HARVARD
Kenneth C. Griffin

GRADUATE SCHOOL
OF ARTS AND SCIENCES

DISSERTATION ACCEPTANCE CERTIFICATE

The undersigned, appointed by the

Division of Medical Sciences

Committee on Virology

have examined a dissertation entitled

Supramolecular complex formation in bacterial anti-phage defense and viral immune evasion

presented by Sadie Priscilla Antine

candidate for the degree of Doctor of Philosophy and hereby certify that it is worthy of acceptance.

Signature: Jonathan Abraham
Jonathan Abraham (Apr 3, 2025 18:11 EDT)

Typed Name: Dr. Jonathan Abraham

Signature: Shira Weingarten-Gabbay
Shira Weingarten-Gabbay (Apr 4, 2025 01:17 GMT+3)

Typed Name: Dr. Shira Weingarten-Gabbay

Signature: Lee Gehrke
Lee Gehrke (Apr 3, 2025 14:05 EDT)

Typed Name: Dr. Lee Gehrke

Signature: Michael Laub
Michael Laub (Apr 3, 2025 13:18 EDT)

Typed Name: Dr. Michael Laub

Date: April 03, 2025

Supramolecular complex formation in bacterial anti-phage defense and viral immune evasion

A dissertation presented

by

Sadie Priscilla Antine

to

The Division of Medical Sciences

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Virology

Harvard University

Cambridge, Massachusetts

April 2025

© 2025 Sadie Priscilla Antine

All rights reserved.

Dissertation Advisor: Philip J. Kranzusch Sadie Priscilla Antine

Supramolecular complex formation in bacterial anti-phage defense and viral immune evasion

Abstract

Bacteria can encode many diverse nuclease-helicase defense systems that protect from viral infection and inhibit phage propagation. An emerging theme in anti-phage defense is the presence of nuclease-helicase operons. However, how nuclease-helicase systems defend against phage infection remains largely unknown. Gabija is one of the most prevalent nuclease-helicase defense systems, occurring in >15% of all sequenced bacterial and archaeal genomes. We discovered that Gabija proteins assemble into an ~500 kDa nuclease-helicase supramolecular complex that degrades phage DNA. Phages evolve diverse immune evasion mechanisms to inhibit host defense systems. We show that a phage-encoded protein, Gabija anti-defense 1 (Gad1), directly binds the Gabija complex by forming an octameric web inhibiting phage DNA recognition and cleavage. Phages can also encode another Gabija inhibitor, Gad2, which does not prevent Gabija DNA targeting *in vitro,* suggesting that Gad2 works upstream of DNA cleavage during infection. To understand nuclease-helicase system diversity, we biochemically screened 14 nuclease-helicase systems and found that they all form protein complexes of varying oligomeric state and have different phage nucleic acid cleavage specificity. Our results define mechanisms by which bacteria use nuclease-helicase anti-phage defense systems to inhibit viral infection and how phages fight back with unique viral immune evasion mechanisms.

Table of Contents

Title i

Copyright ii

Abstract iii

Acknowledgements v

Dedication vi

Figures and Tables vii

Introduction 1

Chapter 1: Structural basis of Gabija anti-phage defence and viral immune evasion 15

Abstract 16

Results and Discussion 16

Methods 41

Acknowledgements 47

Author Contributions 48

Chapter 2: Phages overcome bacterial immunity via diverse anti-defense proteins 49

Abstract 50

Introduction 51

Results and Discussion 52

Methods 77

Acknowledgements 92

Author Contributions 93

Chapter 3: Comparative analyses define distinct classes of nuclease-helicase systems in bacterial anti-phage defense 94

Abstract 95

Introduction 96

Results 97

Discussion 112

Methods 115

Acknowledgements 119

Author Contributions 120

Discussion 121

References 129

Acknowledgements

I would not be here without the community of people who have supported me throughout my Ph.D. journey. I am eternally grateful to every one of you, thank you.

To my mentors who launched my scientific career, Kene Piasta and Amy S.Y. Lee, I am forever indebted to your unwavering mentorship.

To my dissertation advisory committee, Jonathan Abraham, Anne Hochschild, and Sophie Helaine, for your guidance and support throughout my Ph.D.

To my friends and the best collaborators in the Sorek lab, for sharing your scientific brilliance and bridging the worlds of phage biology and biochemistry.

To Philip Kranzusch, for mentoring me, allowing me to do science in your incredible lab, and letting me learn and grow into the person and scientist I am today.

To the Kranzusch Lab, for making the lab a fun place to do science and letting loose on the dance floor at all the retreats and conferences.

To Ady Ragucci, for continuously teaching me how to be a better scientist and human. Going to lab every day, knowing my best friend will be there, is something I will cherish forever.

To Sarah Mooney, my Gabija girly, for all the morning tea talks, biochemistry, and best times.

To Emily Cohen, for being a strong female scientist with the best advice and guidance.

To Tatum Sass and my Virology cohort, for the endless memories and being the best people to go through grad school with.

To Abbey Logan, for being my best friend on the tennis court at age seven to now and for all the life we've laughed through together.

To Tamara Botteri, Sabrina Chow, Mahima Devanahalli, Aishwarya Khanna, Akhila Penumarthy, and the rest of the Brandeis crew, for all the post-grad Boston adventures.

To Víctor Manuel Suárez Casanova, Linda Calleva, and my tennis family, for the mental stability achieved playing matches throughout the years.

To my sister Lucy, and niece and nephew, Harper and Harry, for making Titi remember what life is all about.

To my dad, Peter, for always being there when I need you.

To my stepdad, Mike, for all the love, support, advice, and Moishe, you don't know love until it's puppy love.

To my mom, Joyce, for being my best friend, always answering the phone and showing me what it means to be an independent woman. I would not be here today without you.

To Joyce Grippen, for everything mom

Figures and Tables

Figure I.1 | Human and bacterial cells use conserved protein domains to sense viral infection. 2
Figure I.2 | Restriction Modification and CRISPR-Cas systems target phage nucleic acid to halt phage replication. 4
Figure I.3 | Gabija anti-phage defense. 6
Figure I.4 | OLD nucleases in anti-phage defense. 8
Figure I.5 | Diverse anti-phage defense systems use supramolecular complex formation for defense activity. 10
Figure I.6 | Phages use diverse anti-defense strategies to bind and subvert bacterial anti-phage defense. 13

Figure 1.1 | Structure of the Gabija anti-phage defense complex. 18
Figure 1.2 | GajA and GajB form a supramolecular complex that cleaves phage lambda DNA *in vitro*. 19
Figure 1.3 | Structural characterization of GajA. 21
Figure 1.4 | Structural characterization of GajB. 23
Figure 1.5 | Mechanism of Gabija supramolecular complex assembly. 25
Figure 1.6 | Size comparison of Gad1 to known phage immune evasion proteins and biochemical characterization of Gad1 for binding to the GajAB complex. 28
Figure 1.7 | Structural basis of viral evasion of Gabija defense. 30
Figure 1.8 | Cryo-EM data processing GajAB–Gad1 co-complex. 31
Figure 1.9 | GajAB–Gad1 co-complex Cryo-EM map quality and model to map fitting. 32
Figure 1.10| Biochemical and structural characterization of the GajAB–Gad1 co-complex. 35
Figure 1.11 | Modeling DNA-bound GajA. 38
Figure 1.12 | Inhibition of Gabija DNA binding and cleavage enables viral evasion. 39

Table 1.1 I Summary of X-ray data collection, phasing and refinement statistics 20
Table 1.2 | Cryo-EM data collection, refinement and validation statistics 32

Figure 2.1 | Identification of anti-defense genes based on differential sensitivity to defense systems. 53
Figure 2.2 | Genome comparisons of phages from the SPbeta and SBSphiJ groups. 54
Figure 2.3 | Phages from the same family are differentially sensitive to bacterial defense systems. 56
Figure 2.4 | Gad1 and Gad2 proteins inhibit Gabija defense. 57
Figure 2.5 | Gad1 proteins inhibit Gabija mediated defense. 59
Figure 2.6 | Gad2 inhibits Gabija mediated defense. 61
Figure 2.7 | Tad2 proteins inhibit Thoeris defense. 63
Figure 2.8 | Tad2 proteins inhibit Thoeris mediated defense. 64
Figure 2.9 | Tad2 cancels Thoeris-mediated defense by sequestering 1″–3′ gcADPR. 65
Figure 2.10 | Tad2 binds 1″–3′ gcADPR. 67
Figure 2.11 | Size-exclusion chromatography of Tad2 and various standards. 70
Figure 2.12 | Comparison of Tad2 and Tad1 in the apo and ligand-bound states. 71
Figure 2.13 | Had1 proteins inhibit Hachiman defense. 73
Figure 2.14 | Had1 proteins inhibit Hachiman-mediated defense. 75

Table 2.1: Summary of crystallography data collection, phasing, and refinement statistics. 69

Figure 3.1 | Diverse nuclease-helicase systems form protein complexes. 97
Figure 3.2 | Predicted domain architecture of nuclease-helicase systems. 98
Figure 3.3 | Nuclease-helicase systems form oligomeric complexes of varying stoichiometries. 99
Figure 3.4 | *Ec*PD-T4-4, *Eh*Toprim TA, *Ss*Menshen form complexes of unclear stoichiometry. 101
Figure 3.5 | Nuclease-helicase systems have different phage DNA cleave specificity. 102
Figure 3.6 | Nuclease-helicase systems phage DNA cleavage agarose gel analysis. 105
Figure 3.7 | Nuclease-helicase catalytic active site requirements for phage DNA cleavage. 107
Figure 3.8 | Nuclease-helicase complex formation requirements for phage DNA cleavage. 109
Figure 3.9 | AbpAB, Azaca, and PD-T4-1 cleave DNA at low concentrations. 110

Figure 3.10 | Hachiman, Nhi-like, Olokun, and PD-T4-8 cleave DNA at high concentrations..... 111
Figure 3.11 | A compendium of nuclease-helicase biochemical activity. 113

Introduction

Evolution of Antiviral Responses

To sustain life, all organisms need to protect themselves against viral infection. Human cells mount an antiviral response using innate immune pathways that detect pathogen-associated molecular patterns (PAMPs) through pattern recognition receptors (PRRs) (Figure I.1). For example, human cells detect influenza A viral RNA through the cytosolic PRR, RIG-I leading to the upregulation of interferon-stimulated genes (ISGs)[1]. Cells recognize double-stranded DNA (dsDNA) viruses through PRRs including cGAS, which binds to aberrant dsDNA localized in the cytosol, leading to cGAS production of the nucleotide second messenger, 2′3′-cGAMP, which goes on to bind STING resulting in upregulation of NF-kB responsive genes and an inflammatory antiviral response[2]. Nucleotide-binding oligomerization domain (NOD)-like receptors (NLRs) can sense viral nucleic acid, including ssRNA from influenza, parainfluenza, and RSV to execute an antiviral response[3]. TIR proteins serve as signaling proteins that oligomerize upon PAMP sensing by Toll-like receptors (TLRs)[4]. All forms of innate immunity then inform the adaptive immune system to mount a humoral response where dedicated B and T cells directly bind viral proteins or peptides to prevent viral entry and kill infected cells[5]. The evolution of intricate innate immune systems was thought to be a mammalian trait, but recent work has now shown that bacteria encode innate immune proteins with similar functions to protect against viral infection by bacteriophages, or phages for short[6,7] (Figure I.1).

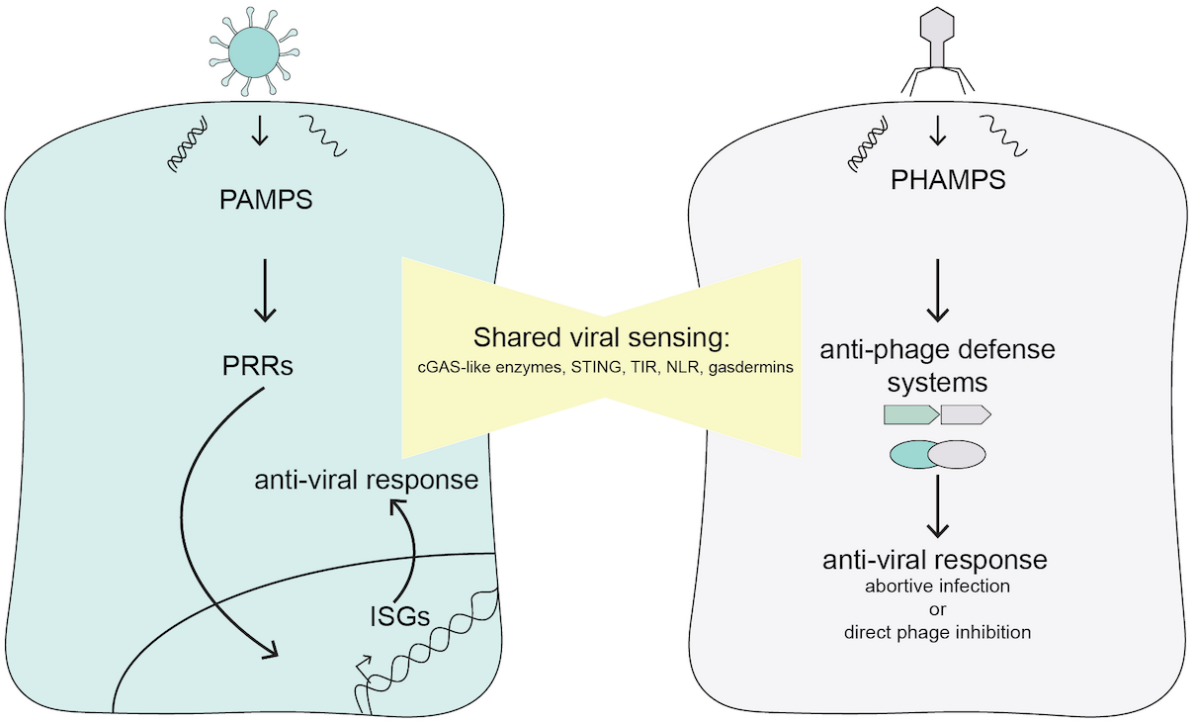


Figure I.1 | Human and bacterial cells use conserved protein domains to sense viral infection.

