Title
Structure, Function, and Assembly of the Jumbo Bacteriophage Nucleus

Permalink
https://escholarship.org/uc/item/1nw3b839

Author
Nieweglowska, Eliza Sara

Publication Date
2023

Peer reviewed|Thesis/dissertation

Structure, Function, and Assembly of the Jumbo Bacteriophage Nucleus

by
Eliza S. Nieweglowska

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

in

Biochemistry and Molecular Biology

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO
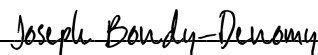
Approved:

DocuSigned by:

Carol Gross
7AFBE9046E0744C...

Carol Gross
_____
Chair
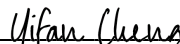
DocuSigned by:

David A. Agard
D2838D07EB746A...

David A. Agard
_____

DocuSigned by:

Joseph Bondy-Denomy
DC0S21g9BF4247C...

Joseph Bondy-Denomy
_____

DocuSigned by:

Yifan Cheng
87A0F4E0B50C4F6...

Yifan Cheng
_____

_____
Committee Members

Dedication

Thank you to my parents—Artur and Elżbieta Niewęgłowski—for being such inspiring, loving, and hardworking people as well as to my aunt—Małgorzata Babiarz—whom I've always looked up to. I would not have gotten here without the endless love, teasing, laughter, stress relief, and overall support that you give me.

Thank you to my fiancé—Simon Swanson—for being the loving, supportive everyday constant through the good and the bad of this journey. I cannot thank you enough for troubleshooting challenges with me, cooking delicious food, encouraging me to develop healthy habits, making the time to looking over my emails and paper, and simply being by my side.

I dedicate this work to the four of you.

Inclusion of Previously Published Work

Chapter 2 is published under the following citation:

Mendoza SD, **Nieweglowska ES**, Govindarajan S, Leon LM, Berry JD, Tiwari A, Chaikeeratisak V, Pogliano J, Agard DA, Bondy-Denomy J. A bacteriophage nucleus-like compartment shields DNA from CRISPR nucleases. Nature. 2020 Jan;577(7789):244-248. doi: 10.1038/s41586-019-1786-y. Epub 2019 Dec 9. PMID: 31819262; PMCID: PMC6949375.

Chapter 3 may be found under the following citation:

Nieweglowska, E.S., Brilot, A.F., Méndez-Moran, M. *et al.* The ɸPA3 phage nucleus is enclosed by a self-assembling 2D crystalline lattice. *Nat Commun* 14, 927 (2023). https://doi.org/10.1038/s41467-023-36526-9

Structure, Function, and Assembly of the Jumbo Bacteriophage Nucleus

Eliza S. Nieweglowska

Abstract

The ΦKZ-like family jumbo bacteriophage ΦPA3 infects the pathogenic bacterium *Pseudomonas aeruginosa*. Upon infection, the bacteriophage forms a dynamic, bipolar spindle within the host using a divergent tubulin family protein known as *ph*age t*u*bulin/FtsZ or PhuZ. This spindle is important for infectivity and centers the phage DNA. The phage DNA, along with the DNA replication and transcription machinery, is found inside of a proteinaceous shell, while the translation and nucleotide synthesis machinery is excluded. This partitioning of proteins according to function suggests a mechanism of selectivity, akin to that of the eukaryotic nucleus. This similarity inspired the naming of this proteinaceous, DNA-containing structure the phage nucleus. The phage nucleus is irregularly shaped and grows over the course of infection, reaching up to half the host cell size or about 0.5 μm in diameter. The dramatic phage nucleus growth is likely driven by DNA replication with which the protein shell must keep up by incorporating new subunits. The proteinaceous exterior of the ΦPA3 phage nucleus is constructed using the most highly expressed, non-virion protein, Gp53, which we name *ph*age n*u*clear e*n*closure protein or PhuN. At the beginning of this research, PhuN had no previously characterized homologues. Since then, PhuN has been studied and observed in a series of jumbo phages spanning bacteria ranging from *Escherichia coli* to *Serratia*. Using an *in vitro* approach coupled with direct visualization using electron microscopy, the work presented in this dissertation begins to uncover the principles and mechanisms underlying the remarkable formation, growth, and function of the phage nucleus. In a series of collaborative work, we show that the phage nucleus shields the bacteriophage DNA from host defense systems, we utilize cryo-EM to show to 3.9 Å that the phage nucleus is enclosed in a proteinaceous 2D crystal, and,

finally, I share unpublished observations worthy of further exploration. This work lays the

foundations for future biochemical and structural investigation probing jumbo phage

mechanisms including but not limited to phage nucleus subunit addition, selectivity, and

capsid packaging.

Table of Contents

Table of Figures

Chapter 3

Chapter 4

Chapter 5

List of Tables

Chapter 1

Introduction

Compartmentalization and Enclosing Space in Biology

Enclosing space is a fundamental biological problem, one that has long distinguished prokaryotic and eukaryotic organisms. Eukaryotic cells have an intricate system of subcellular organization, partitioning space into small organelles like the mitochondria, golgi apparatus, and, of course, nucleus.[1] Within each of these resides highly specialized machinery and an environment catering to the processes occurring within that particular organelle. This environment may require low or high pH, oxidation, or, as recently described for the mitochondria, perhaps even a different temperature.[2,3] Compartmentalizing and enclosing such specialized spaces enables cells to carry out essential reactions both safely and efficiently or to simply protect precious cargo, like DNA in the nucleus.[1]

Biology has evolved an effective toolkit for creating enclosed spaces. Most commonly, as in the aforementioned eukaryotic organelles, phospholipid bilayers, assisted by proteins, wrap around and enclose their cargo. This enclosure relies on the poor miscibility of lipids and water, keeping the phospholipids as a separative layer between two aqueous environments. The contents between the exterior and interior environments may exchange via passage through various protein channels integrated into the bilayer or the merger of smaller vesicles into the bilayer, depositing new contents. A protein called clathrin aides vesicle formation by forming a mesh-like assembly of triskelions on the coat of budding vesicles.[4–7] The curvature and size of these dynamic clathrin assemblies is mediated by accessory proteins.[7]

While clathrin only assists lipid enclosure formation, some biological compartments are built entirely out of proteins, functioning independently of lipids. One common example of a protein-based enclosure is a viral capsid—a rigid protein shell constructed of penta- and hexameric protein assembly units.[8] These capsids are densely packaged with viral DNA and, in

some cases, proteins necessary early in infection. A visually similar compartment can be observed in some bacteria, which otherwise lack canonical means of subcellular organization like phospholipid organelles. These bacterial protein-based structures enclose space by incorporating varied units of protein assembly: triangular and hexameric protein units establish flat surfaces while pentameric units introduce curvature. The best understood such bacterial microcompartment is the carboxysome—a factory for carbon dioxide fixation.[9–14] While visually reminiscent of viral capsids, carboxysomes evolved independently and share little homology with viral capsid proteins.[15] Nanocompartments, on the other hand, are an example of protein-based enclosures that look like capsids and do share an evolutionary history with capsids.[16,17] Both structures exhibit some amount of selectivity, only allowing certain molecules to be enclosed or to enter through small pores at the centers of their small protein unit assemblies.

While common to both eukaryotic and prokaryotic organisms, an only more recently appreciated means of partitioning space is through phase separation.[18] No discrete physical barrier separates the contents of these liquid condensates from the exterior environment. Instead, contents are determined by factors including density, diffusion, and direct protein:protein interactions. Within such structures, proteins interact with each other in a far less organized manner than in carboxysomes and clathrin coats. These liquid-like, membraneless organelles are malleable, exhibit dynamic growth, and readily fuse. The purpose of membraneless organelles often focuses on protecting cells from stress-induced damage through sequestering genetic and protein material until stress dissipates.[19]

Jumbo Phages and the Phage Nucleus

Over a two billion-year coevolutionary[20] time-course, anything is possible but unless we can detect and see it, we will never know. For a long time, bacteriophages—viruses that infect bacteria—were thought and taught to be minimal, no-fuss biological entities. They found and bound to a host bacterium, injected their DNA into the host, hijacked the host machinery,

generated new phage proteins and DNA, packaged particles, and lysed the host cell, only to repeat the process. Most phages accomplished this with fewer than 100 kb of genetic information. In fact, as recently as 2017 the literature only described 93 isolated phages with genomes exceeding 200 kb.[21,22]

In the last few years, an increasing number of bacteriophages with genomes exceeding 200 kb, so called jumbo phages, and even exceeding 500 kb, known as megaphages, have been described.[21] This sharp increase in the identification of huge phages is a result of access to improved tools, like whole genome sequencing, for phage identification.[23] Previous methods relied on phage isolation and were biased against large phage particles, thus they were often filtered out and discarded. While the need for bacteriophages with such large genomes is not immediately obvious, over the last decade we have begun to see that, at this genome scale, infection can follow a fundamentally different, complex progression.[24–27] Thus far, this has been most closely investigated in the ϕKZ-like family of jumbo bacteriophages **(Fig. 1.1)**.

The ϕKZ-like jumbo phage family includes *Pseudomonas aeruginosa* targeting phages ϕKZ and ϕPA3 as well as *Pseudomonas chlororaphis* targeting phage 201ϕ2-1.[28,29] These bacteriophages contain proteins of the divergent tubulin family, *ph*age t*u*bulin/Fts*Z* or PhuZ.[24–26] Upon infection, PhuZ forms a dynamic, bipolar spindle within the host cell. This spindle is important for infectivity, reducing phage burst size by approximately 50% if a GTPase hydrolysis PhuZ mutant is expressed.[25] The replicating phage DNA is enclosed in a proteinaceous shell— distinct from the phage capsid in composition, size, and biophysical properties—that grows from a small punctum at the site of infection to a micron-scale structure brought to the cell center by PhuZ.[27] This growth is likely driven by the generation of new copies of the phage genome within the structure. For packaging newly assembled phage capsids with DNA, the empty phage heads, initially assembled along the bacterial host membrane, are trafficked along PhuZ filaments and deposited at the shell exterior.[30] The factors used for capsid attachment to the PhuZ filaments and shell exterior remain unknown. In addition to housing the replicating phage

DNA, bacteriophage and host proteins are selectively partitioned according to their function relative to the shell—DNA replication and transcription machinery is maintained inside, while translation and nucleotide synthesis machinery is confined to the host cytoplasm.[27] As this structure is analogous to the eukaryotic nucleus in both function and appearance, the shell with its contents is known as the phage nucleus.

The phage nucleus is the first example of a nucleus-like structure established by a bacteriophage and, despite similarities to its eukaryotic counterpart, the phage nucleus is unique. It is a dynamically growing, micron-scale enclosure constructed without the use of a phospholipid membrane. The exterior is instead composed of a protein layer, as initially observed using cryo-electron tomography (cryo-ET).[27] The proteinaceous nature of the phage nucleus makes its dynamic growth from a punctum to a micron-scale structure surprising as most previously described protein enclosures are of a far smaller, defined size. Despite this dynamic growth, fluorescence recovery after photobleaching (FRAP) experiments indicated that subunits in mature phage nuclear shells do not move or diffuse, suggesting stable subunit incorporation.[27] Cryo-ET further showed that phage nuclei vary in shapes and sizes across different infected cells, suggesting a flexible and variable assembly.[27] This property, yet again, contrasts most other protein-based compartments, like the more rigid and uniform carboxysomes and encapsulins. Moreover, the family of proteins known to form phage nuclear shells in the different ϕKZ-like jumbo phages—Gp53 (ϕPA3), Gp54 (ϕKZ), Gp105 (201ϕ2-1), Gp202 (PCH45)—have no previously characterized homologues.[31,32] Combined with its aforementioned role as a selective barrier, this assortment of biophysical properties and lack of homologues brought the phage nuclear shell to our attention as a novel biological phenomenon, ripe for biochemical, structural, and genetic interrogation.

Research Objectives

The phage nucleus is a new kind of biological enclosure that organizes the progression of the jumbo phage infection. The properties of the proteinaceous assembly enclosing the phage nucleus are not encompassed by any previously studied biological enclosures and the primary protein forming the exterior shell was, at the conception of this work, uncharacterized. In this thesis, I pursued biochemical and structural characterization of the primary *ph*age n*u*clear e*n*closure protein, which I have named PhuN, and collaborated with a wonderful team in the Bondy-Denomy lab to implicate this protein assembly in a protective role for the replicating phage DNA inside.

Figures
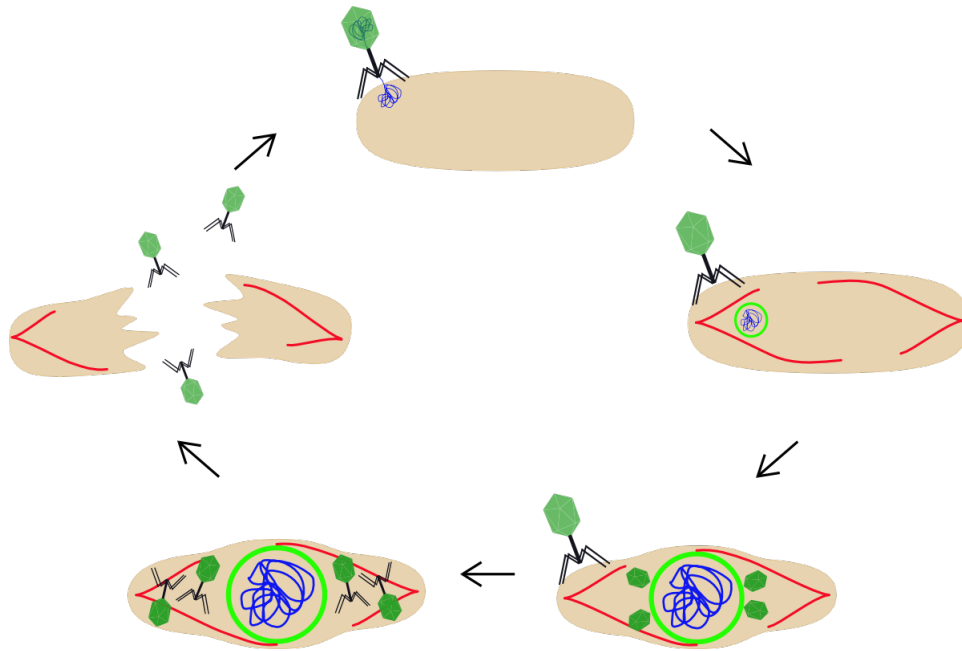
Figure 1.1 | Infection Cycle of Phage Nucleus Forming Jumbo Phages.
PhuZ filaments are shown in red, capsids in dark green, phage DNA in blue, and the phage nuclear shell in neon green.

References

1. Alberts, B. *et al.* Molecular Biology of the Cell. 889–962 (2017) doi:10.1201/9781315735368-16.

2. Chrétien, D. *et al.* Mitochondria are physiologically maintained at close to 50 °C. *Plos Biol* 16, e2003992 (2018).

3. Demaurex, N. pH Homeostasis of Cellular Organelles. *Physiology* 17, 1–5 (2002).

4. Avinoam, O., Schorb, M., Beese, C. J., Briggs, J. A. G. & Kaksonen, M. Endocytic sites mature by continuous bending and remodeling of the clathrin coat. *Science* 348, 1369–1372 (2015).

5. Maib, H., Smythe, E. & Ayscough, K. Forty years on: clathrin-coated pits continue to fascinate. *Mol Biol Cell* 28, 843–847 (2017).

6. McMahon, H. T. & Boucrot, E. Molecular mechanism and physiological functions of clathrin-mediated endocytosis. *Nat Rev Mol Cell Bio* 12, 517–533 (2011).

7. Kirchhausen, T., Owen, D. & Harrison, S. C. Molecular Structure, Function, and Dynamics of Clathrin-Mediated Membrane Traffic. *Csh Perspect Biol* 6, a016725 (2014).

8. Rao, V. B. & Black, L. W. Structure and assembly of bacteriophage T4 head. *Virol J* 7, 356 (2010).

9. Kerfeld, C. A., Aussignargues, C., Zarzycki, J., Cai, F. & Sutter, M. Bacterial microcompartments. *Nat Rev Microbiol* 16, 277–290 (2018).

10. Cameron, J. C., Wilson, S. C., Bernstein, S. L. & Kerfeld, C. A. Biogenesis of a Bacterial Organelle: The Carboxysome Assembly Pathway. *Cell* 155, 1131–1140 (2013).

11. Iancu, C. V. *et al.* Organization, Structure, and Assembly of α-Carboxysomes Determined by Electron Cryotomography of Intact Cells. *J Mol Biol* 396, 105–117 (2010).

12. Yeates, T. O., Tsai, Y., Tanaka, S., Sawaya, M. R. & Kerfeld, C. A. Self-assembly in the carboxysome: a viral capsid-like protein shell in bacterial cells. *Biochem Soc T* 35, 508–511 (2007).

13. Savage, D. F., Afonso, B., Chen, A. H. & Silver, P. A. Spatially Ordered Dynamics of the Bacterial Carbon Fixation Machinery. *Science* 327, 1258–1261 (2010).

14. Sutter, M., Greber, B., Aussignargues, C. & Kerfeld, C. A. Assembly principles and structure of a 6.5-MDa bacterial microcompartment shell. *Science* 356, 1293–1297 (2017).

15. Krupovic, M. & Koonin, E. V. Cellular origin of the viral capsid-like bacterial microcompartments. *Biol Direct* 12, 25 (2017).

16. Nichols, R. J., Cassidy-Amstutz, C., Chaijarasphong, T. & Savage, D. F. Encapsulins: molecular biology of the shell. *Crit Rev Biochem Mol* 52, 1–12 (2017).

17. Giessen, T. W. & Silver, P. A. Widespread distribution of encapsulin nanocompartments reveals functional diversity. *Nat Microbiol* 2, 17029 (2017).

18. Boeynaems, S. *et al.* Protein Phase Separation: A New Phase in Cell Biology. *Trends Cell Biol* 28, 420–435 (2018).

19. Riback, J. A. *et al.* Stress-Triggered Phase Separation Is an Adaptive, Evolutionarily Tuned Response. *Cell* 168, 1028-1040.e19 (2017).

20. Durzyńska, J. & Goździcka-Józefiak, A. Viruses and cells intertwined since the dawn of evolution. *Virol J* 12, 169 (2015).

21. Al-Shayeb, B. *et al.* Clades of huge phages from across Earth's ecosystems. *Nature* 578, 425–431 (2020).

22. Yuan, Y. & Gao, M. Jumbo Bacteriophages: An Overview. *Front Microbiol* 8, 403 (2017).

23. Yuan, A. H. & Hochschild, A. A bacterial global regulator forms a prion. *Science* 355, 198–201 (2017).

24. Zehr, E. A. *et al.* The Structure and Assembly Mechanism of a Novel Three-Stranded Tubulin Filament that Centers Phage DNA. *Structure* 22, 539–548 (2014).

25. Kraemer, J. A. *et al.* A Phage Tubulin Assembles Dynamic Filaments by an Atypical Mechanism to Center Viral DNA within the Host Cell. *Cell* 149, 1488–1499 (2012).

26. Erb, M. L. *et al.* A bacteriophage tubulin harnesses dynamic instability to center DNA in infected cells. *Elife* 3, e03197 (2014).

27. Chaikeeratisak, V. *et al.* Assembly of a nucleus-like structure during viral replication in bacteria. *Science* 355, 194–197 (2017).

28. Monson, R., Foulds, I., Foweraker, J., Welch, M. & Salmond, G. P. C. The Pseudomonas aeruginosa generalized transducing phage φPA3 is a new member of the φKZ-like group of 'jumbo' phages, and infects model laboratory strains and clinical isolates from cystic fibrosis patients. *Microbiology+* 157, 859–867 (2011).

29. Thomas, J. A. *et al.* Characterization of Pseudomonas chlororaphis myovirus 201φ2-1 via genomic sequencing, mass spectrometry, and electron microscopy. *Virology* 376, 330–338 (2008).

30. Chaikeeratisak, V. *et al.* Viral Capsid Trafficking along Treadmilling Tubulin Filaments in Bacteria. *Cell* 177, 1771-1780.e12 (2019).

31. Chaikeeratisak, V. *et al.* The Phage Nucleus and Tubulin Spindle Are Conserved among Large Pseudomonas Phages. *Cell Reports* 20, 1563–1571 (2017).

32. Malone, L. M. *et al.* A jumbo phage that forms a nucleus-like structure evades CRISPR–Cas DNA targeting but is vulnerable to type III RNA-based immunity. *Nat Microbiol* 1–8 (2019) doi:10.1038/s41564-019-0612-5.

Chapter 2

A bacteriophage nucleus-like compartment shields DNA from CRISPR nucleases

Preface

The work presented in this chapter is the result of a collaborative effort spearheaded by Dr. Senén Mendoza during his graduate studies in the Bondy-Denomy lab at UCSF. This project began as a follow-up to the difficulties that Dr. Joseph Bondy-Denomy encountered targeting ΦKZ with CRISPR-Cas in his own graduate work.

My primary role in this project was to visually determine whether the immune systems we tested were excluded from the phage nucleus, focusing on Cas9 and Cas13a. I utilized a published protocol for live fluorescence microscopy, which Dr. Vorrapon Chaikeeratisak (formerly a post-doctoral scholar in the Pogliano lab, UCSD) additionally showed me, as well as adapted a published protocol for immunofluorescence in bacteria for this purpose.

Abstract

All viruses require strategies to inhibit or evade the immunity pathways of cells they infect. The viruses that infect bacteria, bacteriophages (phages), must avoid nucleic-acid targeting immune pathways such as CRISPR-Cas (clustered regularly interspaced short palindromic repeats and CRISPR-associated genes) and restriction-modification (R-M) systems to replicate efficiently. Here, we show that jumbo phage ΦKZ, infecting *Pseudomonas aeruginosa,* segregates its DNA from immunity nucleases by constructing a proteinaceous nucleus-like compartment. ΦKZ resists many DNA-targeting immune systems *in vivo*, including two CRISPR-Cas3 subtypes, Cas9, Cas12a, and the restriction enzymes HsdRMS and EcoRI. Cas and restriction enzymes are unable to access the phage DNA throughout the infection, but engineered re-localization of EcoRI inside the compartment enables phage targeting and cell protection. Moreover, ΦKZ is sensitive to the RNA targeting CRISPR-Cas enzyme, Cas13a,

Chapter 2

A bacteriophage nucleus-like compartment shields DNA from CRISPR nucleases

Preface

The work presented in this chapter is the result of a collaborative effort spearheaded by Dr. Senén Mendoza during his graduate studies in the Bondy-Denomy lab at UCSF. This project began as a follow-up to the difficulties that Dr. Joseph Bondy-Denomy encountered targeting ΦKZ with CRISPR-Cas in his own graduate work.

My primary role in this project was to visually determine whether the immune systems we tested were excluded from the phage nucleus, focusing on Cas9 and Cas13a. I utilized a published protocol for live fluorescence microscopy, which Dr. Vorrapon Chaikeeratisak (formerly a post-doctoral scholar in the Pogliano lab, UCSD) additionally showed me, as well as adapted a published protocol for immunofluorescence in bacteria for this purpose.

Abstract

All viruses require strategies to inhibit or evade the immunity pathways of cells they infect. The viruses that infect bacteria, bacteriophages (phages), must avoid nucleic-acid targeting immune pathways such as CRISPR-Cas (clustered regularly interspaced short palindromic repeats and CRISPR-associated genes) and restriction-modification (R-M) systems to replicate efficiently. Here, we show that jumbo phage ΦKZ, infecting *Pseudomonas aeruginosa,* segregates its DNA from immunity nucleases by constructing a proteinaceous nucleus-like compartment. ΦKZ resists many DNA-targeting immune systems *in vivo*, including two CRISPR-Cas3 subtypes, Cas9, Cas12a, and the restriction enzymes HsdRMS and EcoRI. Cas and restriction enzymes are unable to access the phage DNA throughout the infection, but engineered re-localization of EcoRI inside the compartment enables phage targeting and cell protection. Moreover, ΦKZ is sensitive to the RNA targeting CRISPR-Cas enzyme, Cas13a,

likely due to phage mRNA localizing to the cytoplasm. Collectively, we propose that *Pseudomonas* jumbo phages evade a broad spectrum of DNA-targeting nucleases through the assembly of a protein barrier around their genome.

Jumbo phage ΦKZ Resists CRISPR-Cas targeting

Phages that infect *Pseudomonas aeruginosa* can avoid CRISPR-mediated destruction by encoding "anti-CRISPR" (Acr) proteins that inhibit the Type I-E and I-F CRISPR-Cas systems[2-4]. To determine whether any *P. aeruginosa* phages are resistant to the *P. aeruginosa* Type I-C CRISPR-Cas system[5], a common and understudied variant[6], a strain engineered to express Type I-C *cas3, cas5, cas8, cas7*[7] was provided with a panel of crRNAs targeting phages from five taxonomic groups: JBD30, D3, and JBD68 (distinct temperate siphophages), F8 and ΦKZ (distinct lytic myophages). All phages succumbed to targeting, except ΦKZ **(Fig. 2.1)** and a related phage ΦPA3 **(Fig. 2.5)**. ΦKZ titer did not decrease when challenged using 11 different Type I-C crRNAs **(Fig. 2.1, Fig. 2.6)** nor when exposed to the type I-F CRISPR-Cas system of *P. aeruginosa*[8] **(Fig. 2.6)**.

The ΦKZ genome possesses no homologs of known *acr*[2-4,9,10] or anti-CRISPR associated (*aca*) genes that have previously enabled identification of *acr* genes[4,9,10]. Moreover, gene knockout approaches have not been previously established for ΦKZ. To determine the mechanism for CRISPR evasion, we attempted to utilize the Type II-A CRISPR-Cas9 (SpyCas9) as a genetic tool. SpyCas9 and sgRNAs robustly targeted control phage JBD30 but ΦKZ replication and associated cell lysis was unaffected **(Fig. 2.1)**. An additional six sgRNA sequences also failed to target ΦKZ **(Fig. 2.7)**, as did four against ΦPA3 **(Fig. 2.5)**. Given the ability of this phage to evade unrelated CRISPR systems (Type I and II), including one from a microbe that this phage does not infect (*Streptococcus pyogenes*), we considered that ΦKZ may be generally resistant to CRISPR-Cas immunity, as opposed to relying on specific inhibitor proteins. Type V-A Cas12a (Cpf1) CRISPR-Cas system from *Moraxella bovoculi* was expressed

11

in *P. aeruginosa* and sucessfully targeted the control phage, but not ΦKZ with any of the nine crRNAs tested **(Fig. 2.1, Fig. 2.7)**. The ability of this phage to resist CRISPR systems found in its natural host, *Pseudomonas* (Type I-C and I-F), and those not naturally present (Type II-A and V-A) suggests that this phage possesses a mechanism enabling "pan-CRISPR" resistance.

Restriction-modification systems are the most common form of bacterial immunity in nature and pose a significant impediment to phage replication[1]. Type I R-M (HsdRMS from *P. aeruginosa*) and Type II R-M (EcoRI from *E. coli*) were next tested, possessing 24 and 92 cut sites in the ΦKZ genome, respectively. ΦKZ was propagated on strain PAK, an isolate that generates phages susceptible to PAO1 HsdRMS restriction. When phage JBD30 (5 Type I R-M sites) was assayed in this manner, its titer was reduced by ~5 orders of magnitude, dependent on *hsdR* **(Fig. 2.1)**. In contrast, no restriction was observed for ΦKZ **(Fig. 2.1)** or ΦPA3 (**Fig. 2.5)**. Similarly, the expression of EcoRI reduced JBD30 titer by 3 orders of magnitude (12 EcoRI sites) but had no impact on ΦKZ (92 EcoRI sites) **(Fig. 2.1)**. Together, these experiments demonstrate that ΦKZ is refractory to the *six* selected CRISPR-Cas and restriction endonucleases *in vivo*.

Phage nucleus-like structure excludes immune enzymes

ΦKZ and ΦKZ-like phages infecting *P. aeruginosa* and *P. chlororaphis* were recently shown to construct an elaborate proteinaceous nucleus-like compartment where phage DNA replicates[11,12]. Additionally, a phage-encoded tubulin homologue, PhuZ, centers the compartment within the host cell[11-15]. Proteins involved in DNA replication, transcription, and recombination localize inside the shell, while mRNA and proteins mediating translation localize in the cytoplasmic space[11], akin to the eukaryotic nucleus. Given the apparent exclusion of select proteins, we considered whether this structure was responsible for the pan-resistance of ΦKZ to such distinct immune processes.

P. aeruginosa cells infected with ΦKZ were imaged with immunofluorescence to detect

Cas9 **(Fig. 2.2, Fig. 2.4)**; likewise, Cherry fusions with Cas9, two signature proteins from the

Type I-C and I-F CRISPR-Cas systems (Cas8 and Cas3, **Fig. 2.2**), as well as the restriction

enzyme HsdR **(Fig. 2.2)** were imaged using live cell imaging. These experiments revealed that

the immune enzymes are excluded from the shell during phage infection. DAPI staining

revealed the phage DNA inside the shell, while the host genome was degraded[14]. Proteins

previously shown to be internalized in the shell, phage ORF152 (imaged with anti-myc

immunofluorescence and Cherry fusion) and host Topoisomerase I (Cherry fusion) co-localized

with the DAPI signal, while Cherry was excluded **(Fig. 2.2)**. Although the rules for protein

internalization in the shell are currently unknown, each protein of known function that localizes

inside of the shell interacts with DNA[11,12] (i.e. DNA replication and transcription machinery),

suggesting that the exclusion of DNA-binding Cas and restriction proteins is an adaptive

function of the shell.

To confirm that the ΦKZ phage genome could be a substrate for DNA cleavage, if

accessed, two enzymes that failed *in vivo*, EcoRI and Cas9, were assayed *in vitro*. ΦKZ DNA

was extracted from virions and subjected to restriction digestion reactions with a panel of

restriction enzymes. EcoRI, HindIII, KpnI, and NcoI each cleaved the DNA, while SacI lacks a

sequence recognition motif in the ΦKZ genome and did not **(Fig. 2.3, Fig. 2.9)**. Cas9's ability to

cleave ΦKZ gDNA *in vitro* was next assessed. Due to the large size of the ΦKZ genome (280

kbp), we first subjected purified phage DNA to the restriction enzyme KasI to liberate a 6.98 kbp

product and then cleaved that species with a dual crRNA:tracrRNA-loaded SpyCas9 nuclease *in

vitro*. This reaction depleted the substrate, liberating the expected 4.9 kb and 2.1 kb fragments,

confirming that the phage gDNA is sensitive to Cas9 cleavage **(Fig. 2.3)**. Together these results

demonstrate that immune enzymes are capable of cleaving ΦKZ gDNA when accessed and that

immune evasion is likely not due to an intrinsic feature of the phage DNA, such as base

modifications that can impede Cascade-Cas3, Cas9, and EcoRI[16-19].

Immune enzyme re-localization enables phage targeting

Due to a current inability to mutate, weaken, or knockout the shell structure, we next sought to enable an immunity enzyme to bypass the shell to access phage DNA *in vivo*. The single effector enzyme Cas9 was fused to ORF152, a phage-encoded RecA-like protein that is internalized within the shell[11,12]. Independent fusions to the N- or C- terminus did not affect ΦKZ replication, however **(Fig. 2.10)**. Imaging of one fusion orientation (ORF152-Cas9) revealed peri-shell localization **(Fig. 2.10)**, suggesting that the fusion redirected Cas9 from its previously diffuse state, but perhaps the large Cas9 protein (1,368 amino acids, 158 kDa) was unable to traverse the shell border.

Given the large size (e.g. Cas12a) and complexity (e.g. Cascade, HsdRMS) of each of the other immune systems, we next fused the small, single effector protein EcoRI (278 amino acids, 31.5 kDa) to Cherry-ORF152. This fusion resulted in a notable >2.5 order of magnitude reduction in ΦKZ phage titer and markedly reduced plaque sizes **(Fig. 2.3)**, the first successful *in vivo* DNA targeting observed in this study. Liquid infections revealed that cells expressing EcoRI-Cherry-ORF152 were extremely well protected, gaining >5 orders of magnitude resistance to phage-induced lysis **(Fig. 2.3)**. A catalytic mutant EcoRI(E111G) fused to Cherry-ORF152 displayed no immune activity against ΦKZ in either assay, nor did active EcoRI-Cherry without the ORF152 fusion **(Fig. 2.3)**. Similar results were observed in the absence of the Cherry tag (i.e. EcoRI-ORF152), however ΦKZ targeting was more modest in that case, reducing titers by ~10-fold and protecting cells by a factor of ~10^4 **(Fig. 2.11)**. Imaging of infected cells expressing EcoRI-Cherry or EcoRI(E111G)-Cherry demonstrated that these proteins are excluded from the shell, while EcoRI(E111G)-Cherry-ORF152 is successfully localized inside **(Fig. 2.3)**. EcoRI-Cherry-ORF152 impaired the ability of infected cells to form full shells and proceed through the infection process, and the EcoRI fusion protein was often localized within or adjacent to the DAPI-stained puncta **(Fig. 2.3)**. Some host DNA can be seen

in these cells and infection does not proceed **(Fig. 2.3, Fig. 2.11)**. By bypassing the physical

barrier of the shell using rational engineering, we conclude that the shell is the cause of

resistance to immunity displayed by jumbo phage ΦKZ.

Phage mRNA is sensitive to Cas13a

The nucleus-like structure produced by ΦKZ provides robust resistance to DNA-targeting

immune systems, but natural immune systems may exist that can evade this mechanism. We

envisaged that the mRNA exported for translation in the cytoplasm is likely susceptible to

targeting and that a ribonuclease could provide anti-ΦKZ immunity. To test this, we adapted the

Type VI-A CRISPR RNA-guided RNA nuclease LseCas13a[20,21] for phage targeting in *P.*

aeruginosa. Three LseCas13a spacers (two targeting head gene *gp146* and one targeting the

shell gene *gp054*) decreased ΦKZ plaquing efficiency by >10^6-fold **(Fig. 2.4, Fig. 2.12)**.

Corroborating the plaquing results, LseCas13a was also effective at protecting *P. aeruginosa* in

liquid cultures, with 1-2 orders of magnitude resistance to phage induced lysis **(Fig. 2.4)**. We

suspect that the previously demonstrated collateral damage caused by Cas13a[22] is detrimental

to cell fitness, leading to a modest enhancement to the growth curves compared to EcoRI

targeting. At a frequency of ~10^{-6}–10^{-7}, Cas13 escaper phages were identified **(Fig. 2.13)** and

surprisingly contained genomic deletions at the site of targeting **(Fig. 2.4)**. Notably, all deletions

were in-frame and 5/6 escapers had clear microhomology regions ranging from 7-21 bp flanking

the deletion. One guide targeting the 5' end of *orf146* selected for deletions where the upstream

gene (*orf145*) became fused, in frame, to the targeted gene **(Fig. 2.4)**. These data support the

conclusion that sequence-specific CRISPR-Cas RNA targeting can inhibit this phage.

Prior to shell construction, it is unknown what the state of the phage DNA is. During

Cas13a targeting of the shell mRNA, imaging revealed that infections arrested before the phage

DNA proceeded from its injection site at the poles. Most cells had two polar puncta, even one

hour after infection **(Fig. 2.4)**. The absence of phage DNA diffusion or clearance (e.g. by the

15

endogenous Type I R-M system) suggests that the injected phage genome may be protected by a yet unknown mechanism involving injected, rapidly synthesized, or pre-existing host factors, prior to shell assembly. This is additionally corroborated by observations in the absence of Cas13a arrest, where we have observed DAPI-stained phage DNA adjacent to: i) EcoRI-Cherry-ORF152 puncta **(Fig. 2.3)** or ii) to a nascent shell localizing TopA **(Fig. 2.14)**. The molecular nature and mechanisms underlying this early protection event remain to be elucidated.

Discussion

The assembly of a proteinaceous compartment to house the replicating phage DNA creates a physical protective barrier resulting in the resistance of phage ΦKZ to DNA-cleaving enzymes **(Fig. 2.4)**. Evasion of endogenous *P. aeruginosa* Type I CRISPR-Cas systems by ΦKZ suggests that these jumbo phages are likely Type I CRISPR-Cas resistant in nature. Supporting this hypothesis, our analysis of >4,000 non-redundant *P. aeruginosa* spacers (Type I-C, I-E, and I-F) reported by van Belkum *et al*. (2015) found no spacers against ΦKZ or its jumbo phage relatives ΦPA3, PaBG, KTN4, and PA7 **(Table 2.1)**. This is in contrast to the many spacer matches from each system matching diverse *P. aeruginosa* phages, such as those assayed in our screen **(Fig. 2.1)** and those encoding anti-CRISPR proteins. Additionally, given the efficacy of the RNA-targeting CRISPR-Cas13 system, we propose that Type VI CRISPR systems are well-suited to target the mRNA of DNA phages when the DNA is inaccessible (i.e. due to base modifications or physical segregation).

This phage compartment has only been documented among the jumbo phages of *Pseudomonas*[11,12], however, we consider that physical occlusion of phage genomic DNA through this and other mechanisms may comprise a novel route to immune system evasion by phages. Indeed, recently discovered "mega phages" were reported to encode homologues of phage tubulin, which centers this compartment during infection[23]. The pan-resistance of ΦKZ to DNA-targeting enzymes provides an explanation for the elaborate and impressive nucleus-like

structure. Furthermore, the polar localization of the injected phage DNA during mRNA targeting suggests a poorly understood early protective mechanism. Considering the pronounced resistance to overexpressed immune enzymes and the previously observed docking of capsids at the shell periphery[11,24], we propose that the phage DNA is never exposed to the cytoplasm **(Fig. 2.4)**. Other hypotheses to explain the importance of the shell remain to be addressed, including protection from phage-derived nucleases that degrade the bacterial genome and spatial organization of the large phage genome during replication and packaging. Regardless, we conclude that the phage-assembled nucleus-like structure provides a strong protective barrier to DNA-targeting immune pathways.

