10 bed volumes. Elutions were concentrated using an Amicon centrifugal concentrator to 500 uL, and loaded onto the Superdex 200 Increase 10/300 GL for size exclusion chromatography in 40 mM Tris pH 8, 200 mM NaCl, 1 mM DTT. Sizing fractions were flash frozen in liquid nitrogen and stored at -80°C for later use.

References

[1] Georjon, Héloïse and Bernheim, Aude. "The highly diverse antiphage defence systems of bacteria". *Nature Reviews Microbiology* 21.686-700 (2023). DOI: https://doi.org/10.1038/s41579-023-00934-x.

[2] Murtazalieva1, Khalimat et al. "The growing repertoire of phage anti-defence systems". *Trends in Microbiology* 32.1212-1228 (2024). DOI: https://doi.org/10.1016/j.tim.2024.05.005.

[3] Mendoza, Senén D. et al. "A bacteriophage nucleus-like compartment shields DNA from CRISPR nucleases". *Nature* 577.244-248 (2020). DOI: https://doi.org/10.1038/s41586-019-1786-y.

[4] Chaikeeratisak, Vorrapon et al. "The Phage Nucleus and Tubulin Spindle Are Conserved among Large Pseudomonas Phages". *Cell Reports* 20.1563–1571 (2017). DOI: https://doi.org/10.1016/j.celrep.2017.07.064.

[5] Kokontis, Claire et al. "Multi-interface licensing of protein import into a phage nucleus". *Nature* 639.456–462 (2025). DOI: https://doi.org/10.1038/s41586-024-08547-x.

[6] Ceyssens, Pieter-Jan et al. "Development of Giant Bacteriophage KZ Is Independent of the Host Transcription Apparatus". *Journal of Virology* 88.10501–10510 (2014). DOI: https://doi.org/10.1128/jvi.01347-14.

[7] Enustun, Eray et al. "A phage nucleus-associated RNA-binding protein is required for jumbo phage infection". *Nucleic Acids Research* 52.4440–4455 (2024). DOI: https://doi.org/10.1093/nar/gkae216.

[8] Abramson, Josh et al. "Accurate structure prediction of biomolecular interactions with AlphaFold 3". *Nature* 630.493–500 (2024). DOI: https://doi.org/10.1038/s41586-024-07487-w.

[9] Antonova, Daria et al. “Genomic transfer via membrane vesicle: a strategy of giant phage phiKZ for early infection”. *Journal of Virology* 98.e00205-24 (2024). DOI: https://doi.org/10.1128/jvi.00205-24.

[10] Enustun, Eray et al. “Identification of the bacteriophage nucleus protein interaction network”. *Nature Structural & Molecular Biology* 30.1653–1662 (2023). DOI: https://doi.org/10.1038/s41594-023-01094-5.

[11] Yakunina, Maria et al. “A non-canonical multisubunit RNA polymerase encoded by a giant bacteriophage”. *Nucleic Acids Research* 43.10411–10420 (2015). DOI: https://doi.org/10.1093/nar/gkv1095.

[12] Orekhova, Mariia et al. “The study of the phiKZ phage non-canonical non-virion RNA polymerase”. *Biochemical and Biophysical Research Communications* 511.759–764 (2019). DOI: https://doi.org/10.1016/j.bbrc.2019.02.132.

[13] Mastronarde, David N. “Automated electron microscope tomography using robust prediction of specimen movements”. *Journal of Structural Biology* 152.36–51 (2005). DOI: https://doi.org/10.1016/j.jsb.2005.07.007.

[14] Zheng, Shawn Q et al. “MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy”. *Nature Methods* 14.331-332 (2017). DOI: https://doi.org/10.1038/nmeth.4193.

[15] Punjani, Ali et al. “cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination”. *Nature Methods* 14.290–296 (2017). DOI: https://doi.org/10.1038/nmeth.4169.

[16] Rohou, Alexis and Grigorieff, Nikolaus. “CTFFIND4: Fast and accurate defocus estimation from electron micrographs”. *Journal of Structural Biology* 192.216–221 (2015). DOI: https://doi.org/10.1016/j.jsb.2015.08.008.

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:
F78586ED0BC7433...
Author Signature

10/3/2025
Date