Left, simplified view of viral sensing by human cells through pathogen associated molecular patterns (PAMPs) by pattern recognition receptors (PRRs) that go on to activate an interferon response through the upregulation of interferon stimulated genes (ISGs). Right, simplified view of how bacteria sense phage infection by phage associated molecular patterns (PHAMPs), which include phage proteins, nucleic acid, and phage protein activity. These PHAMPs are sensed by anti-phage defense systems that mount an anti-viral response through abortive infection or direct phage inhibition.

Bacteria encode cGAS, STING, NLR, gasdermin, and TIR-like proteins that all function in mounting an immune response against phage infection, known as anti-phage defense[6]. Cyclic oligonucleotide-based anti-phage signaling systems (CBASS) use cGAS-like enzymes known as CD-NTases to sense phage infection through the detection of viral RNA and other unknown PAMPS[8–10], otherwise known as phage-associate molecular patterns (PHAMPS)[11]. CD-NTases then go on to produce nucleotide second messengers which can range from cyclic di- to tri-nucleotide products[8]. These nucleotide second messengers bind and active effector proteins that execute cell death to save the rest of the bacterial colony[12], which is known as abortive infection. CBASS effectors include proteins that act as membrane disrupters[13], nucleases[14], and TIR proteins that degrade NAD[15]. Bacterial NACHT-domain (bNACHT) containing anti-phage defense systems are sequence and structurally related to human NLR proteins and bind to diverse phage protein triggers to activate N-terminal effector domains to halt phage replication[16,17].

TIR proteins can also be found in Thoeris anti-phage defense systems where they sense viral infection and produce a cyclic ADP-ribose molecule, 1″–3′ gcADPR[18], which then bind effector TIR proteins to degrade all nicotinamide adenine dinucleotide (NAD) in the cell[19]. The origins of innate immunity can be understood by studying anti-phage defense.

Bacterial Anti-Phage Defense

CBASS, bNACHT, and Thoeris are only three examples of the >150 anti-phage defense systems that bacteria can encode[20]. New systems were recently identified because anti-phage defense systems tend to cluster in bacterial genomes, and can be identified by "guilt by association"[21]. 10 of these putative defense systems were determined to inhibit phage infection when expressed in model bacteria challenged with phage[22]. Following this study, hundreds of defense systems have been discovered using guilt by association[23,24], but also, metagenomic mining methods[25], and machine learning approaches[26,27]. On average, a bacteria encodes seven anti-phage defense systems[20]. Defense systems are normally encoded on mobile genetic elements, including integrated prophages, plasmids, and integrons. Therefore, the number and type of anti-phage defense systems can change rapidly.

Restriction-Modification Systems and CRISPR Systems

Restriction-modification (RM) and Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) systems are two of the most common anti-phage systems encoded by bacteria, with 79% and 37% of sequenced bacterial genomes encoding the systems, respectively[20]. RM systems were first discovered in the 1950s when classic experiments showed that bacteria have different susceptibilities to phage infection due to a restriction protein that only cleaves phage DNA and not host DNA[28,29]. Later work found that host-encoded methyltransferases modify the host genome to prevent targeting by the restriction enzyme[30]. There are now four types of restriction-modification systems, I–IV[31]. Type I systems form protein complexes composed of two restriction (REase) enzymes, two methylation (MTase) enzymes, and one specificity (S) protein. The MTase protein produces N6-methyladenine (m6A) on the target DNA sequences and the REase cleaves unmethylated DNA at varying distances away from the target sequence. Type II systems function as separate complexes, with the REase functioning as a monomer, dimer, or tetramer to

cleave DNA at a defined position within or close to the target sequence (Figure I.2). The MTase functions as a monomer to form N4-methylcyotosine (m4C), 5-methylcytosine (m5C), or m6A modifications on host DNA. There are 11 subtypes of Type II REases that all cleave target sequences in different ways. Type III systems encode two proteins, modification (Mod) and restriction (Res). Mod functions to hemimethylate DNA with m6A modifications and Res must bind two copies of a non-palindromic DNA target in the inverse orientation for DNA cleavage. Type IV systems encode one or two proteins and cleave only modified DNA substrates[31]. By methylating bacterial DNA, restriction-modification systems recognize phage DNA as foreign and employ restriction enzymes to cleave phage DNA halting replication.

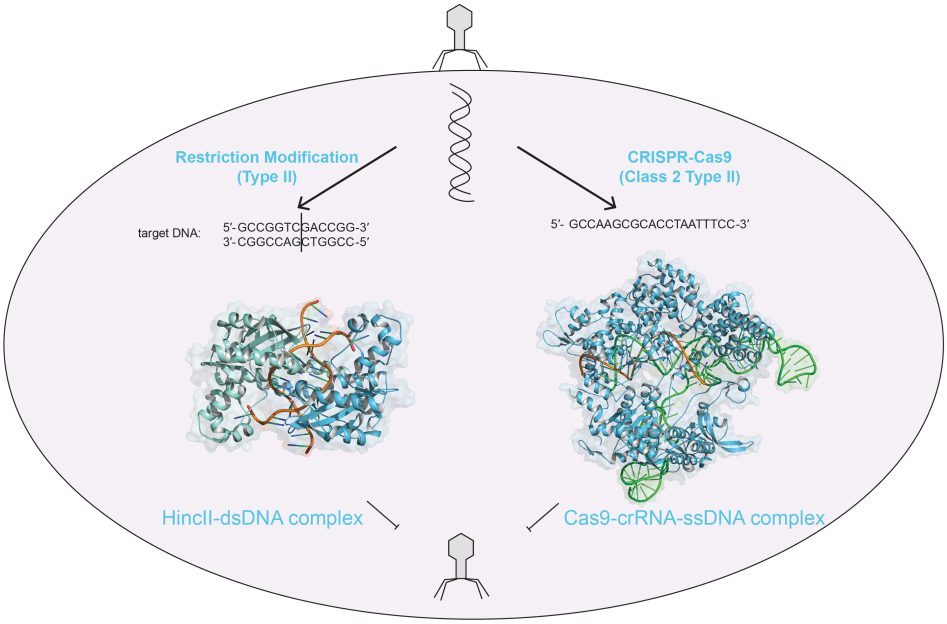


Figure I.2 | Restriction-Modification and CRISPR-Cas systems target phage nucleic acid to halt phage replication. Left, example of a Type II restriction modifications system HincII target DNA cleavage sequence and *Haemophilus influenzae* HincII dimer (blue) bound to target dsDNA (orange) (Protein Data Bank ID: 1TW8). Right, example of a Class 2 Type II CRISR-Cas9 system target ssDNA sequence and *Streptococcus pyogenes* Cas9 (blue) bound to target crRNA (green) and ssDNA (orange) target that is from an unwound dsDNA (PDB ID: 4OO8).

CRISPR and CRISPR-associated proteins (CRISPR-Cas) systems recognize foreign DNA through three stages of immunity: adaptation, CRISPR RNA (cRNA) biogenesis, and interference[32]. There are 2 classes, 6 types and 33 subtypes of CRISPR-Cas systems. All systems utilize Cas1-Cas2 proteins to select foreign DNA, integrate the DNA into CRISPR arrays, and finally process it into crRNAs[33]. CRISPR-Cas systems are then divided into 2 classes depending on how they perform the interference stage of immunity[32]. Class 1 systems utilize multi-protein effector complexes known as CRISPR-associated complex for antiviral defense (Cascade). For Class 1 Type I immunity, Cascade binds crRNAs and targets foreign dsDNA complementary to the cRNA proximal to the protospacer adjacent motif (PAM) site. Cas3 then degrades the non-target strand DNA. For Class 1 Type II immunity, interference occurs on actively transcribing foreign DNA, which leads to nonspecific RNA cleavage[32]. Class 1 Type IV CRISPR systems are highly diverse and likely target plasmids and work together with Type I systems for adaptation[34]. Class 2 CRISPR systems utilize single effector proteins for interference. Class 2 Type II systems utilize Cas9 effector proteins to bind both trans-activating crRNA (tracrRNA) and crRNAs to target dsDNA (Figure I.2). Cas9 then forms an R-loop with the target dsDNA activating the HNH and RuvC domains to cleave both the target and non-target strands[32]. Class 2 Type V utilizes Cas12, while Class 2 Type VI utilizes Cas13 and only requires crRNAs to target foreign DNA[33].

Gabija anti-phage defense

RM systems and CRISPR-Cas systems both have defined rules for targeting foreign nucleic acid. RM systems search for unmethylated target sequences, while CRISPR-Cas systems are guided to foreign DNA by crRNAs. The requirements for DNA cleavage are encoded in the bacterial operon for each system by the methyltransferase and CRISPR arrays. However, this is not the outright case for a newly identified system, Gabija[22]. Gabija is named after the Lithuanian goddess of fire and protects *B. subtilis* cells from *Bacillus* phage infection, but the molecular mechanism for how this occurs was unknown. The Gabija operon contains two genes *GajA*, which was annotated as an ATPase, and *GajB*, a helicase[22] (Figure I.3a). Therefore, unlike RM systems where protein function can now be inferred from gene annotation alone, how Gabija functions to distinguish phage infection from host is unclear. *Bacillus cereus* VD045 GajA was further

shown to have an N-terminal ABC ATPase domain and a C-terminal Topoisomerase-primase (Toprim) domain that acts as a nicking endonuclease on a sequence-specific dsDNA derived from phage lambda DNA[35]. *Bc*GajA can nick DNA C/G ↑ G/AG motifs flanked by A/T bases, which are likely prevalent in both host and phage-derived DNA. *Bc*GajA cleaves this sequence-specific dsDNA *in vitro* at pH 9 and is inhibited by 1 mM ATP. Further work proposed a mechanism where GajA and GajB form a complex and GajA endonuclease activity leads to GajB NTPase activity that depletes NTP levels during infection to halt phage infection[36]. However, *Bacillus cereus* Gabija was tested in *E. coli* against *E. coli* phages, which may not be physiologically relevant for phage-host interactions. Structural analysis found that *Bc*Gabija forms a 4 GajA : 4 GajB complex that is required for anti-phage defense[37–40] (Figure I.3b). The *Bc*Gabija GajA tetramer binds the sequence-specific DNA derived from phage lambda DNA at a 110° angle mirroring how topoisomerases bind DNA[39]. A cryo-EM structure of the GajA tetramer bound to ATP in a hinge region close to the Toprim active site may explain how dsDNA cleavage activity is inhibited by 1 mM ATP[39], but more points to a mechanism where a specific DNA is recognized that leads to ATP hydrolysis and movement in the complex to license DNA cleavage. Overall, how Gabija functions to exert anti-phage defense remains unclear.

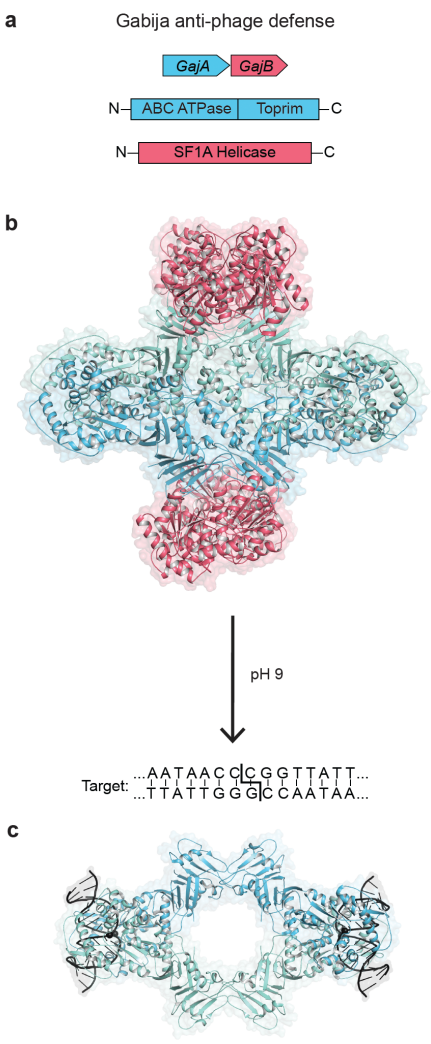


Figure I.3 | Gabija anti-phage defense. a, Gabija operon architecture and protein domains below. GajA is colored in blue and GajB in pink. (Antine et al. 2024) **b,** *Bacillus cereus* VD045 Gabija GajAB complex. The GajA tetramer is highlighted in two shades of blue and GajB helicase dimers on either side are in pink (PDB ID: 8SM3) (Antine et al. 2024). Below, GajA target cleavage sequence derived from phage DNA (Cheng et al. 2021). c, BcGajA tetramer bound to calcium and 21 bp GajA target dsDNA sequence (PDB ID: 8WY5) (Li et al. 2024). DNA is bent at 110° in the Toprim active site.