Methods

Bacterial growth and genetic manipulation

Strains, plasmids, phages, and spacer sequences used in this study are listed in **Tables 2.2-2.5**. *Pseudomonas aeruginosa* strain PAO1 was grown in LB at 37 °C with aeration at 225 RPM. When necessary, plating was performed on LB agar with carbenicillin (250 µg/ml) or gentamicin (50 µg/ml). Gene expression was induced by the addition of L-arabinose (0.1% final) and/or isopropyl β-D-1-thiogalactopyranoside (IPTG, 0.5 mM or 1 mM final). For chromosomal insertions at the attTn7 locus, *P. aeruginosa* cells were electroporated with the integrating vector pUC18T-lac and the transposase expressing helper plasmid pTNS3, and selected on gentamicin. Potential integrants were screened by colony PCR with primers PTn7R and PglmS-down. Electrocompetent cell preparations, transformations, integrations, selections, plasmid curing, and FLP recombinase mediated marker excision with pFLP were performed as described previously[25].

Phage growth and DNA extraction

Phage growth was conducted in LB at 37 °C with PAO1 as a host. Growth curves were conducted in a Biotek Synergy plate reader at 37 °C with orbital shaking set to maximum speed. Phage stocks were diluted and stored in SM buffer[8] and used for routine plaquing assays. Plaque assays were conducted at 37 °C with 20 mL of bottom agar containing 10 mM $MgSO_4$ and 0.35% or 0.7% top agar (often both concentrations were used in parallel) also containing 10 mM $MgSO_4$ and any inducer molecules. 3 µL spots were applied to the top agar after it had been poured and solidified. For high titer lysates to generate phage DNA, plates with a high number of plaques were flooded with SM buffer and collected[8]. The lysates were subsequently DNase treated. Phage DNA was extracted with the Wizard Genomic DNA Purification Kit

(Promega). DNA restriction assays were performed according to standard NEB protocols and restriction fragments were assessed by agarose gel electrophoresis.

Type I-C CRISPR-Cas system assay and expression in *P. aeruginosa* PAO1

Type I-C CRISPR-Cas function was tested by electroporating a strain that naturally contains I-C *cas* genes (strain: F11) with pHERD30T plasmids encoding crRNAs that target phages. To express this system heterologously in PAO1, the four effector *cas* genes (*cas3-5-8-7*) were cloned into pUC18T-lac and inserted in the PAO1 chromosome as described above. After removal of the gentamicin marker, this strain was electroporated with the same pHERD30T crRNA-encoding plasmids to confirm function upon IPTG/arabinose induction.

Type I-F CRISPR-Cas system expression in *P. aeruginosa* PAO1

To express the Type I-F system heterologously in PAO1, all I-F *cas* genes (*cas1, cas3, csy1-4*) were cloned into pMMBHE plasmid and transformed into PAO1. Subsequently, this strain was electroporated with the pHERD30T crRNA-encoding plasmids to confirm function upon IPTG/arabinose induction. To maintain pHERD30T and pMMBHE in the same strain of *P. aeruginosa*, double selection of 30 μg/mL gentamicin and 100 μg/mL carbenicillin was employed.

crRNA cloning and expression

All crRNAs used here, were cloned into established entry vectors in the pHERD30T background. After removing a pre-existing BsaI cut site in the vector by mutagenesis, a pseudo-CRISPR array (i.e. repeat-spacer-repeat for Type I, V, VI, or a sgRNA scaffold for Type II) was then cloned into the vector, where the spacing sequence possessed two inverted BsaI digest sites, to facilitate scarless cloning of new spacers. Desired spacer sequences were chosen randomly across the phage genome (with the correct PAM sequence for the cognate CRISPR-

Cas system) and ordered as two complementary oligonucleotides that generate sticky ends when annealed, to be cloned into the entry vector, which was BsaI digested. Spacer oligonucleotides were PNK treated, annealed, and ligated into the entry vector.

Streptococcus pyogenes (Spy) Cas9 and sgRNA expression in *P. aeruginosa*

The *S. pyogenes Cas9* gene was cloned into a pUC18T-Ara integration vector and then inserted into the attTn7 locus of PAO1. A single guide RNA scaffold was constructed based on a previous design[26] with internal BsaI cut sites to enable insertion of pre-annealed oligos for scarless sgRNA design. This sgRNA scaffold was amplified with primers p30T-gRNA_BsaI and p30T-gRNA_BsaI_rev. The resulting product was inserted into the pHERD30T vector via Gibson assembly following backbone (pJW1) amplification by inverse PCR with primers gRNA_BsaI-p30T and gRNA_BsaI-p30T-rev. The sgRNA scaffold was positioned into pJW1 so that following BsaI cleavage the spacer insert +1 position would conincide with the pBAD TSS +1 position. The resulting plasmid, pJB1, was BsaI digested (NEB) followed by ligation of indicated pre-annealed oligos. **Table 2.5** contains a complete list of all target sequences. The sequence of the sgRNA construct with BsaI site locations is shown in **Table 2.4**.

Cas9 in vitro cleavage

Cas9-based phage genome cleavage *in vitro* was conducted with purified Cas9 protein (NEB #M0386S), and the Cas9-gRNA-tracrRNA based cleavage reaction was then performed according to the manufacturer's (NEB) instructions. Cas9 crRNAs **(Table 2.4)** were ordered as Alt-R CRISPR-Cas9 crRNAs from IDT and utilized without further modification. The tracrRNA was amplified using primers tracrRNA-FOR and tracrRNA-REV from a plasmid (pBR62). The tracrRNA was produced through a T7 RNAP reactions using dsDNA encoding the tracrRNA downstream of a T7 RNAP promoter. Cas9 protein (NEB) was combined with pre-annealed crRNA and tracRNA complex at a 1:1 molar ratio. For targeting KasI-liberated ΦKZ genomic

DNA, 500 ng of ΦKZ DNA was co-incubated with Cas9 RNP, which would cleave at pos. 183,270, and KasI in NEB buffer 3.1 for 1 hour. After stopping the reaction by proteinase K treatment, products were assessed by agarose gel electrophoresis. For Cas9 digestion of whole DNA, the reaction was performed at 37 °C for 4 hrs with 300 ng of ΦKZ or DMS3 genomic DNA and the products were assessed by agarose gel electrophoresis. Two Cas9 guides were selected that would cleave at pos. 158,649 and 168,715 of the ΦKZ genome to liberate an ~10 kb fragment.

Cas12a and crRNA design for PA expression

The humanized allele of the *cpf1* gene of *Moraxella bovoculli* (MBO_03467, KDN25524.1) was sub-cloned from pTE4495 (Addgene) into pUC18T-lac using primers pUC_cpf1_F and pUC_cpf1_R and inserted in the PAO1 chromosome as described above. A Cpf1 repeat-spacer-repeat pseudo-CRISPR array was synthesized as oligonucleotides, annealed, and ligated into a pHERD30T vector, digested with NcoI and HindIII. Spacer sequences were cloned into the resulting vector (pJB2) following BsaI digestion and ligation of pre-annealed spacer oligonucleotide pairs.

Cas13a and crRNA design for PA expression

The wild type allele of the *cas13* gene of *Listeria seeligeri* and *Leptotrichia shahii* were sub-cloned from p2CT-His-MBP-Lse_C2c2_WT and p2CT-His-MBP-Lsh_C2c2_WT (Addgene) into pUC18T-lac. LseCas13 and Lsh Cas13 were inserted in the PAO1 chromosome as described above. An Lse and an Lsh Cas13a repeat-spacer-repeat pseudo-CRISPR array were synthesized as oligonucleotides, annealed, and ligated into a pHERD30T vector, digested with NcoI and EcoRI. Spacer sequences were cloned into the resulting vectors (pSDM057 and pSDM070, respectively) following BsaI digestion and ligation of pre-annealed spacer oligonucleotide pairs. crRNA expression vectors were introduced into PAO1 tn7::*cas13*[Lse] and

PAO1 tn7::*cas13*^{Lsh}. The resulting strains were were assayed for phage sensitivity under standard phage plaing conditions, with induction of both Cas13 and the gRNAs (50 µg/mL gentamicin, 0.1% (L)-arabinose, 1 mM IPTG).

EcoRI expression in *P. aeruginosa*

The wild type allele of the M.*EcoRI* (Methyltransferase) and *R.EcoRI* (Endonuclease) genes were sub-cloned from pSB1A3 EcoRI Methylase-AmilCP and pSB1A3 EcoRI-RTX with EcoRI Methylase-AmilCP (Addgene plasmid # 85166, Addgene plasmid # 85165)p2CT-His-MBP-Lse_C2c2_WT and p2CT-His-MBP-Lsh_C2c2_WT (Addgene) into pHERD30T using Gibson assembly. The resulting plasmids, pSDM160 and pSDM161, were electroporated into PAOI Δ*hsdR* (SDM020). The resulting strains were were assayed for phage sensitivity under standard phage plaing conditions, with induction of both M.EcoRI (for genome protection) and R.EcoRI (50 µg/mL gentamicin, 0.1% (L)-arabinose).

Restriction-Modification Assay

PAO1, PAK, and PAO1 Δ*hsdR* were grown to saturation in LB at 37 °C. 4 mL of 0.7% agar, 10 mM MgSO$_4$ molten top agar were seeded with 100 µL saturated culture and spread on 20 mL 10 mM MgSO$_4$ LB agar plates. 2.5 µL 10-fold serial dilutions of bacteriophage JBD30 and ΦKZ propagated on strain PAO1 and PAK were spotted on plates. Plates were incubated at 37°C overnight and were imaged the following day.

Chromosomal Knockout of *hsdR*

To delete *hsdR* (NP_251422.1) from the PAO1 genome, a PAO1 strain expressing Cas9 from *S. pyogenes* was programmed to express a sgRNA against *hsdR*: GCCCTCATCGAAGAAACCAG. Additionally, the pHERD30T plasmid expressing this sgRNA was engineered to carry a repair template. This repair template consisted of the 500 bp

22

upstream of *hsdR* and the 500 bp downstream of *hsdR* directly enjoined to one another. Induction of Cas9 and sgRNA led to cellular toxicity, as measured by OD_{600} (data not shown). Survivors of Cas9-mediated targeting were isolated. The *hsdR* locus of survivors was then amplified by colony PCR with primers binding outside of the region encompassed by the repair template. Products were resolved by gel electrophoresis in a 1.0% agarose TAE gel at 100 V and visualized using SYBR safe DNA stain. Amplicons of a reduced size were then sequenced, confirming the chromosomal deletion of *hsdR*. A clone with the correct deletion, SDM020, was chosen for downstream experiments.

Construction of fusion proteins

Plasmids expressing cherry alone or cherry tagged with cas3, cas8 (of Type I-C and Type I-F system) and TopA were constructed by Gibson assembly in pHERD30T plasmid digested with SacI and PstI. These fusions have ggaggcggtggagcc (G-G-G-G-A) linker sequence in between them. Cherry was amplified from SF-pSFFV-sfCherryFL1M3_TagBFP (kindly provide by Bo Huang lab, UCSF). cas3, cas8 of the Type I-C and Type I-F systems were amplified from LL77 and PA14 respectively. topA was amplified from gDNA of PAO1. All the primers used for the construction of the plasmids are listed in **Table 2.4**.

Plasmids expressing cherry-Cas9 (pESN28) (primers: prESN74, prESN75, prESN76, prESN77) and Cherry-Cas9-ORF152 (pESN29) (prESN80, prESN81, prESN82, prESN83) were constructed using Gibson Assembly in the pHERD30T plasmid. Cherry-ORF152 (pESN32) (prESN91, prESN92) was constructed by PCR amplification to omit Cas9 and ligated using a KLD reaction. An sgaaaaggsqk linker connects Cherry to the other proteins. A ggggs linker connects Cas9 and ORF152. The plasmid expressing cMyc-ORF152 (prESN153, prESN154) was constructed. All plasmids were designed using SnapGene.

To make translational fusions of proteins, desired gene fragments including Cas9, M.EcoRI, R.EcoRI, ORF152 (NP_803718.1), and sfCherry2 were amplified by PCR from

templates pHERD30T::Cas9, pSB1A3::M.EcoRI-AmilCP, pSB1A3::M.EcoRI-AmilCP;R.EcoRI-

RTX, ΦKZ genomic DNA, and pHERD30T::sfCherry2, respectively. pHERD30T was linearized

by restriction digest with SacI-HF and PstI-HF or by PCR. PCR primers included 15-30 bp

overhangs for Gibson assembly. Overhangs enjoining genes included linkers GGGGS or

GGSGGS. PCRs were treated with DpnI to eliminate template DNA. All PCRs and restriction

digests were purified using a PCR clean up kit (Zymo Research DNA Clean and Concentrator

Kits). Linearized vector and gene fragments treated with Gibson assembly. Gibson assembly

reactions were transformed into competent *E. coli*. Plasmid products were isolated by miniprep

and submitted for sequencing. Correct assemblies were electroporated into appropriate *P.

aeruginosa* strains.

Cas13 escaper isolation

For identifying escapers of Cas13a RNA-targeting, 3 µl of high concentration ɸKZ lysates were

mixed with 150 µl of overnight cultures of SDM078, SDM109 or SDM107 (strains expressing

Cas13a and a gRNA against ɸKZ) . After incubating at 37 °C for 10 minutes, samples were

mixed with 4 mL of 0.7% agar, 10 mM MgSO4, 1 mM IPTG and 0.1% arabinose and plated in

LB agar plates with Gentamicin (50 µg/ml). After overnight incubation, plates were examined for

the presence of escaper plaques. Escapers were formed in SDM078 and SDM109 but not in

SDM107. 10 phages that escaped targeting from SDM078 and SDM109 were purified and the

protospacer locus was amplified using PCR and subsequently sequenced. Six unique outcomes

were identified and shown in Figure 2.4.

Growth curve experiments

Growth curve experiments were carried out in a plate reader as in Borges AL *et al* 2018 *Cell*.

Briefly, cells were diluted 1:100 from a saturated overnight culture with 10 mM MgSO4 and

antibiotics and inducers, as appropriate. 140 µL of diluted culture was added, together with 10

µL of phage to wells in a 96-well plate. This plate was cultured with maximum double orbital

rotation at 37 °C for 24 hours with OD600 measurements every 5 minutes

Immunofluorescence

Sample Growth

5 mL overnight cultures of a strain expressing Cas9 and an sgRNA targeting ΦKZ (SDM065)

and a strain expressing cMyc-ORF152 (bESN27) were grown at 30 °C in LB media with

gentamicin. A 1:30 back-dilution of the overnight culture into LB was grown at 30 °C for 1 h.

Protein and guide expression was induced with 0.1% arabinose and 0.5 mM IPTG, respectively.

After 1 h of expression, an aliquot of uninfected cells was fixed while the remaining cultures

were infected with ΦKZ using MOI 1.5. Infected cell aliquots were collected and fixed at 60 mpi.

Fixation

This protocol was adapted from ref. [27]. Samples were fixed with 5X Fix Solution (12.5%

paraformaldehyde, 150 mM KPO4, pH 7.2) and incubating for 15 minutes at room temperature

followed by 20 minutes on ice. Samples were then washed in PBS 3 times and finally

resuspended in GTE (50 mM glucose, 10 mM EDTA, pH 8.0, 20 mM Tris-HCl, pH 7.65) with 10

ug/mL lysozyme. Resuspended cells were transferred to polylysinated coverslips and dried.

Once dry, coverslips were incubated in cold methanol for 5 minutes followed by cold acetone for

5 minutes. Cells were rehydrated by a rinse in PBS followed by a 3-minute incubation in PBS +

2% BSA blocking solution. Cells were incubated with a 1:50 dilution of primary antibody (Cas9

(7A9-3A3): sc-517386 or cMyc (9E10): sc-40) in PBS + 2% BSA for 1 hr followed by 3, 7 minute

washes in fresh PBS + 2% BSA. Coverslips were then incubated in the dark for 1 hr with

secondary antibody (goat anti-mouse Alexa Fluor 555, Life Technologies A-21424) diluted 1:500

in PBS + 2% BSA. DAPI was added for the final 10 minutes of the incubation. Cells were

washed in PBS 3 times for 7 minutes. Coverslips were then placed on slides using mounting

media (v/v 90% glycerol, v/v 10%Tris pH 8.0 and w/v 0.5% propyl-gallate) and sealed with clear nail polish.

Microscopy and Analysis

Images were collected using a Zeiss Axiovert 200M microscope. Images were later processed using NIS Analysis software at the UCSF NIC. Compartments and cells were manually selected using the Simple ROI Editor. Background subtractions were conducted for each cell and compartment separately. The corrected intensities for the cytoplasmic and compartment regions were averaged over the area of each cell and compartment using Matlab and plotted using Prism 6. Data from two pooled replicates was fitted with a line, showing 95% confidence intervals with dashed lines. The slope is as reported in the plots.

Live cell imaging

For live cell imaging of ɸKZ infection, freshly grown cells from LB plates were picked and resuspended in 100 µl of LB media. 5-10 µL of samples are spotted on 0.85% agarose pads with 1:5 diluted LB. Arabinose (0.01% to 0.05%) and DAPI (2 µg/ml). Samples were then incubated in a humidified chamber and allowed to grow for 3 hours at 30 °C. 5 µL ɸKZ lysate was spotted on top of the agar pad and the samples were grown for an additional 1 hour. Cells were visualized in the microscope after covering them with a cover slip. Nikon Ti2-E inverted microscope equipped with Perfect Focus System (PFS) and Photometrics Prime 95B-25mm camera were used for live cell imaging. Images were processed using NIS Elements-AR software.

Figures
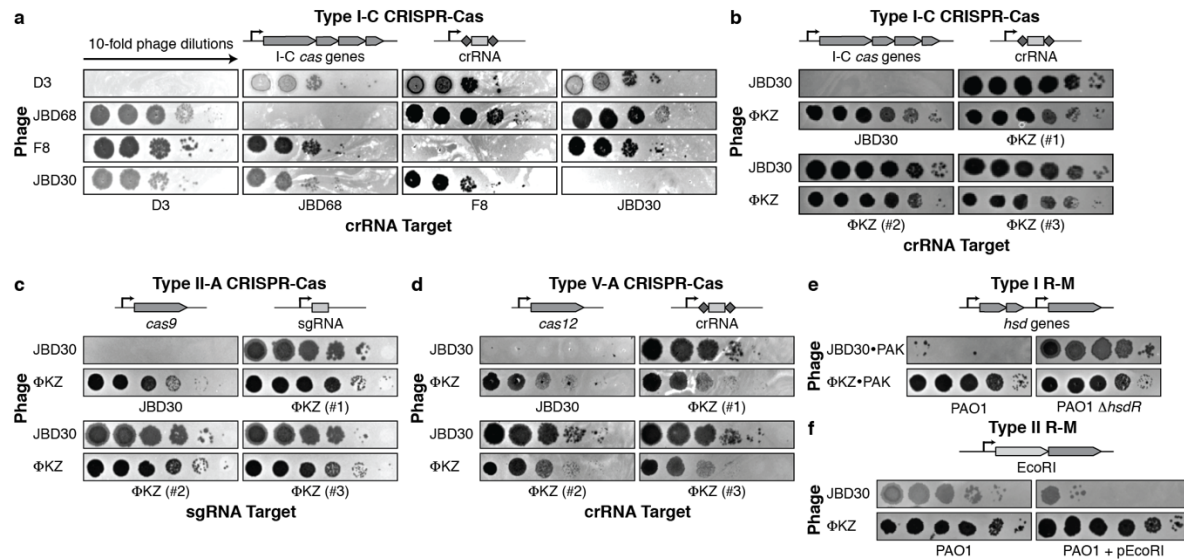
Figure 2.1 | Identification of a phage that resists targeting by diverse CRISPR-Cas and restriction-modification systems.
Plaque assay with the indicated phage spotted in ten-fold serial dilutions on a lawn of P. aeruginosa, dark clearings in the lawn represent phage replication. Strain PAO1 expressing: **a,** Type I-C cas genes (cas3-5-8-7) and crRNAs targeting the indicated phages. **b,** Type I-C cas genes and crRNAs targeting phage JBD30 or distinct phage ΦKZ-targeting crRNAs (#1-#3). **c,** Type II-A cas9 and distinct single guide RNAs (sgRNAs) targeting the indicated phage. **d,** Type V-A cas12a and distinct crRNAs against the indicated phage. **e,** Endogenous Type I R-M system (hsdRSM) in strain PAO1 was assayed using phages propagated on PAK (e.g. JBD30•PAK was first propagated on strain PAK). Together with an isogenic PAO1ΔhsdR knockout. **f,** Type II EcoRI R-M system. Restriction activity was assayed using phages JBD30 and ΦKZ. All plaque assays replicated ≥ 2 times with similar results.
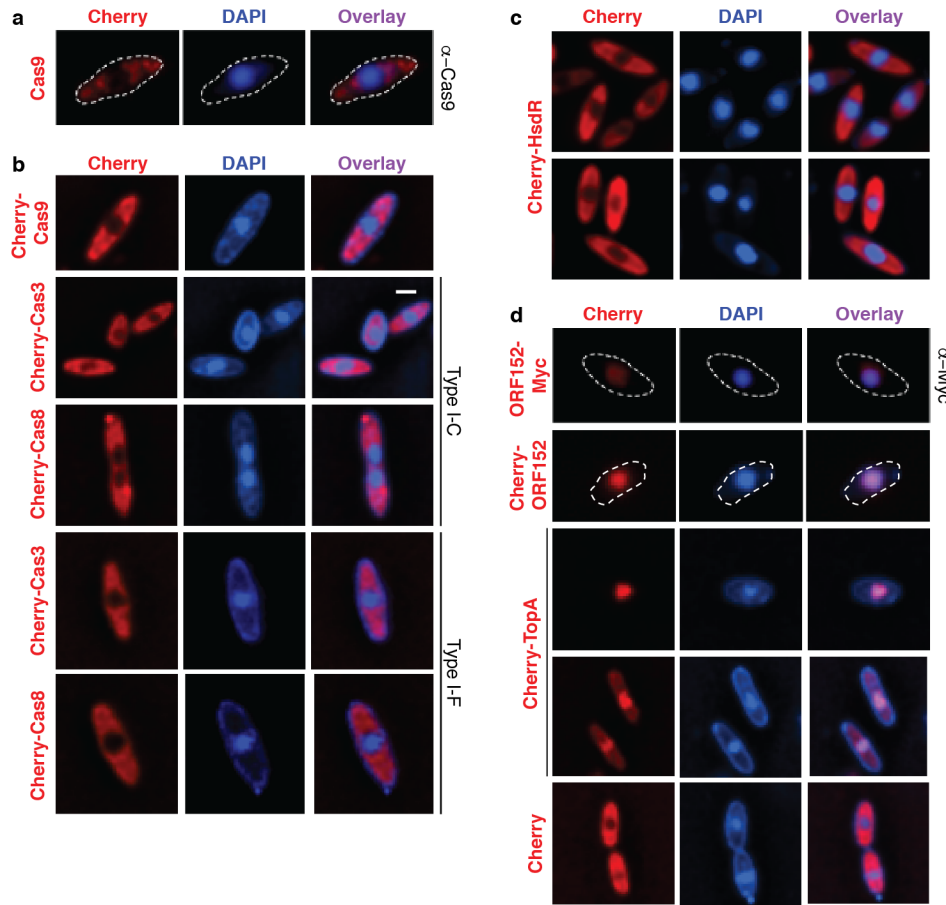
Figure 2.2 | CRISPR-Cas and restriction proteins are excluded from ΦKZ's nucleus-like structure.
a, Fluorescence microscopy of P. aeruginosa immunostained for Cas9, DAPI stain shows the phage DNA within the nucleus-like structure. Live fluorescence microscopy of P. aeruginosa strains engineered to express **b,** II-A Cas9 or I-C or I-F Cas8 or Cas3 proteins fused to Cherry, **c,** a Cherry-HsdR fusion, **d,** Immunostained for Myc-ORF152 (top panels), or live imaging of ORF152 and TopA proteins fused to Cherry, or Cherry alone. All experiments were replicated ≥ 2 times with similar results.
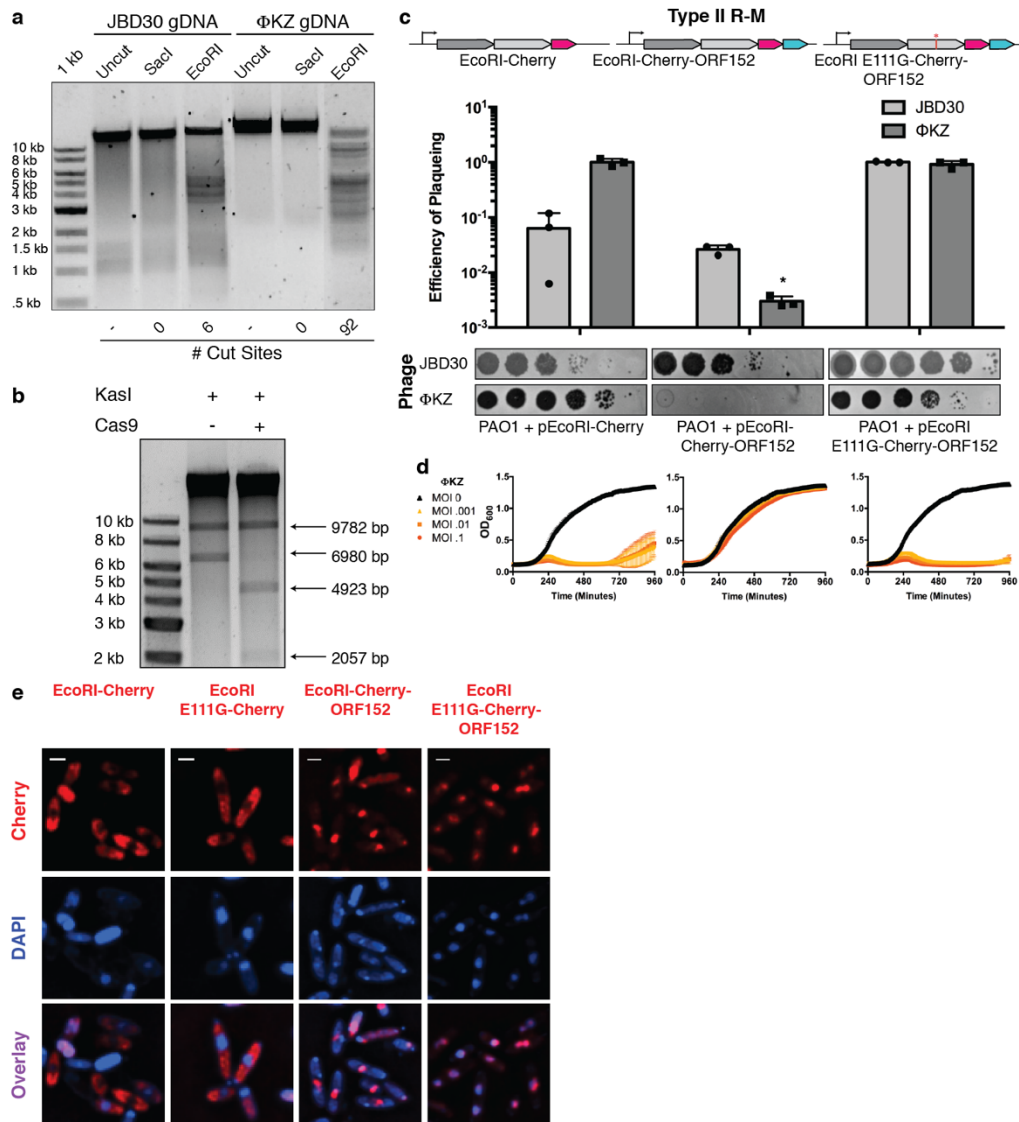
Figure 2.3 | ΦKZ genomic DNA can be cleaved by immune enzymes.
a, ΦKZ and JBD30 genomic DNA digested with the indicated restriction enzymes *in vitro*. The first lane contains a 1 kb DNA ladder. The number of cut sites for each enzyme is shown at the bottom of the gels. Products were visualized on a 0.7% agarose gel, visualized with SYBR Safe nucleic acid stain. **b,** ΦKZ phage genomic DNA digested *in vitro* using KasI, and incubated with and without Cas9 loaded with crRNA:tracrRNA targeting the fragment liberated by KasI. Products were visualized on a 0.7% agarose gel, visualized with SYBR Safe nucleic acid stain. **c,** Strain PAO1 expressing EcoRI-Cherry, EcoRI-Cherry-ORF152, or EcoRI E111G-Cherry-ORF152 fusion protein. Plaque assays were conducted as in Fig. 2.1 and quantified (n=3). Mean values are plotted as bars with error bars representing standard deviation. A t-test comparing ΦKZ EOP on PAO1 pEcoRI-Cherry to PAO1 pEcoRI-Cherry-ORF152 yielded a p-value of 2.88×10^{-4}. **d,** Growth curves monitoring the OD_{600} of PAO1 cells infected with the indicated ΦKZ multiplicity of infection (MOI). **e,** Live fluorescence imaging of infected *P. aeruginosa* strains engineered to express EcoRI-Cherry, EcoRI E111G-Cherry, EcoRI-Cherry-ORF152 or EcoRI E111G-Cherry-ORF152. DAPI stain shows the phage DNA. Cherry shows EcoRI fusion protein. *In vitro* digestion experiments a and b replicated twice with similar results. Plaque assays l, growth curves (d), and microscopy (e) were replicated ≥ 3 times.

29
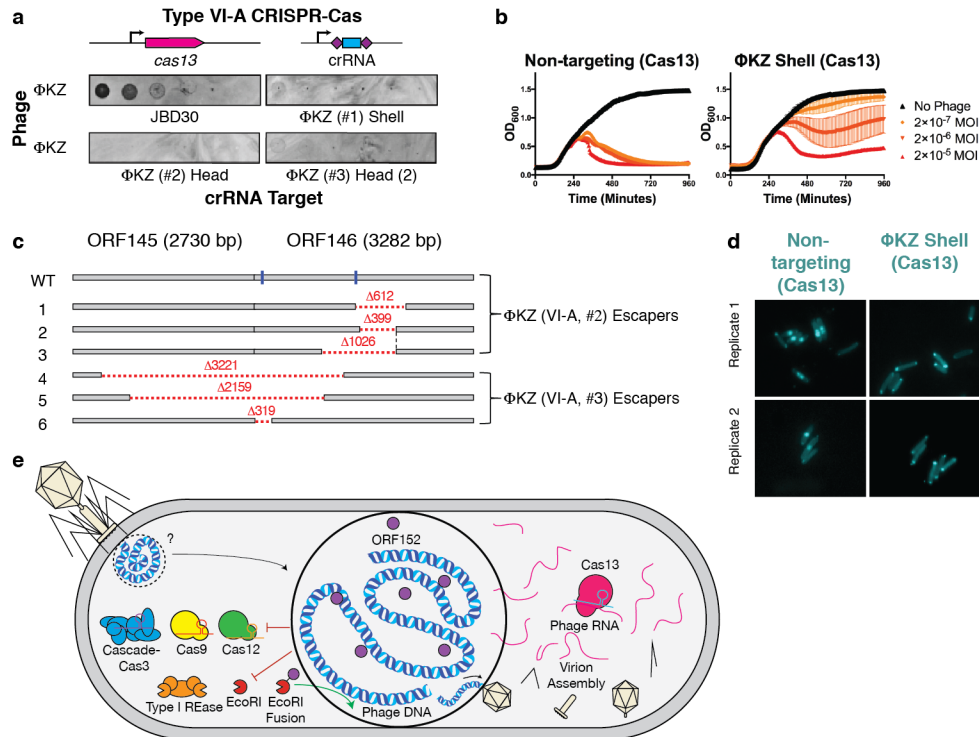
Figure 2.4 | Phage ΦKZ DNA is sensitive to RNA-targeting Cas13.
a, PAO1 expressing LseCas13a and a crRNA targeting phage ΦKZ. Plaque assays were conducted as in Fig. 2.1. **b,** Growth curves measuring the OD_{600} of PAO1 infected with the indicated ΦKZ MOI. **c,** Cas13 escaper ΦKZ phage mutant deletions (red dashed lines) with target sites indicated (blue line). **d,** Live fluorescence imaging of *P. aeruginosa* strains engineered to express LseCas13a and crRNAs targeting ΦKZ. Cyan stain shows the phage DNA. **e,** Model summarizing the ΦKZ nucleus-like structure excluding (flat arrow) Cas9, Cas12, Cascade-Cas3 (Type I-C, Type I-F) and Type I restriction endonucleases (REase) and Type II REase, while the mRNA (red) is exported and can be targeted by Cas13. The nucleus-like structure is resistant to the indicated nucleases, but EcoRI fusion (to internal protein ORF152) enables targeting. Cas13 plaque assays a were replicated >3 times with similar results. Growth curve experiments b were replicated twice with similar results. Escaper mutants c were isolated once and verified by PCR, sequencing, and plaque assays. Microscopy d was replicated twice with similar results.

30
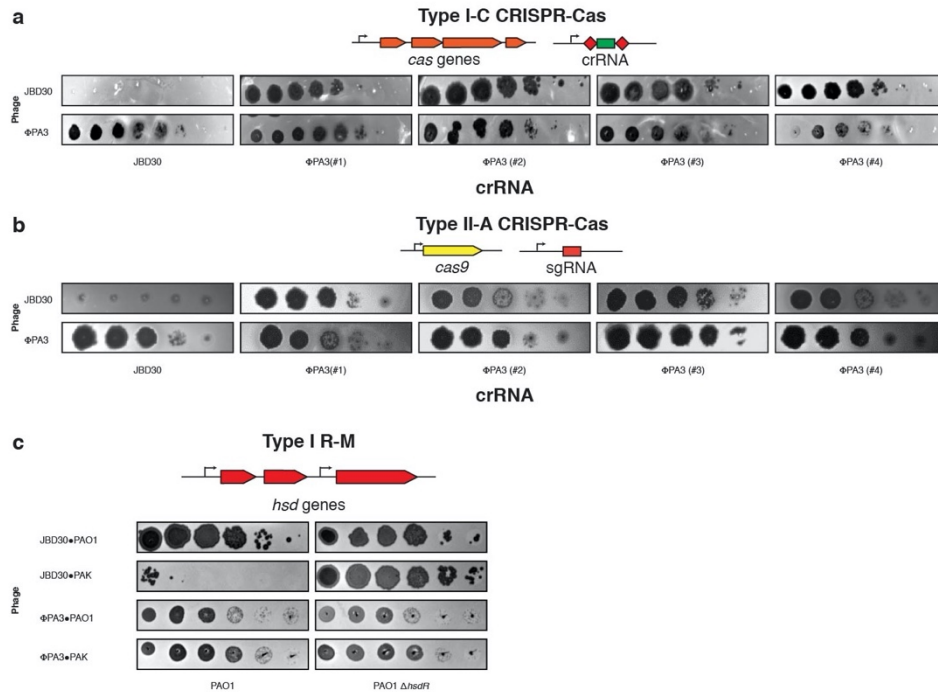
Figure 2.5 | Jumbo phage ΦPA3 resists targeting by CRISPR Cas and a restriction-modification system.

a, Strain PAO1 was engineered to express the I-C *cas* genes and distinct crRNAs targeting the indicated phages, and plaque assays were conducted as in Fig. 2.1. **b,** Strain PAO1 was engineered to express the Type II-A Cas9 protein and distinct single guide RNAs (sgRNAs) targeting the indicated phage. Plaque assays were conducted as in Figure 2.1. **c,** The endogenous Type I R-M system (*hsdRSM*) in strain PAO1 was assayed using phages propagated on PAO1 or PAK as indicated (e.g. JBD30•PAO1 was first propagated on strain PAO1). Together with an isogenic PAO1Δ*hsdR* knockout, all strains were subjected to a plaque assay as in Figure 2.1. All plaque assays were replicated twice with similar results.
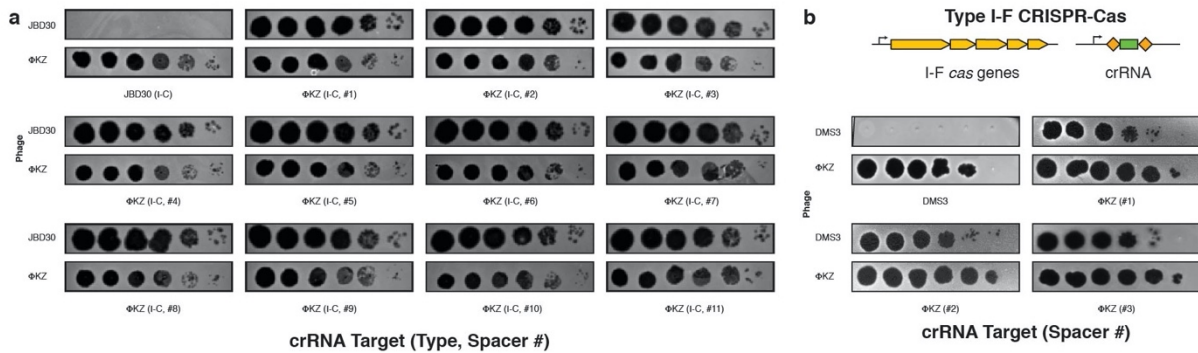
Figure 2.6 | Phage ΦKZ resists P. aeruginosa Type I-C and Type I-F CRISPR-Cas immunity.
a, Strain PAO1 was engineered to express the I-C *cas* genes and distinct crRNAs targeting phage JBD30 and phage ΦKZ, and plaque assays were conducted as in Fig. 2.1. **b,** Strain PAO1 was engineered to express the I-F *cas* genes and distinct crRNAs targeting phage JBD30 and phage ΦKZ, and plaque assays were conducted as in Fig. 2.1. All plaque assays replicated ≥ 2 times with similar results.
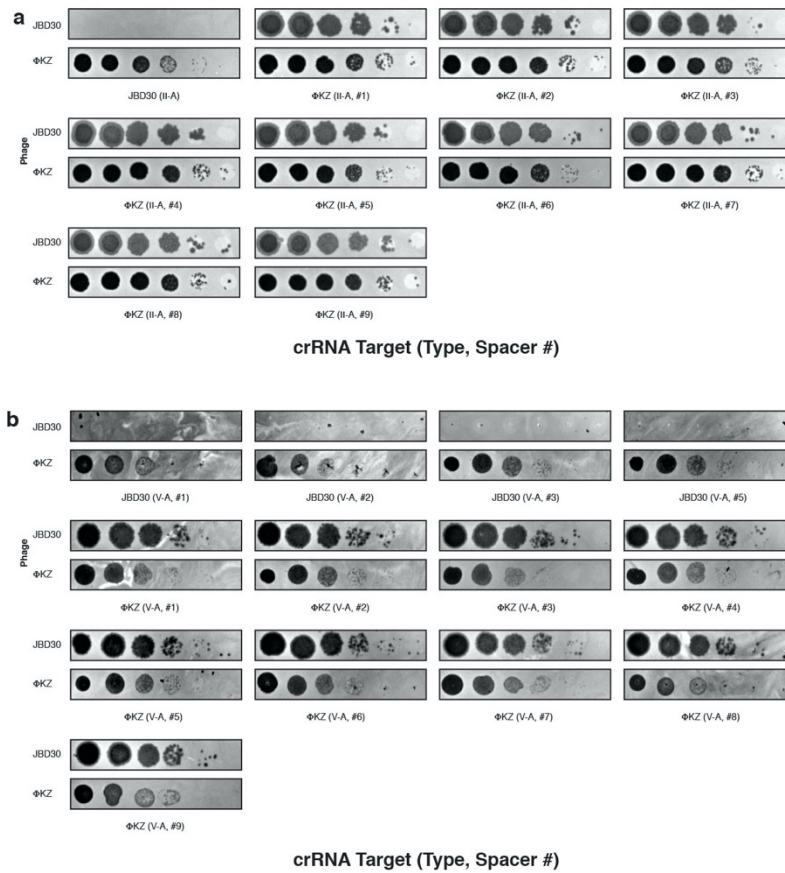
Figure 2.7 | Phage ΦKZ resists targeting by heterologous Type II-A and V-A CRISPR-Cas systems.
a, Strain PAO1 was engineered to express the Type II-A Cas9 protein and distinct single guide RNAs (sgRNAs) targeting the indicated phage. Plaque assays were conducted as in Figure 2.1.
b, Strain PAO1 was engineered to express the Type V-A Cas12a protein and distinct crRNAs against the indicated phage. Plaque assays were conducted as in Figure 2.1. All plaque assays replicated ≥ 2 times with similar results.
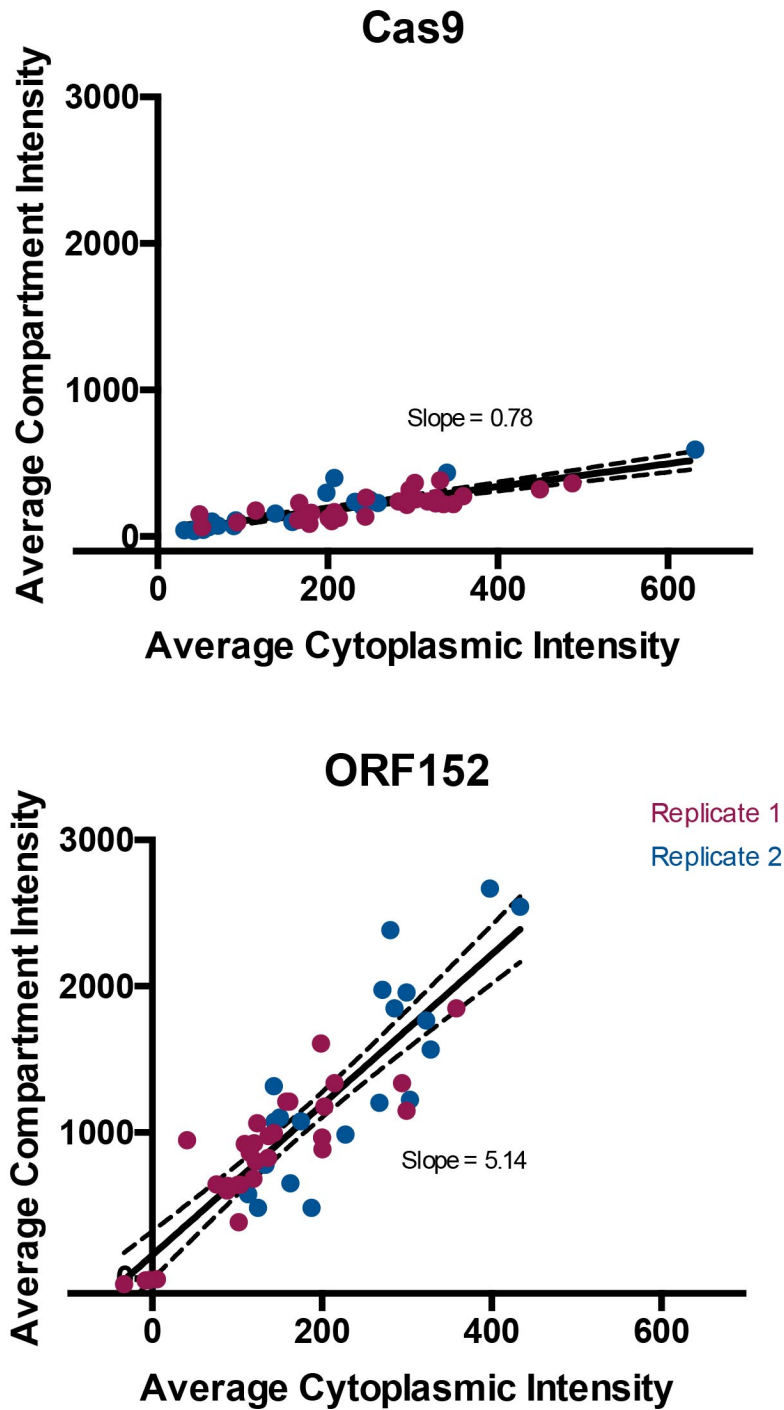
Figure 2.8 | Quantification of Cas9 localization during phage ΦKZ infection of P. aeruginosa.
Localization of Cas9 and ORF152 in the cytoplasm and shell during ΦKZ infection were quantified. Data points (individual cells) from two pooled replicate experiments were fitted with a line, showing 95% confidence intervals with dashed lines. The slope is as reported in the plots.
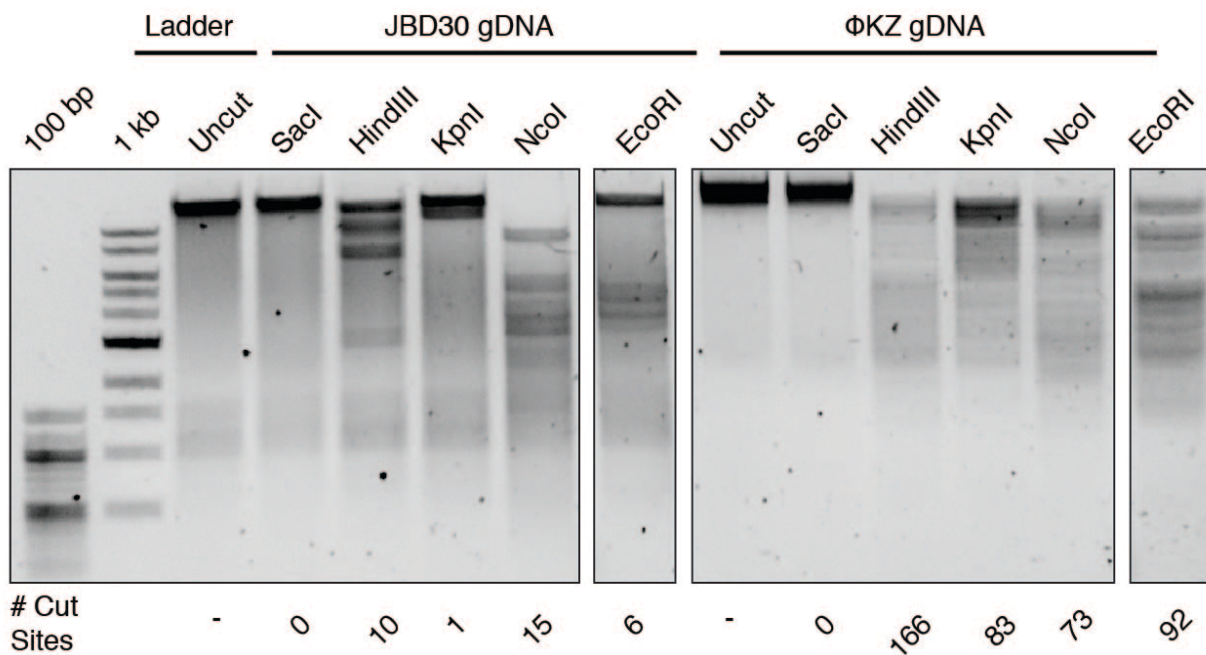
Figure 2.9 | Phage ΦKZ genomic DNA is susceptible in vitro to cleavage by Restriction Endonucleases.
Genomic DNA was isolated from phages ΦKZ and JBD30 and was subjected to digestion with the indicated restriction enzymes *in vitro*. The number of cut sites for each enzyme is shown at the bottom of the gels. Products were visualized on a 0.7 % agarose gel, visualized with SYBR Safe nucleic acid stain. *In vitro* digestion experiment was replicated twice with similar results.
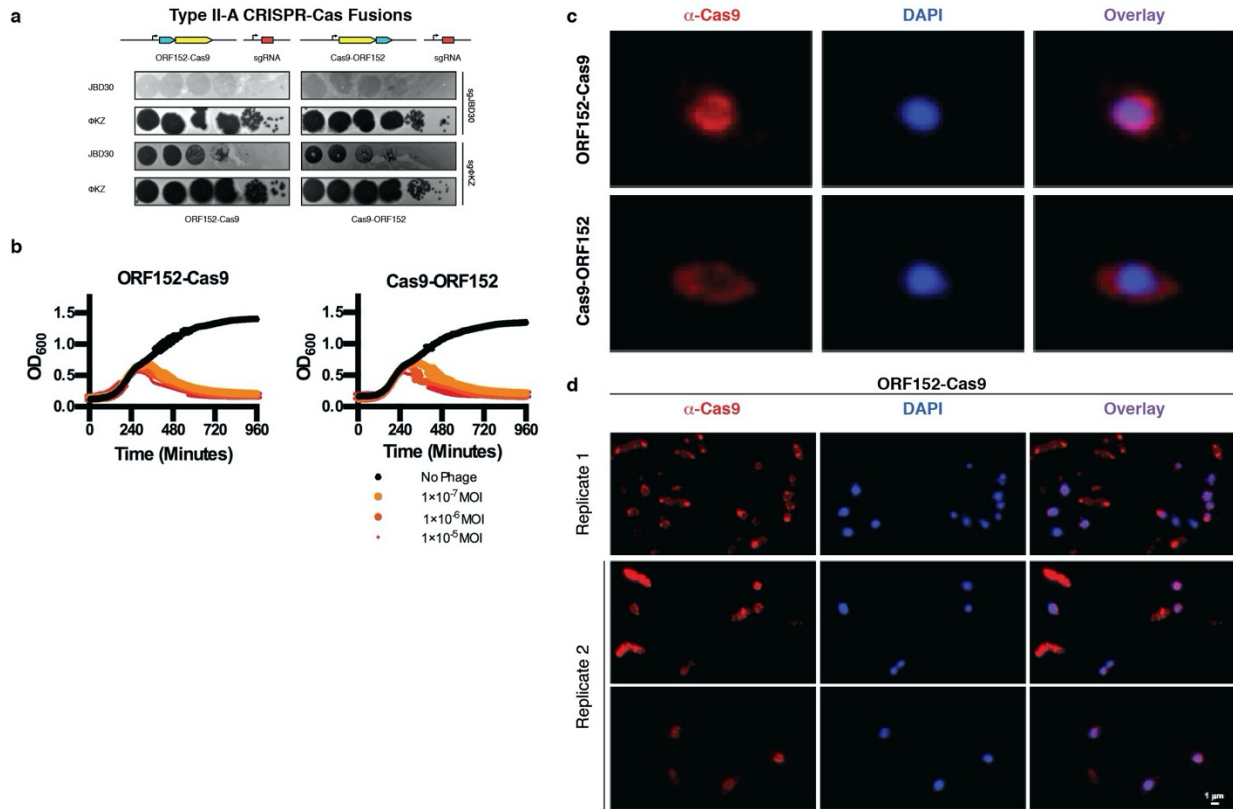
Figure 2.10 | Cas9 fusion to ORF152 localizes to periphery of shell, but does not enable immune activity against ΦKZ.
a, Strain PAO1 was engineered to express Cas9 fused to ORF152 at either the N- or C-terminus, with single guide RNAs (sgRNAs) targeting the indicated phage. Plaque assays were conducted as in Fig. 2.1, **b,** growth curves were conducted, monitoring the OD_{600} of PAO1 cells infected with the indicated ΦKZ multiplicity of infection (MOI). **c,** Fluorescence microscopy of PAO1 fusion strains, immunostained for Cas9, in cells expressing ORF152-Cas9 or Cas9-ORF152. DAPI stain shows the phage DNA within the shell. **d,** Fluorescence microscopy of *P. aeruginosa*, immunostained for Cas9, in cells expressing ORF152-Cas9. DAPI stain shows the phage DNA within the shell. All plaque assays were replicated 4 times with similar results. Growth curve experiments were replicated three times with similar results. Microscopy was replicated twice with similar results.
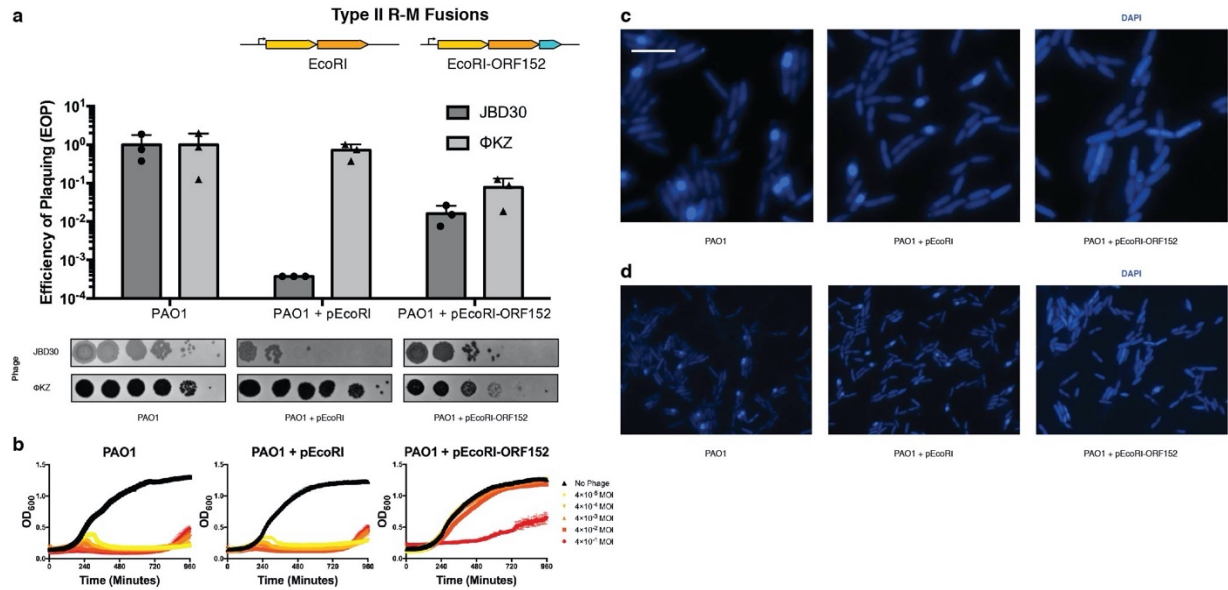
Figure 2.11 | Fusion of EcoRI restriction enzymes to ORF152 enables immune activity.
a, Strain PAO1 was engineered to express EcoRI or EcoRI-ORF152 fusion protein. Plaque assays were conducted as in Fig. 2.1 and quantified (n=3). **b,** Growth curves were conducted, monitoring the OD_{600} of PAO1 cells infected with the indicated ΦKZ multiplicity of infection (MOI). **c,** Live fluorescence imaging of *P. aeruginosa* strains engineered to express EcoRI or EcoRI-ORF152. DAPI stain shows the phage DNA. **d,** Live fluorescence imaging of *P. aeruginosa* strains engineered to express EcoRI, or EcoRI-ORF152. DAPI stain shows the phage DNA within the shell. Wide field of view. All plaque assays were replicated 3 times with similar results. Growth curve experiments were replicated twice with similar results. Microscopy was replicated twice with similar results.
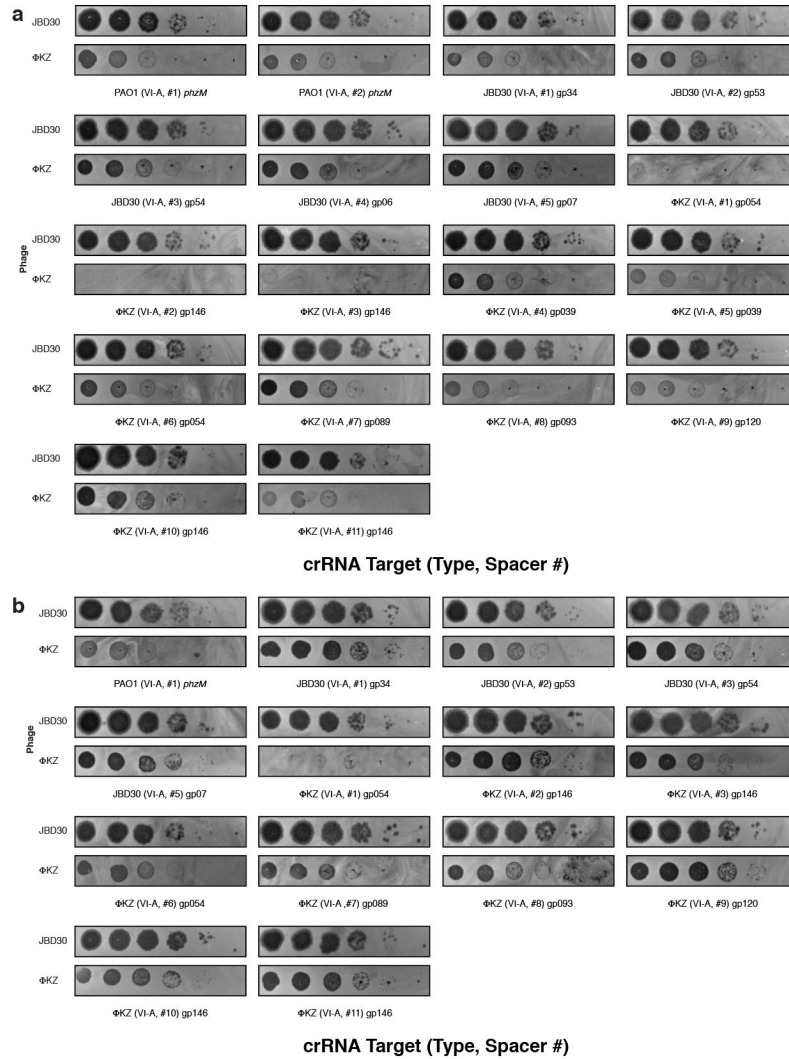
Figure 2.12 | Phage ΦKZ DNA is sensitive to RNA-targeting Cas13.
a, Strain PAO1 expressing LseCas13a and distinct crRNAs targeting the indicated gene. Plaque assays conducted as in Figure 2.1. **b,** Strain PAO1 expressing LshCas13a and distinct crRNAs targeting the indicated gene. Plaque assays conducted as in Figure 2.1. All plaque assays were replicated twice with similar results.
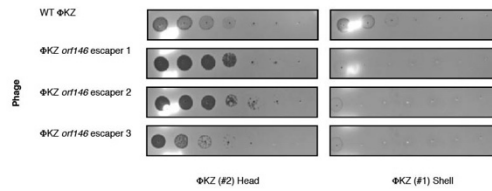
Figure 2.13 | Phage ΦKZ escaper mutants are selected by Cas13-mediated RNA targeting. Strain PAO1 expressing LseCas13a and a crRNA targeting the indicated gene. Plaque assays conducted as in Figure 2.1 using wild type and escaper mutant ΦKZ. WT ΦKZ is targeted by both strains and the faint clearings observed here are not observed as plaques in a full-plate assay. All plaque assays were replicated twice with similar results.
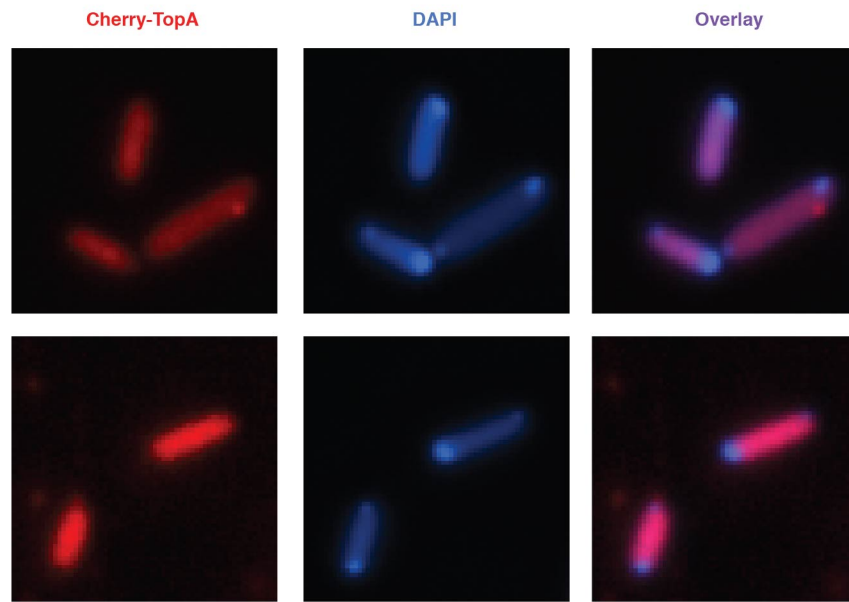
Figure 2.14 | Observation of DAPI-stained phage DNA adjacent to a Cherry-TopA nascent shell.
Live fluorescence imaging of *P. aeruginosa* strains engineered to express Cherry-TopA infected with ΦKZ. DAPI stain labels DNA. Microscopy was replicated three times with similar results.

Tables

Table 2.1 | ΦKZ and ΦKZ-like phages have no natural spacers matching their genomes from a natural collection of >4000 *P. aeruginosa* spacers.

The total number of Type I CRISPR spacers with a perfect match to the indicated P. aeruginosa phages assayed in this study and previous CRISPR-Cas studies. The experimental sensitivity of each phage to the indicated subtypes are shown. AcrIE3 and AcrIF1 are I-E and I-F anti-CRISPR proteins, respectively. * indicates that all spacers have mismatches (≤4) to the F8 genome.

Phage	# spacers	Type I CRISPR sensitivity	Reference
DMS3m	75	I-C: sensitive I-E: resistant (AcrIE3) I-F: sensitive	This study, ref. 3,8
JBD30	51	I-C: sensitive I-E: sensitive I-F: resistant (AcrIF1)	This study, ref. 2,3
JBD18	51	I-F: sensitive	Ref. 8
JBD25	46	I-F: sensitive	Ref. 8
JBD68	28	I-C: sensitive	This study
D3	49	I-C: sensitive	This study
F8	3*	I-C: sensitive	This study
ΦKZ	0	I-C: resistant I-F: resistant	This study
phiPA3	0	I-C: resistant	This study
PaBG	0	not assayed	
KTN4	0	not assayed	
PA7	0	not assayed	

Table 2.2 | Phage and Strains

Name	Source	Reference
ΦKZ	Davidson Lab	[28]
ΦPA3	Agard Lab	[12]
DMS3m	O'Toole Lab	[8]
JBD30	Davidson Lab	[29]
F8	Davidson Lab	Unpublished Accession: DQ163917
D3	Davidson Lab	[30]
JBD68	Davidson Lab	[29]
	Genotype	**Reference**
F11	Wild-type isolate (Native I-C system)	Unpublished (Davidson Lab)
JW31	PAO1 tn7::I-C/Cas3^{PA}	This study
JB10	PAO1 tn7::cas9^{Spy}	This study
JB90	PAO1 tn7::cas12a^{Mb}	This study
SDM084	PAO1 tn7::cas13^{Lse}	This study
SDM020	PAO1 Δ*hsdR*	This study

Table 2.3 | Plasmids

Name	Information	Ref.
pHERD30T	Arabinose inducible, gentamicin resistant shuttle vector	[31]
pUC18T-Lac	IPTG inducible, Amp^R/Gent^R, Tn7 integrative plasmid with FRT sites flanking gentamicin cassette	[32]
pTNS3	Expresses tnsABCD for flipping out cassette flanked by FRT sites	[33]
pBR62	Contains Spy Cas9 tracrRNA sequence	This study
pJW1	pHERD30T with BsaI site at pos. 235 GAGACC mutated to GTGACC	This study
pJW13	pJW1 with Type I-C pseudo-CRISPR array for spacer cloning	This study
pJB1	pJW1 with Type II-A sgRNA backbone at the +1 TSS of pBAD	This study
pJB2	pJW1 with Type V-A pseudo-CRISPR array for spacer cloning	This study
pSDM057	pJW1 with Type VI-A (LseCas13a) pseudo-CRISPR array for spacer cloning	This study
pSDM070	pJW1 with Type VI-A (LshCas13a) pseudo-CRISPR array for spacer cloning	This study
pSDM046	pHERD30T with ORF152-Cas9 Fusion	This study
pSDM051	pHERD30T with Cas9-ORF152	This study
pSDM160	pHERD30T with M.EcoRI-AmilCP	This study
pSDM161	pHERD30T with M.EcoRI-AmilCP;R.EcoRI	This study
pSDM164	pHERD30T with M.EcoRI-AmilCP;R.EcoRI-ORF152	This study
pSDM166	pHERD30T with M.EcoRI-AmilCP;R.EcoRI-sfCherry2-ORF152	This study
pSDM168	pHERD30T with M.EcoRI-AmilCP;R.EcoRI E111G-ORF152	This study
pSDM169	pHERD30T with M.EcoRI-AmilCP;R.EcoRI E111G-sfCherry-ORF152	This study
pSDM170	pHERD30T with M.EcoRI-AmilCP;R.EcoRI E111G-sfCherry2	This study
pSDM171	pHERD30T with M.EcoRI-AmilCP;R.EcoRI-sfCherry2	This study
pSG-IF-cas	IPTG inducible, Carb^R pMMBHE plasmid with all Type I-F cas genes (cas1,cas3,csy1-4)	This study
pAB04	pHERD30T with Type I-F pseudo-CRISPR array for spacer cloning	This study
pSG30T-cherry	Expresses cherry from arabinose inducible pHERD30T vector	This study
pSG30T-cherry-cas3(IC)	Expresses cherry-tagged Cas3 (Type I-C) from arabinose inducible pHERD30T vector	This study
pSG30T-cherry-cas8(IC)	Expresses cherry-tagged Cas8 (Type I-C) from arabinose inducible pHERD30T vector	This study
pSG30T-cherry-cas3(IF)	Expresses cherry-tagged Cas3 (Type I-F) from arabinose inducible pHERD30T vector	This study
pSG30T-cherry-cas8(IF)	Expresses cherry-tagged Cas8 (Type I-C) from arabinose inducible pHERD30T vector	This study
pSG30T-cherry-TopA	Expresses cherry-tagged TopA (PA01) from arabinose inducible pHERD30T vector	This study
pESN28	Expresses cherry-Cas9	This study
pESN29	Expresses Cherry-Cas9-ORF152	This study
pESN32	Expresses Cherry-ORF152	This study
pESN34	Expresses cMyc-ORF152	This study

Table 2.4 | Oligonucleotides, g-blocks, crRNAs

Name	Sequence	Notes
I-C_BsaI_for	gatccGTCGCGCCCCGCACGGGCGCGTGGATTGAAACgagaccTCTCTGGACAAAggtctcGTCGCGCCCCGCACGGGCGCGTGGATTGAAACa	Underlined regions indicate location of BsaI sites
I-C_BsaI_rev	agcttGTTTCAATCCACGCGCCCGTGCGGGGCGCGACgagaccTTTGTCCAGAGAggtctcGTTTCAATCCACGCGCCCGTGCGGGGCGCGACg	Underlined regions indicate location of BsaI sites
sgRNA scaffold sequence	ccatagagaccACGTACGTACggtctcAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT	Underlined regions indicate location of BsaI sites
p30T-gRNA_BsaI	ctctactgtttctccatccatagagaccacgtacgtacg	
p30T-gRNA_BsaI_rev	gcccaaaaaaacgggtccgggcaggataggtgaag	
gRNA_BsaI-p30T	atggagaaacagtagagagttgc	
gRNA_BsaI-p30T-rev	acccgtttttttgggctag	
Cpf1_crRNA_top	catgaaatttctactgtttgtagatGgagaccTCTCTGGACAAAggtctcGaaatttctactgtttgtagat	Underlined regions indicate location of BsaI sites
Cpf1_crRNA_bottom	agctatctacaaacagtagaaatttCgagaccTTTGTCCAGAGAggtctcCatctacaaacagtagaaatt	Underlined regions indicate location of BsaI sites
tracrRNA-FOR	gaaattaatacgactcaactatagaaaacagcatagcaagttaaaata	For T7 RNAP in vitro rxn
tracrRNA-REV	aaaaaaagcacccgactcggtgccac	For T7 RNAP in vitro rxn
PTn7R	cacagcataactggactgaatttc	Ref. [32]
PglmS-down	gcacatcgggcgacgtgctctc	Ref. [32]
LseCas13a crRNA Top	catggggtaagagactacctctatatgaaagaggactaaaaccgagaccacgtacgtacggtctccggtaagagactacctctatatgaaagaggactaaaacg	Underlined regions indicate location of BsaI sites
LseCas13a crRNA Bottom	aattcgttttagtcctctttcatatagaggtagtctcttaccggagaccgtacgtacgtggtctcggttttagtcctctttcatatagaggtagtctcttaccc	Underlined regions indicate location of BsaI sites
LshCas13a crRNA Top	catggggatttagaccaccccaatatcgaaggggactaaaaccgagaccacgtacgtacggtctccggatttagaccaccccaatatcgaaggggactaaaacg	Underlined regions indicate location of BsaI sites
LshCas13a crRNA Bottom	aattcgttttagtccccttcgatattggggtggtctaaatccggagaccgtacgtacgtggtctcggttttagtcccttcgatattggggtggtctaaatccc	Underlined regions indicate location of BsaI sites
Alt-R CRISPR-Cas9 crRNA		
KZ (IV) – KasI	rArUrC rUrUrG rCrGrC rCrArG rArUrC rArCrG rUrGrG rUrUrU rUrArG rArGrC rUrArU rGrCrU	Pos. 183,270
KZ (IV) – 1	rGrArArUrCrUrGrCrUrArArUrArArGrGrU rUrCrArGrUrUrUrUrArGrArGrCrUrArUrGrC rU	Pos. 158,649
KZ (IV) – 2	rUrCrArCrCrArCrGrCrArUrUrArCrArUr rCrArGrGrUrUrUrUrArGrArGrCrUrArUrGrC rU	Pos. 168,715
DMS3m (IV) - 1	rGrCrC rGrArC rArUrU rUrUrC rCrArG rUrUrG rGrCrG rUrUrU rUrArG rArGrC rUrArU rGrCrU	Pos. 17,751
DMS3m (IV) - 2	rUrCrA rCrGrA rCrGrA rCrCrC rArGrA rArGrC rGrUrG rUrUrU rUrArG rArGrC rUrArU rGrCrU	Pos. 28,033
Fusion primers		
prESN80	CTAAAGCAAAAGCTtaaAAGCTTGGCACTGGCCG	Cherry-Cas9-ORF152
prESN81	AGCCACCGCCACCgtcacctcctagctgactcaaatcaat	
prESN82	ctaggaggtgacGGTGGCGGTGGC	
prESN83	AGTGCCAAGCTTttaAGCTTTTGCTTTAGCTAGTTCAAGAG	
prESN74	tgagtcagctaggaggtgactaaAAGCTTGGCACTGGCC	Cherry-Cas9

Name	Sequence	Notes
prESN75	attgagtatttcttatccatTTTTTGGCTGCCTCCTGC	Cherry-Cas9
prESN76	CTGCAGGAGGCAGCCAAAAatggataagaaatactcaataggcttagatatcgg	
prESN77	CGGCCAGTGCCAAGCTTttagtcacctcctagctgactcaaatcaatgcgt	
prESN91	ATGTTTGGTAAACATTTCGAACGTCCA	
prESN92	TTTTTGGCTGCCTCCTGCAG	
prESN153	AGACCTGGGTGGCGGTGGCTCG	cMyc-ORF152
prESN154	TCTTCGCTAATCAGTTTCTGTTCCATAGAGCTCGAATTCTTATCAG	
F-30T-sfCh_P29	tacccatgggatctgataagaattcgagctATGGAGGAGGACAACATGGCC	pSG30T-cherry
R-csy1-link-sfCh(TAA-)_P30	GTAGGCGTTGGGAGGGGAGAGGTggctccaccgcctccGCCGCCGGTGCTGTGTCTGGC	
F-link-csy1 (ATG-)_P31	ggaggcggtggagccACCTCTCCCCTCCCAACGCCTAC	
R-30T-csy1_P32	gacggccagtgccaagcttgcatgcctgcaTCAGTCACGCTCATCTTCGAG	
F-30T-sfCh_P29	tacccatgggatctgataagaattcgagctATGGAGGAGGACAACATGGCC	pSG30T-cherry-cas3(IC)
R-Cas3(IC)-link-chry_P101	TGAGTATCGCTAGCCTCCGCGTCggctccaccgcctccGCCGCCGGTGCTGTGTCTGG	
F-link-Cas3(IC)_P102	ggaggcggtggagccGACGCGGAGGCTAGCGATACTCA	
R-30T-Cas3(IC)_P103	gacggccagtgccaagcttgcatgcctgcaCTACCAACATAGCCGCTCGC	
F-30T-sfCh_P29	tacccatgggatctgataagaattcgagctATGGAGGAGGACAACATGGCC	pSG30T-cherry-cas8(IC)
R-Cas8c-link-Cherry_P104	TAGTCATTGAGGGCCGAAAGGATggctccaccgcctccGCCGCCGGTGCTGTGTCTGG	
F-link-Cas8c_P105	ggaggcggtggagccATCCTTTCGGCCCTCAATGACTA	
R-30T-Cas8c_P106	gacggccagtgccaagcttgcatgcctgcaTTACTCGACGGAATCGGGGC	
F-30T-sfCh_P29	tacccatgggatctgataagaattcgagctATGGAGGAGGACAACATGGCC	pSG30T-cherry-cas3(IF)
R-csy1-link-sfCh(TAA-)_P30	GTAGGCGTTGGGAGGGGAGAGGTggctccaccgcctccGCCGCCGGTGCTGTGTCTGGC	
F-link-csy1 (ATG-)_P31	ggaggcggtggagccACCTCTCCCCTCCCAACGCCTAC	
R-30T-csy1_P32	gacggccagtgccaagcttgcatgcctgcaTCAGTCACGCTCATCTTCGAG	
F-30T-sfCh_P29	tacccatgggatctgataagaattcgagctATGGAGGAGGACAACATGGCC	pSG30T-cherry-cas8(IF)
R-cas3 (ATG-)-link-sfCh_P42	CATTGCGACACCAGCAGGATGTTggctccaccgcctccGCCGCCGGTGCTGTGTCTGG	
F-link-cas3_P43	ggaggcggtggagccAACATCCTGCTGGTGTCGCAATG	
R-30T-cas3_P44	gacggccagtgccaagcttgcatgcctgcaTCAGTTGTATTTCTTGAACC	
F-30T-sfCh_P29	tacccatgggatctgataagaattcgagctATGGAGGAGGACAACATGGCC	pSG30T-cherry-TopA
R-topA-link-ch_P107	TCCACGATGACCAGCGATTTACCggctccaccgcctccGCCGCCGGTGCTGTGTCTGG	
F-link-topA_P108	ggaggcggtggagccGGTAAATCGCTGGTCATCGTGGA	
R-30T-topA_P109	gacggccagtgccaagcttgcatgcctgcatcaGCGCTTGTCTTCGACCTTC	
M.EcoRI-AmilCP_fwd	CTGATAAGAGGAGGACAGCTATGGCCGGCGCACGCAAC	pSDM160
M.EcoRI-AmilCP_rev	TAA AAC GAC GGC CAG TGC CAT TAA CCG GTG GCG ACC ACA GG	
M.EcoRI-AmilCP;R.EcoRI-rev	TAAAACGACGGCCAGTGCCATTAACCGGTTTTGCTAGTCAGCTGCTC	pSDM161

Name	Sequence	Notes
M.EcoRI-AmilCP;R.EcoRI-rev	TAAAACGACGGCCAGTGCCATTAACCGGTTTTGCTAGTCAGCTGCTC	pSDM164
ORF152_fwd	GGTGGCGGTGGCTCGATGTTTGGTAAACATTTCGAAC	
ORF152_rev	TAAAACGACGGCCAGTGCCATTAAGCTTTTGCTTTAGCTAG	
GGS(2)-sfCherry2_fwd	GGCGGGTCGGGCGGCTCGATGGAGGAGGACAACATG	pSDM166
sfCherry2_rev	TGGGTTTCATGCCGCCGGTGCTGTGTCT	
EcoRI-mut Rev	ctggtgctttgctccggcaactaac	pSDM168, pSDM169, pSDM170
EcoRI-mut Fwd	gttagttgccggagcaaagcaccag	
GS-Cas9 F	GGTGGCGGTGGCTCGatggataagaaatactcaataggc	pSDM046
Cas9-p30T R2	ccagtgccaagcttgcatgcctgcaTTAGTCACCTCCTAGCTGACTC	
pHERD30T-ORF152 F	GGGATCTGATAAGAATTCGAGCTatgtttggtaaacatttcgaacg	
ORF152-GS R	CGAGCCACCGCCACCagcttttgctttagctagttc	
p30T-Cas9 F2	atgggatctgataagaattcgagctATGGATAAGAAATACTCAATAGGC	pSDM051
Cas9-p30T R2	ccagtgccaagcttgcatgcctgcaTTAGTCACCTCCTAGCTGACTC	