ABC ATPase and Toprim domains in anti-phage defense

ABC ATPase domain and Toprim domain architecture of GajA mirrors another type of protein involved in phage-host interactions (Figure I.4). Overcoming lysogeny

defect (OLD) proteins were first discovered in the *E. coli* phage P2. Phage P2 is a lysogenic phage, which usually integrates into the *E. coli* genome and becomes lytic upon DNA damage[41]. *E. coli* cells with mutations in *recBCD* do not become P2 lysogens when infected, and instead, cell death occurs[42,43]. RecBCD is an exonuclease complex that distinguishes between foreign DNA and host DNA by the presence of a Chi sequence in dsDNA breaks[44]. Once a Chi sequence is encountered, RecBCD initiates double-stranded DNA break repair. Phage and mobile genetic element DNA that do not contain Chi sequences are completely degraded by RecBCD[45]. Phage P2 mutants in the *Old* gene allow phages to lysogenize *recBCD* mutant *E.coli* indicating that OLD is activated by DNA damage that RecBCD usually recognizes and degrades[46] (Figure I.4c). OLD proteins were later characterized structurally and contain an N-terminal ABC ATPase domain interrupted by a dimerization domain, followed by a C-terminal Toprim domain. OLD proteins exist as dimers or tetramers in solution and act as metal-dependent DNA exo- and endonucleases[47,48]. Class 1 OLD proteins exist as single genes in bacterial genomes[48], while Class 2 OLD proteins are encoded in operons next to a UvrD/PcrA/Rep-like helicase[47], which are Gabija anti-phage defense systems. OLD proteins are found in a wide variety of anti-phage defense systems, including Menshen, PARIS, and Retron systems[24,49] (Figure I.4a).

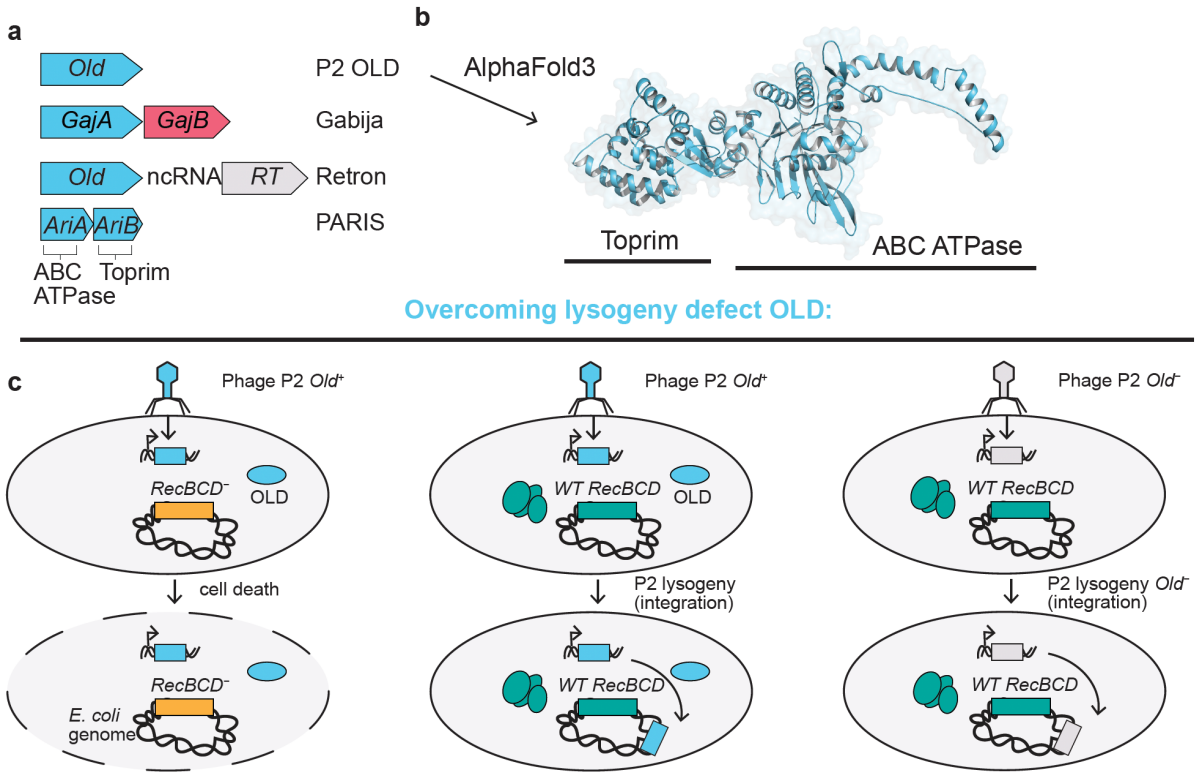


Figure I.4 | OLD nucleases in anti-phage defense. a, Operon schematic of anti-phage defense systems that contain an OLD protein or domain. OLD is highlighted in blue. **b,** AlphaFold3 model of *E. coli* Phage P2 OLD protein with indicated N-terminal ABC ATPase and Toprim domains. **c,** Model of P2 OLD cellular toxicity in *E. coli* with *RecBCD* mutations. OLD proteins sense disruption of RecBCD activity, however, the mechanism remains unknown.

The PARIS anti-phage defense system encodes two proteins AriA an ABC ATPase and AriB a Toprim nuclease[50–52]. A phage P4 satellite PARIS system forms a supramolecular 425 kDa complex of 6 AriA subunits in a propeller shape that is then docked by 3 AriB Toprim domains. PARIS is activated upon detection of DNA-like phage proteins, including phage T7 overcome classical restriction (Ocr), which is an inhibitor of restriction-modification systems. AriB TOPRIM proteins then dimerize to cleave host transfer RNAs (tRNAs)[50,51,53]. The required catalytic motifs for Toprim domain proteins include a conserved glutamate residue and two conserved aspartates (DxD)[54]. Toprim domain-containing proteins, including topoisomerases, have been implicated in cleaving both DNA and RNA substrates[54]. ABC ATPase proteins are a distinct family of P-loop NTPases that hydrolyze ATP in response to a signal[55]. ABC ATPase that are fused to Toprim domain proteins usually contain a coiled-coil insert, lack a Zn-hook domain, and have deviations in the ATPase D-loop and Q-loop motifs[55]. OLD nuclease proteins represent a diverse family of enzymes that all likely cleave phage DNA during infection.

Superfamily 1 Helicases

Gabija protein GajB has homology to superfamily 1 (SF1) dsDNA break repair helicases including Rep and UvrD[56]. Rep and UvrD unwind DNA at both ssDNA and dsDNA junctions and translocate on DNA in the 3′–5′ direction[57]. SF1 helicases have four subdomains (1A, 1B, 2A, 2B) that rotate relative to each other upon ATP hydrolysis and ssDNA translocation[58]. Dimerization and interaction with other proteins increase the processivity of DNA unwinding. For example, during methyl mismatch guided repair, UvrD dimerizes and is activated by binding a dimer of the DNA repair protein MutL[59]. DNA damage during phage infection may trigger GajB helicase activity.

Supramolecular Complexes in Antiviral Immunity

The formation of oligomeric complexes allows for control of activation, signal transduction, and prevention of autoimmunity in antiviral signaling. In human immunity, NLR proteins oligomerize in response to many PAMP signals to form large structures, including inflammasome disks[60]. Additionally, gasdermin proteins in both human and bacterial immunity are proteolytically activated to form large oligomeric pores that cause cell death[60,61]. In CBASS and Thoeris immunity, TIR and SLOG domain proteins oligomerize into long filaments upon binding nucleotide second messages to exert antiviral defense[15,62] (Figure I.5a). Pre-formed supramolecular complexes also function to exert anti-phage defense. Restriction by an adenosine deaminase acting on RNA (RADAR) systems forms 10 mDa complex of a 7-mer AAA+ ATPase and 12-mer deaminase[63,64] (Figure I.5b). PARIS also forms a supramolecular complex 6 ABC ATPase and 3 Toprim proteins that then unleashes the Toprim dimers to cleave host tRNA[50,53] (Figure I.5c). Gabija forms a 500 kDa complex of an OLD nuclease core docked by helicase dimers on either side of the complex[37–40] (Figure I.5d). Widespread use of supramolecular complexes in diverse anti-phage defense systems demonstrate a theme where bacteria use large complexes to overcome viral infection.

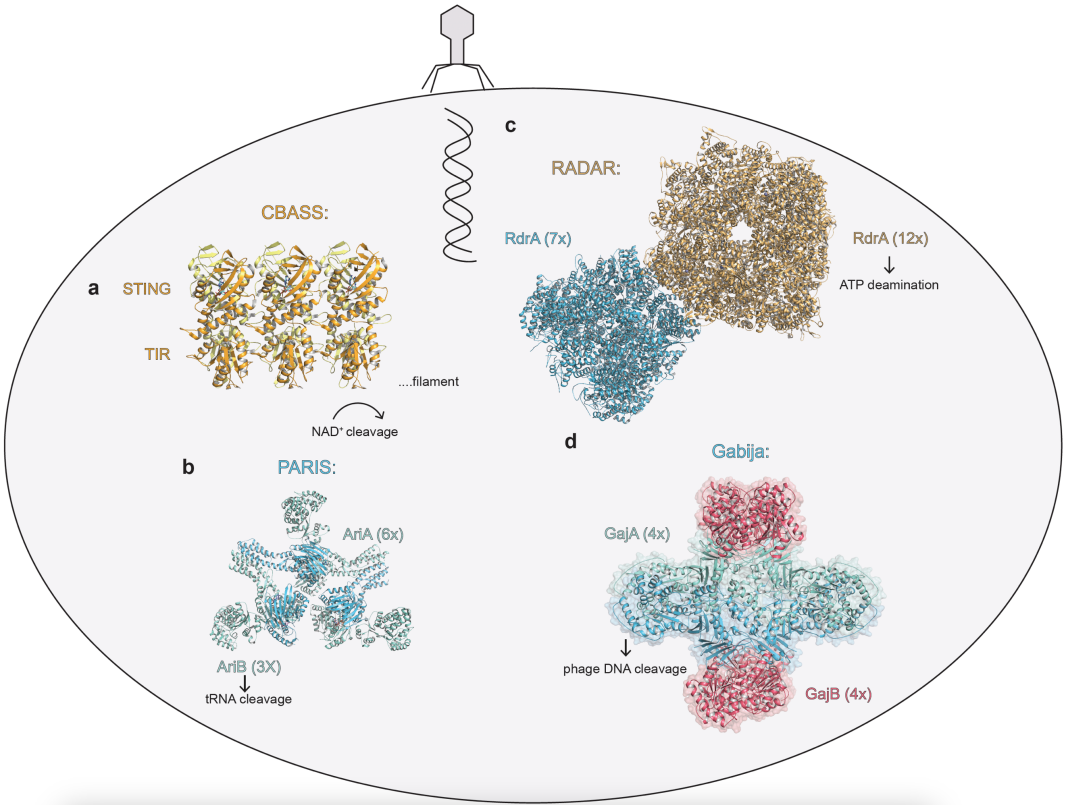


Figure I.5 | Diverse anti-phage defense systems use supramolecular complex formation for defense activity.

a, *Sphingobacterium faecium* TIR-STING filament induced by cyclic di-GMP binding, which leads to NAD^+ hydrolysis (PDB ID: 7UN8). **b,** 10 MDa *E. coli* RADAR 12:7 RdrAB complex that deaminates ATP for anti-phage defense (PDB ID: 8FNW). **c,** *E. coli* B185 PARIS 6:3 AriAB complex that is activated by phage DNA-mimic binding to activate AriB trRNA cleavage (PDB ID: 8V48). **d,** *Bc*Gabija GajAB supramolecular complex that degrades phage DNA (PDB ID: 8SM3) (Antine et al. 2024).

Phage anti-defense

While bacteria encode anti-phage defense systems, phages evolve diverse strategies to defeat immunity and effectively replicate known as anti-defense. Phages can mutate the target of the anti-phage defense systems or encode dedicated proteins that overcome defense. Phages can evade restriction-modification and CRISPR systems by mutating target DNA sequences in phage genomes[65,66]. However, for CRISPR, this can then lead to the acquisition of new spacers, leading to a never-ending evolutionary battle[66]. Phages can also mask the target sites of nucleic acid targeting defense systems. For example, phage T4 encodes an N6-adenine methyltransferase (Dam), which methylates cytosines in restriction-modification target sites[67]. Furthermore, T4 phage DNA is then glucosylated by dedicated glucosyltransferases, which further inhibit restriction-modification and CRISPR systems[68].

Many anti-phage defense systems are activated by the activity of or physically binding to a phage protein[69]. To escape Retron-Eco8, ShosTA, Hachiman, Hna, AbpAB, systems, phages mutate ssDNA binding proteins (SSBs)[69–71]. SSBs bind ssDNA produced during DNA replication and help with DNA polymerase processivity[72]. Retron-Eco8 physically binds to WT SECphi4 SSB *in vitro*, but not the escaper phage point mutant form[69]. However, other systems do not directly bind phage SSBs, indicating that phage DNA replication and SSB activity activates defense. CapRel^{SJ46}, SEFIR, and Thoeris are activated by phage capsid proteins[69,73–75]. CapRel^{SJ46} and bNACHT11 bind phage capsid proteins but also structurally unrelated phage proteins at the same interaction interface[17,73]. Recognizing distinct PAMPs highlights shared mechanisms between human and bacterial immunity.

Instead of mutating activating proteins, phages can encode dedicated anti-defense proteins that block host immunity. Phages evade nucleotide second messenger-based systems CBASS, Thoeris, and Pycsar systems by anti-defense proteins that sequester or degrade the nucleotide signals[18,76–81] (Figure I.5a–c). Anti-CBASS 2 (Acb2) forms a hexamer that binds three cyclic dinucleotides in each of the corners of the hexamer[76] (Figure I.5b). Intriguingly, Acb2 can also bind cyclic trinucleotides at the same time in a different pocket from where cyclic dinucleotide bind[78]. Phage Thoeris anti-defense 1 (Tad1) also forms a hexamer and can sequester 1″–3′-glycocyclic ADP-ribose (1″–3′-gcADPR) produced by Thoeris systems along with cyclic di- and trinucleotides from CBASS systems[18,80] (Figure I.5a). Anti-CBASS 4 (Acb4) blocks CBASS defense by forming a tetramer and binding cyclic dinucleotides between protomer interfaces[79]. Acb1 acts as a monomer to degrade di- and trinucleotides from CBASS immunity that contain a 3′–5′ adenosine linkage[81] (Figure I.5c). Overall, sponges and nucleases are a diverse family of anti-defense proteins that enable phages to evade cyclic nucleotide-based defense systems.

Phages can also encode anti-defense proteins that directly bind and inactivate their cognate defense system. The most well-studied anti-defense binders are phage proteins that mimic DNA and bind to the active sites of restriction-modification and CRISPR systems[82,83] (Figure I.6 d–e). Anti-CRISPR DNA mimics have a negative charge distribution that mirrors the pattern of phosphates on the target DNA that the specific CRISPR system targets[82]. For example, AcrIIA2 binds to *Streptococcus pyogenes* Cas9 in a sgRNA-bound state and occupies the PAM dsDNA binding site of Cas9 preventing a conformational change required for target dsDNA binding and cleavage[82,84] (Figure I.6d). Phage T7 Ocr is the first protein to be

expressed during infection and mimics bent B-form DNA to target the active sites of both the modification and restriction enzymes of Type I RM systems[85] (Figure I.6e). The phage Lambda protein, Gam binds as a dimer to mimic dsDNA and blocks the DNA-binding channel of the RecB subunit of RecBCD[86] (Figure I.6f). Phages encode a wide diversity of DNA mimic proteins to target all the differing ways nucleic acid interact with anti-phage defense systems.

Phages also encode proteins that directly bind anti-phage defense systems without mimicking DNA. A recent AlphaFold2-pulldown screen revealed anti-defense proteins that bind and block CBASS Thoeris defense[87]. Anti-CBASS 3 (Acb3) likely binds to both the nucleotidyltransferases active site and ligand binding surface of a type III CBASS system from *E. coli*, effectively blocking both phage cue binding site and the generation of a nucleotide second messenger. Thoeris anti-defense 3 (Tad3) binds as a dimer to two Thoeris B monomers (ThsB) (Figure I.6g). Tad3 then extends into the TIR-domain active site of ThsB and prevents 1″–3′-gcADPR second messenger production[87]. OLD anti-defense 1 (Oad1) blocks *Vibrio cholerae* Class I OLD anti-phage defense by directly binding the defense system *in vivo*, however, how this occurs structurally has not been investigated[88]. One of the key characteristics of anti-defense proteins is that they have no structural or sequence similarity to well defined proteins making the discovery of new inhibitors with wildly different structures an exciting hallmark of phage anti-defense[89] (Figure I.6h).

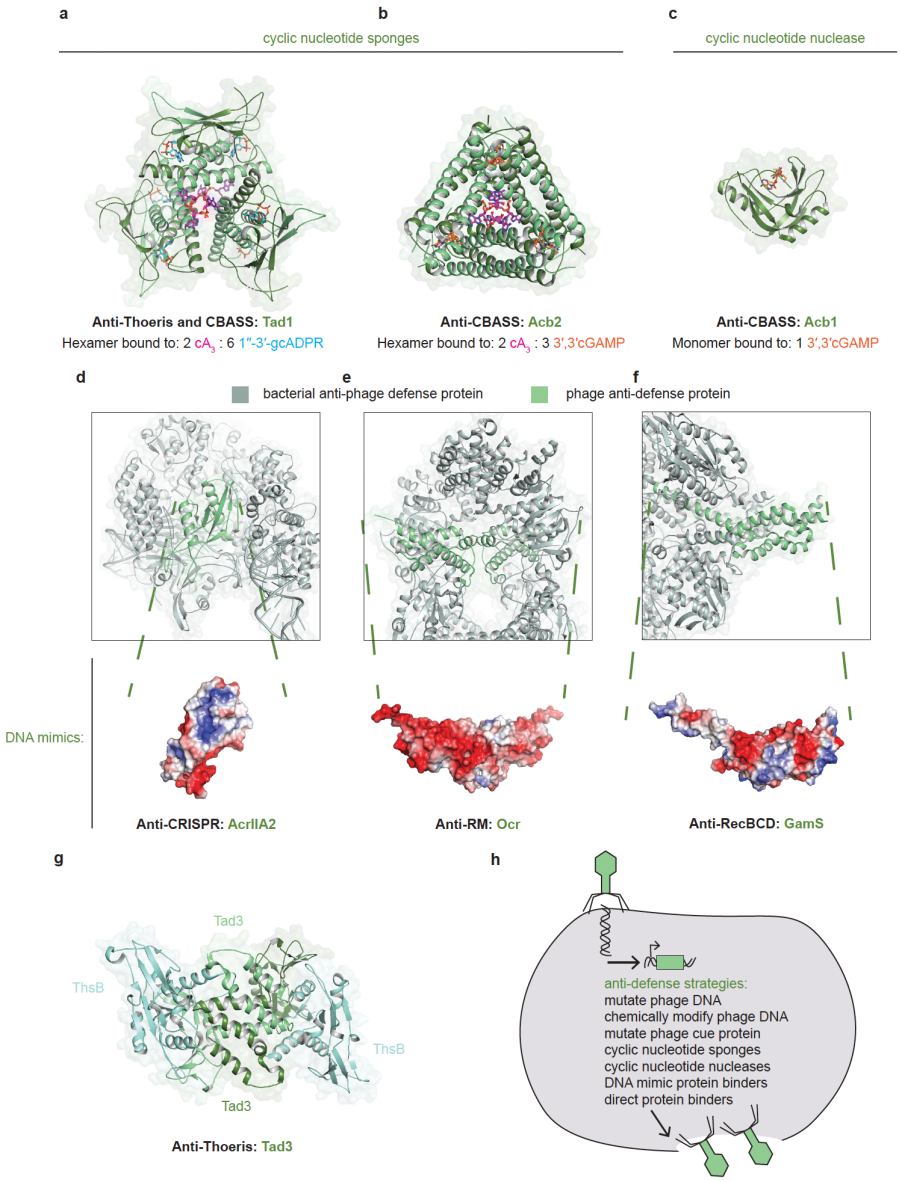


Figure I.6 | Phages use diverse anti-defense strategies to bind and subvert bacterial anti-phage defense.

a, *Clostridioides mangenotii* prophage Tad1 hexamer bound to 2 molecules of cA_3 (magenta) and two 6 molecules of 1″-3′-gcADPR (cyan) (PDB ID: 8KBF). Tad1 protomers are highlighted in two shades of green. **b,** *Pseudomonas phage* PaP2 Acb2 hexamer bound to two molecules of cA_3 (magenta) and 3 molecules of 3′3′-cGAMP (orange) (PDB ID: 8IY2). **c,** *Erwinia* phage FBB1 Acb1 in complex with nonhydrolyzable 3′3′-cGAMP (orange) (PDB ID: 7T26). **d,** *Listeria* phage LP-101 AcrIIA2 (green) in complex with *Streptococcus pyogenes* Cas9 and crRNA (grey) (PDB ID: 6MCC). **e,** *Escherichia* phage T7 Ocr dimer bound to Type I RM EcoKI (PDB ID: 2Y7C). **f,** Phage Lambda GamS dimer complex with *E. coli* RecBCD (PBD ID: 5MBV). **g,** Tad3 dimer from a metagenomic sample (green) bridging two Thoeris ThsB monomers (light blue) (PDB ID: 9B7D). **h,** Summary of the diverse strategies phages can employ to circumvent bacterial anti-phage defense.

Phage nucleic acid targeting for anti-phage defense

The diverse strategies by which phages can overcome nucleic acid-based anti-phage defense has likely led to the evolution of many sensors and diverse nuclease effectors. Newly discovered anti-phage phage defense systems encode many nuclease proteins with phospholipase D (PLD), OLD, (P)D-(D/E)-XK domains, and domains of unknown function (DUF4297, DUF1937, DUF2075) that all likely function through nucleic acid cleavage. Phospholipase D nucleases are part of a large group of proteins in both eukaryotes and prokaryotes that are involved in phospholipid metabolism and recognize phosphodiester bonds to cleave substrates using a $H(X)K(X)_4D$ motif[90]. (P)D-(D/E)-XK, named after their catalytic motif, are involved in a wide variety of processes, including replication, DNA damage repair, intron splicing, and are the main nuclease found in restriction enzymes[91]. Like Gabija, these nucleases are usually accompanied by a helicase domain/protein, including DEAD-box, HerA, and RNA helicases. Hachiman is another nuclease-helicase system that forms a 1:1 complex that degrades both phage and host DNA in response to DNA damage[92]. These systems are annotated as individual systems as they have very little sequence similarity or structure, indicating diverse ways in which they are activated. Understanding how different nuclease-helicase systems exert phage defense will lead to a broader picture of antiviral immunity as many of these proteins are also involved in eukaryotic immunity.

Chapter 1: Structural basis of Gabija anti-phage defence and viral immune evasion

Sadie P. Antine[1,2], Alex G. Johnson[1,2], Sarah E. Mooney[1,2], Azita Leavitt[3], Megan L. Mayer[4], Erez Yirmiya[3], Gil Amitai[3], Rotem Sorek[3], Philip J. Kranzusch[1,2,5*]

[1] Department of Microbiology, Harvard Medical School, Boston, MA 02115, USA

[2] Department of Cancer Immunology and Virology, Dana-Farber Cancer Institute, Boston, MA 02115, USA

[3] Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel

[4] Harvard Center for Cryo-Electron Microscopy, Harvard Medical School, Boston, MA 02115, USA

[5] Parker Institute for Cancer Immunotherapy at Dana-Farber Cancer Institute, Boston, MA 02115, USA

The published manuscript can be accessed as follows:

Nature. (2024) **625**, 360–365.
https://doi.org/10.1038/s41586-023-06855-2

Abstract

Bacteria encode hundreds of diverse defence systems that protect them from viral infection and inhibit phage propagation[22–25,52]. Gabija is one of the most prevalent anti-phage defence systems, occurring in more than 15% of all sequenced bacterial and archaeal genomes[20,22,93], but the molecular basis of how Gabija defends cells from viral infection remains poorly understood. Here we use X-ray crystallography and cryo-electron microscopy (cryo-EM) to define how Gabija proteins assemble into a supramolecular complex of around 500 kDa that degrades phage DNA. Gabija protein A (GajA) is a DNA endonuclease that tetramerizes to form the core of the anti-phage defence complex. Two sets of Gabija protein B (GajB) dimers dock at opposite sides of the complex and create a 4:4 GajA–GajB assembly (hereafter, GajAB) that is essential for phage resistance *in vivo*. We show that a phage-encoded protein, Gabija antidefence 1 (Gad1), directly binds to the Gabija GajAB complex and inactivates defence. A cryo-EM structure of the virally inhibited state shows that Gad1 forms an octameric web that encases the GajAB complex and inhibits DNA recognition and cleavage. Our results reveal the structural basis of assembly of the Gabija anti-phage defence complex and define a unique mechanism of viral immune evasion.