Table 2.5 | Spacer sequences

Phage (Type, #)	Sequence
JBD30 (I-C)	AACCTCGCGGCATCCGCAACAACAACCCCGGCAA
D3 (I-C)	ACGATTGCGGACATGGCAGGCTGCCGCTGCTGGA
F8 (I-C)	GCCAATCGGCCGATAGATGAAGCTGTGGAGGGTC
JBD68 (I-C)	AGCGGCGTGAGGTTGGACCTTGCTGCCGACCATT
ΦKZ (I-C, #1)	TGGACTAACAAATACGCCTATATATTCCGATCCT
ΦKZ (I-C, #2)	GAACTTGTATTTAACCCAAAGGTTTTTAAATGGT
ΦKZ (I-C, #3)	CCCATTTATTATTTTCTTTATTTATCCAACCGTA
ΦKZ (I-C, #4)	TAAAAAGAAAATTATATAAATAGTATATTATGTG
ΦKZ (I-C, #5)	TTTACATTCTTCTAAACTAATATTTAATTCATCT
ΦKZ (I-C, #6)	CATTATCATCTACCTCTTTTAATTTATCTTTAAT
ΦKZ (I-C, #7)	CAAAAGGATTATTTGATGTTGTGGTGAAAGAAAA
ΦKZ (I-C, #8)	GCCATTTTCTTTCACCACAACATCAAATAATCCT
ΦKZ (I-C, #9)	AGATAATGGGGATATTTTGTATTTTGATAACAAG
ΦKZ (I-C, #10)	TACTGGTACCCATAGAAGTTATATTTTACCAGCT
ΦKZ (I-C, #11)	ATTGAAATCAAGTAAAATCTCAAATGGAATCTGT
JBD30 (II-A)	GGCATCCGCAACAACAACCC
ΦKZ (II-A, #1)	GAATCTGCTAATAAGGTTCA
ΦKZ (II-A, #2)	TCACCACGCATTACATCAGG
ΦKZ (II-A, #3)	AAATTATATTAATCACAATG
ΦKZ (II-A, #4)	ATATATTCCGATCCTTATCC
ΦKZ (II-A, #5)	AAACATCCTCATGATAACCA
ΦKZ (II-A, #6)	AGCAGTAGCTTGAGTTTGAA
ΦKZ (II-A, #7)	TTTTAGATGAAGTAAAAAAG
ΦKZ (II-A, #8)	CTATTACCATTTTCGTCAAT
ΦKZ (II-A, #9)	CTCATTTTTTATTCCTACGT
ΦKZ (II-A, #10)	AGAGAACTGTTTAAACACAA
JBD30 (V-A, #1)	CGTTTTACCCACTGAGCGAACGC
JBD30 (V-A, #2)	GCCCGTTTCGATACCGCACATA
JBD30 (V-A, #3)	TACCGCGCCGCCCTTCTGGAGGA
JBD30 (V-A, #4)	AAGGTGCCGCACGGTGTCCCAGT
JBD30 (V-A, #5)	CGCATGACTCTCTATATGGGGCC
ΦKZ (V-A, #1)	GAATCTGCTAATAAGGTTCATGG
ΦKZ (V-A, #2)	AATAGGAATATAGCTATGCTAAT
ΦKZ (V-A, #3)	TGTGTCTCTTTTCCAAATGCTTT
ΦKZ (V-A, #4)	GGGATTCAACTATTGGAAAAGCA
ΦKZ (V-A, #5)	GCTAGTTGTTCATCAAATGATGA
ΦKZ (V-A, #6)	CTCATTAATAACAGATACTTTGT
ΦKZ (V-A, #7)	CTGGGCATTAATGACGATATATC
ΦKZ (V-A, #8)	TTACAGCCTCGTCAGACAGGTAA
ΦKZ (V-A, #9)	CCTAATGCATTCCATTTAAATAC
PAO1 (VI-A, #1)	TCAAATTACGCGCAGCAGCAAGAT
PAO1 (VI-A, #2)	CGGCCTGCAGGATGGCCTTGGTCA
JBD30 (VI-A, #1)	GCGGCCAGCCCGGCCCCTCGTCCA
JBD30 (VI-A, #2)	GGCGCGTTGATGCGGACCTGGCCA
JBD30 (VI-A, #3)	AATGCCCTTCTCCCGGCAAACCGT
JBD30 (VI-A, #4)	TACCCGCAAGTTGTTGAGGGCTGA
JBD30 (VI-A, #5)	GGTGCCGGCCGGCTTGATGCCCAT
ΦKZ (VI-A, #1)	GCAGGAGCAGTAGCTTGAGTTTGA
ΦKZ (VI-A, #2)	TCATTAGTTTCAACCCAGTATGAA
ΦKZ (VI-A, #3)	ACATAATCTTCAAATGCAGAAGCC
ΦKZ (VI-A, #4)	AACCAGCACCACCACAAAAGTAAA
ΦKZ (VI-A, #5)	AGAATTACTAGTGCATTTAGTACT
ΦKZ (VI-A, #6)	AACGATGTAAGGAGTGAACTGCGG
ΦKZ (VI-A, #7)	TAACAGCTTGTAGATAATAACCAG
ΦKZ (VI-A, #8)	GTCTGCAATAGTCTTCGGATTATT
ΦKZ (VI-A, #9)	AGGTTAGCATTGGAGTTACCCATT

Phage (Type, #)	Sequence
ΦKZ (VI-A, #10)	CAGCACCTTTGGAAACTACCCAAT
ΦKZ (VI-A, #11)	GCTTCCATAGTAACTAAGTATGCT
ΦPA3 (II-A, #1)	TCGTATGTCTCGTGCATCTT
ΦPA3 (II-A, #2)	ATGTGTAAAGTTCTCTGATG
ΦPA3 (II-A, #3)	CTCCCACTTATTCCATCTCA
ΦPA3 (II-A, #4)	GCTGTACCAGTCCACTTGTC
ΦPA3 (II-A, #5)	CTCACTCATTAGACAAGTTT
DMS3 (I-F)	AACGGCCGACGCTTCTGGGTCGTCGTGAAAGT
ΦKZ (I-F, #1)	GGCATTATGAAGTGGACCTAATGTATCAATAA
ΦKZ (I-F, #2)	GATGGCTACCATCATTAATACTCTTAAGACTT
ΦKZ (I-F, #3)	AGGAATTGATGATGACATTAAAGTAATTGATT
ΦKZ (I-F, #4)	ATTCTACTAATAAATGCCGGTGACAGAAATTA
ΦKZ (I-F, #5)	TCGTGGATTAAGTGATAGTGGATATTTACAAT

References

1. Koonin, E. V., Makarova, K. S. & Wolf, Y. I. Evolutionary Genomics of Defense Systems in Archaea and Bacteria. *Annu Rev Microbiol* 71, annurev–micro–090816–093830 (2017).

2. Bondy-Denomy, J., Pawluk, A., Maxwell, K. L. & Davidson, A. R. Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature* 493, 429–432 (2013).

3. Pawluk, A., Bondy-Denomy, J., Cheung, V. H. W., Maxwell, K. L. & Davidson, A. R. A new group of phage anti-CRISPR genes inhibits the type I-E CRISPR-Cas system of Pseudomonas aeruginosa. *mBio* 5, e00896–e00896–14 (2014).

4. Pawluk, A. *et al.* Inactivation of CRISPR-Cas systems by anti-CRISPR proteins in diverse bacterial species. *Nature Microbiology* 1, 1–6 (2016).

5. van Belkum, A. *et al.* Phylogenetic Distribution of CRISPR-Cas Systems in Antibiotic-Resistant Pseudomonas aeruginosa. *mBio* 6, e01796–15 (2015).

6. Makarova, K. S. *et al.* An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Micro* 13, 722–736 (2015).

7. Marino, N. D. *et al.* Discovery of widespread Type I and Type V CRISPR-Cas inhibitors. *Science* 362, eaau5174–242 (2018).

8. Cady, K. C., Bondy-Denomy, J., Heussler, G. E., Davidson, A. R. & O'Toole, G. A. The CRISPR/Cas adaptive immune system of Pseudomonas aeruginosa mediates resistance to naturally occurring and engineered phages. 194, 5728–5738 (2012).

9. Pawluk, A. *et al.* Naturally Occurring Off-Switches for CRISPR-Cas9. *Cell* 167, 1829–1838.e9 (2016).

10. Rauch, B. J. *et al.* Inhibition of CRISPR-Cas9 with Bacteriophage Proteins. *Cell* 168, 150–158.e10 (2017).

11. Chaikeeratisak, V. *et al.* Assembly of a nucleus-like structure during viral replication in bacteria. *Science* 355, 194–197 (2017).

12. Chaikeeratisak, V. *et al.* The Phage Nucleus and Tubulin Spindle Are Conserved among Large Pseudomonas Phages. *Cell Rep* 20, 1563–1571 (2017).

13. Kraemer, J. A. *et al.* A phage tubulin assembles dynamic filaments by an atypical mechanism to center viral DNA within the host cell. *Cell* 149, 1488–1499 (2012).

14. Erb, M. L. *et al.* A bacteriophage tubulin harnesses dynamic instability to center DNA in infected cells. *Elife* 3, e03197 (2014).

15. Zehr, E. A. *et al.* The structure and assembly mechanism of a novel three-stranded tubulin filament that centers phage DNA. *Structure* 22, 539–548 (2014).

16. Bryson, A. L. *et al.* Covalent Modification of Bacteriophage T4 DNA Inhibits CRISPR-Cas9. *mBio* 6, e00648–15–9 (2015).

17. Strotskaya, A. *et al.* The action of Escherichia coli CRISPR-Cas system on lytic bacteriophages with different lifestyles and development strategies. *Nucleic Acids Research* 45, 1946–1957 (2017).

18. Vlot, M. *et al.* Bacteriophage DNA glucosylation impairs target DNA binding by type I and II but not by type V CRISPR-Cas effector complexes. *Nucleic Acids Research* 46, 873–885 (2018).

19. Huang, L. H., Farnet, C. M., Ehrlich, K. C. & Ehrlich, M. Digestion of highly modified bacteriophage DNA by restriction endonucleases. *Nucleic Acids Research* 10, 1579–1591 (1982).

20. Abudayyeh, O. O. *et al.* C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* 353, aaf5573 (2016).

21. Gootenberg, J. S. *et al.* Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science* 356, eaam9321–442 (2017).

22. Meeske, A. J. & Marraffini, L. A. RNA Guide Complementarity Prevents Self-Targeting in Type VI CRISPR Systems. *Mol Cell* 71, 791–801.e3 (2018).

23. Al-Shayeb, B. *et al.* Clades of huge phage from across Earth's ecosystems. *bioRxiv* 3, 572362 (2019).

24. Chaikeeratisak, V. *et al.* Viral Capsid Trafficking along Treadmilling Tubulin Filaments in Bacteria. *Cell* 177, 1771–1780.e12 (2019).

25. Choi, K.-H. & Schweizer, H. P. mini-Tn7 insertion in bacteria with single attTn7 sites: example Pseudomonas aeruginosa. *Nat Protoc* 1, 153–161 (2006).

26. Jinek, M. *et al.* A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* 337, 86–821 (2012).

27. Cowles, K. N. *et al.* The putative Poc complex controls two distinct Pseudomonas aeruginosa polar motility mechanisms. *Molecular Microbiology* 90, 923–938 (2013).

28. Mesyanzhinov, V. V. *et al.* The genome of bacteriophage phi KZ of Pseudomonas aeruginosa. *Journal of Molecular Biology* 317, 1–19 (2002).

29. Bondy-Denomy, J. *et al.* Prophages mediate defense against phage infection through diverse mechanisms. *The ISME Journal* 10, 2854–2866 (2016).

30. Kropinski, A. M. Sequence of the genome of the temperate, serotype-converting, Pseudomonas aeruginosa bacteriophage D3. 182, 6066–6074 (2000).

31. Qiu, D., Damron, F. H., Mima, T., Schweizer, H. P. & Yu, H. D. PBAD-Based Shuttle Vectors for Functional Analysis of Toxic and Highly Regulated Genes in Pseudomonas and Burkholderia spp. and Other Bacteria. *Applied and Environmental Microbiology* 74, 7422–7426 (2008).

32. Choi, K.-H. & Schweizer, H. P. mini-Tn7 insertion in bacteria with single attTn7 sites: example Pseudomonas aeruginosa. *Nat Protoc* 1, 153–161 (2006).

33. Choi, K.-H. *et al.* Genetic tools for select-agent-compliant manipulation of Burkholderia pseudomallei. *Applied and Environmental Microbiology* 74, 1064–1075 (2008).

Chapter 3

The φPA3 phage nucleus is Enclosed by a Self-Assembling 2D Crystalline Lattice

Abstract

To protect themselves from host attack, numerous jumbo bacteriophages establish a phage nucleus—a micron-scale, proteinaceous structure encompassing the replicating phage DNA. Bacteriophage and host proteins associated with replication and transcription are concentrated inside the phage nucleus while other phage and host proteins are excluded, including CRISPR-Cas and restriction endonuclease host defense systems. Here, we show that nucleus fragments isolated from φPA3 infected *Pseudomonas aeruginosa* form a 2-dimensional lattice, having p2 or p4 symmetry. We further demonstrate that recombinantly purified primary *Ph*age N*u*clear E*n*closure (PhuN) protein spontaneously assembles into similar 2D sheets with p2 and p4 symmetry. We resolve the dominant p2 symmetric state to 3.9 Å by cryo-EM. Our structure reveals a two-domain core, organized into quasi-symmetric tetramers. Flexible loops and termini mediate adaptable inter-tetramer contacts that drive subunit assembly into a lattice and enable the adoption of different symmetric states. While the interfaces between subunits are mostly well packed, two are open, forming channels that likely have functional implications for the transport of proteins, mRNA, and small molecules.

Introduction

There is a constant evolutionary pressure for bacteria to develop defense mechanisms against invading bacteriophages and for the phages to develop effective countermeasures.[1] To that end, restriction-modification and numerous CRISPR systems are widespread amongst bacterial hosts while phages have developed their own DNA modification and anti-CRISPR systems.[2] A subset of so-called jumbo bacteriophages, defined by having genomes exceeding 200 kb, have recently been shown to encode an elaborate system for sequestering the phage

genome away from host nucleolytic attack, conveying broad resistance to DNA targeting by the host.[3] This is accomplished via the assembly of a selectively permeable protein shell that encompasses the phage genome.[4,5] The shell with its contents is referred to as the phage nucleus for its remarkable functional similarity to the eukaryotic nucleus: this structure forms a selectively permeable compartment that is centered within the host bacteria by a dynamic bipolar spindle formed from a phage encoded, divergent tubulin called PhuZ.[4,6–8] The bacteriophage and host proteins involved in phage replication and transcription are concentrated within the phage nucleus shell while translation and nucleotide synthesis machinery, the aforementioned CRISPR-Cas and restriction endonucleases, as well as other host and exogenous proteins are effectively excluded.[3,4] Unlike other proteinaceous prokaryotic compartments, such as carboxysomes or viral capsids, this shell can grow significantly throughout infection to reach nearly micron-scale, a process likely driven by genome replication.[4,9–11] In contrast to some other membraneless compartment forming systems, such as protein condensates, subunits do not diffuse throughout the shell of the phage nucleus, indicative of unique assembly properties.[4,12]

While the total protein composition of the phage nucleus shell is unknown for any jumbo phage, Gp105 from the ϕKZ family phage 201ϕ2-1 was shown to be a marker for the shell.[4] It is the most highly expressed protein that is not part of the mature phage particle.[4] Here we formally introduce this protein and its related family members as *ph*age n*u*cleus e*n*closure or PhuN proteins. The PhuN family currently includes Gp53 (ϕPA3), Gp54 (ϕKZ), Gp105 (201ϕ2-1), Gp202 (PCH45), and numerous putative homologous proteins from newly sequenced jumbo bacteriophages.[4,5,13,14] Beyond these other phage proteins, PhuN has no clear previously characterized homologous relatives. This unique assortment of biophysical and biological properties combined with the mystery about the tertiary structure of PhuN established it as an exciting target for structural and biochemical analysis.

In this work, we demonstrate that phage nucleus fragments isolated from ϕPA3 infected

P. aeruginosa cells form a quasi-square lattice. We further show that Gp53, the PhuN family member from bacteriophage ɸPA3, readily assembles into large 2D lattices *in vitro*. We utilized a limited tilt, cryo-electron microscopy data collection scheme paired with single particle processing to determine a 3.9 Å map of PhuN assemblies. Assisted by RoseTTAFold[15] and AlphaFold,[16] we present an atomic model and analysis of the assembly enclosing the phage nucleus. The structure reveals a two-domain core that assembles into tetramers. Flexible termini and large loops mediate adaptable inter-tetramer contacts that drive shell assembly into a largely p2 symmetric lattice. While the interfaces between subunits are mostly well packed, two of the interfaces are open, forming clear channels that likely have important functional implications. Despite the lack of detectable sequence homology, analysis of the atomic structure reveals that the basic architecture is derived from a fusion of acetyltransferase and nuclease domains.

Results

Isolation of Phage Nucleus Fragments

 P. aeruginosa were infected with ɸPA3 and lysed after 60 minutes of infection. The lysates were separated using differential centrifugation to isolate large fragments of the phage nucleus. The resulting isolates resemble full or fragmented shells and have a clearly defined lattice structure **(Fig. 1a)**. Two-dimensional averaging of the isolated fragments further reveals a lattice with subunits of both p2 and p4 symmetry with unit cell lengths and angles of 120 Å, 120° and 110 Å, 90°, respectively **(Fig. 1a)**. The subunits forming the lattice appear to be a single, approximately 60 by 40 Å protein species. Recent low resolution *in situ* tomography of phage nuclear shells from phages 201ɸ2-1 and Goslar reported p4 symmetry but is otherwise consistent with this observation, demonstrating a lattice with similarly sized subunits.[17] The isolated shell fragments often appear folded over and the lattice can vary in directionality, suggestive of both structural plasticity and imperfections as well as locally altered symmetry

54

within the lattice of endogenous mature phage nuclear shells **(Fig. 1a)**.

PhuN Self-Assembles During Purification

ϕPA3 PhuN is a 66.6 kDa protein with marginal solubility upon expression in *E. coli,* a problem surmounted by the addition of an N-terminal maltose-binding protein (MBP) solubility tag. Importantly, the addition of MBP does not prevent PhuN from incorporating into the phage nuclear shell during ϕPA3 infection and fragments isolated from MBP-PhuN expressing *P. aeruginosa* after ϕPA3 infection exhibit a lattice similar to that composed of only wild type subunits **(Supplementary Fig. 1a-c)**. All *in vitro* data presented here utilize the MBP-PhuN fusion and, for simplicity, will generally be referred to as PhuN.

During purification, exposure to an anion exchange column induced the formation of a broad spectrum of PhuN assemblies observed upon elution **(Fig. 1b)**. While the exact reason for this preferential on-column assembly is not known, it likely occurs as a result of increased local PhuN concentration during sample application, binding, and elution. Exposure to the high surface charge density of the column resin may also play a role.

These assemblies were grouped into three broad classes based on their appearance in negative stain election microscopy. Class I primarily contains monomeric PhuN **(Fig. 1b)**. Class II includes rounder species approximately 25 nm in diameter with distinct edges. Despite efforts to optimize the production of these for structure analysis, they proved to be an elusive species, often outnumbered by more irregular assemblies like those found in Class III **(Fig. 1b)**. Class III consists of highly variable 50-130 nm long, elongated assemblies **(Fig. 1b)**. The Class III species revealed a striated organization highlighting the striking diversity in both the shape and size of PhuN assemblies. In cryo-EM, although heterogeneously shaped, Class III assemblies appear to have a more regular, lattice-like organization and showed a density around the exterior consistent with the presence of the N-terminally fused MBP **(Supplementary Fig. 1d,e)**.

55

Growing and Imaging PhuN 2D Lattices

We initially set out to determine conditions that would allow us to control the assembly rate and state of PhuN. Using the anion exchange column as an assay to monitor shifts in assembly states, we instead found that at pH 6.5 monomeric PhuN forms large, 2D polycrystalline assemblies when applied to negatively charged grids for negative stain EM **(Supplementary Fig. 2b)**. The low pH was also necessary to maintain the integrity of shell fragments isolated from infected cells, suggesting that pH may be an important factor to consider in shell assembly *in vivo.* Stress induced by antibiotic exposure has been observed to decrease the cytosolic pH to 6.7 in *P. aeruginosa* and other bacteria.[18] Perhaps phage infection may act as a similar metabolic stressor and have an analogous impact on pH.

These 2-dimensional crystals were further optimized using a 2D surface crystallization method as often used in 2D electron crystallography.[19] Purified His-tagged PhuN was diluted into a buffer droplet in Teflon wells that had a pre-formed lipid monolayer containing 21% Ni-NTA modified lipids **(Supplementary Fig. 3)**. This approach yielded larger, better ordered crystalline assembly monolayers than both those observed in negative stain and those isolated from infected cells. These arrays were then prepared for and imaged by cryo-EM. On conventional grids, few lattice patches could be found and those that did adsorb were only visible in very thick ice. We overcame this challenge by utilizing amino functionalized graphene oxide grid supports[20] for the bulk of our data acquisition as they stabilized the lattices in somewhat thinner ice and greatly increased the number of lattice patches adsorbed. Despite optimizing lattice formation, the 2D sheets generally included many local regions of distinct lattice orientation as well as numerous bends and waves **(Fig. 1c and Supplementary Fig. 4)**. Consequently, we chose a predominantly real-space approach, analogous to single particle analysis,[21] instead of the more traditional 2D electron crystallography strategy **(Supplementary Fig. 5)**. To obtain the necessary views, data were collected at multiple tilt angles: 0°, 15°, 30°,

35°, 40°, 45°, 50°, 60° **(Supplementary Fig. 4 and 5)**. In order to compensate for the high degree of beam induced motion and lower image quality resulting from increased sample thickness in our tilted samples, data were collected at a high frame rate and processed using a tilt-optimized version of MotionCor2.[22] Using this approach, we were able to resolve PhuN assemblies in 2D arrays to ~3.9 Å and build an atomic model starting with predictions from RoseTTAFold[15] and AlphaFold[16] **(Fig. 2b-d, Supplementary Fig. 5 and 6, Supplementary Table 1 and 2)**.

PhuN Lattices: Four Unique Interfaces

The ɸPA3 PhuN monomer is comprised of two connected domains. Despite the lack of convincing sequence homology to other proteins, structural comparison of our model with coordinates in the PDB using the Dali Structure Comparison Server[23] revealed that the large domain is structurally homologous to acetyltransferases while the smaller domain has a less clear fold homology, sharing similarities with proteins including GTPases, transport proteins, as well as several restriction endonucleases **(Supplementary Table 3)**. As other proteins forming nanocompartments or encapsulins often share structural similarities with capsids,[24] the PhuN homologies suggest a unique evolutionary origin. Since no related binding sites or catalytic residues could be identified in PhuN, it is unlikely that either domain has retained any ancestral catalytic activity. It remains to be seen whether these homologies have functional implications beyond a purely structural role.

PhuN assembles into a tetrameric lattice with a dimeric asymmetric unit, giving rise to p2 symmetry with a unit cell length and angle of 120 Å and 100°, respectively **(Fig. 2a)**. At the 2D classification stage, we observed a minor subset of p4 symmetric particles (unit cell length and angle: 120 Å, 90°), as in the isolated phage nuclear fragments **(Supplementary Fig. 7a)**. Alignment of the p2 asymmetric subunit models using the acetyltransferase-like domain results in an RMSD of ~ 2 Å for all Cα and 0.77 Å for the core Cα. The bulk of the differences between

the structured domains of asymmetric subunits appear to arise from small shifts in helices as well as interconnecting loops. Each core tetramer is held together by tight packing of four monomers with both domains contributing, while interactions between the tetramers are looser and predominantly coupled via highly flexible elements (loops, termini, etc.). In keeping with the p2 symmetry, each PhuN tetramer forms four unique interfaces, two of which form channels: a Diamond channel at the center of the core tetramer, an Open Hairpin channel and Closed Hairpin interface forming the lateral contacts between core tetramers, and a Loop interface where the outer corners of four tetramers meet **(Fig. 2a,e)**.

The central Diamond channel is lined by the four smaller PhuN domains of the core tetramer. The resulting diamond-shaped opening has a highly negative charge **(Fig. 3b)** and measures 32 Å by 33 Å across the corners. In an elegant domain swap that holds the subunits tightly packed around the channel, a small positively charged helix within the N-terminal 37 residues of the nuclease domain threads into the neighboring subunit, nestling at the top within a negatively charged pocket **(Fig. 3a)**.

The Open Hairpin channel and Closed Hairpin interface are located at the lateral interface of two core tetramers. These lateral interactions are established by the same β-hairpins making contacts with different residues of the neighboring tetramers, approximately 20 Å apart **(Fig. 3c and Supplementary Fig. 8a-d)**. At the Open interface, the tips of opposing β-hairpins (residues 111-126) interact with residues near where the C-terminus leaves the structured acetyltransferase-like domain of the neighbors **(Supplementary Fig. 8a-b,d)**. Keeping these hairpins apart creates a somewhat narrow channel, measuring 31 Å by 41 Å across the corners of the channel. At the Closed interface, these opposing β-hairpins form direct lateral interactions that close the channel and move the tips of the β-hairpins closer to a structured loop (residues 245-258) **(Fig. 3c and Supplementary Fig. 8c-d)**. As we do not observe any large conformational changes between the asymmetric subunits, this suggests that the 20 Å shift between the β-hairpin positioning at the Open versus Closed Hairpin interfaces

arises from a largely rigid motion of the entire core tetramer and thus, as a result, introduces the 10° skew in the lattice network **(Fig. 2a and Supplementary Fig. 8e,f)**. One conjecture is that perhaps these are not fixed interfaces, but rather represent Open and Closed states that the lattice can adopt for a functional purpose such as relieving tension in the nearly micron-scale assemblies or adjusting the channel created at the Open interface **(Supplementary Fig. 8e,f and Movie 1)**. Going forward, we will refer to the asymmetric subunits by the positioning of their β-hairpins such that the β-hairpin from PhuN-O forms the Open channel while the PhuN-C β-hairpin forms the Closed interface.

The Loop interface appears to be, in part, held together by interactions between the large flexible loops (residues 272-291) from at least two different subunits. While the full densities of these loops remain unresolved in our 3D map, they are clearly visible in 2D classes **(Fig. 2a)**. The loops from two PhuN-O subunits from opposing core tetramers interact laterally **(Fig. 2a)**. The loops belonging to the PhuN-C subunits at those interfaces are not visible suggesting they are unstructured or in an alternative, less visible conformation in the 2D class. These loops account for the largest conformational differences we observed between asymmetric partners.

The final 12 residues of the C-terminal tail reside in a groove between the two domains of PhuN, right over the β-hairpin **(Fig. 2e)**. We also see strong densities in our high-resolution map for what is likely the C-terminal strands at the back of the acetyltransferase-like domain **(Fig. 2e)**. These densities differ in the two asymmetric subunits. Unfortunately, there was insufficient local resolution to trace how the C-terminus crosses between tetramers. This leaves some ambiguity as to which exact paths the long termini take to cross the Open Hairpin channel and Closed Hairpin interface before resting their final 12 residues in the positively charged groove **(Fig. 2e and Fig. 6b)**. One possibility is that the unresolved C-terminal tail regions may be free and unstructured to allow flexibility for the tetramers to accommodate alternating between Open and Closed Hairpin states, shell curvature, or different lattice orientations and

symmetry states.

PhuN Symmetry and Assembly

Since we observe the Loop, β-hairpin, as well as the N- and C-terminal tails at the assembly interfaces, we tested deletions of these regions for their ability to self-assemble *in vitro* and their ability to integrate into phage nuclear shells *in vivo* during ϕPA3 infection, as compared to full-length PhuN **(Fig. 4a)**. The *in vitro* samples were fused to MBP as before and the *in vivo* deletions were fused to mNeonGreen for live fluorescence microscopy. Deleting the 37 N-terminal residues abrogated PhuN incorporation into the phage nuclear shell *in vivo* and resulted in no clear assembly *in vitro* **(Fig. 4b and Supplementary Fig. 9b)**. This suggests that the N-terminal residues are necessary for the formation of the core tetramer and further assembly. *In vitro*, deleting the C-terminal tail severely compromised both assembly as seen by anion exchange **(Fig. 4f)** and lattice formation as seen by negative stain EM **(Fig. 4d and Supplementary Fig. 9d)**. In contrast, incorporation into the phage nuclear shell *in vivo* was minimally affected **(Fig. 4d and Supplementary Fig. 9d)**, perhaps due to the presence of wild-type protein and multivalent C-terminal interactions of PhuN in the lattices. In a rather surprising result, deletion of the β-hairpin removed the ability of PhuN to integrate into the phage nuclear shell during infection, yet the purified deletion still formed 2D lattices as observed in negative stain EM **(Fig. 4e and Supplementary Fig. 9e)**. This suggests that, while PhuN does not necessarily require the β-hairpin to form lattices, assemblies lacking the β-hairpin are biophysically, biologically, or otherwise unfavorable in the cellular context in the presence of wild-type protein. The deletion of the large Loop **(Fig. 4c and Supplementary Fig. 9c)** significantly decreased but did not abolish PhuN integration into the phage nuclear shell during infection. The Loop deletion similarly showed a mild propensity for both lattice-like assembly and enrichment of the small 25 nm Class II species we initially observed with the full-length protein but no long-range order in negative stain **(Fig. 4c and Supplementary Fig. 9c)**. These

deletions together show that the Loop, β-hairpin, as well as the N- and C-terminal tails we observe at the interfaces of intra- and inter-tetramer interactions do, in fact, play critical roles in the ability of PhuN to assemble into lattices and form the phage nuclear shell *in vivo*.

During the review of this manuscript, a publication reported *in vitro* assembled four-fold symmetric cuboids of the homologous PhuN protein from bacteriophage 201ϕ2-1.[17] Despite the clear difference in observed symmetries, the core domains of the 201ϕ2-1 monomers closely match the ϕPA3 monomers we describe in this work **(Fig. 5a)**. By contrast the quaternary organizations of monomers within the tetramer differ by an out of plane rotation of ~ 64°, corresponding to the large difference between the flat sheet subunit arrangement in our arrays and the cuboid organization described for 201ϕ2-1 **(Fig. 5c-e)**. Accommodating this different organization requires alterations in a loop (residues 242-259 ϕPA3 and 271-289 201ϕ2-1), near the β-hairpin (residues 106-129 ϕPA3 and 134-157 201ϕ2-1), and the N-terminal tail (residues 19-36 ϕPA3 and 48-64 201ϕ2-1). While the N-terminal helix and binding pocket are analogous to that of our ϕPA3 PhuN, the 201ϕ2-1 N-terminal tail is extended further away from the core domains of the protein to form the faces of the *in vitro* cuboids **(Fig. 5a,b)**. Moreover, the tetrameric ϕPA3 model from our *in vitro* assembled lattices has an excellent fit into the 24 Å cryo-electron tomography density[17] obtained from 201ϕ2-1 infected cells **(Fig. 5c-e)** (masked correlation coefficients: 0.69 ϕPA3 tetramer; 0.49 calculated and 0.56 reported[17] for the 201ϕ2-1 cuboid tetramer). This observation, along with the agreement between our isolated and *in vitro* assembled lattice 2D classes, indicates that the subunit arrangement in our tetrameric model is an accurate representation of the lattices formed *in vivo* by PhuN family proteins. These two distinctive assemblies—highly constrained p4 cuboids[17] and p2/p4-containing flat sheets— formed by proteins with nearly identical tertiary structure highlight the remarkable adaptability of the PhuN protein family and invite future investigation of the p2 and p4 symmetries.

To better understand where the different symmetries originate in our samples, we mapped our p2 and p4 particles back to their micrographs and found they often reside in distinct

clusters on the lattices **(Fig. 6a)**. These clusters suggest that the different symmetries may occur in regions distributed around the phage nucleus in infected cells, perhaps serving functional roles such as enabling shell curvature, flexibility, and assembly or enabling the passage of small molecules. Given the similarity of the p2 and p4 core domains, honing in on the unresolved Loop, C-terminal tail, and interfaces between tetramers will likely be key to understanding the functional role of the different symmetries and ultimately delineating the assembly mechanism of the phage nuclear shell.

While we do not observe the N-terminal MBP tag in our maps, the MBP appears on the exterior of our small *in vitro* assemblies **(Supplementary Fig. 1d)**. From this, we speculated that the lattice surface with the exposed N-terminus faces the cytosol. This is the same orientation suggested by the excellent fit of our tetrameric model into the 201φ2-1 nuclear shell map[17] determined by *in situ* tomography. Given this orientation, the exterior of the phage nuclear shell is sporadically positively charged while the interior surface is quite negatively charged **(Fig. 3b)**. This could be an elegant strategy for keeping the DNA away from the surface, preventing it from inadvertently leaving the protection of the lattice network.

Discussion

Once thought to be a defining feature of eukaryotic organisms, the phage nucleus proves to be a protein-based, membrane-free structure accomplishing a similar task. In this work we describe the self-assembly of the 66.6 kDa PhuN protein from bacteriophage φPA3 into a predominantly p2 symmetric lattice that closely matches fragments of the endogenous phage nucleus isolated from φPA3 infected *P. aeruginosa* as well as *in situ* reconstructions from the closely related bacteriophage 201φ2-1.[17] This tetrameric lattice features four distinct interfaces, two of which form clear channels (Diamond, Open Hairpin) that could allow the passage of small molecules, unfolded proteins, or nucleotides. We further show that the lattice assembles through interactions between flexible loops and establishes a negatively charged interior surface

for the phage nucleus.

These lattices and channels provide a framework with which to approach further mechanistic questions in this system ranging from growth and adaptability of the lattice, to DNA protection and packaging, to selective passage in and out of the phage nucleus. The channels we identify may allow for diffusion of small molecules as is or perhaps can accommodate larger molecules following conformational changes induced at specific sites or in response to binding of other phage factors. The different interactions of the large flexible loops and possibility for varied C-terminal tail binding observed here could help accommodate such changes. This idea is supported by both our observation of a minor population of p4 symmetric particles in our lattices **(Supplementary Fig. 7a)** and the recent report of p4 symmetric cuboids assembled *in vitro* using 201φ2-1 PhuN.[17] The fact that the same monomeric subunit can readily form distinct PhuN symmetry states and assemble into structures ranging from flat sheets to highly constrained cuboids, as seen in 201φ2-1, implies an unexpectedly flexible, dynamic assembly far different from the highly stereotyped, rigid lattices found in viral capsids or compartments like the carboxysome.[10,11,25,26] *In vivo*, scattered throughout this remarkable lattice may be other protein-based channels to allow for capsid docking, efficient DNA extrusion and packaging into empty capsids, as well as the selective passage of other large molecules.

In many other protein-based assemblies, such as eukaryotic and prokaryotic viral capsids and bacterial microcompartments, space is enclosed through combinations of triangular or hexametric units forming flat faces and pentameric units providing curvature.[10,11,26] At the molecular level, the ability of the same proteins to pack in these different modes arises from quasi-equivalent interactions amongst flexible domains.[26] Long tails are often used around symmetry points to stabilize the assemblies.[26] The observations reported here suggest that quasi-equivalence may also be mediated directly by flexible tails and loops, not only forming quasi-symmetric tetramers but also flexibly linking adjacent tetramers together.

Given the extensive interactions between tetramers mediated by these long tails and

loops, adding new subunits during growth while maintaining a permeability barrier may require the transient formation of dislocations or perhaps the aid of helper host proteins, such as chaperones, as occurs during the disassembly of clathrin lattices.[27] The observation of small assemblies, similar to those found *in vitro,* attached to an isolated shell **(Supplementary Fig. 1a)** also raises the possibility of a phage nuclear shell growth mechanism driven by the incorporation of small, pre-formed assemblies rather than individual monomers. The regular spacing and close proximity of these assemblies may also point to the existence of specialized regions along the phage nucleus where subunit addition occurs. This could allow for rapid and efficient growth while minimizing the number of disruptions made to the phage nuclear shell lattice. Alternatively, these small assemblies may correspond to distinct protein complexes with quite different functional roles.

The flexibility of the C-terminal tail, availability of some variation in its binding to neighboring subunits, and potential for toggling between symmetric states is perhaps what enables the astounding diversity in both shape and size of the bacteriophage nuclei,[4] especially considering they appear to be largely constructed from a single ~ 70 kDa protein across various jumbo phages.[5] Despite the high degree of structural conservation, PhuN proteins from different species do not appear to interact during co-infections,[28] suggesting that there may be localized differences altering their rates of lattice incorporation or prohibiting interactions altogether. As differing N- and C-terminal tail lengths account for the primary structural differences between AlphaFold predictions of known PhuN proteins **(Supplementary Fig. 10)**, a close analysis of PhuN evolution from diverse jumbo bacteriophages will be instrumental in delineating mechanistic similarities and differences that underlay the ability of PhuN proteins to assemble into the micron-scale phage nuclei observed *in vivo*. Investigating these questions will provide a more thorough understanding of where these structures came from, how they nucleate and grow, as well as how we can co-opt them for both experimental and practical applications going forward.

Methods

6XHisMBPGp53 Construct and Expression

An N-terminal maltose binding protein (MBP) tag and *E. coli* codon optimized φPA3 *gp53* gene were inserted into a pET15 backbone to create pESN6. pESN6 was transformed into BL21 Star (DE3) pLysS cells for expression. 20 mL of an overnight in LB was grown with carbenicillin and chloramphenicol. 1 L of Terrific Broth (TB) growth media was inoculated with around 2% of the overnight or until the starting OD_{600} reached 0.001. Cells were grown with 250 rpm nutation at 37°C until an OD_{600} of 0.5. Cells and the incubator were then cooled to 16°C. Expression was induced with 1 mM IPTG and proceeded overnight at 16°C with 250 rpm nutation. Cells were spun down the following morning at 3,000 x g for 8 minutes and pellets were transferred to 50 mL conical tubes. These were flash frozen and stored at -80°C or used immediately for purification.

Phage Shell Isolation

P. aeruginosa PA01 expressing MBP-gp53 were lysed at 60 min post φPA3 infection in NP40 Lysis Buffer pH 6.5 (50 mM Bis-TrisPropane, 150 mM NaCl, 0.5% NP40, 5% glycerol, 5 mM DTT, 20 ng/µl Lysozyme, 1 mM EDTA, 1mM EGTA) by dounce homogenizer. The lysate was clarified at 16,000 x g for 5 min and the insoluble fraction resuspended in Wash/Shell buffer pH 6.5 (20 mM Bis-TrisPropane, 150 mM NaCl, 1 mM DTT, 1 mM EDTA, 2 mM $MgCl_2$). This was subject to further 500 x g (5 min) and 15,000 x g (10 min) centrifugation with the insoluble fraction isolated and resuspended in Wash/Shell buffer each time. The final 15,000 x g pellet resuspension was used directly for downstream analysis by negative stain or cryo-EM.

Immunofluorescence Microscopy

Immunofluorescence microscopy samples were prepared as previously published and reiterated below.[3] 5 mL overnight cultures of bESN30 (PA01 expressing 6XHisMBPGp53) were grown at

30°C in LB media with gentamicin. A 1:30 back-dilution of the overnight culture:LB was grown at 30°C for 1 hr. Protein expression was induced with 0.1% arabinose. After 1 hr of expression, an aliquot of uninfected cells was fixed while the remaining cultures were infected with φPA3 using MOI 1.5. Infected cell aliquots were collected and fixed at 60 mpi.

Samples were fixed with a 15 minute incubation at room temperature in 5X Fix Solution (12.5% paraformaldehyde, 150 mM KPO4 pH 7.2) followed by an additional 20 minutes on ice. Samples were then washed in PBS 3 times and finally resuspended in GTE pH 7.65 (50 mM glucose, 10 mM EDTA, pH 8.0, 20 mM Tris-HCl) with 10 µg/mL lysozyme. Resuspended cells were transferred to polylysinated coverslips and dried. The dry coverslips were incubated for 5 minutes each in cold methanol then cold acetone. Cells were rehydrated by rinsing in PBS and incubating in PBS + 2% BSA blocking solution for 3 minutes. Cells were incubated with a 2 µg/mL dilution of primary antibody (Maltose Binding Protein Epitope Tag Antibody, Rockland Inc. Rabbit Polyclonal, Product # 200-401-385) in PBS + 2% BSA for 1 hr and rinsed with 3, 7 minute washes in fresh PBS + 2% BSA. Coverslips were then incubated in the dark for 1 hr with secondary antibody (ThermoFisher Goat Anti-rabbit IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor555, Catalog #A-21429) diluted to 4 µg/mL in PBS +2% BSA. DAPI was added for the final 10 minutes of the incubation. Samples were washed in PBS 3 times for 7 minutes. Coverslips were then placed on slides using mounting media (v/v 90% glycerol, v/v 10%Tris pH 8.0, and w/v 0.5% propyl-gallate), sealed with clear nail polish, and imaged using a Zeiss Axiovert 200M microscope.

Fluorescence Microscopy

Bacteria were prepared for imaging as previously described and reiterated below.[4] 1.2% agar pads prepared on concavity slides were supplemented with 0.05% arabinose to induce protein expression and 1 µg/mL DAPI for phage nucleus staining. 5 µL of resuspended *P. aeruginosa*

strain PAO1 colonies expressing PhuN-mNeonGreen truncations were grown on the agar pad at 30°C for 3 hrs in a humid chamber. 3 µL of ϕPA3 phage lysate (10^9 PFU/mL) was spread on the agar pad and allowed to infect for 30 min at 30°C in the humid chamber before imaging.

Microscopy was performed on an inverted epifluorescence (Ti2-E, Nikon, Tokyo, Japan) equipped with a Photometrics Prime 95B 25-mm camera and the Perfect Focus System (PFS). Images were acquired using Nikon Elements AR software (version 5.02.00). Cells were imaged through channels of phase contrast (200 ms exposure, for cell recognition), blue (DAPI, 50 ms exposure, for phage DNA), and green (GFP, 200 ms exposure, for PhuN-mNeonGreen constructs) at 100x objective magnification. For each condition, cells were imaged in 17 stacks in the Z-axis centered on the middle focal plane with a step size of 0.26 µm. Images were deconvolved with Huygens Essential 20.10 (Scientific Volume Imaging, the Netherlands, http://svi.nl) using the CMLE algorithm in the Deconvolution Wizard, and final figure images were prepared in Fiji.[29]

Purification of 6XHisMBPGp53 and All Deletions
A cell pellet from 1 L of culture was resuspended in 12 mL of Lysis Buffer pH 7.42 (20 mM HEPES, 500 mM NaCl, 0.2% TritonX100, 1 mM DTT, 5% Glycerol, 5 mM $MgCl_2$) per each gram of pellet with 1 cOmplete Protease Inhibitor Tablet and benzonase. This mixture was homogenized using a manual glass homogenizer and lysed using an Avestin EmulsiflexC3. The sample was clarified with a 20 minute spin at 18,500 x g. The soluble fraction was nutated with equilibrated Ni·NTA resin for 1-2 hours in the cold room. The sample was transferred to a gravity column and washed 2X with ATP Wash Buffer pH 7.42 (5 mM ATP, 5 mM $MgCl_2$, 20 mM HEPES, 500 mM NaCl, 20 mM Imidazole, 1 mM DTT, 5% Glycerol) and 2X with Wash Buffer pH 7.42 (20 mM HEPES, 500 mM NaCl, 20 mM Imidazole, 1 mM DTT, 5% Glycerol). Finally, the sample was eluted and collected in the Elution Buffer pH 7.42 (20 mM HEPES, 150 mM

NaCl, 600 mM Imidazole, 1 mM DTT, 5% Glycerol). The eluate was aliquoted and flash frozen the same day and stored at -80°C.

Anion Exchange

The sample (thawed or fresh) was buffer exchanged into MonoQ Binding Buffer (pH 7.65: 20 mM Tris, 50 mM NaCl, 1 mM DTT, 0.5 cOmplete Protease Inhibitor Tablet; pH 6.5: 20 mM Bis-TrisPropane, 50 mM NaCl, 1 mM DTT, 1 mM EDTA, 0.5 cOmplete Protease Inhibitor Tablet) using 3.5K MWCO Snakeskin dialysis tubing or a 7K MWCO Zeba Spin Desalting Column. The sample was loaded on the 10/100 or 5/50 GL MonoQ Anion Exchange Column, washed, and eluted over a linear gradient from 0 to 65% over 13 mL (5/50 gL) using the Elution Buffer (pH 7.65: 20 mM Tris, 1 M NaCl, 1 mM DTT, 0.5 cOmplete Protease Inhibitor Tablet; pH 6.5: 20 mM Bis-TrisPropane, 1 M NaCl, 1 mM DTT, 1 mM EDTA, 0.5 cOmplete Protease Inhibitor Tablet).

Negative Stain Electron Microscopy Preparation and Imaging

3 μL of sample were applied to a glow discharged carbon coated grid (30 seconds, 20 mA). After 30 seconds, this was flushed away with deionized water followed by 30 μL Uranyl Formate (UF). ~5 μL UF were left on the grid to stain for 30 seconds. This stain wash and incubation was repeated 2 more times. The grid was finally dried by filter paper side-blotting and gentle wafting through the air. The grids were imaged using a 120 kV FEI Tecnai T12 (LaB6 filament, Gatan UltraScan 895 4k CCD) or 200 kV FEI Tecnai T20 (FEG electron source, TVIPS 8K x 8K camera). Images were collected at 62 kx (defocus ~ -0.981 nm) or 145 kx (defocus ~ -0.295 nm).

Cryo-EM Imaging of Isolated Shell Fragments

6XHisMBP-PhuN isolated shell fragments were imaged using the 200 kV Glacios microscope. 34 micrographs were collected manually, 6 of which were used to pick and extract 9,568

particles. After 3 rounds of classification, 376 particles remained in the class shown in the **Supplementary Fig. 1b** inset. The WT PhuN isolated shell fragments were imaged using the 200 kV Talos Arctica (GATAN K3 camera). 10,864 particles were selected from 36 micrographs and processed to a final p2 class with 3,758 particles and p4 class with 1,538 particles **(Fig. 1a)**.

2D Crystal Preparation and Sample Freezing

The most concentrated monomeric fraction was selected from the Anion Exchange purification. The concentration was used to calculate the volume of buffer and sample needed to reach a working concentration of 0.01 mg/mL 6XHisMBPgp53 in a 30 µL crystallization drop. A humidity chamber was made by wetting filter paper inside of a petri dish. A teflon block with approximately 5 mm wide and 13 mm deep wells was placed in this chamber. The calculated volume of 2D Crystallization Buffer pH 6.5 (20 mM Bis-TrisPropane, 150 mM NaCl, 1 mM DTT, 1 mM EDTA, 2 mM $MgCl_2$, 1 mM ATP, 0.25 cOmplete Protease Inhibitor Tablet) was pipetted into each well. 1 µL of Lipid Mix (0.73 mg/mL EGG PC, 0.27 mg/mL DGS-NTA(Ni)) was dispensed onto each buffer droplet. The volume of concentrated protein was then gently added to each droplet. After a 1 hour incubation at room temperature, the lattices were adsorbed by directly touching the grid surface to the lipid monolayer using either amino-functionalized graphene oxide[20] or QUANTIFOIL R1.2/1.3 400-mesh copper holey carbon grids without glow-discharging. These were then plunge-frozen using a Vitrobot Mark IV (FEI) at 10°C with 100% humidity using a 4 second blot time, 0 second wait time, and blot force of 0 for the functionalized grids and a 6 second blot time, 0 second wait time, and blot force of 3 for the QUANTIFOIL grids.

Acquisition of Cryo-Electron Microscopy Data

As the data is composed of 15 collections varying tilt, grid type, and frame rates, the following description and **Supplementary Table 2** summarize the ranges of settings for all of the

collections combined. The micrographs were collected using two 300 kV FEI Titan Krios electron microscopes equipped with GATAN K3 direct electron detectors and operated using SerialEM software. A nominal defocus range of 0.8-2.5 μm under focus was used for collection. A nominal magnification of 105 kx, corresponding to a physical pixel size of 0.835 Å or 0.834 Å, was used. A total dose of 66 or 67 e-/Å^2 was maintained for each collection. The samples were tilted from 0° to 60° for static dataset collection to obtain side-views of the 2D crystalline sample. To mitigate beam-induced motion, we utilized GO/NH$_2$ coated R1.2/1.3 300 mesh gold grids[20] and collected data using a higher frame rate to obtain a finer sampling of the beam-induced motion.

Processing Cryo-Electron Microscopy Data

All micrographs were motion corrected using UCSF MotionCor2[22] with a 9 by 7 patch. Each tilt, frame rate, and grid type was imported and processed separately through 2D classification. The first datasets, collected on Quantifoil grids, were used to generate the first 3D model. After importing the motion corrected micrographs into cryoSPARC,[30] CTF estimation was done using cryoSPARC Patch CTF estimation (multi). cryoSPARC Blob Picker was then used to initially select particles. Two rounds of 2D classification were used to clean up picks. The selected particles from each tilt (0° 2,123 ptcls, 40° 308,819 ptcls, 50° 194,876 ptcls) were merged in a cryoSPARC Ab-Initio Reconstruction with C1 symmetry. A 176,652 particle class from the Ab-Initio Reconstruction gave a clear lattice volume with an apparent p2 symmetry. This volume was further used as a template for cryoSPARC Heterogenous Refinement.

Additional datasets underwent the same processing through 2D classification. Selected particles were added to the best particle subset for cryoSPARC Heterogenous Refinements with the best resolved volume as a template. The template was ultimately recentered on the diamond shaped channel at which point p2 symmetry was also imposed going forward. The centered template was also used to create a particle picking template for each tilted dataset. The template was

generated using the relion project feature in RELION v.3.0.8,[31] saving a projection ±5° from the

tilt at data collection every 3° and repeating this every 3° around the z axis from 0-180°. This

yielded 244 templates for each tilted dataset. These templates were imported into cryoSPARC

and all datasets were repicked using their respective template set. The picks were cleaned up

with two rounds of 2D Classification resulting in ~ 600,000 particles. These better centered

particles were classified using cryoSPARC Heterogenous Refinement and the resulting best

class was processed using cryoSPARC Non-Uniform Refinement[32] to a reported 4.11 Å. These

particles were exported for further processing in cisTEM.[33,34] In cisTEM, a mask was applied to

the central tetramer for several rounds of Manual 3D Refinement.[33] The reported resolution

estimates did not always correspond to changes in map quality, thus each map was visually

inspected for improvement. The particles were further classified in 3D using cisTEM, resulting in

a final class containing 29,303 particles.[33] This final volume was deconvolved to mitigate

stretching along the z-axis due to the missing cone of information resulting from tilted data

collection.

Deconvolution of a Map with Preferred Orientation

To help minimize the blurring due to limited tilt angles and subsequent missing cone of Fourier

data, deconvolution was performed using ER-Decon II. This software was initially developed to

deconvolve 4D wide-field light microscopy volumes with extremely low signal-to-noise ratio,[35]

but has subsequently applied to STEM[36] and TEM[37] cryo-EM data. ER-Decon II is implemented

in the program PRIISM.[38] To apply ER-Decon II on a cryo-EM density map, an optical transfer

function (OTF) that characterizes effects of preferred orientation and is equivalent to the point

spread function (PSF) is generated from the directional FSC of the cryo-EM map which provides

a representation of preferred orientation.[39] The smoothing and nonlinearity parameters for

deconvolution were 0.5 and 10000, respectively. The detailed procedure and full evaluation of

cryo-EM volume deconvolution will be described in detail separately, but the general principles have been addressed previously by Croxford, M., *et. al.*[37]

Model Building

RoseTTAFold[15] and AlphaFold,[16] were utilized to generate initial models. RoseTTAFold provided us with a first look at PhuN Gp53 while AlphaFold ultimately gave us a model which fit into the density slightly better, particularly after processing with RosettaRelax.[40] The N-terminus was positioned manually in ChimeraX Isolde[41] and further refined using RosettaRefine.[40] Residues 1-18, 276-287, and 556-602 did not have corresponding densities and were deleted for fitting. Invading C-terminal tails were similarly placed using Chimera,[42] relaxed using ChimeraX Isolde,[41,43] and refined using RosettaRefine.[40] The final model fitting was done using our deconvolved map (low pass filter 4.25 Å, B-factor 60 Å^2 applied in Relion), fitting each asymmetric monomer within the dimer density and doing further refinements using both RosettaRefine[40] and ChimeraX Isolde.[41] The dimer was used to create a p2 symmetric tetramer and the interface between the dimers was refined using ChimeraX Isolde.[41] The two subunits making this interface were duplicated and "fit in map" in ChimeraX to create the final p2 symmetric tetramer model.[42,43] Structure validation was done using Phenix MolProbity.[44,45]

The larger, 16-mer assembly layer was constructed using a map processed with a larger mask (EMD-29451 [https://www.ebi.ac.uk/emdb/entry/EMD-29451], Tracing p2 phiPA3 PhuN Tetramer Interfaces), resulting in a lower resolution, as a guide. This map included neighboring subunits around the core tetramer. Copies of the higher resolution PhuN tetramer map (EMD-29310 [https://www.ebi.ac.uk/emdb/entry/EMD-29310], Deconvolved phiPA3 PhuN Tetramer, p2) were fit into the lower resolution map using the "Fit in Map" tool in ChimeraX. The

corresponding 16-mer model was built by rigid body fitting copies of tetramer models into the newly positioned higher resolution tetramer maps with the ChimeraX "Fit in Map" tool.

Data Analysis and Figure Preparation

Figures were created using UCSF Chimera[42] and ChimeraX[43] using the deconvolved tetramer map (low pass filter 4 Å, B-factor 80 Å^2 applied in Relion).[43] 2D class panels were prepared in 3dmod[46] Version 4.11.20 using the Low-frequency sigma (0.050) as well as the high-frequency cutoff (0.450) and falloff (0.027) parameters. EM micrograph panels were also prepared in 2dmod.[46] Anion exchange traces were plotted using MatLab (MATLAB. (2021,2022). *version 9.11.0 and 9.12.0 (R2021b, R2022a).* Natick, Massachusetts: The MathWorks Inc). Figures were assembled in Affinity Designer (1.10.5). Electrostatic potential coloration was completed using the APBS and PDB2PQR Biomolecular Solvation Software web server at pH 6.5.[47]

Data Availability The cryo-EM maps generated in this study have been deposited to the Electron Microscopy Data Bank (EMDB) under accession codes EMD-29550 [https://www.ebi.ac.uk/emdb/entry/EMD-29550] (phiPA3 PhuN Tetramer, p2), EMD-29310 [https://www.ebi.ac.uk/emdb/entry/EMD-29310] (Deconvolved phiPA3 PhuN Tetramer, p2), and EMD-29451 [https://www.ebi.ac.uk/emdb/entry/EMD-29451] (Tracing p2 phiPA3 PhuN Tetramer Interfaces). The atomic coordinates have been deposited to the PDB under accession codes 8FNE [http://doi.org/10.2210/pdb8fne/pdb] (phiPA3 PhuN Tetramer, p2) and 8FV5 [http://doi.org/10.2210/pdb8fv5/pdb] (Representation of 16-mer phiPA3 PhuN Lattice, p2). Publicly available entries used in this study are PDB 7SQR and EMDB EMD-25221. All other data, bacterial strains, and plasmids are available from the corresponding author on a reasonable request.

Code Availability PRIISM and ER-Decon II are both available on request from agard@msg.ucsf.edu as the terms of the license established prior to this study with the University of California and HHMI do not allow uploading the code in a public repository.

Author Contributions

ESN performed cloning, experimental design, *in vitro* purification, sample preparation, EM data collection and analysis, model fitting, coordinated experimental design for fragment isolations with MM, trained MM in negative stain EM, and prepared this manuscript all under the supervision of DAA. AFB conducted motion correction of collected cryo-EM data and advised EM processing, experimental design, and interpretation. MM isolated shell fragments. CK

carried out live fluorescence microscopy experiments from cloning through data analysis under the supervision of JBD. MB and DB provided early access to RoseTTAFold. JL and YC conducted EM map deconvolution and resolution estimation.

Competing Interests

Figures
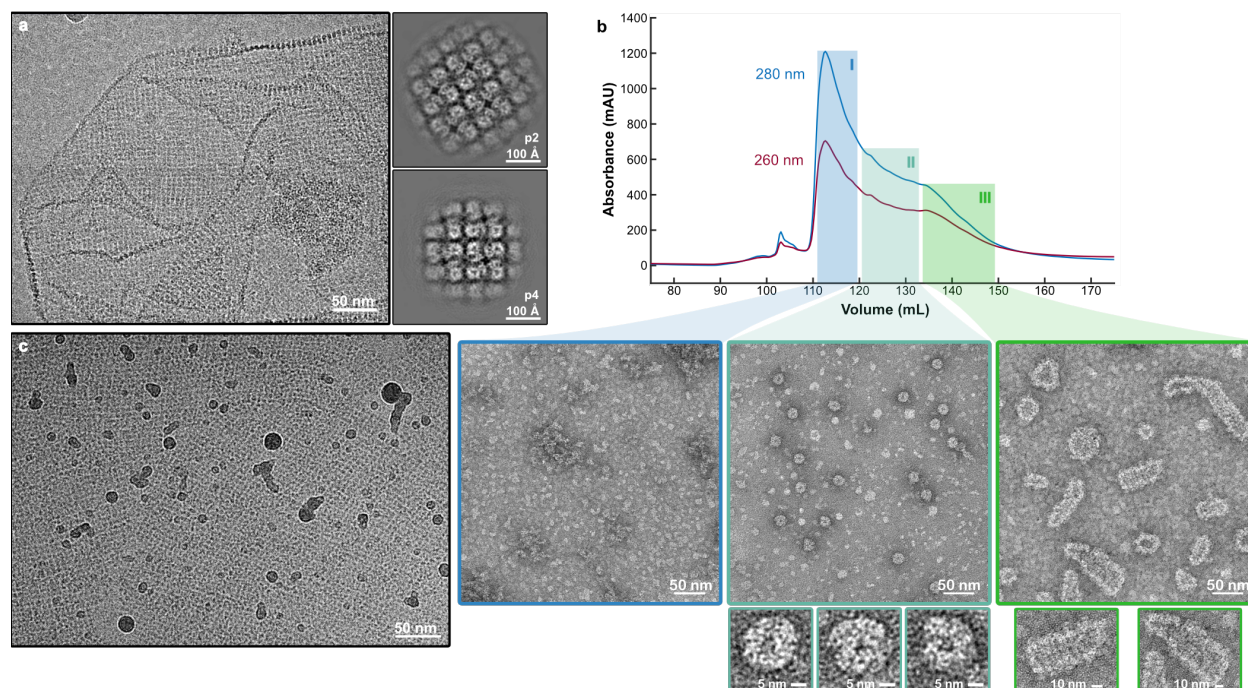
Figure 3.1 | PhuN forms a broad range of assemblies.
a, Micrograph showing an unlabeled phage nuclear shell fragment isolated from φPA3 infected wild-type *Pseudomonas aeruginosa* alongside the subsequent p2 (120 Å, 120° unit cell length, angle) and p4 (110 Å, 90° unit cell length, angle) symmetries observed upon 2D classification. Folds and differences in lattice orientation are visible. (n = 1 unlabeled shell isolation experiment). **b,** Anion exchange trace showing extended elution profile and corresponding negative stain EM micrographs. Class I (blue) shows concentrated monomeric species. Class II (turquoise) shows medium sized assemblies in the 25 nm range. Class III (green) reveals assemblies with elongated shapes and striated texture. **c,** *in vitro* assembled 2D crystalline arrays at 0° tilt showing three main patches with distinct lattice orientations and gentle curvature. Borders between different crystal patch orientations appear smooth. (n = 6 grids imaged for this publication).
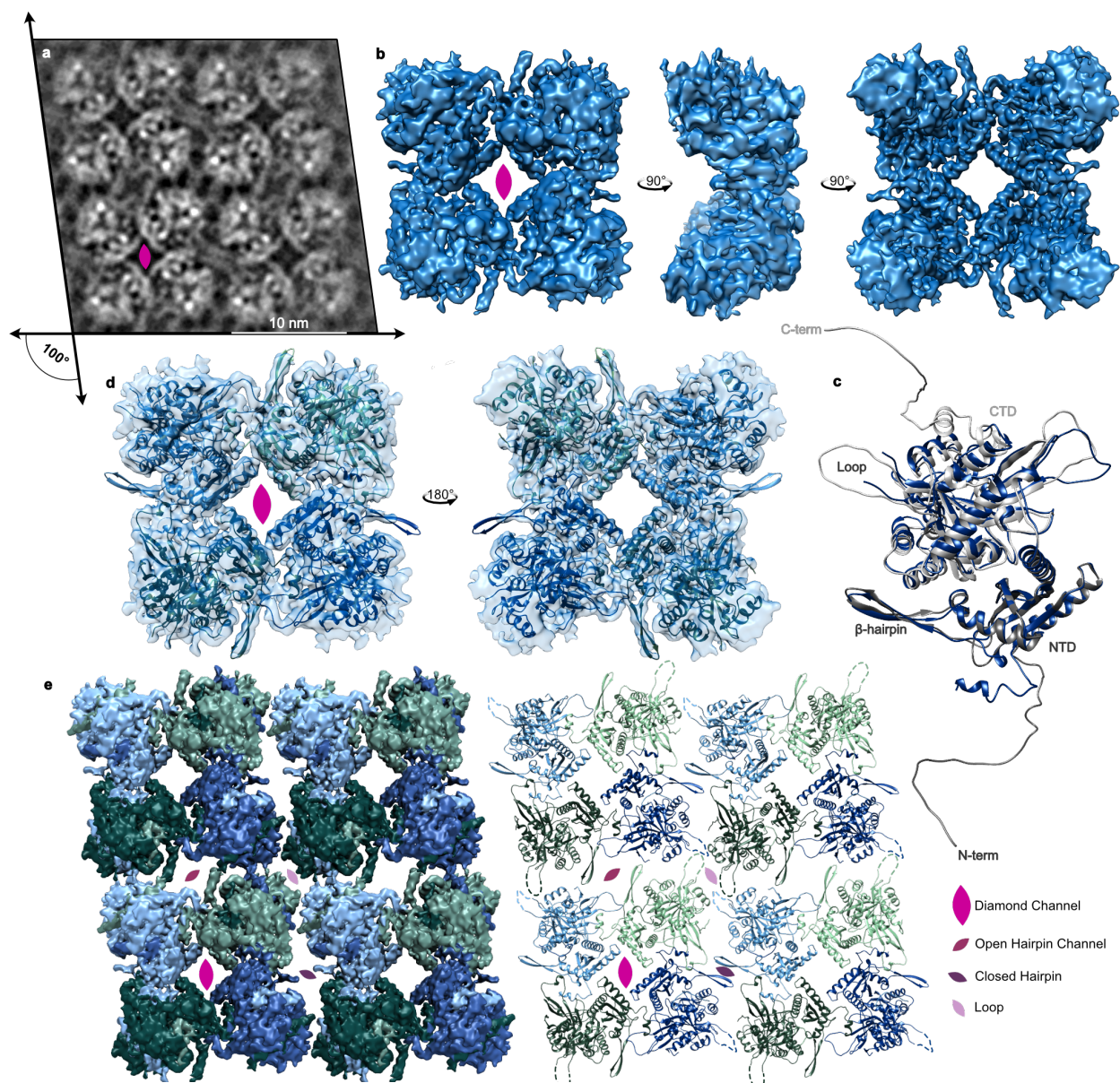
Figure 3.2 | PhuN assembles into a p2 symmetric lattice.
a, Example 0° tilt 2D class average highlighting the p2 symmetry and resulting four unique interfaces. The lattice is skewed to an approximately 100° angle. The positioning of the flexible loops that are difficult to resolve in 3D is visible. **b,** Highest resolution map of the core tetramer centered on the Diamond interface as viewed from different angles. **c,** The final PhuN-C subunit model (blue) overlaid with the AlphaFold prediction used as our starting model (grey). Residues 1-18, 276-287, and 556-602 had no corresponding density in our map and were deleted from the model. **d,** The tetrameric model of PhuN-C and PhuN-O fit into the map in b. **e,** The map from b segmented and positioned to recreate the channels observed in 2D classes (left) paired alongside the corresponding 16mer-model (right). In both the segmented maps and models, PhuN-O is represented in green while PhuN-C subunits are represented in blue. Unresolved and unmodeled loops are represented with dashed lines to reflect the 2D class while the C-terminal tails are excluded.
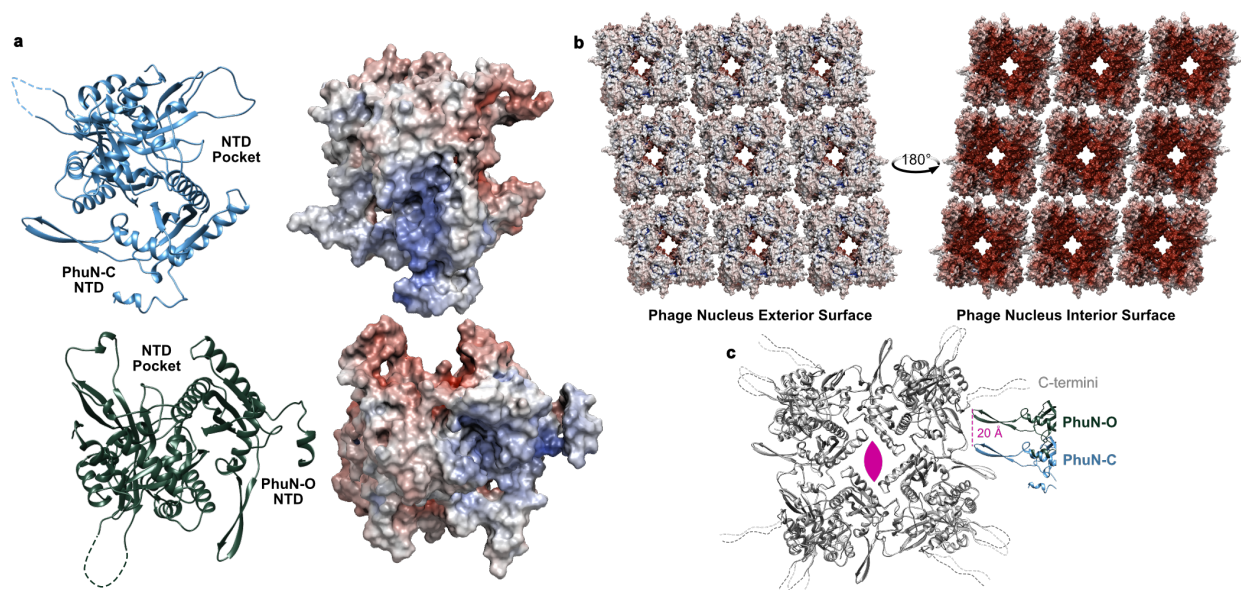
Figure 3.3 | A closer look at PhuN interactions and electrostatic potential.
a, Surfaces of PhuN-C (top) and PhuN-O (bottom) colored by electrostatic potential with their corresponding models on the left. **b,** Electrostatic potential coloration of the interior and exterior phage nucleus surfaces created by PhuN assemblies. Surface charge calculation was done at pH 6.5 and includes the C-terminal tail tips. Red corresponds to negative and blue corresponds to positive charge. **c,** Direct comparison of PhuN-C and PhuN-O. The asymmetric subunits in the tetramer are overlaid showing minor differences (light and dark grey) while the interaction with the neighboring β-hairpins differs dramatically, shifting by approximately 20 Å from near the C-terminus (PhuN-O) towards a flexible loop (PhuN-C).
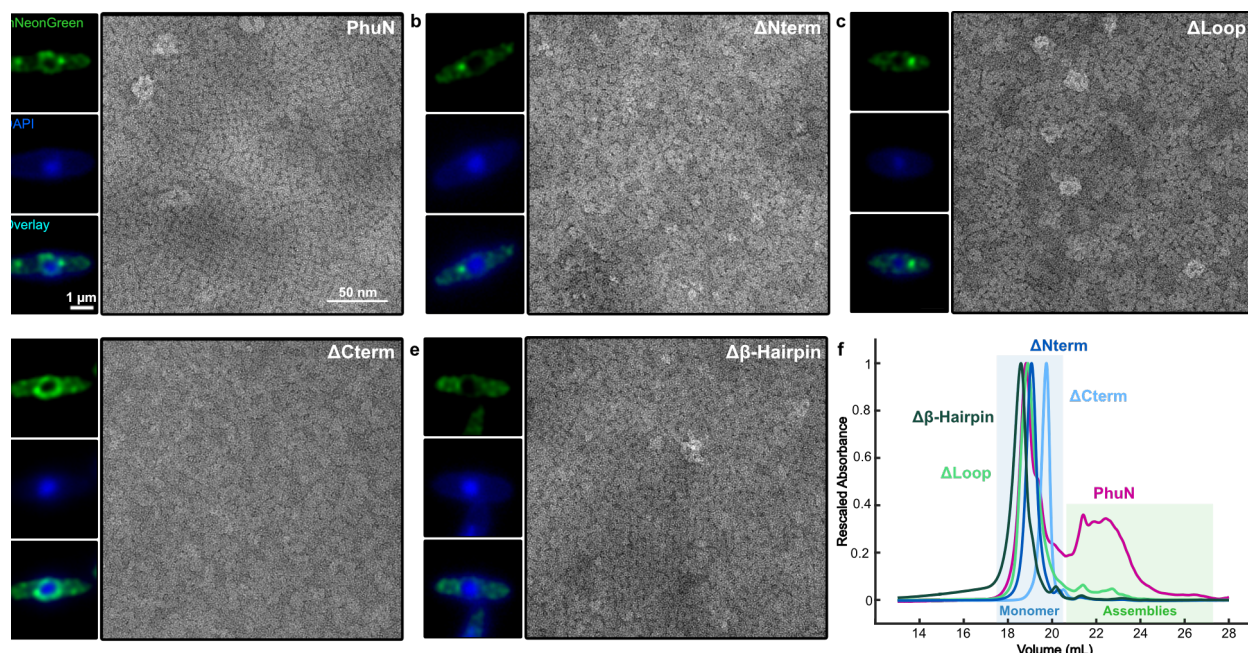
Figure 3.4 | φPA3 PhuN truncations show defects in shell integration in vivo and self-assembly in vitro.