Results and Discussion

Bacterial Gabija defence operons encode the proteins GajA and GajB, which together protect cells against diverse phages[22]. To define the structural basis of Gabija anti-phage defence, we co-expressed *Bacillus cereus* VD045 GajA and GajB and determined a 3.0 Å X-ray crystal structure of the protein complex (Figure 1.1a,b, Figure 1.2a,b and Table 1.1). The structure of the GajAB complex reveals an intricate 4:4 assembly with a tetrameric core of GajA subunits braced on either end by dimers of GajB (Figure 1.1b). We focused our analysis first on individual Gabija protein subunits. GajA contains an N-terminal ATPase domain that is divided into two halves by the insertion of a protein dimerization interface (discussed further below) (Figure 1.1c). The GajA ATPase domain consists of an eleven-stranded β-sheet (β1^{ABC}, β2^{ABC}, β4–6^{ABC} and β3^{ABC}, β7–11^{ABC}) that folds around the central α1^{ABC} helix (Figure 1.1c and Figure 1.3a). Sequence analysis of diverse GajA homologues shows that the GajA ATPase domain contains a highly conserved ATP-binding site that is shared with canonical ABC ATPase proteins[94] (Figure 1.3a). The GajA C terminus contains a four-stranded parallel β-sheet β1–β4T surrounded by three α-helices α3^{T}, α4^{T} and α12^{T} that form a Toprim (topoisomerase-primase) domain associated with proteins that catalyse double-stranded DNA (dsDNA) breaks[54] (Figure 1.1c and Figure 1.3a). Consistent with a role in dsDNA cleavage, the structure of GajA confirms previous predictions of overall shared homology between GajA and a class of DNA endonucleases named OLD (overcoming lysogenization defect) nucleases[47,48]. Discovered at first as an *Escherichia coli* phage P2 protein responsible for cell toxicity in *recB* and *recC* mutant cells[42,43,46], OLD nucleases occur in diverse bacterial genomes, either as single proteins (class 1) or associated with partner UvrD/PcrA/Rep-like helicase proteins (class 2), but the specific function of most OLD nuclease proteins is unknown[47,48]. GajA is a class 2 OLD nuclease, with the Toprim domain containing a complete active site composed of DxD after β3^{T} (D432 and D434), an invariant glutamate after β2^{T} (E379) and an invariant glycine between α1^{T} and β1^{T} (G409). This is similar to the active site of *Burkholderia pseudomallei* (BpOLD), which was previously shown to be essential for a two-metal-dependent mechanism of DNA cleavage[45] (Figure 1.1d and Figure 1.3a).

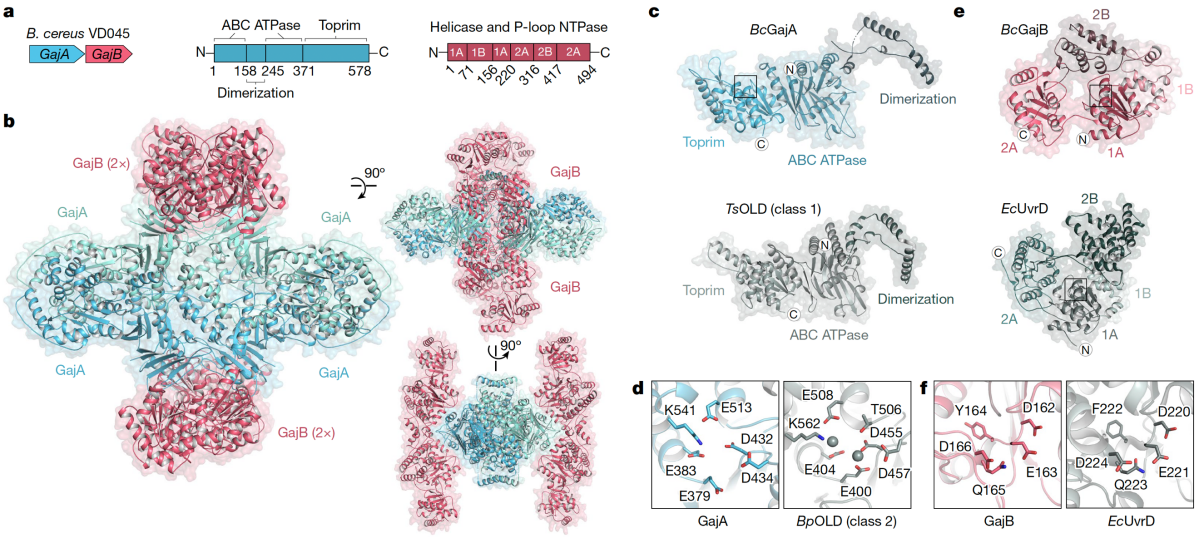


Figure 1.1 | Structure of the Gabija anti-phage defense complex.

a, Schematic of *B. cereus* Gabija defense operon and domain organization of GajA and GajB. **b,** Overview of the GajAB X-ray crystal structure shown in three orientations. GajA protomers are depicted in two shades of blue and GajB protomers are in red. **c,** Isolated GajA monomer (top) and comparison with a *Ts*OLD nuclease monomer (bottom) (PDB ID 6P74) (Schiltz et al. 2020). **d,** Close-up view of GajA (left) and *Bp*OLD (right) (PDB ID 6NK8) (Schiltz et al. 2019). Toprim catalytic residues. Location of GajA cutaway image is indicated with a box in (c) and magnesium ions are depicted as grey spheres. **e,** Isolated GajB monomer (top) and comparison with *Ec*UvrD (bottom) (PDB ID 2IS2) (Lee & Yang 2006). **f,** Close-up view of GajB (left) and *Ec*UvrD (right) DEXQD-box motif. Location of GajB and UvrD cutaway image is indicated with a box in (e).

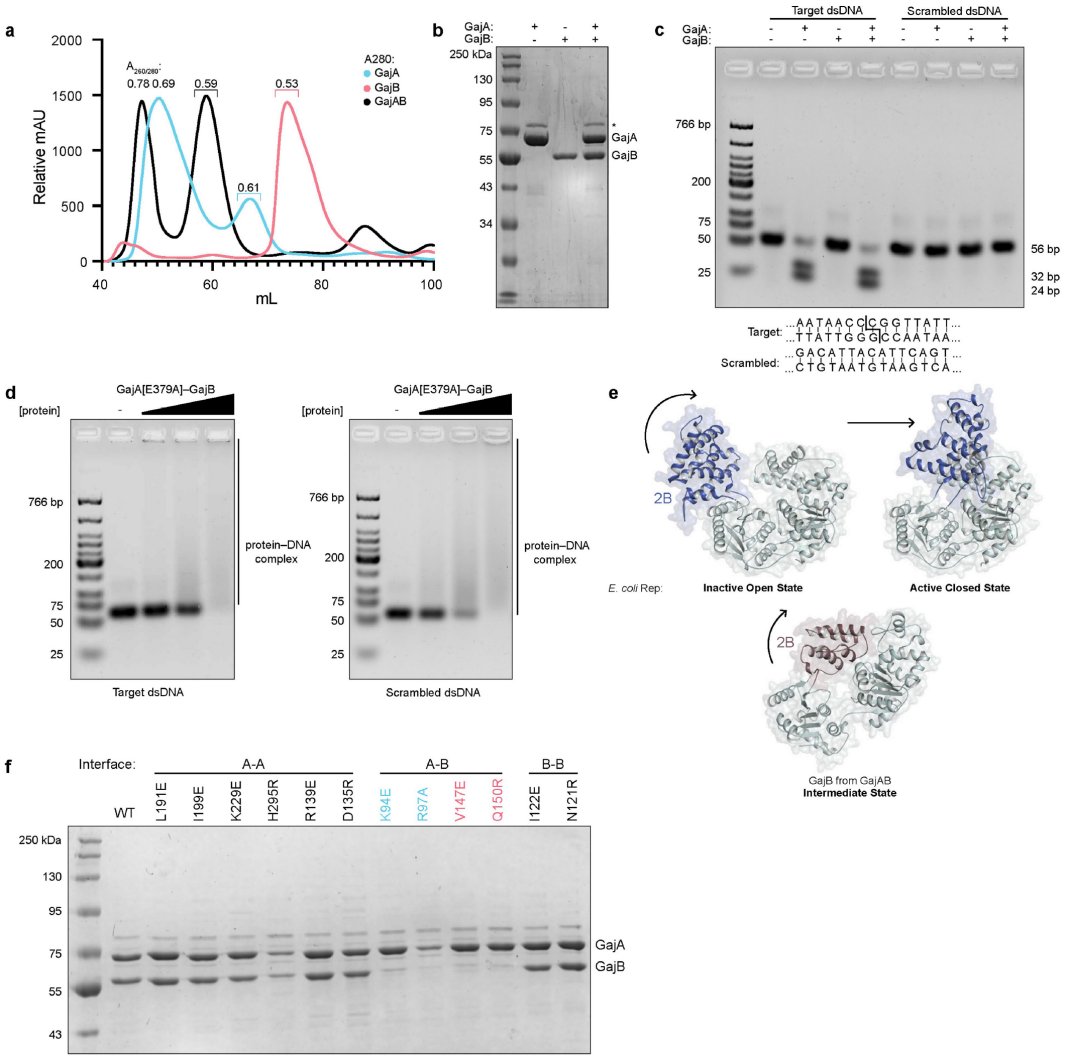


Figure 1.2 | GajA and GajB form a supramolecular complex that cleaves phage lambda DNA *in vitro*.

a, Size-exclusion chromatography (16/600 S200) analysis of recombinant *Bc*GajA and *Bc*GajB proteins, and the co-expressed *Bc*GajAB complex. Brackets indicate fractions collected for biochemical and structural analysis with $A_{260/280}$ of the final purified proteins listed above. **b** SDS-PAGE**,** analysis of purified GajA, GajB, and GajAB. Asterisk indicates minor contamination with the *E. coli* protein ArnA. Data are representative of at least 3 independent experiments. **c,** Agarose gel analysis of the ability of GajA, GajB, and GajAB to cleave a 56-bp target and scrambled dsDNA demonstrates that GajA alone and the GajAB complex can cleave target DNA only. The sequence-specific GajA target dsDNA with cleavage site described in Cheng et al. 2021 and the scrambled 15-bp sequence are shown below. **d,** Catalytic dead GajA[E379A]–GajB complex binding to target dsDNA (left) and scrambled dsDNA (right). **e,** Structural comparison of GajB and *Ec*Rep (PDB ID 1UAA) (Korolev et al. 1997) demonstrates the GajB 2B domain is rotated in a partially active intermediate position in the GajAB complex structure. **f,** SDS-PAGE analysis of *Bc*GajAB mutant protein complex formation after co-expression and Ni-NTA pull-down demonstrates that mutations to the GajA–GajB interface disrupt complex formation. The GajA and GajB homo-oligomerization interfaces are not required for GajA–GajB interaction, but it is not known if these mutants remain competent at forming the wildtype 4:4 complex. Data in panels c, d, f are representative of 3 independent experiments.

Table 1.1 | Summary of X-ray data collection, phasing and refinement statistics

	Gabija GajA–GajB **(8SM3)**
Data collection	
Space group	P 6_2 2 2
Cell dimensions	
a, *b*, *c* (Å)	215.79 215.79 173.81
α, β, γ (°)	90.0, 90.0, 120.0
Resolution (Å)	49.24–3.00 (3.10–3.00)
R_{pim}	4.0 (80.5)
$I / \sigma(I)$	15.4 (1.4)
Completeness (%)	100.0 (100.0)
Redundancy	18.1 (16.1)
Refinement	
Resolution (Å)	49.24–3.00
No. reflections	
Total	872109
Unique	48144
Free	2000
R_{work} / R_{free}	23.69 / 26.60
No. atoms	
Protein	8501
Ligand / ion	5
Water	–
B-factors	
Protein	130.19
Ligand / ion	174.06
Water	–
R.m.s. deviations	
Bond lengths (Å)	0.002
Bond angles (°)	0.442

Dataset was collected from an individual crystal. Values in parentheses are for the highest resolution shell.