Live fluorescence microscopy and negative stain EM micrographs of **a,** full-length PhuN as well as the following PhuN deletions: **b,** N-terminal tail (residues 1-37) **c,** Loop (residues 272-291) **d,** C-terminal tail (residues 556-602), and **e,** β-Hairpin (residues 111-126). Fluorescence microscopy images of PhuN or PhuN deletions were collected at 40 minutes post infection, deconvolved, and are displayed in green while the DAPI stained DNA is shown in blue. The 1 μm and 50 nm scale bars apply to all fluorescence and EM panels, respectively. (Independent sample preparation and imaging: fluorescence microscopy n = 3, negative stain EM n = 1). **f,** Rescaled anion exchange traces for the deletions shown in panels a-e. The decreased propensity for PhuN assembly on the column is evidenced by the sharp increase in the primary peak corresponding to the monomeric species (blue box) relative to the assembly peaks (green box). Samples from monomeric peaks were used for negative stain in panels a-e.
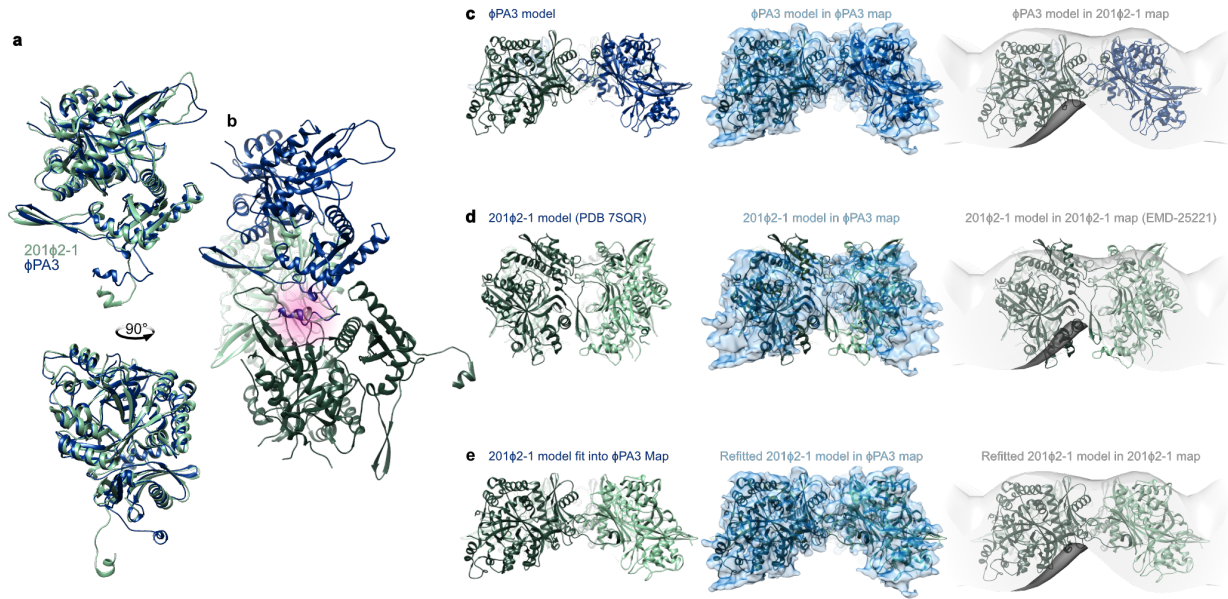
Figure 3.5 | PhuN proteins from ɸPA3 and 201ɸ2-1 share similar structures.
a, Top and side views of an overlay of the ɸPA3 (blue) and 201ɸ2-1[17] (green) monomeric
subunits. The overlay was determined by fitting the monomers into the blue ɸPA3 EM density
shown in panels c-e. The greatest backbone differences between the models reside in the
extended positioning of the N-terminal residues, β-Hairpin, and a resolved flexible loop. **b,** The
binding pockets utilized by both ɸPA3 and 201ɸ2-1 proteins are the same, as highlighted in
pink. This was determined by fitting the dark green 201ɸ2-1 subunit into the ɸPA3 map while
retaining the light green subunit to preserve the N-terminal tail interaction. The N-terminal helix
is positioned in the same pocket as that of the blue ɸPA3 subunit. **c,** p2 ɸPA3 model with the
corresponding ɸPA3 density (blue) and compared to the 201ɸ2-1 tomography density (grey,
EMD-25221[17]). **d,** Rigid body fitting of the published p4 201ɸ2-1 tetrameric model (PDB
7SQR[17]) into the ɸPA3 density (blue) and 201ɸ2-1 tomography volume (grey, EMD-25221[17]).
The model protrudes outside of the tomography volume. **e,** 201ɸ2-1 model (PDB 7SQR[17]) after
independently rigid body fitting each monomer into the ɸPA3 density alongside the improved fit
of the resulting model into the 201ɸ2-1 tomography density (EMD-25221[17]).
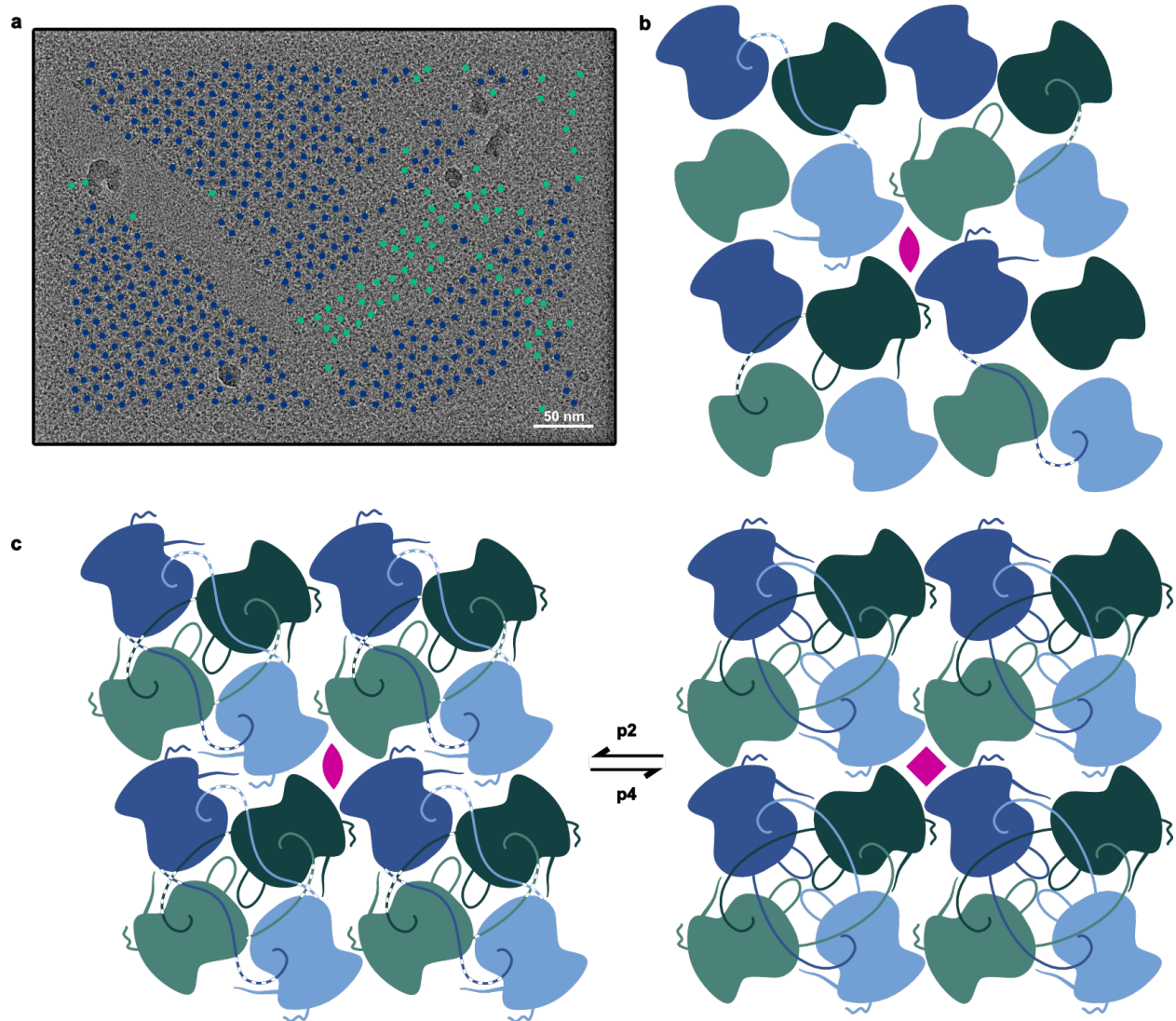
Figure 3.6 | PhuN assembles through a series of complex and likely adaptive C-terminal tail exchanges exhibiting both p2 and p4 symmetries.

a, Micrograph of *in vitro* assembled PhuN lattices with the p2 (blue) and p4 (green) particles displayed. (n = 8). **b,** Cartoon showing core tetramer with N- and C-terminal tail interactions as well as the loops visible clearly in 2D classes. **c,** Cartoon model tracing the N- and C-terminal tail interactions at all four unique interfaces compared to the p4 equivalent.
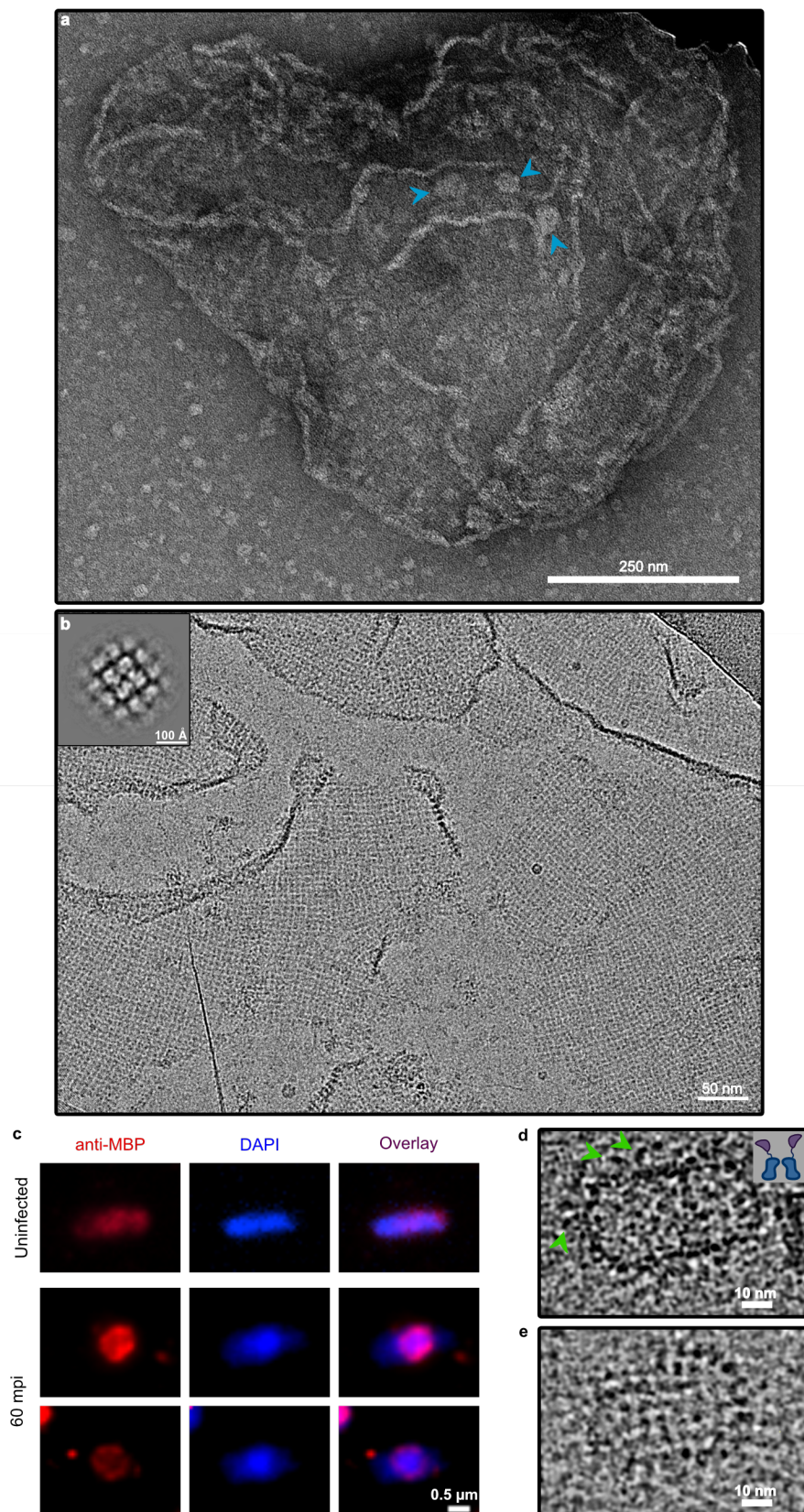
Supplementary Figure 3.1 | φPA3 PhuN assemblies.
a, Negative Stain micrograph of a phage nuclear shell fragment isolated from *P. aeruginosa* expressing 6XHisMBP-PhuN. The isolated species is about the size expected for an intact compartment. The lattice texture is visible as well as many wrinkles and folds. Smaller, round species are present on the surface (blue arrows), similar in appearance to Class II assemblies *in vitro*. (n = 2 for shell fragments, n = 1 for likely complete shell.) **b,** CryoEM micrograph of a shell fragment isolated from *P. aeruginosa* expressing 6XHisMBP-PhuN showing many shell fragments with a lattice texture, comparable to the lattices observed in main Figure 1. (n = 1). **c,** Immunofluorescence microscopy of *P. aeruginosa* cells expressing the 6XHisMBPGp53 fusion. The fusion protein integrates into the shell and localizes at the cell center around the phage DNA 60 minutes post infection. (n = 1). **d,** CryoEM images of *in vitro* PhuN assemblies. Densities consistent with the presence of MBP are visible on the exterior (green arrows). The inset shows a cartoon representation of a side-view of PhuN (blue) with MBP (purple). (n = 1). **e,** The assemblies show a lattice like surface. (n = 1).
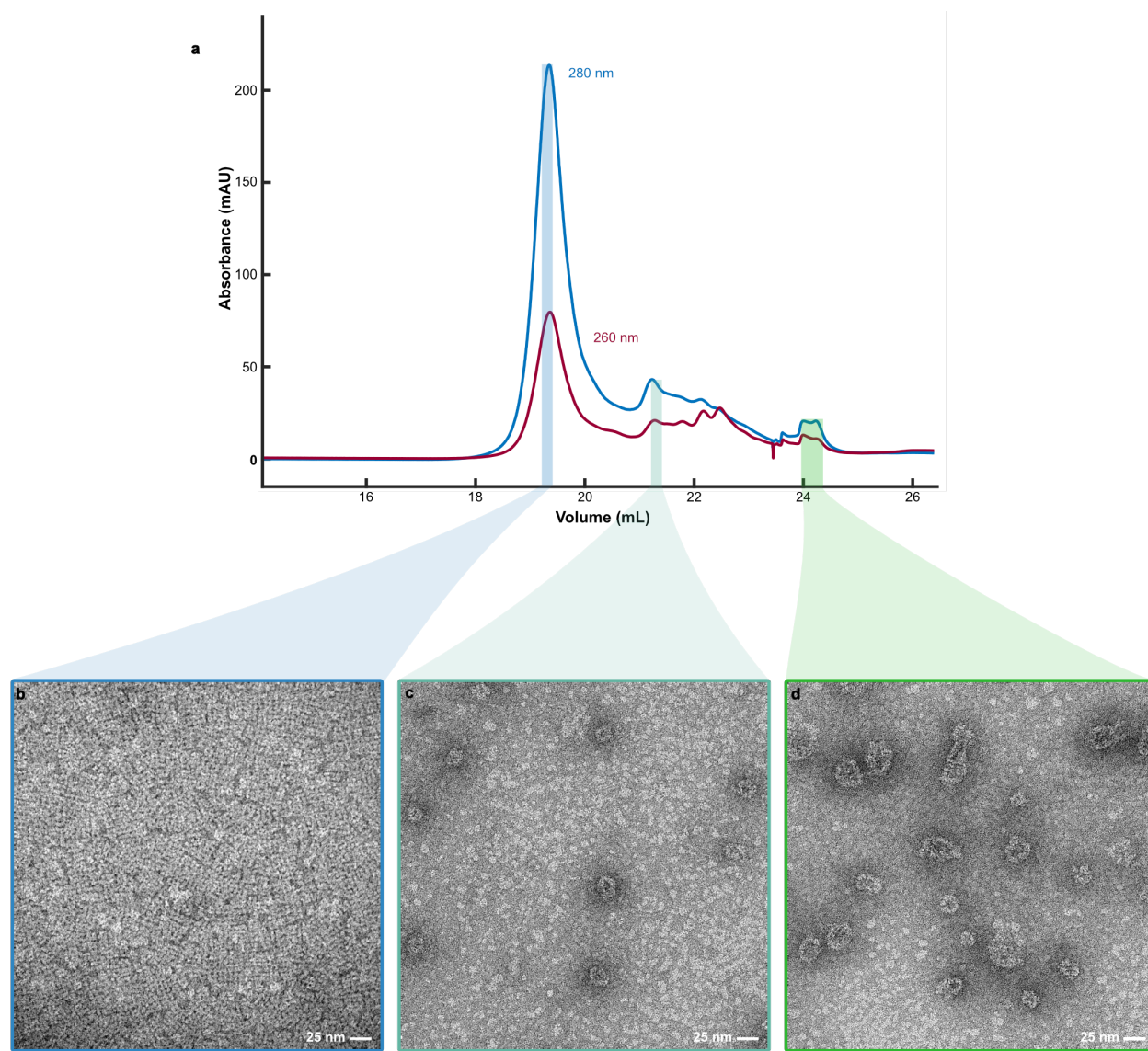
Supplementary Figure 3.2 | PhuN crystalizes at pH 6.5.
a. Anion exchange purification using a 1 mL GE MonoQ column at pH 6.5. Highlighted fractions correspond to micrographs b-d. **b,** Monomeric species in Class I form 2D polycrystalline assemblies when applied to negatively charged grids for negative stain EM. **c,** 25 nm mid-sized species present in Class II. **d,** Assemblies 25 nm and larger present in later elution in Class III. Samples prepared and imaged n = 3 times.
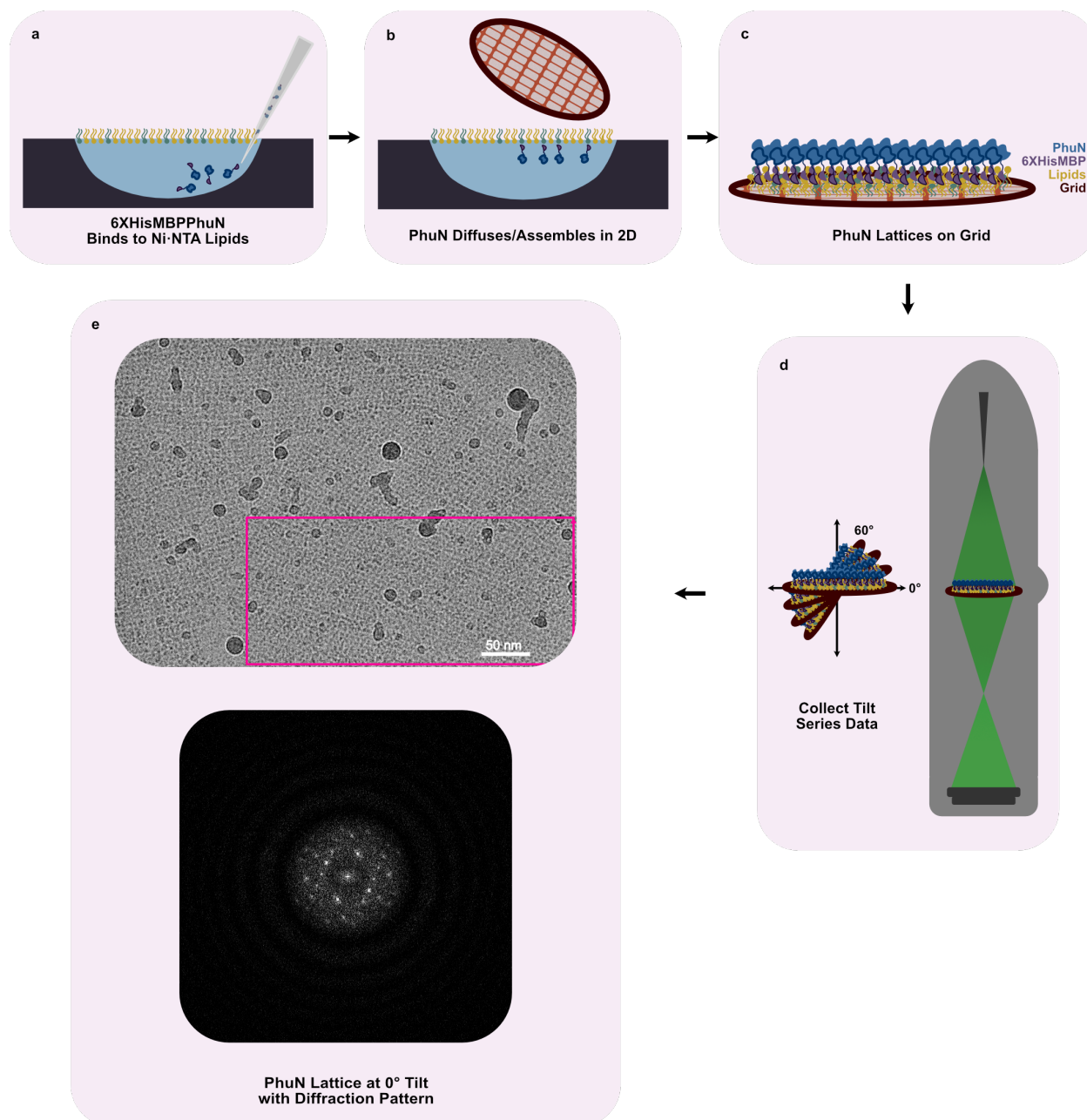
Supplementary Figure 3.3 | in vitro 2D crystal preparation.
a, 6XHisMBP-PhuN is injected into a buffer droplet with a pre-formed 21% Ni·NTA Lipid Monolayer. **b-c,** After incubation to allow for assembly, the PhuN arrays are adsorbed to a grid by touching the carbon side of the grid to the top of the lipid-containing droplet. These grids are then plunge frozen for CryoEM. **d,** Data is collected at fixed tilts ranging from 0-60°. **e,** 0° image from Fig. 1c shown with the diffraction patten corresponding to the region denoted with the magenta rectangle. (n = 1).
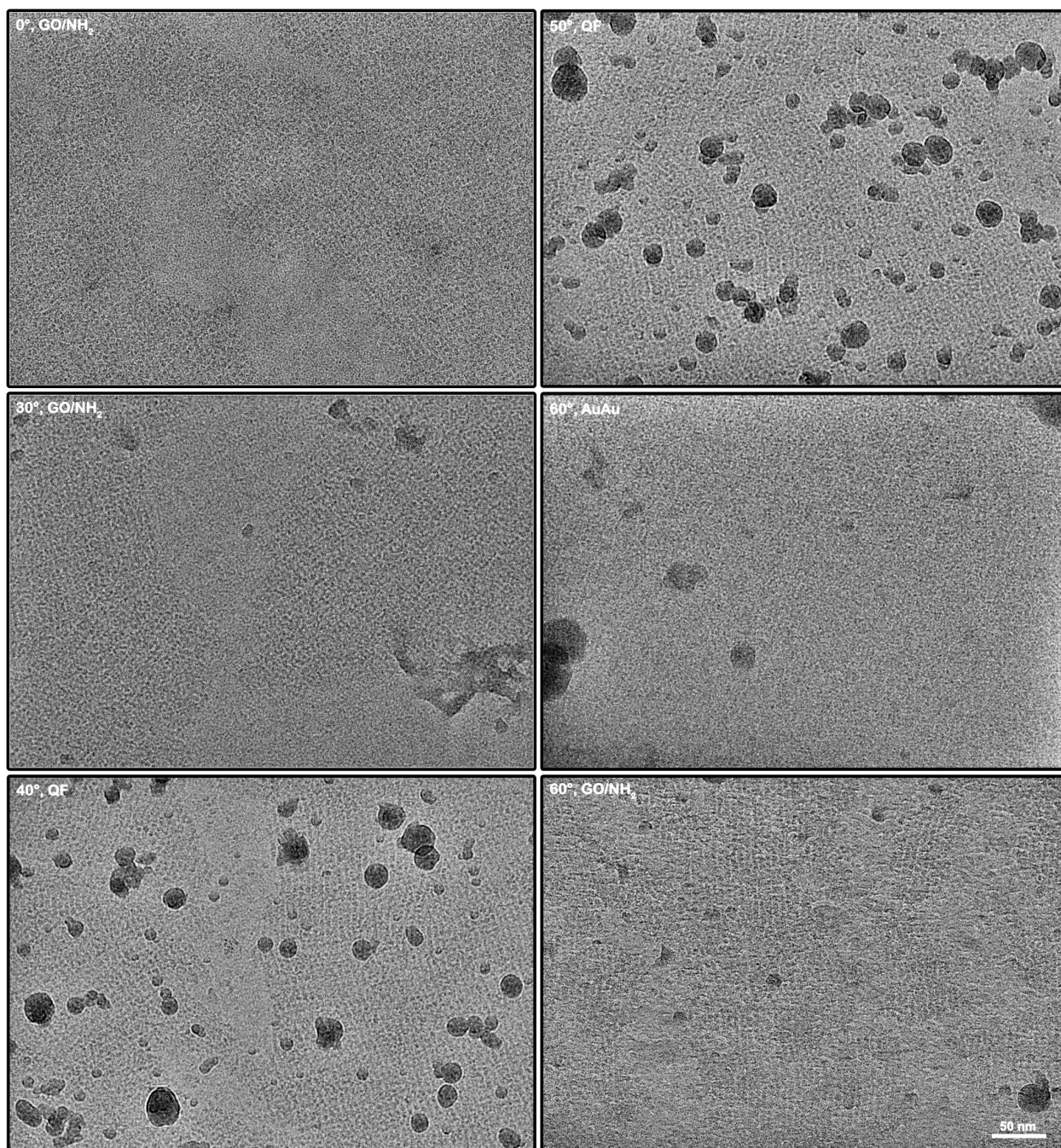
Supplementary Figure 3.4 | Example micrographs collected at the specified tilt angle.
Tilt angle is listed alongside the grid support used in each micrograph: Traditional Quantifoil
(QF), Amino-functionalized graphene-oxide (GO/NH$_2$), UltrAuFoil (AuAu). The lattices often
appear with bends and breaks as seen throughout these micrographs. Number of datasets
collected: n(0°, GO/NH$_2$) = 3; n(30°, GO/NH$_2$) = 1; n(40°, QF) = 1; n(50°, QF) = 1; n(60°, AuAu)
= 1; n(60°, GO/NH$_2$) = 3.

Supplementary Figure 3.5 | CryoEM processing pipeline.
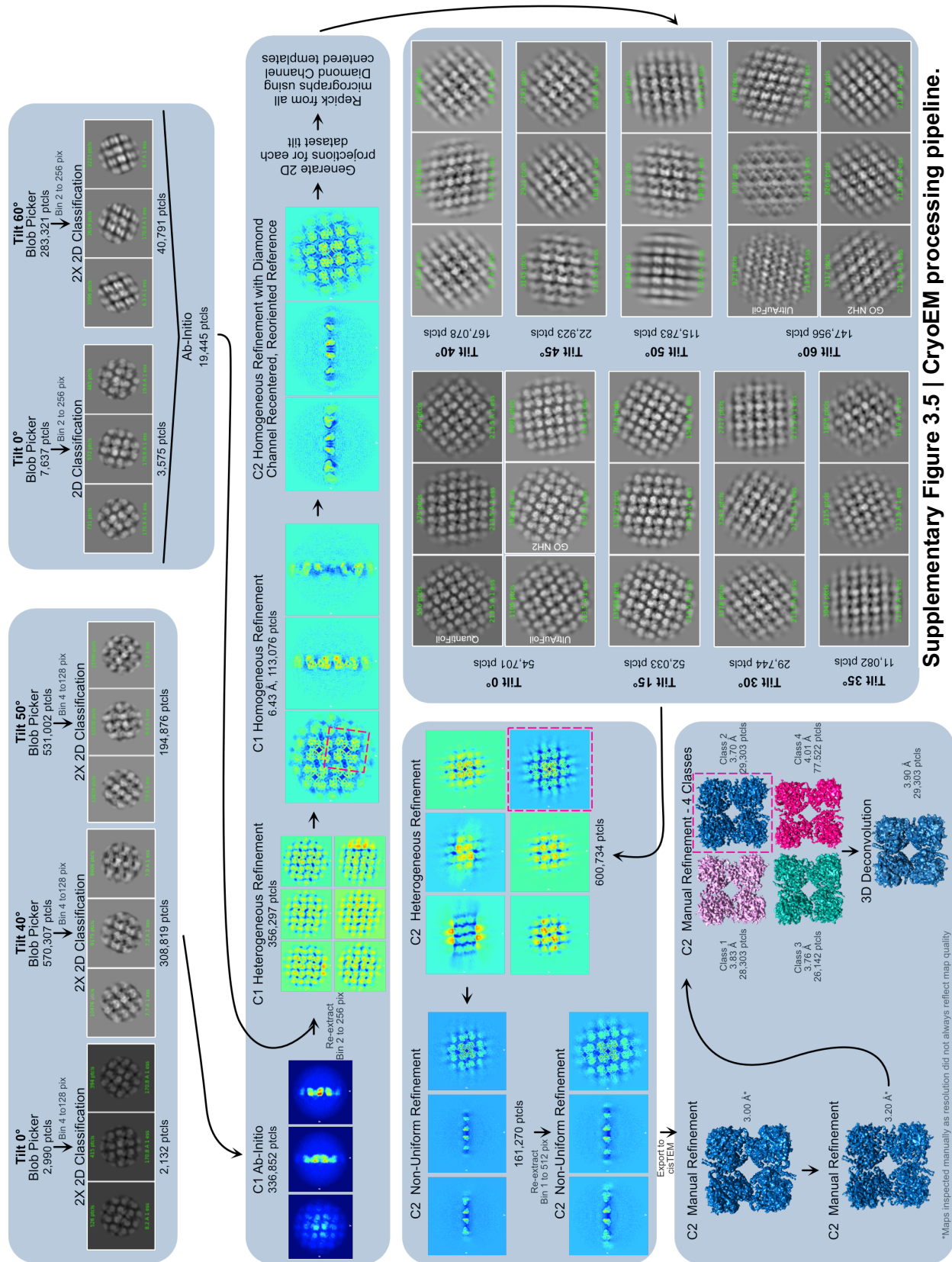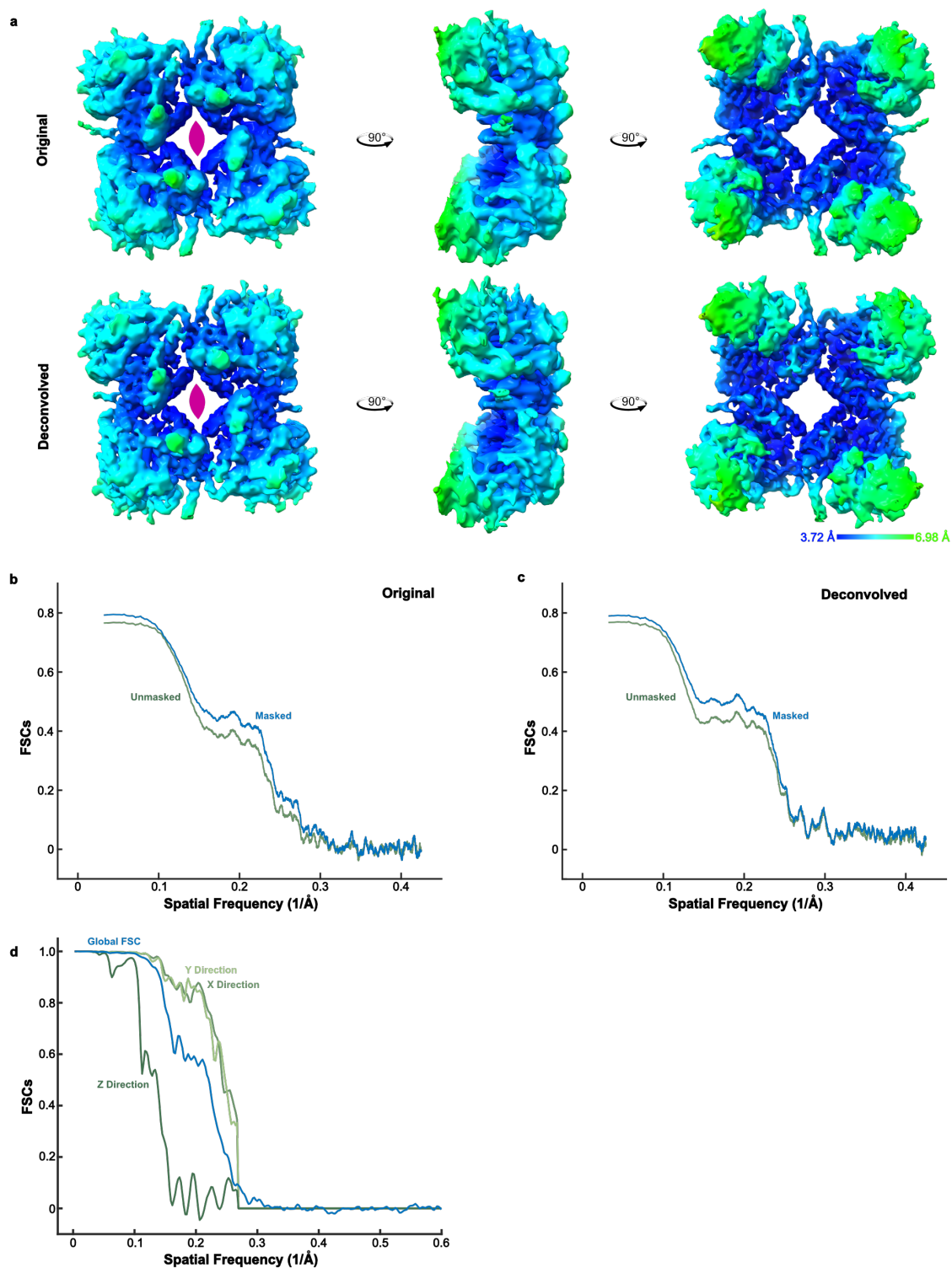
Supplementary Figure 3.6 | FSC curves corresponding to the original and deconvolved PhuN cryoEM maps.
a, Original and deconvolved maps colored by the local resolution of the original map. **b,c,** Map to model FSC for the original and deconvolved maps, respectively. **d.** 3DFSC calculated for the original map.
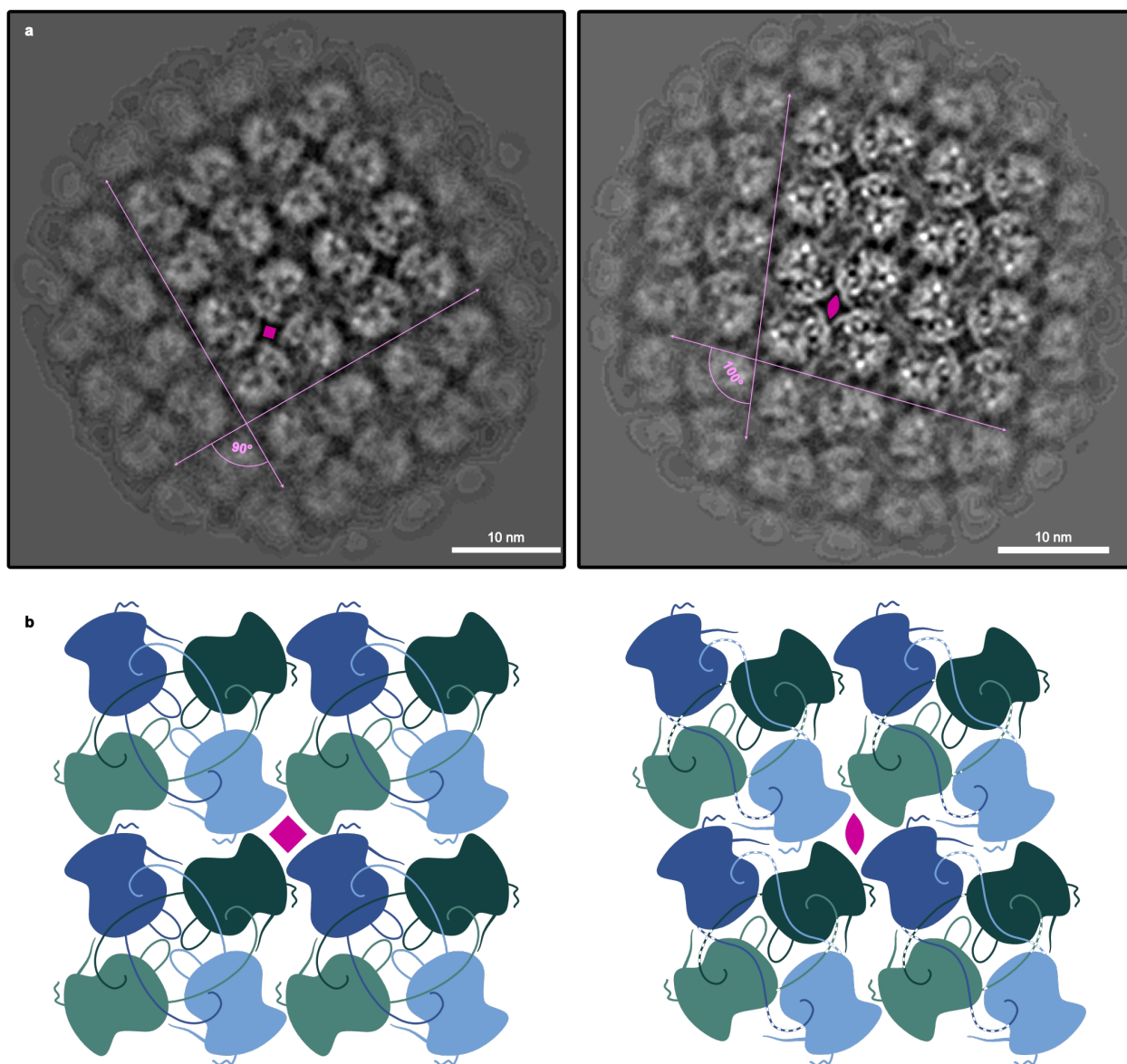
Supplementary Figure 3.7 | Comparison of p2 and p4 lattice symmetries.
a, A small population with p4 symmetry (left) is observed in 2D classification. The flexible loops at the Loop interface appear to all face the center while the remaining two interfaces are replaced by one that is most similar to the Open Hairpin interface in our p2 classes (right, full view of the class shown in Fig. 2a). **b,** Cartoon depiction of how the p4 and p2 lattices may compare.

b

c

180°

180°

a

Diamond Channel

Open Hairpin Channel

Closed Hairpin

Loop

d

PhuN-O

PhuN-C

20 Å

C-termini

Loops

e

10 nm

90°

90°

See Movie 1
for Overlay

Overlay

f
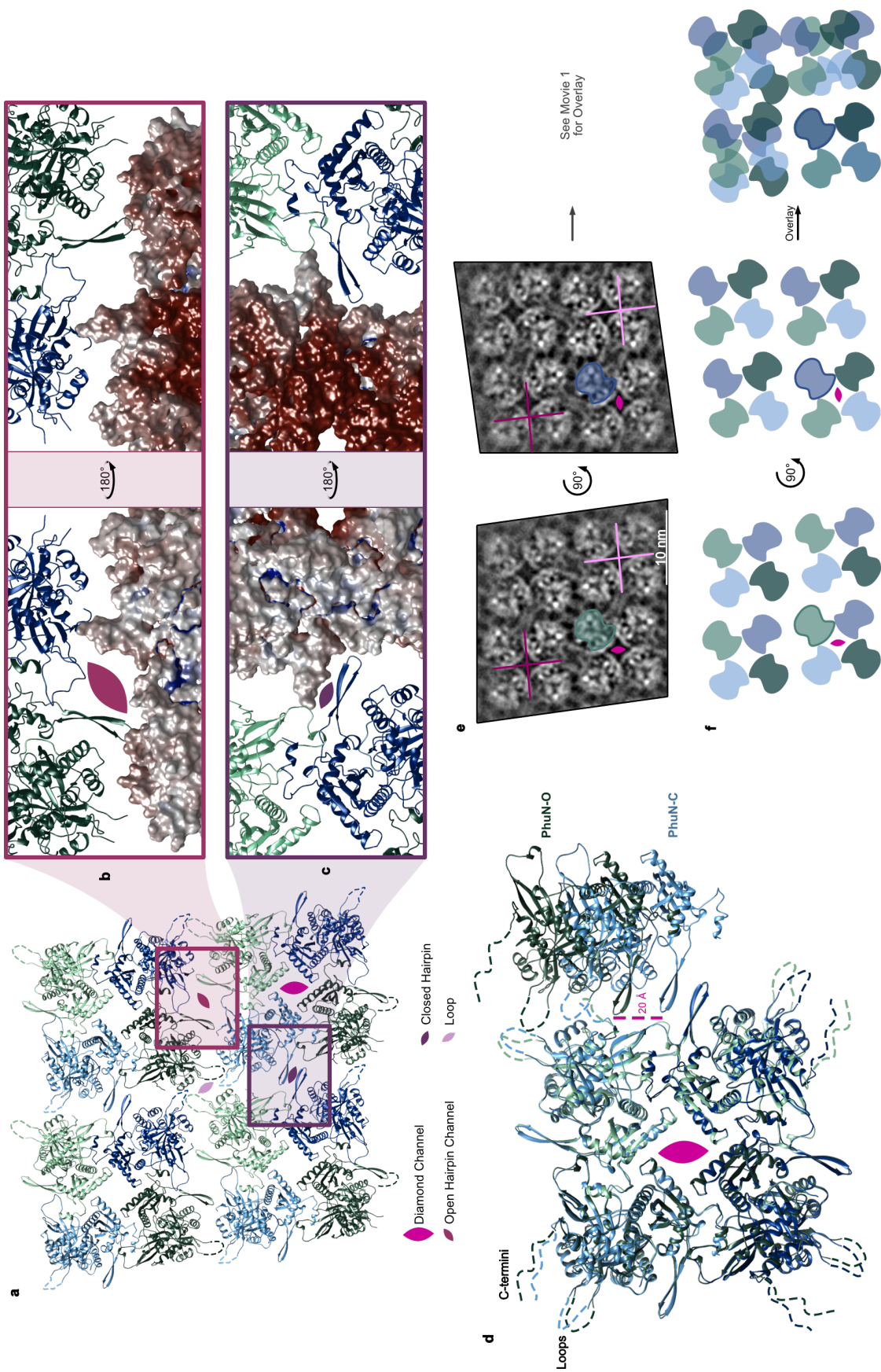
Supplementary Figure 3.8 | Closer look at the positioning of the β-Hairpin at the Open Channel and Closed Interface.
a, 16-mer model as in Fig. 2e. The two channels and two interfaces are labeled. **b,** The Open Channel is shown as viewed from the cytosol (left) and inside the phage nucleus (right). **c,** The Closed interface is similarly represented with the phage nucleus exterior on the left and interior on the right. The surfaces of two of the four subunits at each interface are shown and colored by electrostatic potential. **d,** Two tetramers aligned on the asymmetric subunits with a neighboring PhuN-O and PhuN-C subunit. The β-hairpin of PhuN-O is near the C-terminus of its neighbor while the PhuN-C β-hairpin is closer to a structured loop. **e,** 2D class, as shown in figure 2a, compared to a 90° rotation of the same 2D class after alignment on the highlighted asymmetric PhuN-O (green) and PhuN-C (blue) subunits. The centers of two Diamond Channels are shown with a light and dark pink cross (left) and are kept in the same position in the rotated image (right) to show the ~ 20 Å lattice shift. See Movie1 for a morph between the aligned subunits. **f,** Cartoon depiction of panel (e) with a direct overlay of the rotated lattice showing the new positioning of the shifted subunits. The alignment was done using the bolded PhuN-O (green) and PhuN-C (blue) as in panel (e).

91
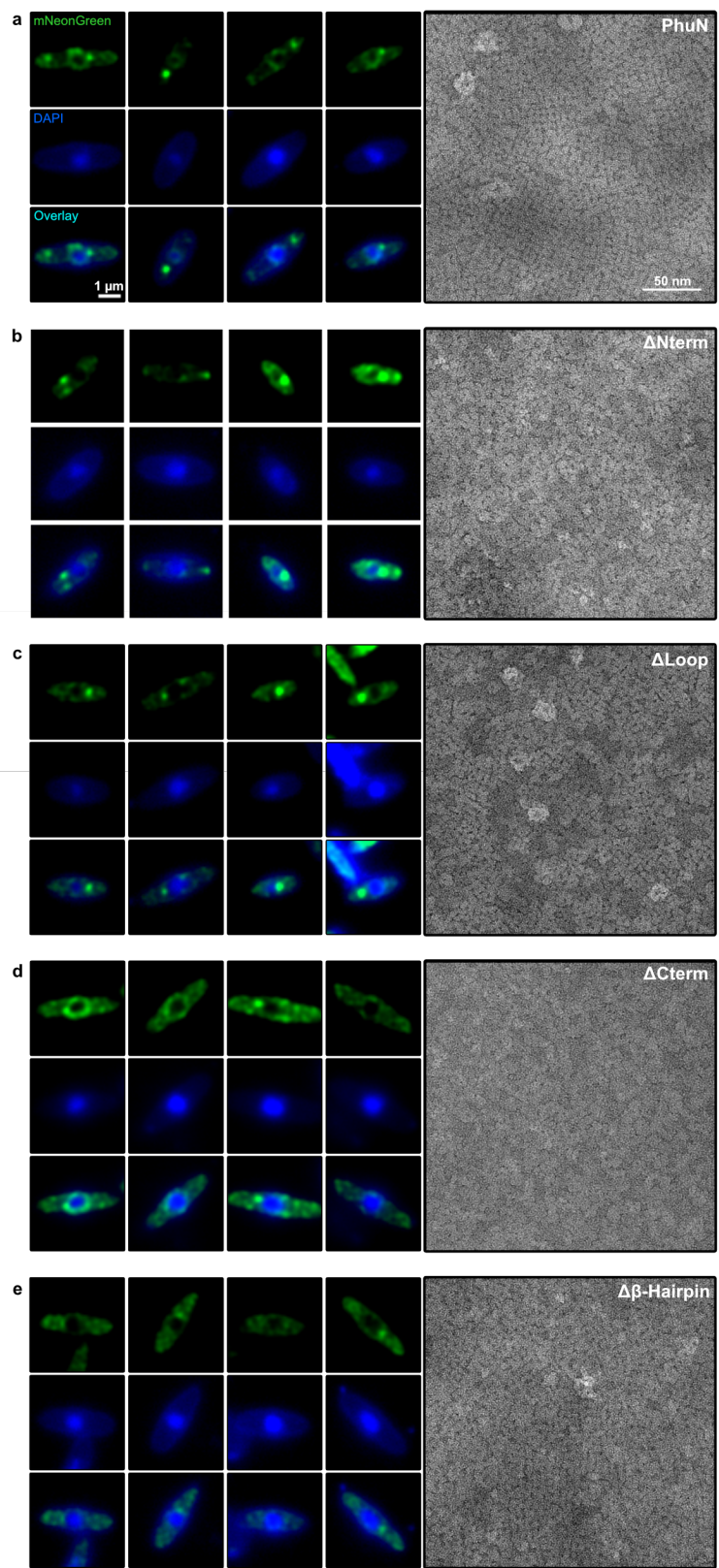
Supplementary Figure 3.9 | Additional images of ɸPA3 PhuN truncations showing defects in shell integration in vivo and self-assembly in vitro.
Live fluorescence microscopy and negative stain EM micrographs of **a,** full-length PhuN as well as the following PhuN deletions: **b,** N-terminal tail (residues 1-37), **c,** Loop (residues 272-291), **d,** C-terminal tail (residues 556-602), and **e,** β -Hairpin (residues 111-126). Fluorescence microscopy images of PhuN or PhuN deletions were collected at 40 minutes post infection, deconvolved, and are displayed in green while the the DAPI stained DNA is shown in blue. The 1 μm and 50 nm scale bars apply to all fluorescence microscopy and EM panels, respectively. (Independent sample preparation and imaging: fluorescence microscopy n = 3, negative stain EM n = 1).
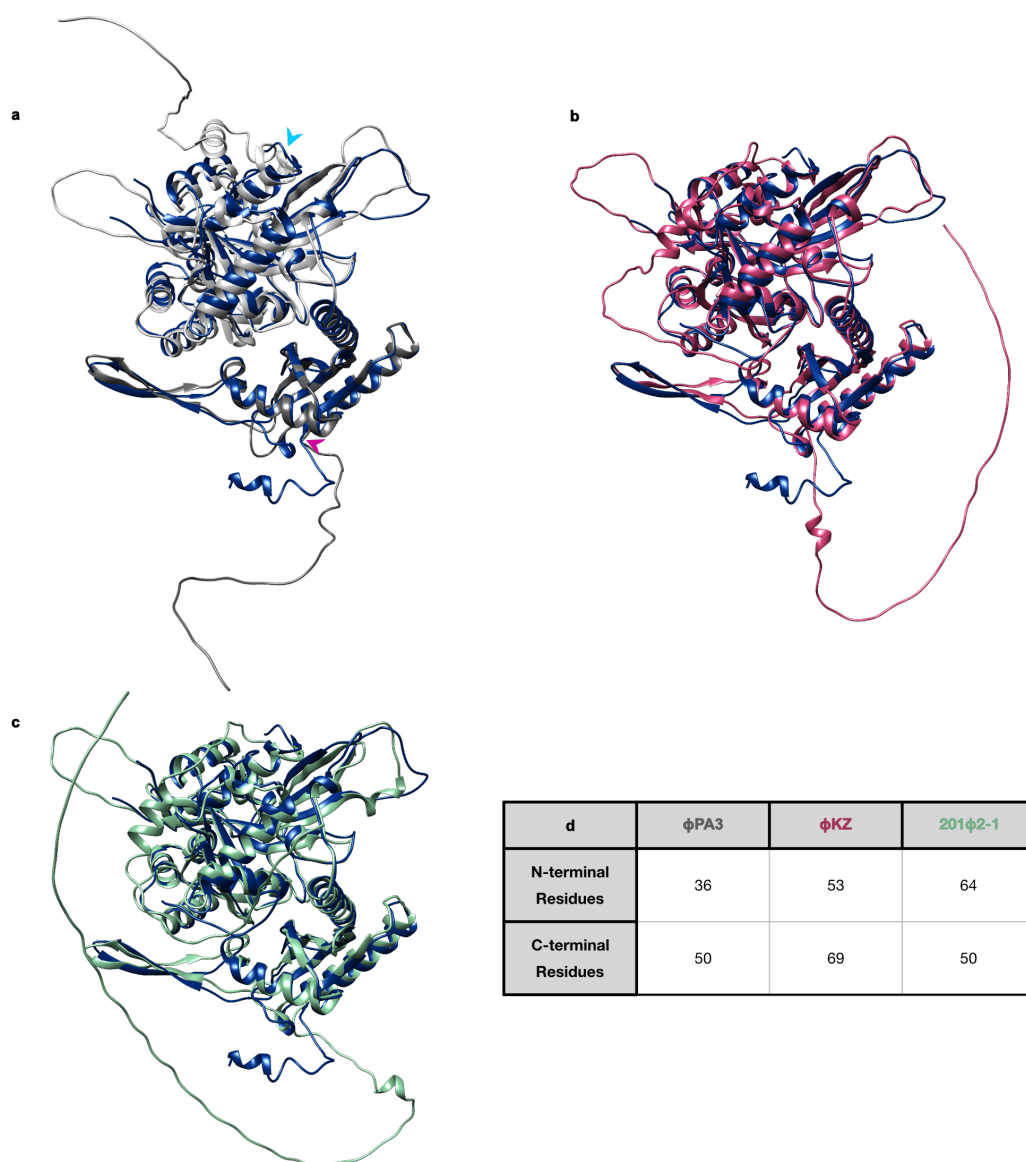
Supplementary Figure 3.10 | AlphaFold PhuN predictions compared to the final φPA3 PhuN model.
Final model of φPA3 gp53 (dark blue) overlaid with **a,** the starting φPA3 AlphaFold Prediction (grey), **b,** φKZ AlphaFold prediction (pink), and **c,** the 201φ2-1 AlphaFold prediction (green). In (b) and (c), the C-terminal tails are positioned across the back of the large, acetyltransferase-like domain. **d,** A summary of differing N- and C-terminal tail lengths as measured to the start of the helix lining the Diamond channel (pink arrow in panel a) for the N-term and from the end of the β-strand for the C-term (blue arrow in panel a).

d	φPA3	φKZ	201φ2-1
N-terminal Residues	36	53	64
C-terminal Residues	50	69	50

Tables

Supplementary Table 3.1 | Cryo-EM data collection, refinement, and validation statistics.

	phiPA3 PhuN Tetramer, p2 (EMDB-29550)	Deconvolved phiPA3 PhuN Tetramer, p2 (EMD-29310); phiPA3 PhuN Tetramer, p2 (PDB 8FNE)	Tracing p2 phiPA3 PhuN Tetramer Interfaces (EMD-29451); Representation of 16-mer phiPA3 PhuN Lattice, p2 (PDB 8FV5)
Data collection and processing			
Magnification	105 kx		
Voltage (kV)	300		
Electron exposure (e–/Å^2)	66 or 67		
Defocus range (μm)	0.8-2.5		
Pixel size (Å)	0.834-0.835		
Symmetry imposed	C2		
Initial particle images (no.)	600,734		
Final particle images (no.)	29,303		161,270
Map resolution (Å) FSC threshold	3.9 0.143	-	7.0 0.143
Map resolution range (Å)	3.72-6.98		4.52-7.95
Refinement			
Initial model used	AlphaFold Prediction		-
Model resolution (Å) FSC threshold	6.8 0.5	6.9 0.5	-
Map sharpening B factor (Å^2)	-	80	-
Model composition Non-hydrogen atoms Protein residues		16,864 2,148	-
R.m.s. deviations Bond lengths (Å) Bond angles (°)		0.018 1.748	-
Validation MolProbity score Clashscore Poor rotamers (%)		0.60 0.27 0.00	-
Ramachandran plot Favored (%) Allowed (%) Disallowed (%)		98.96 1.04 0.00	-

Supplementary Table 3.2 | Cryo-EM data collection parameters.

Collected Datasets by Grid Type, Tilt, and Frame Rate						
Grid Type	**Quantifoil**	**GO/NH₂**	**GO/NH₂**	**GO/NH₂**	**UltrAuFoil**	**UltrAuFoil GO/NH₂**
Tilt Angle	0°, 40°, 50°	0°, 60°	0°, 15°	0°, 30°, 45°, 60°	0°, 60°	35°
Frames	120	118	246	444	444	444
Magnification	105,000					
Voltage (kV)	300					
Electron Dose (e-/Å²)	66	67	67	67	67	67
Defocus Range (µm)	1.5-2.5	0.8-1.6	0.8-1.6	0.8-1.6	0.8-1.6	0.8-1.6
Pixel Size (Å)	0.834	0.835	0.835	0.835	0.835	0.835

Supplementary Table 3.3 | Dali Server searches of the PhuN NTD and CTD.

					NTD

PDB ID-Chain	Z	rmsd	lali	nres	%ID	PDB Description
3s1s-A	5.5	3.1	96	871	7	RESTRICTION ENDONUCLEASE BPUSI
3nat-A	5.2	3.4	98	156	14	UNCHARACTERIZED PROTEIN
2w8m-B	4.7	4.5	101	168	11	ORF D212
6em5-b	4.6	3.7	98	421	15	5.8S RIBOSOMAL RNA
3nbm-A	4.5	3.1	80	104	6	PTS SYSTEM, LACTOSE-SPECIFIC IIBC COMPONENTS
1yra-B	4.5	3.3	101	261	9	ATP(GTP)BINDING PROTEIN
6vda-A	4.5	3.1	67	86	10	ZT_DIMER DOMAIN-CONTAINING PROTEIN
2eo0-A	4.5	3.4	89	130	7	HYPOTHETICAL PROTEIN ST1444
2ayx-A	4.4	3.1	84	254	10	SENSOR KINASE PROTEIN RCSC
1d02-B	4.4	3.9	96	200	11	DNA (5'-D(*GP*CP*CP*AP*AP*TP*TP*GP*GP*C)-3')
6xpd-A	4.3	3.7	76	303	3	ZINC TRANSPORTER 8
6vd8-B	4.3	3.7	63	78	6	PROBABLE CATION EFFLUX SYSTEM PROTEIN
2vyc-A	4.3	3.8	96	755	7	BIODEGRADATIVE ARGININE DECARBOXYLASE
1s28-A	4.2	3.2	79	130	6	ORF1
5fr7-B	4.1	2.9	77	146	12	AMYR
4ccs-A	4.1	3.5	93	226	6	CBIX
7p82-C	4.1	5.0	89	408	3	S-ADENOSYLMETHIONINE SYNTHASE
5u4u-A	4.0	3.5	91	157	5	MGC81300 PROTEIN
6jcx-F	4.0	3.0	66	186	9	DNA-DIRECTED RNA POLYMERASE SUBUNIT ALPHA
1hrk-A	4.0	4.4	100	359	7	FERROCHELATASE

					CTD

PDB ID-Chain	Dali Z-score	rmsd	lali	nres	%id	PDB Description
7byy-A	5.7	3.9	134	180	9	ACETYLTRANSFERASE
1y9k-B	5.2	3.6	118	154	8	IAA ACETYLTRANSFERASE
6c30-A	5.1	3.4	129	213	8	GNAT FAMILY ACETYLTRANSFERASE
3f5b-A	5.1	3.0	120	172	6	AMINOGLYCOSIDE N(6')ACETYLTRANSFERASE
1kzf-A	5.0	4.3	127	198	9	ACYL-HOMOSERINELACTONE SYNTHASE ESAI
1xeb-A	5.0	3.9	112	149	8	HYPOTHETICAL PROTEIN PA0115
2vzz-A	4.6	3.5	126	210	6	RV0802C
7kd7-A	4.6	3.9	124	204	9	N-ALPHA-ACETYLTRANSFERASE 40
3owc-A	4.5	3.8	120	172	10	PROBABLE ACETYLTRANSFERASE
7cs1-A	4.5	4.5	121	191	10	AMINOGLYCOSIDE 2'-N-ACETYLTRANSFERASE
5hgz-A	4.4	4.2	124	211	6	N-ALPHA-ACETYLTRANSFERASE 60
6edw-B	4.3	3.8	122	746	11	ISOCITRATE LYASE 2
5ktc-A	4.3	3.5	119	187	10	FDHC
3efa-A	4.2	3.5	108	146	3	PUTATIVE ACETYLTRANSFERASE
5o7o-A	4.1	3.0	108	146	6	DESFERRIOXAMINE SIDEROPHORE BIOSYNTHESIS PROTEIN
2wpw-A	4.0	4.3	129	328	5	ORF14
5gnc-A	3.9	12.1	72	597	6	AVH146
3fnc-B	3.8	3.5	111	161	6	PUTATIVE ACETYLTRANSFERASE
4rs2-B	3.8	3.9	112	183	6	PREDICTED ACYLTRANSFERASE WITH ACYL-COA N-ACYLTRA
3t9y-A	3.8	3.1	95	135	9	ACETYLTRANSFERASE, GNAT FAMILY

Supplementary Table 3 | Dali Server searches of the PhuN NTD and CTD.
20 highest scoring Dali Server outputs with the NTD and CTD as the input.

References

1. Durzyńska, J. & Goździcka-Józefiak, A. Viruses and cells intertwined since the dawn of evolution. *Virol J* **12**, 169 (2015).

2. Borges, A. L., Davidson, A. R. & Bondy-Denomy, J. The Discovery, Mechanisms, and Evolutionary Impact of Anti-CRISPRs. *Ann Rev Virol* **4**, 1–23 (2016).

3. Mendoza, S. D. *et al.* A bacteriophage nucleus-like compartment shields DNA from CRISPR nucleases. *Nature* 1–5 (2019) doi:10.1038/s41586-019-1786-y.

4. Chaikeeratisak, V. *et al.* Assembly of a nucleus-like structure during viral replication in bacteria. *Science* **355**, 194–197 (2017).

5. Chaikeeratisak, V. *et al.* The Phage Nucleus and Tubulin Spindle Are Conserved among Large Pseudomonas Phages. *Cell Reports* **20**, 1563–1571 (2017).

6. Kraemer, J. A. *et al.* A Phage Tubulin Assembles Dynamic Filaments by an Atypical Mechanism to Center Viral DNA within the Host Cell. *Cell* **149**, 1488–1499 (2012).

7. Erb, M. L. *et al.* A bacteriophage tubulin harnesses dynamic instability to center DNA in infected cells. *Elife* **3**, e03197 (2014).

8. Zehr, E. A. *et al.* The Structure and Assembly Mechanism of a Novel Three-Stranded Tubulin Filament that Centers Phage DNA. *Structure* **22**, 539–548 (2014).

9. Chowdhury, C., Sinha, S., Chun, S., Yeates, T. O. & Bobik, T. A. Diverse Bacterial Microcompartment Organelles. *Microbiol Mol Biol R* **78**, 438–468 (2014).

10. Yeates, T. O., Tsai, Y., Tanaka, S., Sawaya, M. R. & Kerfeld, C. A. Self-assembly in the carboxysome: a viral capsid-like protein shell in bacterial cells. *Biochem Soc T* **35**, 508–511 (2007).

11. Kerfeld, C. A., Aussignargues, C., Zarzycki, J., Cai, F. & Sutter, M. Bacterial microcompartments. *Nat Rev Microbiol* **16**, 277–290 (2018).

12. Wang, Z., Lou, J. & Zhang, H. Essence determines phenomenon: Assaying the material properties of biological condensates. *J Biol Chem* **298**, 101782 (2022).

13. Birkholz, E. A. *et al.* A Cytoskeletal Vortex Drives Phage Nucleus Rotation During Jumbo Phage Replication in E. coli. *Biorxiv* 2021.10.25.465362 (2021) doi:10.1101/2021.10.25.465362.

14. Al-Shayeb, B. *et al.* Clades of huge phages from across Earth's ecosystems. *Nature* **578**, 425–431 (2020).

15. Baek, M. *et al.* Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876 (2021).

16. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

17. Laughlin, T. G. *et al.* Architecture and self-assembly of the jumbo bacteriophage nuclear shell. *Nature* **608**, 429–435 (2022).

18. Arce-Rodríguez, A., Volke, D. C., Bense, S., Häussler, S. & Nikel, P. I. Non-invasive, ratiometric determination of intracellular pH in Pseudomonas species using a novel genetically encoded indicator. *Microb Biotechnol* **12**, 799–813 (2019).

19. Kelly, D. F., Dukovski, D. & Walz, T. A practical guide to the use of monolayer purification and affinity grids. *Methods Enzymol* **481**, 83–107 (2010).

20. Wang, F. *et al.* Amino and PEG-amino graphene oxide grids enrich and protect samples for high-resolution single particle cryo-electron microscopy. *J Struct Biol* **209**, 107437 (2020).

21. Righetto, R. D., Biyani, N., Kowal, J., Chami, M. & Stahlberg, H. Retrieving high-resolution information from disordered 2D crystals by single-particle cryo-EM. *Nature Communications* **10**, (2019).

22. Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14**, 331–332 (2017).

23. Holm, L. & Rosenström, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res* **38**, W545–W549 (2010).

24. Nichols, R. J., Cassidy-Amstutz, C., Chaijarasphong, T. & Savage, D. F. Encapsulins: molecular biology of the shell. *Crit Rev Biochem Mol* **52**, 1–12 (2017).

25. Tanaka, S., Sawaya, M. R. & Yeates, T. O. Structure and Mechanisms of a Protein-Based Organelle in Escherichia coli. *Science* **327**, 81–84 (2010).

26. Johnson, J. E. & Speir, J. A. Quasi-equivalent viruses: a paradigm for protein assemblies. *J Mol Biol* **269**, 665–75 (1997).

27. Kirchhausen, T., Owen, D. & Harrison, S. C. Molecular Structure, Function, and Dynamics of Clathrin-Mediated Membrane Traffic. *Csh Perspect Biol* **6**, a016725 (2014).

28. Chaikeeratisak, V. *et al.* Viral speciation through subcellular genetic isolation and virogenesis incompatibility. *Nature Communications* **12**, (2021).

29. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**, 676–682 (2012).

30. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods* **14**, (2017).

31. Scheres, S. H. W. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* **180**, 519–530 (2012).

32. Punjani, A., Zhang, H. & Fleet, D. J. Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nat Methods* **17**, 1214–1221 (2020).

33. Grant, T., Rohou, A. & Grigorieff, N. cisTEM, user-friendly software for single-particle image processing. *eLife* **7**, (2018).

34. Asarnow, D., Palovcak, E. & Cheng, Y. UCSF pyem v0.5. *Zenodo* (2019) doi:10.5281/zenodo.3576630.

35. Arigovindan, M. *et al.* High-resolution restoration of 3D structures from widefield images with extreme low signal-to-noise-ratio. *Proc National Acad Sci* **110**, 17344–17349 (2013).

36. Waugh, B. *et al.* Three-dimensional deconvolution processing for STEM cryotomography. *Proc National Acad Sci* **117**, 27374–27380 (2020).

37. Croxford, M. *et al.* Entropy-regularized deconvolution of cellular cryotransmission electron tomograms. *Proc National Acad Sci* **118**, e2108738118 (2021).

38. Chen, H., Clyborne, W. K., Sedat, J. W. & Agard, D. A. PRIISM: an integrated system for display and analysis of 3-D microscope images. *Biomed Image Process Three-dimensional Microsc* 784–790 (1992) doi:10.1117/12.59604.

39. Dang, S. *et al.* Cryo-EM structures of the TMEM16A calcium-activated chloride channel. *Nature* **552**, 426–429 (2017).

40. Wang, R. Y.-R. *et al.* Automated structure refinement of macromolecular assemblies from cryo-EM maps using Rosetta. *Elife* **5**, e17219 (2016).

41. Croll, T. I. ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr Sect D* **74**, 519–530 (2018).

42. Pettersen, E. F. *et al.* UCSF Chimera—A visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605–1612 (2004).

43. Pettersen, E. F. *et al.* UCSF ChimeraX : Structure visualization for researchers, educators, and developers. *Protein Sci* **30**, 70–82 (2020).

44. Afonine, P. V. *et al.* New tools for the analysis and validation of cryo-EM maps and atomic models. *Acta Crystallogr Sect D Struct Biology* **74**, 814–840 (2018).

45. Williams, C. J. *et al.* MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci* **27**, 293–315 (2018).

46. Kremer, J. R., Mastronarde, D. N. & McIntosh, J. R. Computer visualization of three-dimensional image data using IMOD. *J Struct Biol* **116**, 71–6 (1996).

47. Jurrus, E. *et al.* Improvements to the APBS biomolecular solvation software suite. *Protein Sci* **27**, 112–128 (2018).

Chapter 4

Disrupting PhuN Interactions to Build an Understanding of the Phage Nuclear Lattice Assembly Mechanism

Preface

I have always been interested in not only seeing what these incredible phage structures look like but also determining how such a large, protein-based organelle nucleates, grows, and adds subunits. While we do not have a mechanism quite yet, the results presented in this chapter demonstrate that the *in vitro* 2D lattice assembly protocol is a powerful tool for future work in that direction. Since the acceptance of the publication included in Chapter 3, I have pursued the deletions described therein via cryo-EM to assay for lattice defects.

Introduction

The phage nucleus is a remarkable proteinaceous assembly that appears to be constructed of just one protein—PhuN. While we now have a better idea of what these assemblies look like as well as some hints about the flexibility and motion that PhuN assemblies are capable of as tetramers (p2, p4 symmetry) and as protomers (out of plane rotation in cuboids[1]), we have very little understanding of how these assemblies nucleate, grow, and maintain selectivity.

Chapter 3 identified the N-terminus, β-hairpin, Loop, and C-terminus as components of PhuN important for assembly. Deletion of these features resulted in abrogation of lattice formation *in vitro*, as tracked by negative stain EM, for the N-terminus and C-terminus deletions. In contrast, the Loop deletion retained some and the β-hairpin retained all propensity for spontaneous 2D crystallization in negative stain EM. *In vivo* testing of these same deletions in the context of full-length, wild-type PhuN resulted in no incorporation of the N-terminal and β-hairpin deletions; the Loop was mildly and the C-terminal tail deletion seemed strongly

incorporated into the phage nuclear shell. The lattice defects introduced with these deletions and causing these phenotypes were not known.

Due to their continued formation of 2D crystals *in vitro* while having severely diminished or completely abrogated *in vivo* incorporation phenotypes, the Loop and β-hairpin results necessitated further experiments. While all four deletions were investigated here, lattices were only observed to form using the Loop and β-hairpin deletions **(Fig. 4.1a, Fig. 4.2a,b)**. These results are discussed below, providing a first look at how these flexible PhuN elements contribute to lattice formation.

β-hairpin Deletion Enables Tighter Tetramer Particle Jamming

Micrographs of the β-hairpin deleted lattices appeared slightly more compact than those obtained with full-length PhuN **(Fig. 4.1a)**. After processing, a broad assortment of 2D classes emerged showing that the lattices have been strongly perturbed **(Fig. 4.1b)**. Only one class emerged showing a p2 symmetric class similar to those observed with full-length PhuN, except the hairpins are missing at what would be the Open channel and Closed interface. The remaining classes revealed tetramers that are closely packed. The straight lines between rows and columns of tetramers became wavy with the tetramers rotated, jamming together more tightly. The Loop interface was also impacted in two different ways. In some classes, the loops are visible as lines going into the center of the interface; while in others, the center is empty and a star shaped periphery appears, possibly representing a different orientation of the loops. In some of the 2D classes, the protomers within tetramers appear different, as well. One possibility is that they are rotating out of plane with the hairpin gone.

While the full impact of these changes is not yet clear, it seems that the β-hairpin is important for keeping the tetramers spaced and for establishing the p2 and p4 symmetries described in Chapter 3. Without the β-hairpin, the tetramers seem to experience particle jamming, shifting in the 2D lattice plane and pushing together tightly. If the subunits additionally

rotate out of plane, this could be detrimental to the integrity of the lattices and establishing an

effective, selective barrier between the interior and exterior of the phage nucleus. The observed

defects are likely why, despite forming 2D crystalline lattices *in vitro*, the β-hairpin deletion is

excluded from the phage nuclear shell *in vivo*. Further experiments will be necessary to

determine the specific biological, biophysical, or other factors making these defects detrimental

to the phage infection.

Loop Deletion Enriches p4 Symmetry and Class II Assemblies

While the lattices look similar to those grown with full-length PhuN, deleting the large

flexible loop at the Loop interface results in 2D classes of primarily p4 particles **(Fig. 4.2a-c)**.

This suggests the Loop plays a large role in mediating PhuN lattice symmetry. How this occurs

is not clear, but given that the long, flexible C-terminal tail crosses over between tetramers in

that region of PhuN, it is possible there is an interaction between them.

In Chapter 3, we note that by negative stain EM this deletion seemed to have enriched

for the ~25 nm in diameter Class II assemblies originally observed during full-length PhuN

MonoQ purification. This finding is confirmed in the lipid-assembled lattices described here,

resulting in the recovery of a clear 2D class of Class II assemblies **(Fig. 4.2c)**. These Class II

assemblies correspond to the cuboids recently characterized from *in vitro* assemblies of PhuN

from bacteriophage 201ϕ2-1.[1] What stands out about these cuboids is the location of the

individual particles forming this 2D class **(Fig. 4.2b)**. The particles appear in a somewhat regular

pattern across lattices. Moreover, the simple presence of these assemblies within this sample

suggests they may have been stably attached to the lattices as they were adsorbed to grids for

freezing. The exciting possibility from such stable attachment is that the Loop could play an

important role in subunit addition and lattice growth. The cuboids may be attached because they

are the means for subunit addition into assembled lattices but, without the Loop, they cannot

integrate or it becomes a slow process, allowing them to accumulate along the lattice sheets.

Alternatively, without the Loop, the lattices may not be able to receive new subunits. If either of these is the case, the observed lattices grew largely by subunit addition at the edges. While it is unclear what the exact mechanistic defect is, it is clear that deleting the Loop encourages the buildup of Class II cuboids and enriches for p4 symmetric lattices.

Going forward, it will be exciting to determine the exact mechanistic role of this Loop. One approach to addressing this question is collecting tomography data on lattices decorated with Class II cuboids to determine how the Class II cuboids associate with the lattices. Another approach is to grow full-length PhuN lattices and add Loop deleted cuboids to test for enrichment along the lattices. Vice versa, Loop deleted lattices can be assembled and full-length cuboids can be introduced, purified using the recently published protocol, to see whether cuboids, sheets, or both need the loops for successful subunit addition. It is, of course, possible that something entirely different is happening with this Loop deletion, but this is currently the most simple hypothesis.

Deletions and Looking Ahead

The ability to grow PhuN lattices *in vitro* and monitor the impact of various deletions is a powerful tool to determine the mechanism for PhuN lattice growth and subunit addition. *In vivo*, this process may be aided by other factors, a subject briefly touched upon in Chapter 5.

We have only explored these concepts within the φPA3 bacteriophage system. At the start of this project, I briefly worked with PhuN from φKZ as well as 201φ2-1. I saw similar Class I, II, and III assemblies using PhuN 201φ2-1 **(Fig. 4.3)** as those observed with φPA3 PhuN, but that sample proved to be less amenable to *in vitro* work in my hands. With that said, repeating these experiments and testing these deletions in the homologous bacteriophage systems will deepen our understanding of the fundamental principles behind phage nucleus formation. One particularly interesting question is whether pH 6.5 is the correct condition for all PhuN proteins to form lattices or whether this will need to be fine-tuned for each system separately.

Methods

2D Crystal Preparation

      PhuN deletions were purified and 2D crystals were grown and frozen following the protocols detailed in Chapter 3 using QUANTIFOIL R1.2/1.3 400-mesh copper holey carbon grids. The data presented here uses protein from the same preparations used in Chapter 3. Fresh lipids were ordered for these experiments. In samples forming 2D crystals, they were easier to find in holes and populated most of the grids imaged. I recommend replacing lipids after no more than 10 months.

Cryo-EM Data Collection

      Images were collected at 0° tilt at 200 keV on the Talos-Arctica equipped with a Gatan K3 direct detector camera. A nominal magnification of 45,000x (physical pixel size of 0.865 Å) was used with a total dose of 64.1 e-/Å^2.

Cryo-EM Data Processing

      For the β-hairpin deletion, 236/421 micrographs had lattices and were of sufficient quality (not images of carbon) for processing. Particles were picked using the 0° PhuN template set used in Chapter 3 and 155,379 particles were extracted. Two rounds of 2D classification were done and representative classes are shown in Fig 4.1b.

      For the Loop deletion, 38/213 exposures contained high quality lattices in holes. From these, particles were picked using the same 0° PhuN template set used in Chapter 3. 24,371 particles were extracted and subjected to one round of 2D classification. Example 2D classes are shown in Fig. 4.2b.

Negative Stain EM

Samples prepared as described in Chapter 3 for full-length PhuN from φPA3.

Figures
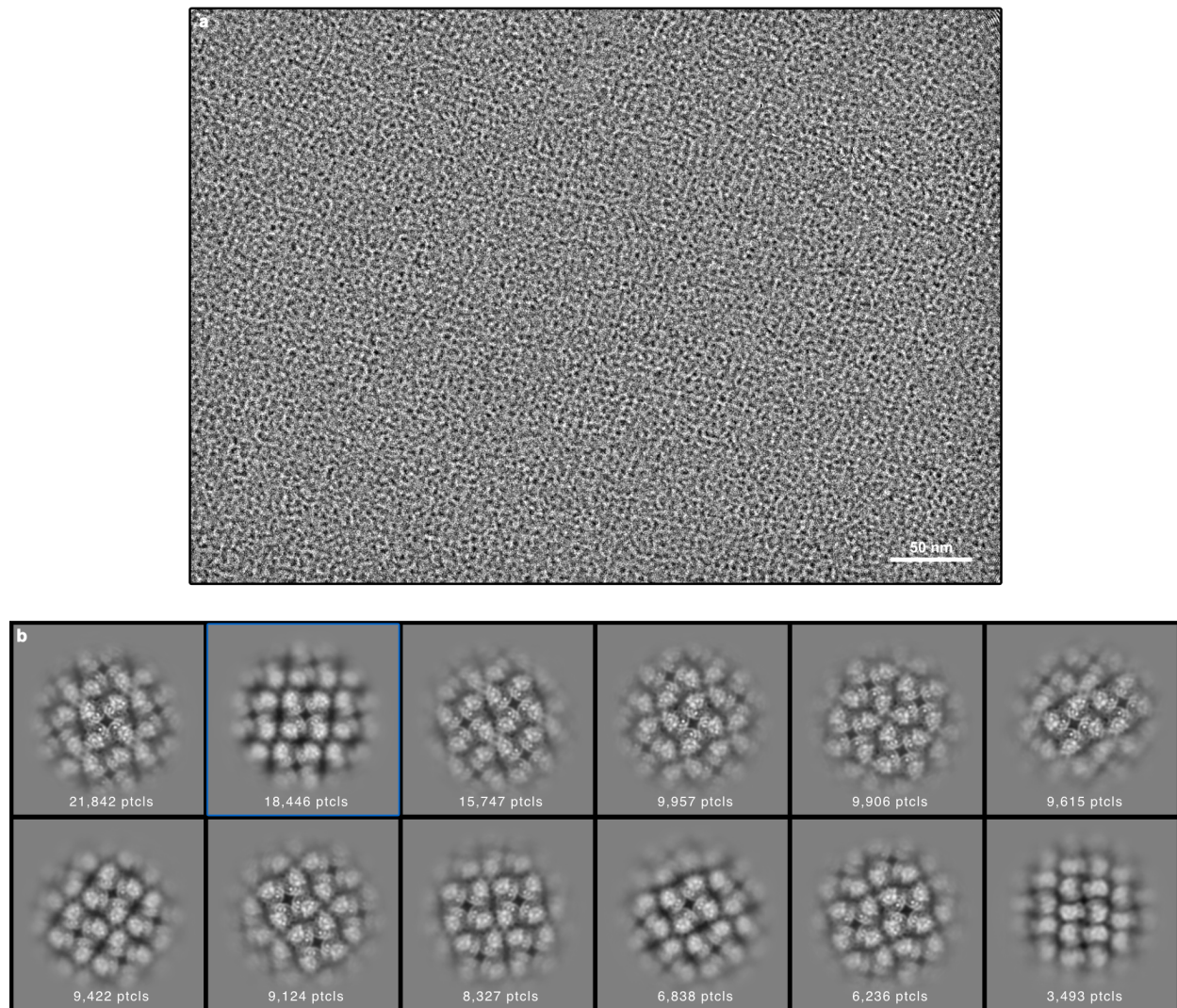
Figure 4.1 | PhuN β-Hairpin Deletion.
a, Representative micrograph of a lattice grown using β-Hairpin-deleted PhuN. Lattice appears slightly more densely packed than full-length PhuN lattices. **b,** 2D classes obtained from the dataset represented in panel a. Most classes show rotated and tightly packed tetramers. Apart from the missing β-hairpins, the class outlined in blue looks like a p2 class obtained from full-length PhuN 2D classification.

Figure 4.2 | PhuN Loop Deletion.
a, Example image of a lattice grown with the Loop deletion of PhuN. **b,** Micrograph with an enrichment of Class II cuboids, appearing as dark patches. Five example Class II cuboids are circled in pink. **c,** Example 2D classes obtained from processing the dataset represented in panels a and b. The classes have p4 symmetry, no clear p2 class emerged. The class boxed in blue corresponds to a Class II cuboid.

108

Figure 4.3 | PhuN from bacteriophage 201φ2-1 forms assemblies on the MonoQ.
MonoQ anion exchange purification of PhuN from 201φ2-1 forms a broad MonoQ peak, slightly different from φPA3 PhuN. 280 nm signal is shown in blue, 260 nm signal is shown in red. Class I-IV assembly peaks are highlighted and shown with their corresponding negative stain micrographs. Classes III and IV were quite similar.

References

1. Laughlin, T. G. *et al.* Architecture and self-assembly of the jumbo bacteriophage nuclear

     shell. *Nature* 608, 429–435 (2022).

Chapter 5

Loose Ends and Next Steps: DNA, Nucleotides, and Gp2, oh my!

Preface

The work in this chapter is incomplete and includes a summary of curious observations that I think are worth following up on. Some were accidental observations, while others were probed for directly.

The DNA binding experiments were done with University of California, Berkeley undergraduate Gayathri Sundaram under my supervision and have been described in her own words in her undergraduate thesis.

Gp2 is, to my knowledge, the only other protein identified as a marker for the phage nucleus. Its identity and relevant *Pseudomonas* expression plasmids were shared by Dr. Katrina Nguyen during her time in the Pogliano Lab at UCSD within the context of our, at that time, collaboration. Dr. Chari Noddings worked with Gp2 during her rotation with me.