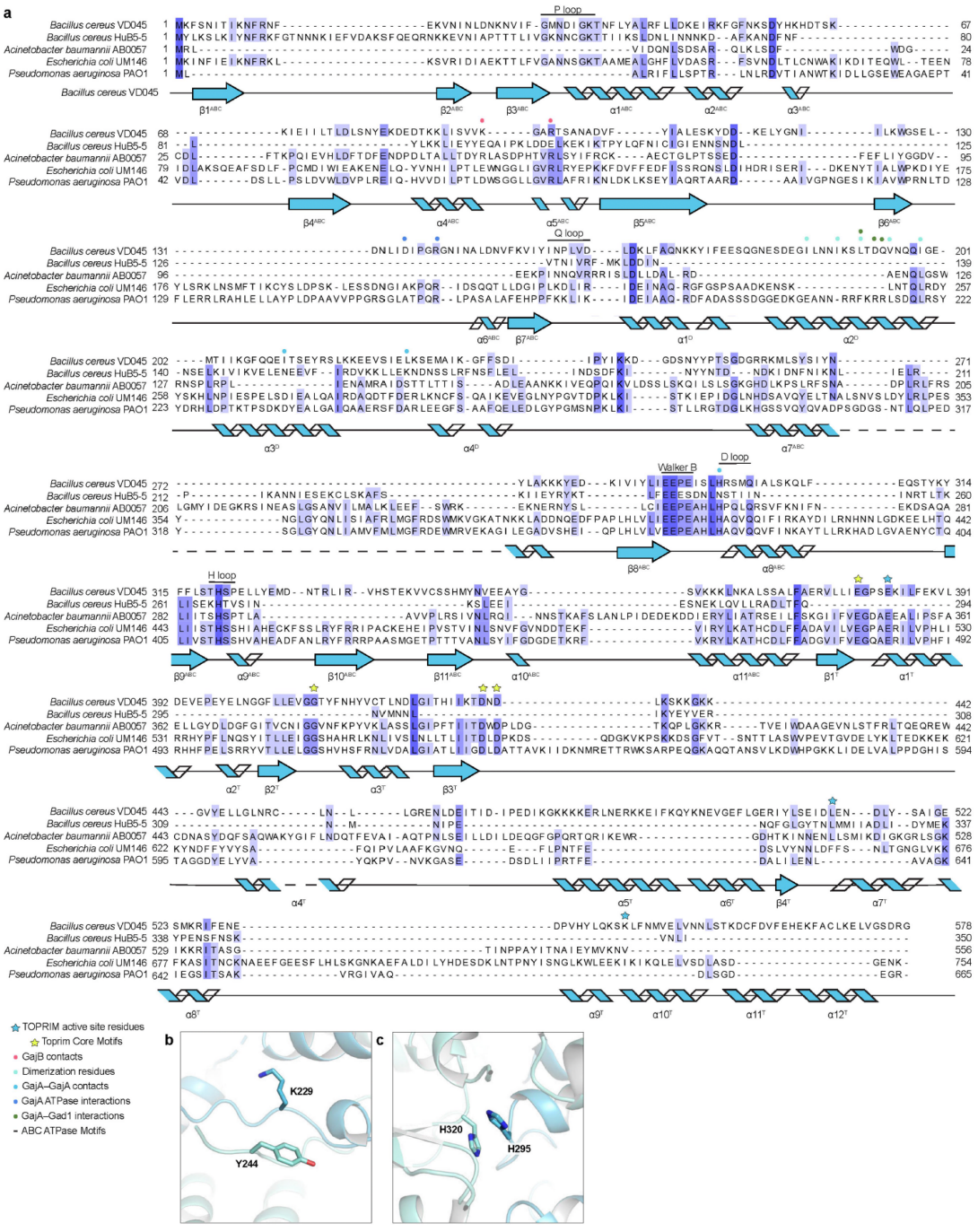


Figure 1.3 | Structural characterization of GajA.

a, Structure-guided alignment of GajA proteins from indicated bacteria colored according to amino acid conservation. The determined *Bacillus cereus* VD045 GajA secondary structure is displayed, and active-site and oligomerization interface residues are annotated according to the key below. Secondary structure abbreviations include ABC ATPase domain (ABC), dimerization domain (D), and Toprim domain (T). **b,c,** Zoomed-in views of GajA–GajA oligomerization interactions including dimerization domain interactions (b) and ABC ATPase domain interactions (c).

The structure of GajB reveals a superfamily 1A DNA helicase domain. Bacterial DNA helicases belonging to this superfamily typically have a role in DNA repair[57] (Figure 1.1e). Superfamily 1A helicase proteins such as UvrD, Rep and PcrA translocate along single-stranded DNA in the 3′ to 5′ direction, and are architecturally divided into four subdomains—1A, 1B, 2A and 2B—that reposition relative to each other during helicase function[95]. GajB contains all of the conserved helicase motifs that are required for ATP hydrolysis and nucleic acid unwinding, including a Walker A motif Gx(4)GK-[TT] and a UvrD-like DEXQD-box Walker B motif that is responsible for the hydrolysis of nucleoside triphosphate[95–97] (Figure 1.1f and Figure 1.4a). Activation of superfamily 1A DNA helicase proteins such as UvrD and Rep is known to require protein dimerization and the rotation of the 2B subdomain[58,59,98]. Comparisons with UvrD and Rep show that GajB protomers in the GajAB complex exhibit a partial rotation of the 2B domain relative to 2A–1A–1B, consistent with a partially active conformation that is poised to interact with phage DNA (Figure 1.2e).

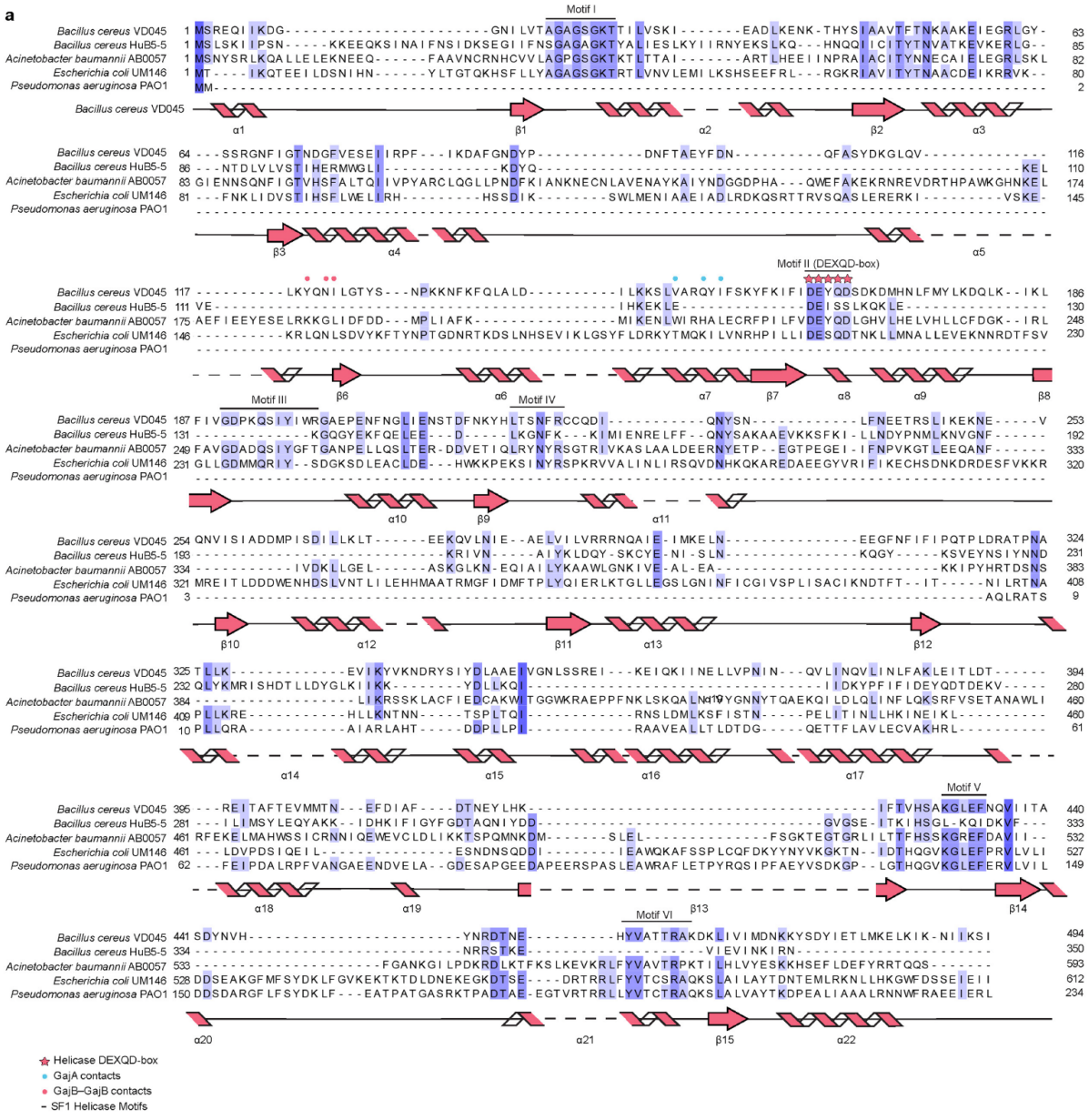


Figure 1.4 | Structural characterization of GajB.

a, Structure-guided alignment of GajB proteins from indicated bacteria colored according to amino acid conservation. The determined *Bacillus cereus* VD045 GajB secondary structure is displayed, and active-site and oligomerization interface residues are annotated according to the key below.

To define the mechanism by which the Gabija complex assembles, we analysed oligomerization interfaces within the GajAB structure. Purification of individual Gabija proteins shows that GajA alone is sufficient to oligomerize into a homo-tetrameric assembly (Figure 1.2a). GajB migrates as a monomer on size-exclusion chromatography, supporting a stepwise model of GajAB assembly (Figure 1.5a and Figure 1.2a). GajA tetramers form through two highly conserved oligomerization interfaces (Figure 1.5b,c and Figure 1.3). First, the GajA N-terminal ATPase domain contains an insertion between $\beta7^{ABC}$ and $\beta8^{ABC}$ that consists of four α-helices ($\alpha1–\alpha4^{D}$) that zip up against a partnering GajA protomer to form a hydrophobic interface along the α2D helix (Figure 1.5b). A similar α1–α4D dimerization interface exists in the structure of the *Thermus scotoductus* class 1 OLD (*Ts*OLD) protein, which shows that this interface is conserved within divergent OLD nucleases[48] (Figure 1.1c and Figure 1.5c). The GajA ATPase domain contains a second oligomerization interface in a loop between $\beta6^{ABC}$ and $\alpha6^{ABC}$, in which hydrogen-bond contacts between D135 and R139 interlock two GajA dimers to form the tetrameric core assembly (Figure 1.5c). Compared to GajA, the GajB–GajB dimerization interface is minimal and consists of a hydrophobic surface in the GajB helicase 1B domain centered at Y119 and I122 (Figure 1.5d). Major GajA–GajB contacts also occur in the GajB helicase 1B domain, in which GajA R97 in a loop between $\alpha4^{ABC}$ and $\beta5^{ABC}$ forms hydrogen-bond contacts with Q150 in GajB α7 along with hydrophobic packing interactions centred at GajB V147 (Figure 1.5e and Figure 1.4a). Notably, the GajAB structure shows that the GajB helicase 1A subdomain, which includes the DEXQD-box active site, is positioned adjacent to the GajA ATPase domain, suggesting that GajB ATP hydrolysis and DNA unwinding activity might regulate the activation of the GajA ATPase domain (Figure 1.5e). In addition to the major GajAB interface contacts, Gabija supramolecular complex assembly is driven by extensive protomer interactions that result in around 31,000 $Å^2$ of surface area buried for the GajA tetramer and around 1,800 $Å^2$ of surface area buried for each GajB subunit.

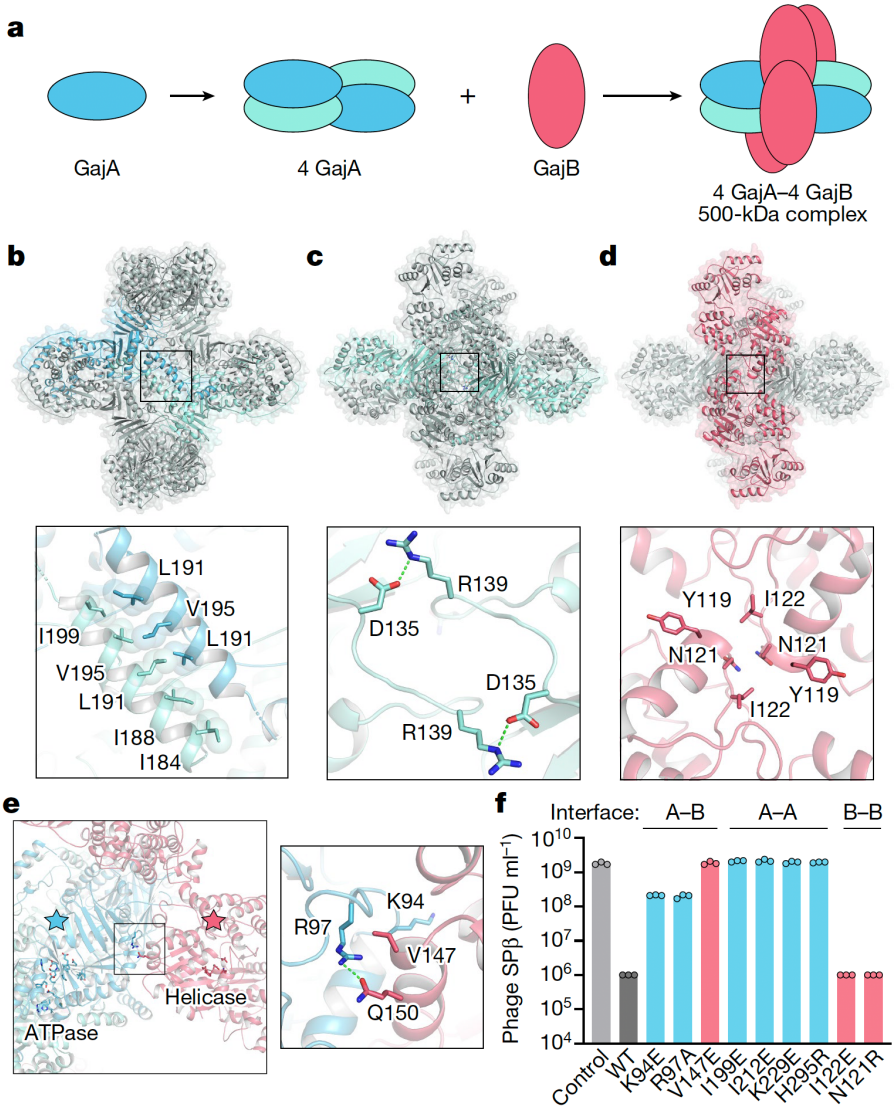


Figure 1.5 | Mechanism of Gabija supramolecular complex assembly.