Putative PhuN:DNA Interactions

The primary cargo of the phage nucleus is DNA.[1] As we've seen in earlier chapters, the 2D crystalline PhuN assembly enclosing the phage DNA forms a protective boundary, excluding various DNA-targeting endonucleases and CRISPR systems.[2] From this, we know that at the start of infection the phage genome must be protected. In Chapter 2, we showed that early stalled jumbo phage infections have DNA puncta at the poles.[2] Unpublished work from our lab and a recent publication point to early compartmentalization of the phage DNA in what is likely a membrane-based structure.[3] The suggested name for this structure is the phajectosome, per Dr. Deepto Mozumdar. Eventually the DNA must copy or transfer into the PhuN enclosed phage nucleus. Likewise, each newly synthesized phage genome must exit the phage nucleus to package the nascent, empty capsids that dock around its exterior.[1,4] The mechanism for capsid

111

attachment to the phage nucleus is unknown. Due to the abundance of DNA and a need for it to become enclosed and eventually exit the phage nucleus, we wondered whether PhuN, as the primary nuclear enclosure protein, can directly interact with DNA.

We used fluorescence polarization to ask whether PhuN from bacteriophage ϕPA3 interacts with DNA. For these experiments, I first surveyed the ϕPA3 genome for high, low, and average GC content; using 30 base pair windows for analysis, the genome ranges from 10% to 73% with an average GC content of 47% **(Fig. 5.1a)**. I designed fluorescein labeled oligos from snippets of the ϕPA3 genome with representative GC contents. These were then annealed to their unlabeled complementary strands and used for fluorescence polarization experiments.

From initial experiments, it appears that PhuN does have an affinity for DNA. Moreover, it has a higher affinity for the lower GC content oligo **(Fig. 5.1b)**. This would be an interesting mechanism for perhaps finding a place along the DNA to start packaging capsids, as these low GC content regions are sparsely distributed throughout the genome. However, from running equivalent samples in native gels, the fluorescent DNA band does not migrate at the expected molecular weight of PhuN **(Fig. 5.1c)** and the DNA band does not have a corresponding protein band in the native gel. This suggests that, despite additional purification steps, there may be a low concentration contaminant with a high DNA affinity present in the sample. These experiments could be cleaner and should be redone before any conclusions are drawn. My greatest concern is that the benzonase added during Ni·NTA purification, despite many washes, remains in the sample at low concentrations and binds the DNA in our experiments.

For future attempts, I suggest running additional purification steps like anion exchange followed by size exclusion chromatography. The benefit to using the anion exchange purification is that the sample will be further cleaned up and the experiment can be expanded to screening Class I, II, and III assemblies to assay the impact of assembly state on PhuN:DNA interactions. The caveat is that native gels have always been challenging due to the propensity for PhuN to

assemble and possibly degrade. The large assemblies can be particularly difficult to see as crisp bands since they vary in shape and size.

In addition to these binding assays, I collected a small dataset of *in vitro* assembled PhuN 2D crystals incubated with low GC content DNA. The results from the 0° tilt dataset are shown **(Fig. 5.2a)**. Processing these micrographs revealed a new 2D class **(Fig. 5.2b)**. The monomers within the tetramer 2D class appeared in an altered conformation, enlarging the Diamond channel, but no DNA was visible. As the classes never went to high resolution they could, perhaps, be a result of misalignments. If this is the case, this could nevertheless point to more heterogeneity in the sample after exposure to DNA, prompting further investigation. Since the oligos in this experiment are not visible in 2D classes, it's possible they have passed through one of the PhuN lattice channels or had other brief interactions with the arrays that altered PhuN conformation. The new PhuN 2D class could result from a mixed population of protein recovering from DNA exposure. To resolve the concern of short oligo-induced heterogeity, 2D crystals should be grown and incubated with gently sheared DNA isolated from phage particles. These longer strands should take longer to pass through or have longer interactions with the arrays, providing ample time to capture them.

Verifying and characterizing these PhuN:DNA interactions will help build an understanding of the mechanisms for sequestering DNA inside the phage nucleus and packaging it into nascent capsids. Given that DNA and the interior of the phage nucleus are both negatively charged, as discussed in Chapter 3, reconciling these charges and delineating any interactions between these two molecules is an important step towards fully understanding the phage nucleus.

Putative PhuN:ATP interaction

An ATP wash step was introduced to the Ni·NTA purification of PhuN, as included in the methods of Chapter 3, to eliminate any trace ATPase contaminants. Dr. Yanxin Liu, an Agard

Lab post-doc at the time, ran purified PhuN through a rough ATP-hydrolysis assay on my behalf and suggested that an efficient ATPase was present in the sample. After adding an ATP wash step, this ATPase activity disappeared. It was not surprising that an assembly-prone protein like PhuN would attract ATPases. This result was also consistent with the occasional observation of particles resembling GroEL in negative stain EM. The corresponding ATPase data has been lost, so anyone interested is encouraged to replicate it and perhaps explore different assays.

With the ATP wash added to the PhuN sample purification, the anion exchange trace was unexpectedly altered. The 260/280 ratio of the sample was elevated and the MonoQ 260 and 280 traces resembled a "dampened oscillation" curve **(Fig 5.3a)**. The oscillating peaks had 260/280 values greater than one, suggesting an enrichment of nucleotides. This experiment was replicated splitting an Ni·NTA purification in half, proceeding with a wash step in one half and without one in the other **(Fig. 5.3b)**. This experiment confirmed that the oscillating MonoQ trace is a result of the additional ATP wash step. Surprisingly, the corresponding SDS-PAGE gel showed a new protein species at approximately 70 kDa in the post-ATP wash sample **(Fig. 5.3c)**. However unlikely, it should be noted that this is roughly the same molecular weight of Gp53 without the MBP tag. There is no reason for MBP to have been cleaved off since Tev protease was not added, but an occasional enrichment of a species at this molecular weight after freeze/thaw or bench-top incubation of the 6XHisMBPGp53 sample has been observed in the past. An altogether different explanation for the new gel band is that providing ATP during the wash may encourage the binding of an *E. coli* chaperone, such as DnaK, to PhuN. Since ATP exposure results in a "dampened oscillation" pattern with extreme 260 signal enrichment at the corresponding 280 peaks in the MonoQ trace, this may point to a PhuN assembly state (tetramer, octamer, etc) specific interaction between PhuN and ATP, possibly via an ATP dependent chaperone. Given the molecular weight of the new gel band observed after ATP exposure, much like Hsp70 assists with clathrin disassembly,[5] one can imagine a mechanism where a chaperone like DnaK aids PhuN tetramers assembling into larger lattices. While highly

114

speculative, Hsp70 could hold or untangle the flexible PhuN C-terminal tails or Loops in cuboids or tetramers to allow for their addition into larger assemblies.

Between these two results, it seems that ATP may have some relationship with PhuN, whether directly or via another protein. No nucleotide binding sites were identified within the PhuN model nor was there a clear density for ATP in the density described in Chapter 3. The later could be resolution limited.

Replicating these experiments will be an important first step in describing the influence of nucleotides on PhuN activity and mechanism of assembly. Confirming the presence of and conducting mass spectrometry on the additional band observed post ATP wash and adding the resulting protein to *in vitro* assembled PhuN lattices for cryo-EM would be a great place to start follow-up experiments. It may also be worth exploring alternative mononucleotides, like GTP. A PhuN:mononucleotide or PhuN:ATP-dependant chaperone interaction would be an exciting discovery, expanding our understanding of what PhuN does and interacts with during infection.

Another protein marking the phage nucleus, Gp2

Based on the shell fragment isolations described in Chapter 3, the phage nuclear shell lattice appears to be composed entirely of PhuN. However, the Pogliano Lab used GFP labeling to show that bacteriophage protein Gp2, conserved across various nucleus-forming jumbo phages, forms a ring around the replicating phage DNA, much like PhuN **(Fig. 5.4a)**. This localization suggests that Gp2 may be a sparsely lattice integrated or a phage nucleus associated protein. ɸPA3 Gp2 is not identified as a phage nucleus-related protein via mass spectrometry of isolated phage nuclear shell fragments, suggestive of a more decorative, looser association between Gp2 and the phage nucleus.[6] Curiously, ɸKZ Gp2 did co-elute with ɸKZ PhuN during SEC-MS experiments, but no further experiments have been done with Gp2 from ɸKZ.

Gp2 is a 197 amino acid (22 kDa) protein. To better understand its activity and function, as a rotation student, Dr. Chari Noddings began an *in vitro* characterization of Gp2. Using the same protocol developed for PhuN, a largely clean Gp2 sample was obtained after Ni·NTA purification **(Fig. 5.4b)**. The 6XHis tag was cleaved off Gp2 using Tev protease after which Gp2 was further purified using Ni·NTA to remove uncleaved PhuN and Tev followed by size exclusion chromatography. The column purification yielded one large, sharp peak corresponding to a monomer of Gp2.

Next, crystal trays were set with the hope that Gp2, being a smaller protein, would readily crystallize. Trays set with protein from two different purifications of Gp2 showed a few conditions with tiny crystals. JCSG III E5 (0.16 M calcium acetate, 0.08 M sodium cacodylate pH 6.5, 14.4% PEG 8 K, 20% Glycerol) was the only condition to eventually grow a large crystal. Unfortunately, due to being a multiple crystal, it was not usable for structure determination **(Fig. 5.4c)**. This crystal growth was not successfully replicated.

After Chari's rotation, undergraduate Gayathri Sundaram helped purify Gp2 again, aiming to improve on previous crystal attempts. During the size exclusion chromatography step of this purification, Gp2 ran where a 66 kDa protein would be expected **(Fig. 5.4d,e)**. This suggests that we isolated a Gp2 trimer. The species size was consistent with a trimer in negative stain EM, as well **(Fig. 5.4e)**. It is not clear what was different in our purification of Gp2 from the original purifications carried out by Chari. We were not able to obtain crystals of this trimer, however, given recent advances in cryo-EM and protein prediction[7,8] **(Fig. 5.4f)**, these small assemblies could be investigated using single particle analysis. It is not known whether the monomer, trimer, another assembly state, or all possible assembly states are biologically relevant states of Gp2 *in vivo*.

While the function of Gp2 remains unknown, its localization to the phage nuclear shell is clear **(Fig. 5.4a)**. Since there are only two proteins observed to mark the phage nuclear shell, Chari and I briefly probed for PhuN:Gp2 interactions through mixing purified proteins and,

separately, co-expressing PhuN with Gp2. Mixing monomeric Gp2 with PhuN did not result in an obvious change in assembly at pH 7.65, as monitored by dynamic light scattering (DLS). Due to the heterogeneity of PhuN assemblies, DLS proved to be a poor method to reliably assay changes in PhuN assembly. Thus, it would be worth following up on these interactions using other techniques. Running size exclusion chromatography (SEC) or pairing SEC with multi-angle light scattering (MALS) to see if the proteins co-elute is a good next step. Another approach could be growing PhuN lattices, adding purified mono- or trimeric Gp2, collecting a small cryo-EM dataset, and using 2D classification to check for the presence of Gp2 in/on the lattices or changes in PhuN assembly.

In parallel to these mixing experiments, Chari co-expressed 6XHisMBPPhuN with unlabeled Gp2 in *E. coli*. While there were bands at molecular weights corresponding to both proteins in gels, it is unclear whether Gp2 was actually present in the Ni·NTA pulldowns attempted since the pulldowns were not clean. A better approach may be to capitalize on the MBP tag and pull-down using amylose resin, which gives fewer contaminants than nickel. Adding a GFP or other tag to track Gp2 via western blots will also be crucial to confirming a PhuN:Gp2 interaction. These experiments were all carried out prior to discovering that pH 6.5 is important for PhuN lattice assembly. Perhaps decreasing pH in these experiments will be illuminating, especially since the positive crystal screen hits for Gp2 were often around pH 6.5.

The function of Gp2 and how it interacts with the phage nuclear lattice are entirely unclear. While it is currently not known whether Gp2 integrates into the PhuN lattice or whether it decorates the exterior or interior of the phage nucleus, the absence of Gp2 in mass spectrometry analysis of isolated phage nuclear lattice fragments suggests Gp2 is likely loosely associated and washed away during fragment purification. Following up on these experiments will give insight into the role of this unknown protein during ϕPA3 infection and perhaps begin to shed light onto, as of yet, poorly understood aspects of the phage nucleus like the mechanism of subunit addition and selectivity. Since Gp2 localizes around the phage nucleus but is not

recovered in mass spectrometry, there is room for the discovery of other small and/or loosely associated proteins decorating or interacting with the lattice of the phage nucleus that simply have not yet been identified.
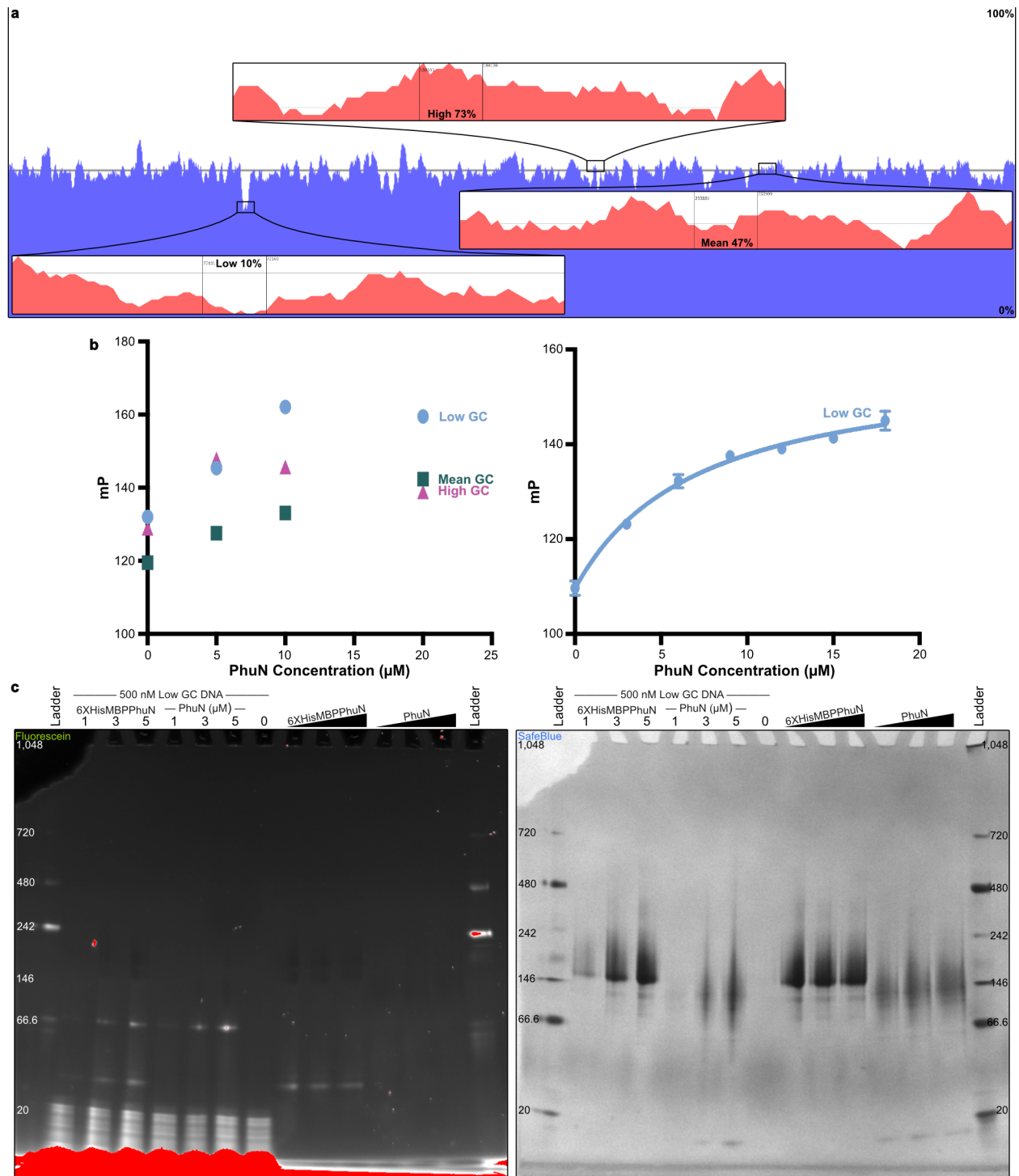
Figures

Figure 5.1 | Investigating PhuN:DNA Interactions.
a, Analysis of the φPA3 genome GC content. Purple is a 1,000 bp window GC% analysis, salmon is a 30 bp window analysis. **b,** Fluorescence polarization shows PhuN sample preferentially binds low GC content DNA (left) with a relatively weak binding affinity (right). **c,** Native gel of 6XHisMBPPhuN and Tev cleaved PhuN after incubation with low GC content DNA. DNA runs at the same molecular weight with and without Tev cleavage and not with PhuN.
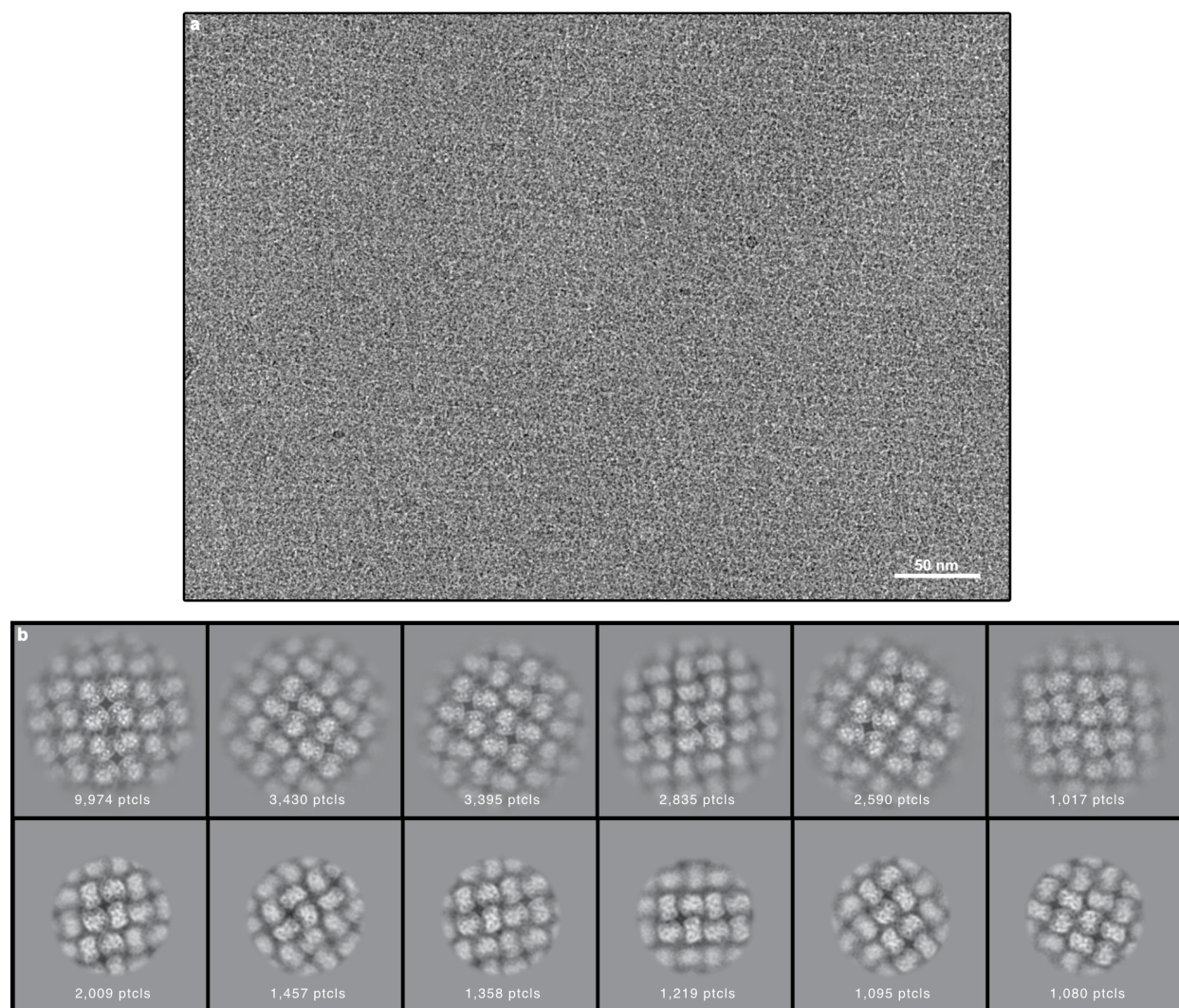
Figure 5.2 | Cryo-EM of PhuN after DNA exposure.
a, Example micrograph of PhuN lattice after exposure to low GC content DNA. **b,** 2D classes recovered from the dataset represented in panel a. The top row shows p2 and p4 PhuN 2D classes, as described in Chapter 3. The bottom row represents a new series of classes enriched in further processing after removing the classes shown in the top row.
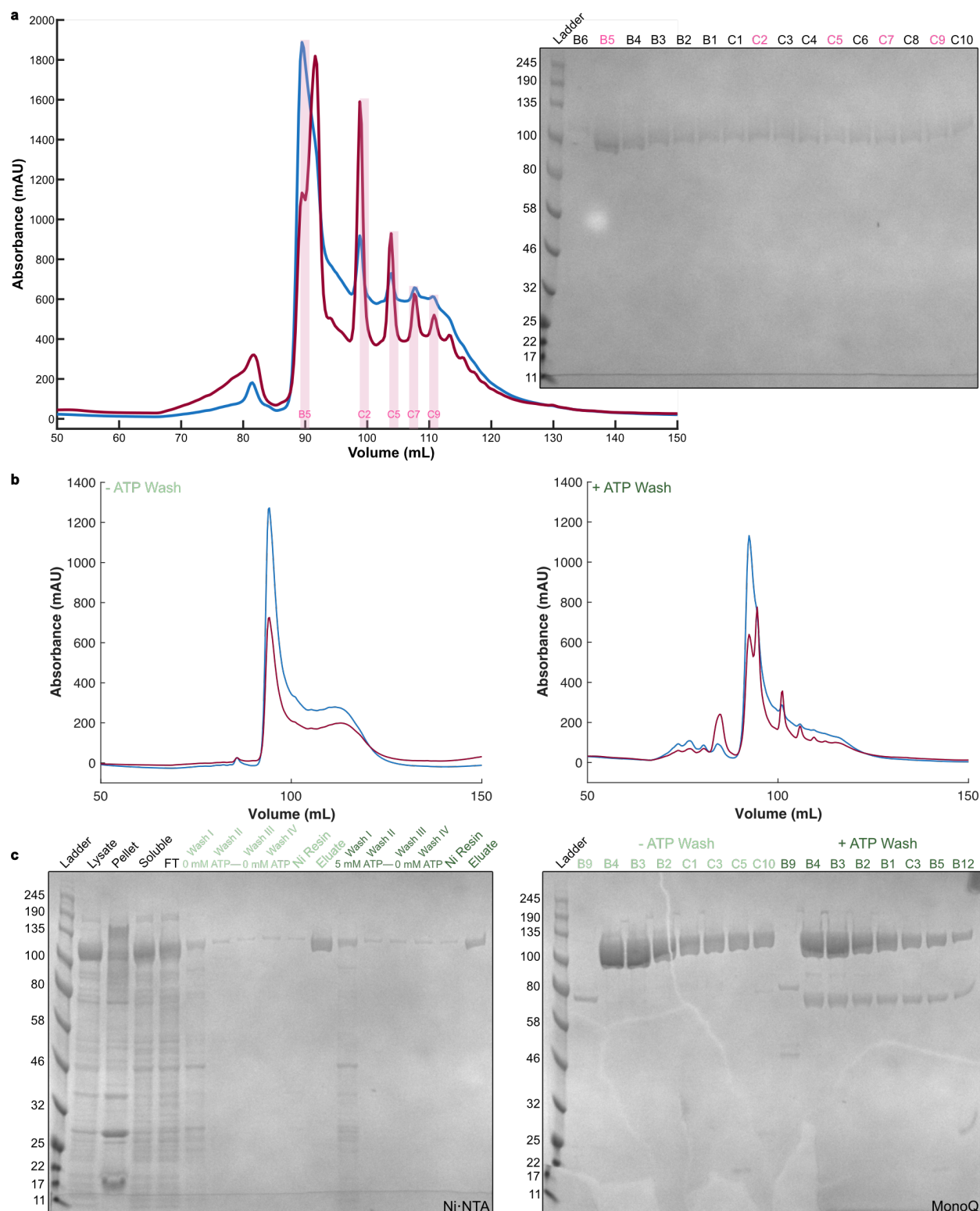
Figure 5.3 | Investigating PhuN:ATP Interactions.
a, MonoQ run and gel of 6XHisMBPPhuN purified on Ni·NTA with an ATP wash. **b,** MonoQ runs of split Ni·NTA purification into +/- ATP wash samples. **c,** Gels for the Ni·NTA and MonoQ purification steps from panel b. An unexpected 70 kDa species appears in the MonoQ run gel. In all MonoQ traces, 280 nm and 260 nm traces are shown in blue and red, respectively.
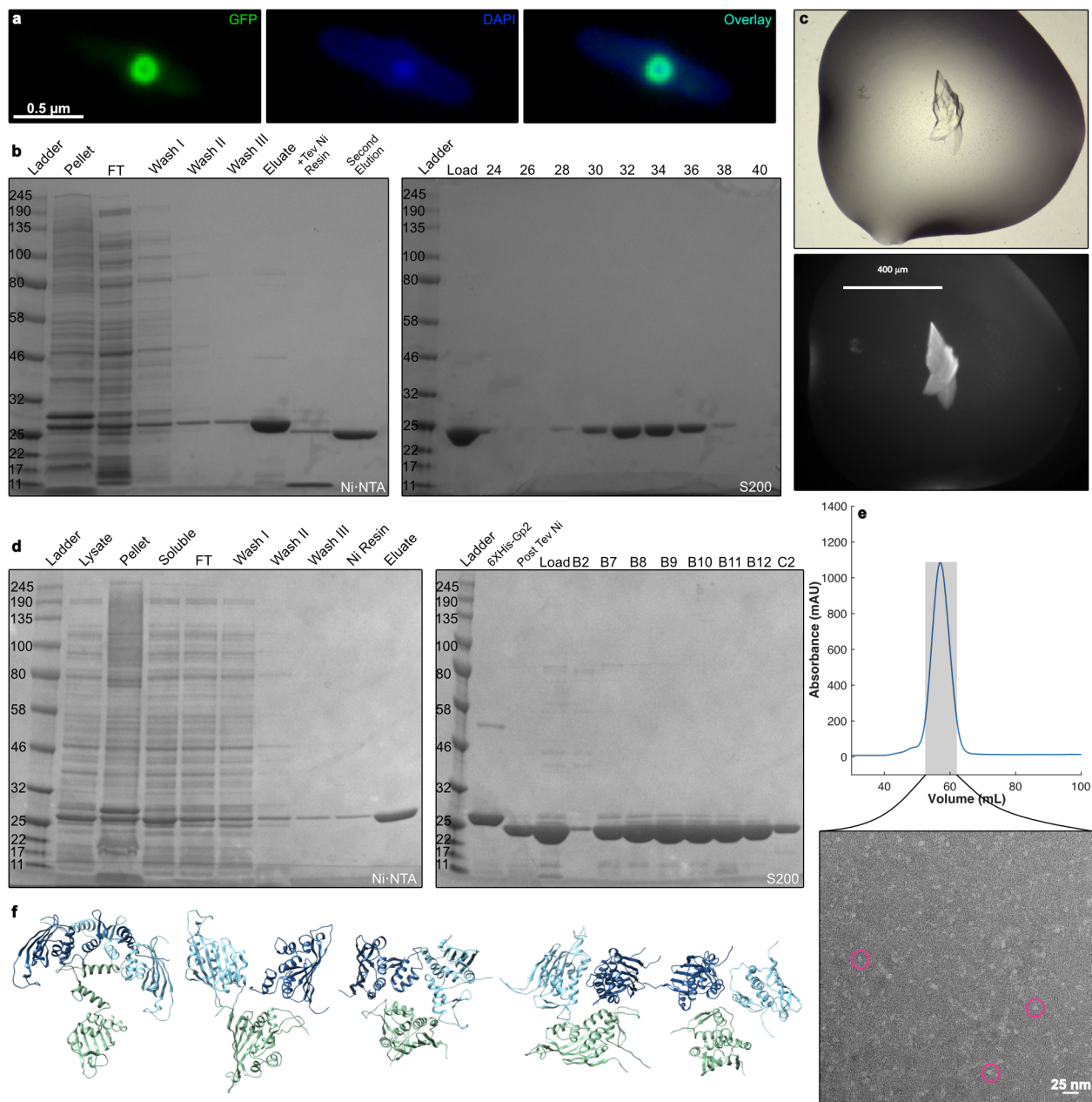
Figure 5.4 | Characterization of Gp2.
a, Fluorescence microscopy of Gp2-GFP localized around the phage nucleus 46 minutes post infection, as visualized with DNA DAPI staining. **b,** First high yield Gp2 purification (see Chari Lab Notebook #1). Ni·NTA purification was followed by Tev cleavage, Ni·NTA to remove Tev and uncleaved product, and finally the sample was concentrated and purified via size exclusion chromatography (S200 column). **c,** Promising Gp2 crystal obtained using protein from the panel b purification. **d,** Replication of the panel b purification. Sample is less clean. **e,** S75 trace corresponding to the gel shown in panel d with a negative stain EM image of the highlighted pooled fractions. Gp2 elutes at the expected molecular weight of a 66 kDa protein. Particles of a size consistent with a 66 kDa protein are visible in negative stain EM, some circled in pink. **f,** AlphaFold2 predictions of a Gp2 trimer.[7]

Methods

Fluorescence Polarization

Fluorescence polarization of fluorescein labeled DNA oligonucleotides with PhuN was measured using a SpectraMax M5 plate reader (Molecular Devices) using excitation and emission wavelengths of 490 and 510 nm, respectively. Experiments testing GC content preference were done with 100 µM of 30 bp oligonucleotides. The GC content of the oligos was 10%, 47%, and 73%, based on the phage genome. PhuN concentration was varied from 0-20 µM in 5 µM increments in Fig. 5.1b. 40 µM reactions were set up in Reaction Buffer (pH 7.42: 20 mM Tris, 250 mM NaCl, 1 mM DTT, cOmplete Protease Inhibitor Tablet) at 25°C with 0.02% Tween-20 and 0.20 mg/mL BSA. Data collection was done over 15 minutes. The finer sampled Low GC DNA:PhuN binding experiment was done using the above conditions with PhuN concentrations ranging from 1-21 uM in 3 uM increments.

Native Gel

PhuN samples with and without Tev were purified using size exclusion chromatography. These were then incubated with DNA (10% GC, forward sequence: AATGATCTAAAATTATTAATATCTATTTTT) and ran on a native gel.

Sample Preparation, Freezing, and Cryo-EM Data Collection of DNA Exposed PhuN Lattices

Samples were prepared exactly as described in Chapter 3. Immediately after protein addition, annealed 10% GC oligos were added to the 2D crystallization well. The sample was incubated at room temperature for 1 hour. Next, samples were adsorbed by touching the QUANTIFOIL grid surface to the lipid monolayer and plunge-frozen, as before, using a Vitrobot Mark IV (FEI) at 10°C with 100% humidity using a 6 second blot time, 0 second wait time, and blot force of 3. Image collection was done using an FEI Titan Krios electron microscope (Thermo Fisher Scientific) at 300 keV equipped with a K3 direct electron detector (Gatan) and Bioquantum

energy filter (Gatan) set to a slit width of 20 eV. Images were recorded using a nominal magnification of 105,000x (0.835 Å physical pixel) with a total dose of 67 e-/Å^2. Data were collected at 0°, 40°, and 50°. The 0° data were most informative and are shown in Fig. 5.2.

Data Processing

94/393 micrographs collected contained lattices. These were processed exactly as described in Chapter 3, using the same 0° particle picking templates. 63,738 particles were extracted and used in 2D classification. After the first round of 2D classification, only the new 2D classes were selected for another round of classification. Representative images of these and their particle counts are shown in Fig. 5.2b.

PhuN ATP MonoQ

bESN9 was grown and 6XHisMBPPhuN was purified using Ni·NTA with an ATP wash step, exactly as described in Chapter 3. The sample was then applied to a GE 10/100 MonoQ GL column using MonoQ Binding Buffer (pH 7.65: 20 mM Tris, 50 mM NaCl, 1 mM DTT, 0.5 cOmplete Protease Inhibitor Tablet) and MonoQ Elution Buffer (pH 7.65: 20 mM Tris, 1 M NaCl, 1 mM DTT, 0.5 cOmplete Protease Inhibitor Tablet). 6XHisMBPPhuN from 1 L of bESN9 was used for panel 5.3a while an equivalent liter was split in half for panels b and c, each experiment getting protein from 500 mL of bacterial growth.

Live Fluorescence Microscopy of Gp2-GFP

Sample preparation and microscopy were done as described in Chapter 3. In Fig. 5.3a, 0.1% arabinose was used in the agar pads to induce Gp2-GFP expression and the image was collected at 46 minutes post infection.

Gp2 Growth

6XHisGp2 was expressed in Rosetta(DE3)pLysS *E. coli* cells. 10 mL of an overnight in LB was grown with carbenicillin and chloramphenicol. 1 L of Terrific Broth (TB) growth media was inoculated with 20 mL of the overnight. Cells were grown with 250 rpm nutation at 37°C until an OD_{600} of 0.5. Cells and the incubator were then cooled to 16°C. Expression was induced with 1 mM IPTG and proceeded overnight (~ 16 hrs) at 16°C with 250 rpm nutation. Cells were spun down the following morning at 4,000 x g for 15 minutes and pellets were transferred to 50 mL conical tubes. These were flash frozen and stored at -80°C.

Gp2 Purification

A cell pellet from 1 L of culture was resuspended in 6 mL of Lysis Buffer (pH 7.42: 20 mM HEPES, 500 mM NaCl, 0.2% TritonX100, 1 mM DTT, 5% Glycerol, 5 mM MgCl2) per each gram of pellet with 1 cOmplete Protease Inhibitor Tablet and benzonase. This mixture was homogenized using a manual glass homogenizer and lysed using an Avestin EmulsiflexC3. The sample was clarified with a 15 minute spin at 12,000 x g. The soluble fraction was nutated with equilibrated Ni·NTA resin for 1-2 hours in the cold room. The sample was transferred to a gravity column and washed 3X with Wash Buffer (pH 7.42: 20 mM HEPES, 500 mM NaCl, 20 mM Imidazole, 1 mM DTT, 5% Glycerol). 6XHisGp2 was eluted and collected in the Elution Buffer (pH 7.42: 20 mM HEPES, 150 mM NaCl, 600 mM Imidazole, 1 mM DTT, 5% Glycerol). The eluate was cleaved with Tev protease at 4°C overnight in 10,000 MW dialysis tubing while dialyzing into Size Exclusion Buffer (pH 7.42: 20 mM Tris, 150 mM NaCl, 1 mM DTT, cOmplete protease inhibitor tab). After cleavage, the sample was incubated with Ni·NTA for 30 minutes to remove uncleaved product and Tev. The cleaved Gp2 was concentrated to ~ 5 mL and purified via S200 16/60 (monomer run, Fig. 5.4b) or S75 16/60 (trimer run, Fig. 5.4d-e). Approximately 25 mg of Gp2 were recovered from the purification shown in Fig. 5.4b. The sample was

concentrated to 2.2 mL at 7 mg/mL and was frozen with 10% glycerol, resulting in a final

concentration of 5.6 mg/mL Gp2. See Dr. Chari Nodding's "Notebook #1" for additional details.

Negative Stain EM

The pooled fractions from the S75 trace in Fig. 5.4e were prepared for negative stain EM

exactly as described in Chapter 3.

Crystal Screens

Crystallization conditions were screened using the JCSG I, JCSG II, JCSG III, JCSG IV, and

Classic L plates using hanging drops of Gp2 at a concentration of 5.3 mg/mL. Small crystals

formed in JCSG I F5 (0.2 M magnesium chloride, 20% PEG 3350) and JCSG IV E9 (0.2 M

Calcium Acetate, 0.1 M sodium cacodylate pH 6.5, 40% PEG 600), while the large crystal in Fig.

5.4c formed in JCSG III E5 (0.16 M calcium acetate, 0.08 M sodium cacodylate pH 6.5, 14.4%

PEG 8 K, 20% Glycerol).

Preparation of Figures

All figures were prepared in IMOD,[9] Chimera,[10] MatLab (MATLAB. (2021,2022). *version 9.11.0

and 9.12.0 (R2021b, R2022a).* Natick, Massachusetts: The MathWorks Inc), and assembled in

AffinityDesigner, (v2.0.3) as in Chapter 3.

References

1. Chaikeeratisak, V. *et al.* Assembly of a nucleus-like structure during viral replication in bacteria. *Science* 355, 194–197 (2017).

2. Mendoza, S. D. *et al.* A bacteriophage nucleus-like compartment shields DNA from CRISPR nucleases. *Nature* 1–5 (2019) doi:10.1038/s41586-019-1786-y.

3. Laughlin, T. G. *et al.* Architecture and self-assembly of the jumbo bacteriophage nuclear shell. *Nature* 608, 429–435 (2022).

4. Chaikeeratisak, V. *et al.* Viral Capsid Trafficking along Treadmilling Tubulin Filaments in Bacteria. *Cell* 177, 1771-1780.e12 (2019).

5. Sousa, R. & Lafer, E. M. The role of molecular chaperones in clathrin mediated vesicular trafficking. *Frontiers Mol Biosci* 2, 26 (2015).

6. Fossati, A. *et al.* Next-generation interaction proteomics for quantitative Jumbophage-bacteria interaction mapping. *Biorxiv* 2023.01.13.523954 (2023) doi:10.1101/2023.01.13.523954.

7. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589 (2021).

8. Baek, M. *et al.* Accurate prediction of protein structures and interactions using a three-track neural network. *Science* 373, 871–876 (2021).

9. Kremer, J. R., Mastronarde, D. N. & McIntosh, J. R. Computer visualization of three-dimensional image data using IMOD. *J Struct Biol* **116**, 71–6 (1996).

10. Pettersen, E. F. *et al.* UCSF Chimera—A visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605–1612 (2004).

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:

_____   3/13/2023
6F2505033988456...   Author Signature   Date