a, Schematic model of GajAB complex formation by GajA tetramerization and GajB docking. **b,** Overview of the GajA $\alpha2^{D}$–$\alpha2^{D}$ dimerization interface and detailed view of interacting residues. For clarity, each GajA monomer is depicted in two shades of blue. **c,** Overview of the GajA–GajA ATPase interaction and detailed view of inter-subunit D135–R139 interaction. **d,** Overview of the minimal GajB–GajB dimer interface and detailed view of GajB–GajB hydrophobic interactions centered around Y119, N121, and I122. **e,** Overview of the GajA–GajB interface highlighting proximity of GajA ABC ATPase and GajB helicase active site residues (left) with box indicating location of GajA R97 and GajB V147 and Q150 interaction (right). **f,** Analysis of mutations in the GajA–GajB (A–B), GajA–GajA (A–A), and GajB–GajB (B–B) multimerization interfaces. GajA and GajB mutations were selected by identifying central residues with well-defined protein–protein contacts within each multimerization interface and tested to determine the impact on the ability of the *B. cereus* Gabija operon to defend cells against phage infection. Data represent the phage SPβ average plaque-forming units (PFU) mL^{-1} of three biological replicates with individual data points shown.

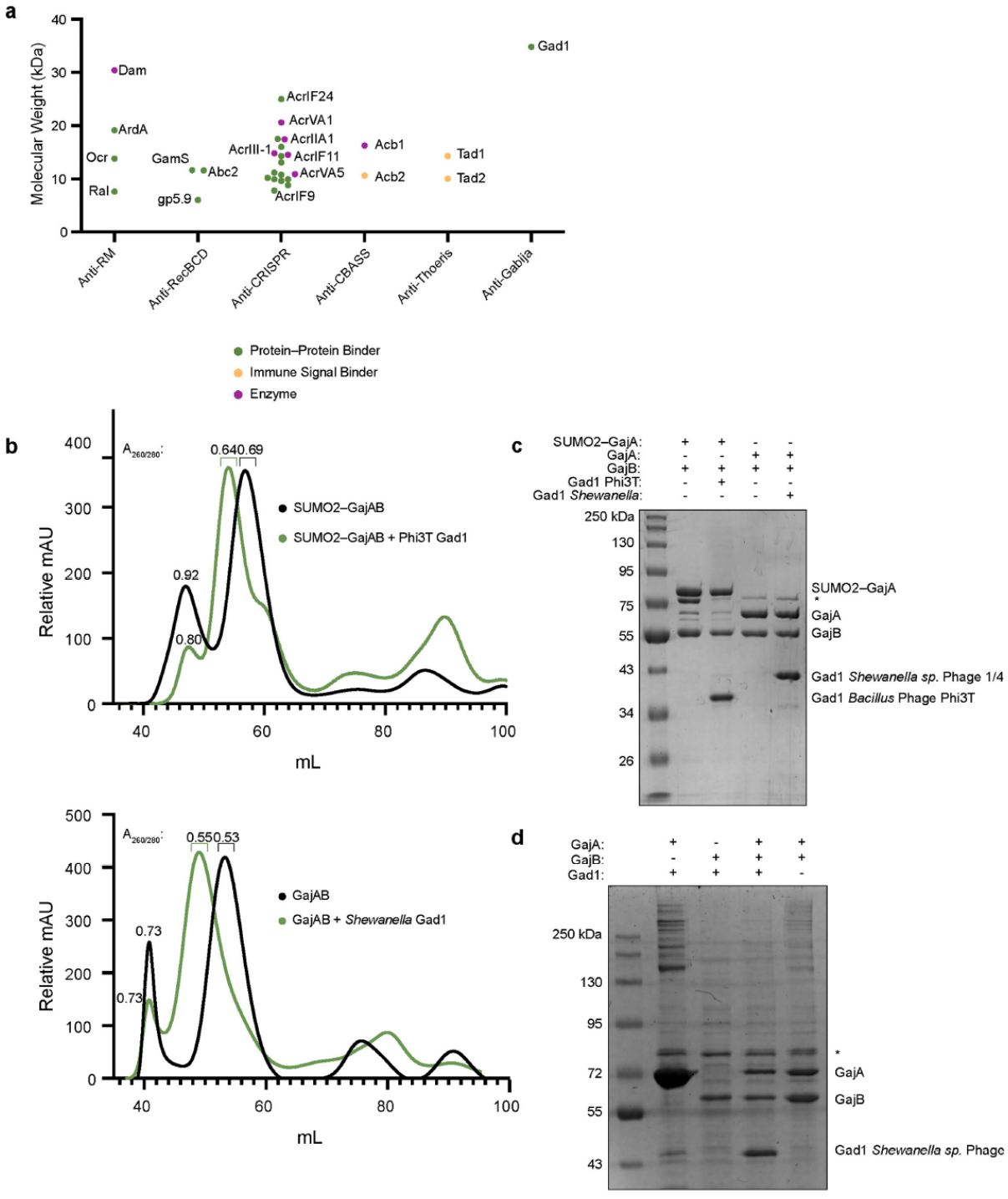
We reconstituted Gabija activity *in vitro* and observed that the GajAB complex binds to and rapidly cleaves a previously characterized 56-bp dsDNA substrate that contains a sequence-specific motif derived from phage lambda DNA[35] (Figure 1.2c). The GajAB complex can interact with a scrambled DNA sequence but is unable to cleave this target DNA (Figure 1.2c,d). GajA and GajB proteins are each essential for phage defence *in vivo*[22,35], but we observed, in agreement with previous results, that GajA is alone sufficient to cleave target DNA and does not require GajB i*n vitro*[35,36] (Figure 1.2c). These results suggest that GajAB complex formation could have a specific role in controlling substrate recognition or nuclease activation during phage infection. To confirm these findings, we analysed protein interaction interfaces in the GajAB complex structure and tested the effects of a panel of mutations on the assembly of the Gabija complex *in vitro* and the ability of Gabija to defend Bacillus subtilis cells from phage SPβ infection *in vivo*. Mutations to the GajA–GajB hetero-oligomerization interface, including GajA K94E and R97A and GajB V147E and Q150R, disrupted complex formation, indicating that these regions are crucial for Gabija complex assembly (Figure 1.2f). Likewise, these substitutions to the GajA–GajB interface markedly reduced the ability of Gabija to inhibit phage replication in *B. subtilis*. Substitutions to the GajA–GajA dimerization interface including I199E, I212E and K229E also resulted in the complete loss of phage resistance (Figure 1.5f). By contrast, phage resistance was tolerant to mutations in the GajB–GajB interface, which suggests that this minimal interaction surface is not strictly essential for anti-phage defence. Together, these results define the structural basis of GajA and GajB interaction and show that the formation of the GajAB supramolecular complex is crucial for Gabija anti-phage defence.

To overcome host immunity, phages encode evasion proteins that specifically inactivate anti-phage defence[76,77,81,99–101]. In a companion study, Yirmiya et al. report the discovery of a viral inhibitor of Gabija anti-phage defence[102], and we reasoned that defining the mechanism of immune evasion would provide further insight into the function of the Gabija complex. Gad1 is a Bacillus phage Phi3T protein that is atypically large (35 kDa) compared to other characterized phage immune-evasion proteins (Figure 1.6a). Protein interaction analysis showed that Gad1 binds directly to GajAB (Figure 1.6b,c), and we used cryo-EM to determine a 2.6 Å structure of the GajAB–Gad1 co-complex assembly (Figure 1.7a,b, Figure 1.8, Figure 1.9a–g and Table 1.2). The GajAB–Gad1 co-complex structure reveals a mechanism of inhibition in which Gad1 proteins form an oligomeric web that wraps 360 Å around the host defence complex. Eight

copies of phage Gad1 encircle the GajAB assembly, forming a 4:4:8 GajAB–Gad1 complex that is around 775 kDa in size (Figure 1.7b,c). Gad1 mainly recognizes the GajA nuclease core, forming extensive contacts along the surface of the GajA dimerization domain (Figure 1.7c,d). Key GajAB–Gad1 contacts include hydrogen-bond interactions from a Gad1 positively charged loop located between β5 and β6 with GajA $\alpha2^{D}$ (Figure 1.7e and Figure 1.10a–c), and hydrophobic packing interactions between Gad1 Y190 and F192 with GajA α2D (Figure 1.7f and Figure 1.10a). Although the contacts between Gad1 and GajB are limited, both GajA and GajB are necessary for Gad1 interaction, indicating that Gad1 specifically targets the fully assembled GajAB complex to inactivate host anti-phage defence (Figure 1.6d).

Figure 1.6 | Size comparison of Gad1 to known phage immune evasion proteins and biochemical characterization of Gad1 for binding to the GajAB complex.

a, Analysis of known phage immune evasion proteins according to function and molecular weight demonstrates that Gad1 is atypically large for an evasion protein that functions through protein–protein interactions with a host anti-phage defense system. Phage immune evasion proteins are categorized and exhibited as colored dots colored according to the key below. Notable evasion proteins are indicated with text labels (Bondy-Denomy et al. 2013; Stanley et al. 2018; Hampton et al. 2020; Hobbs et al. 2022; Huiting et al. 2023; Leavitt et al. 2022; Yin et al. 2023; Tock et al. 2005; Wilkinson et al. 2022; Athukoralage et al. 2020). **b,** Top, size-exclusion chromatography analysis (16/600 S200) of SUMO2-tagged *Bc*GajAB with or without phage Phi3T Gad1 used for cryo-EM structural studies. Bottom, size-exclusion chromatography analysis (16/600 S300) of *Bc*GajAB with or without *Shewanella* phage 1/4 Gad1 used for biochemical studies. Brackets indicate fractions collected and the $A_{260/280}$ of the final purified proteins is indicated above. *Shewanella* phage 1/4 Gad1 was used preferentially for biochemical studies due to less toxicity during *E. coli* expression. **c,** SDS-PAGE analysis of purified SUMO2-tagged GajAB, SUMO2-tagged GajAB in complex with phage Phi3T Gad1, untagged GajAB, and untagged GajB in complex with *Shewanella* phage 1/4 Gad1. **d,** SDS-PAGE analysis of Ni-NTA co-purified GajA, GajB, and GajAB with *Shewanella* phage 1/4 Gad1 indicates that Gad1 only binds the fully assembled GajAB complex. Asterisk indicates minor contamination with the *E. coli* protein ArnA. Data in panels b, c, d are representative of 3 independent experiments.


a
Molecular Weight (kDa)
Dam
ArdA
Ocr
Ral
GamS
Abc2
gp5.9
AcrIF24
AcrVA1
AcrIIA1
AcrIII-1
AcrIF11
AcrVA5
AcrIF9
Acb1
Acb2
Tad1
Tad2
Gad1
Anti-RM
Anti-RecBCD
Anti-CRISPR
Anti-CBASS
Anti-Thoeris
Anti-Gabija
Protein–Protein Binder
Immune Signal Binder
Enzyme
b
Relative mAU
mL
SUMO2–GajAB
SUMO2–GajAB + Phi3T Gad1
0.92
0.80
0.64
0.69
GajAB
GajAB + Shewanella Gad1
0.73
0.73
0.55
0.53
c
SUMO2–GajA:
GajA:
GajB:
Gad1 Phi3T:
Gad1 Shewanella:
250 kDa
SUMO2–GajA
GajA
GajB
Gad1 Shewanella sp. Phage 1/4
Gad1 Bacillus Phage Phi3T
d
GajA:
GajB:
Gad1:
250 kDa
GajA
GajB
Gad1 Shewanella sp. Phage

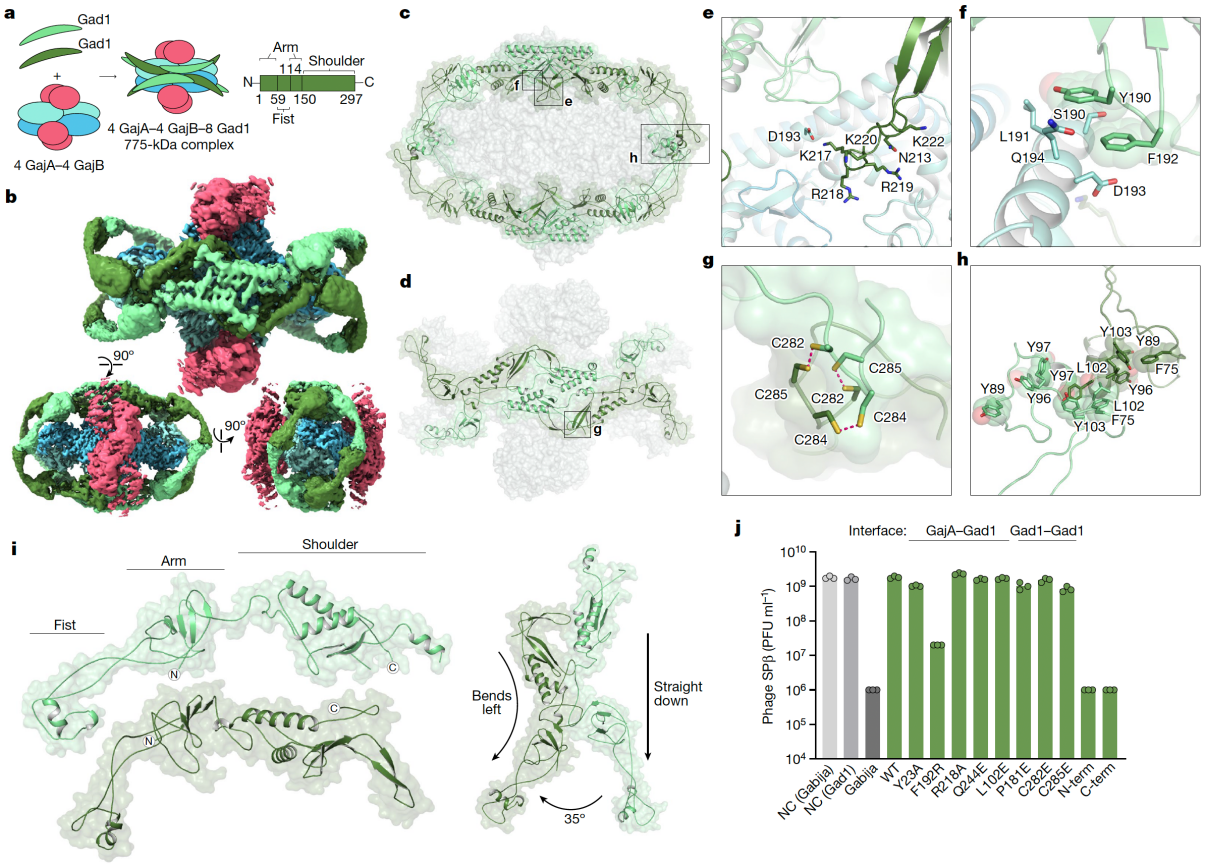


Figure 1.7 | Structural basis of viral evasion of Gabija defense.

a, Schematic model of GajAB–Gad1 co-complex formation and domain organization of phage Phi3T Gad1. **b,** Cryo-EM density map of *Bc*GajAB in complex with Phi3T Gad1 shown in three different orientations. The map is colored by the model, with Gad1 monomers depicted in two shades of green. **c,** Side-view of the complete Gad1 octameric complex and **d,** top-down view of the Gad1 tetrameric interface with boxes highlighting close-up views in (e–h). **e,f,** Zoomed-in views of major Gad1–GajA interface contacts including a Gad1 positively charged loop (e) and hydrophobic interactions with GajA $\alpha2^{D}$ (f). **g,h,** Zoomed-in views of major Gad1–Gad1 oligomerization interactions including disulfide bonds in the C-terminal shoulder domain (g) and fist–fist domain contacts modeled by rigid-body placement of an AlphaFold2 fist domain structure prediction into the cryo-EM map (h). **i,** Two distinct conformations of Gad1 observed in the GajAB–Gad1 co-complex structure. Differences in Gad1 arm domain rotation are highlighted on the right. **j,** Analysis of Gad1 mutations in the GajA–Gad1 and Gad1–Gad1 multimerization interfaces on the ability Gad1 to enable evasion of Gabija defense. Data represent PFU mL^{-1} of phage SPβ infecting cells expressing *Bc*Gabija and *Shewanella sp.* phage 1/4 Gad1, or negative control (NC) cells expressing either plasmid empty vector. *Shewanella sp.* phage 1/4 Gad1 residues are numbered according to the Phi3T Gad1 structure. *Shewanella sp.* phage 1/4 Gad1 N-term and C-term truncations are M1–L152 and V153–E348, respectively. Data are the average of three biological replicates with individual data points shown.

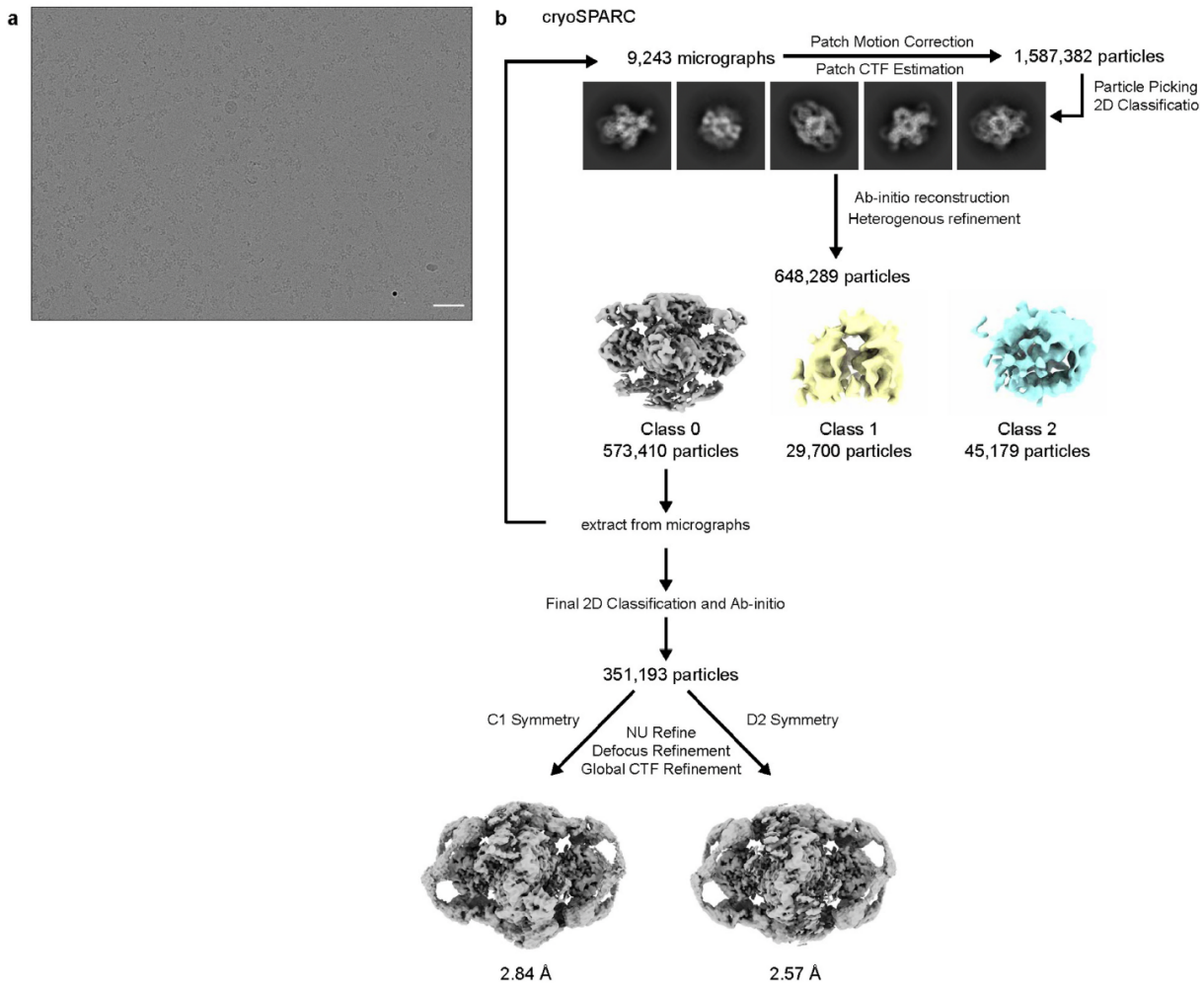


Figure 1. 8 | Cryo-EM data processing GajAB–Gad1 co-complex.

a, Section of a representative electron micrograph (n = 9,243) of SUMO2–GajAB in complex with phage Phi3T Gad1. Scale bar is 50 nm. **b,** Data processing scheme used to generate the final 2.57 Å map.

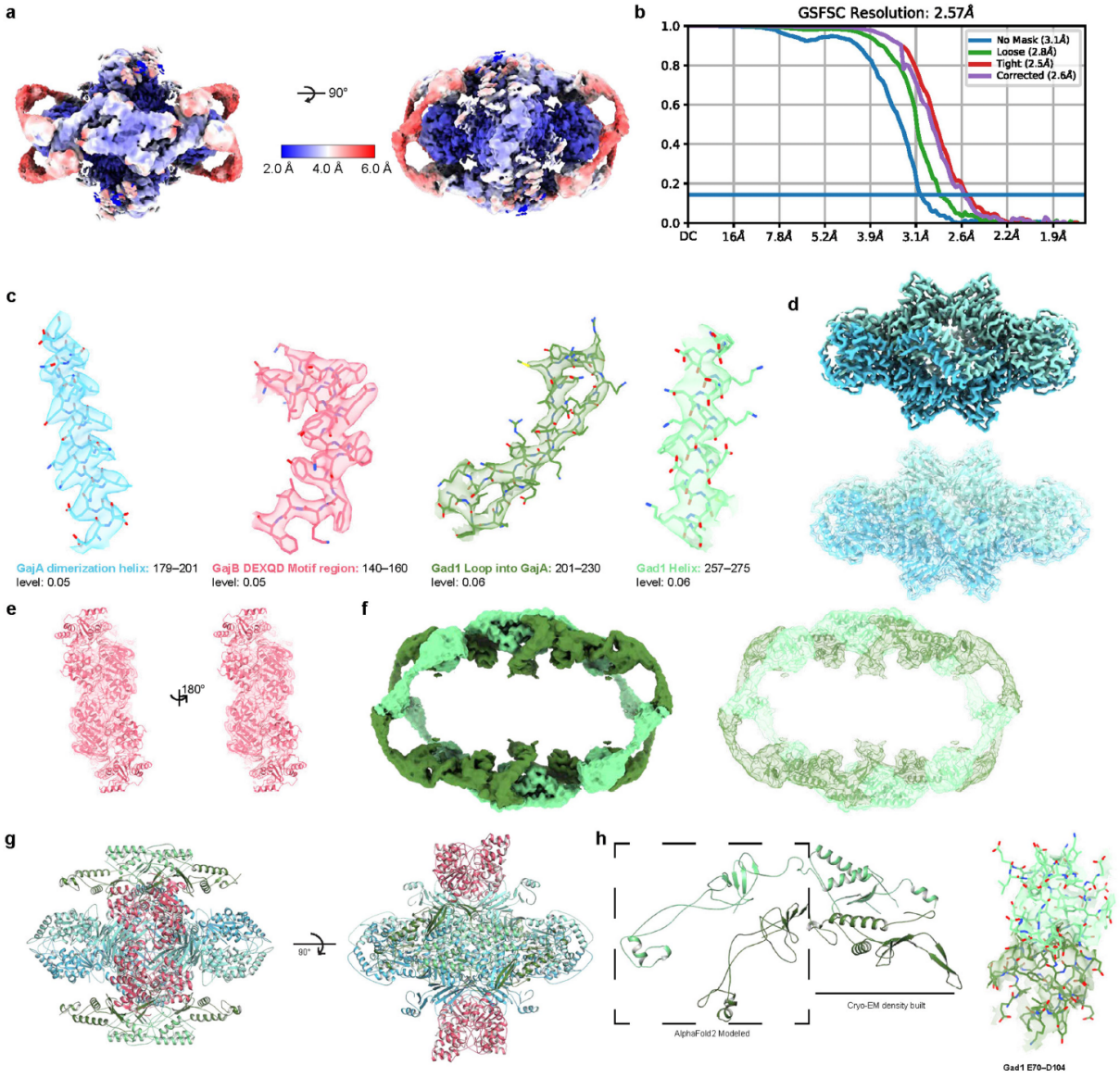


Figure 1.9 | GajAB–Gad1 co-complex Cryo-EM map quality and model to map fitting.

a, Reconstruction of the GajAB–Gad1 co-complex colored by local resolution. **b,** Fourier shell correlation (FSC) of the EM map. **c,** GajA, GajB, and Gad1 map to model fit for designated regions. **d,e,f,** Isolated GajA, GajB, Gad1 density maps with model fitting. **g,** GajAB–Gad1 model that was used for refining the cryo-EM map for Extended Data Table 2. **h,** Left, sections of Gad1 chains that were built *de novo* from the cryo-EM density and built using rigid-body placement of AlphaFold2 modeled residues. Right, cryo-EM density used to fit placement of Gad1 fist–fist domain contacts that complete protomer interactions.

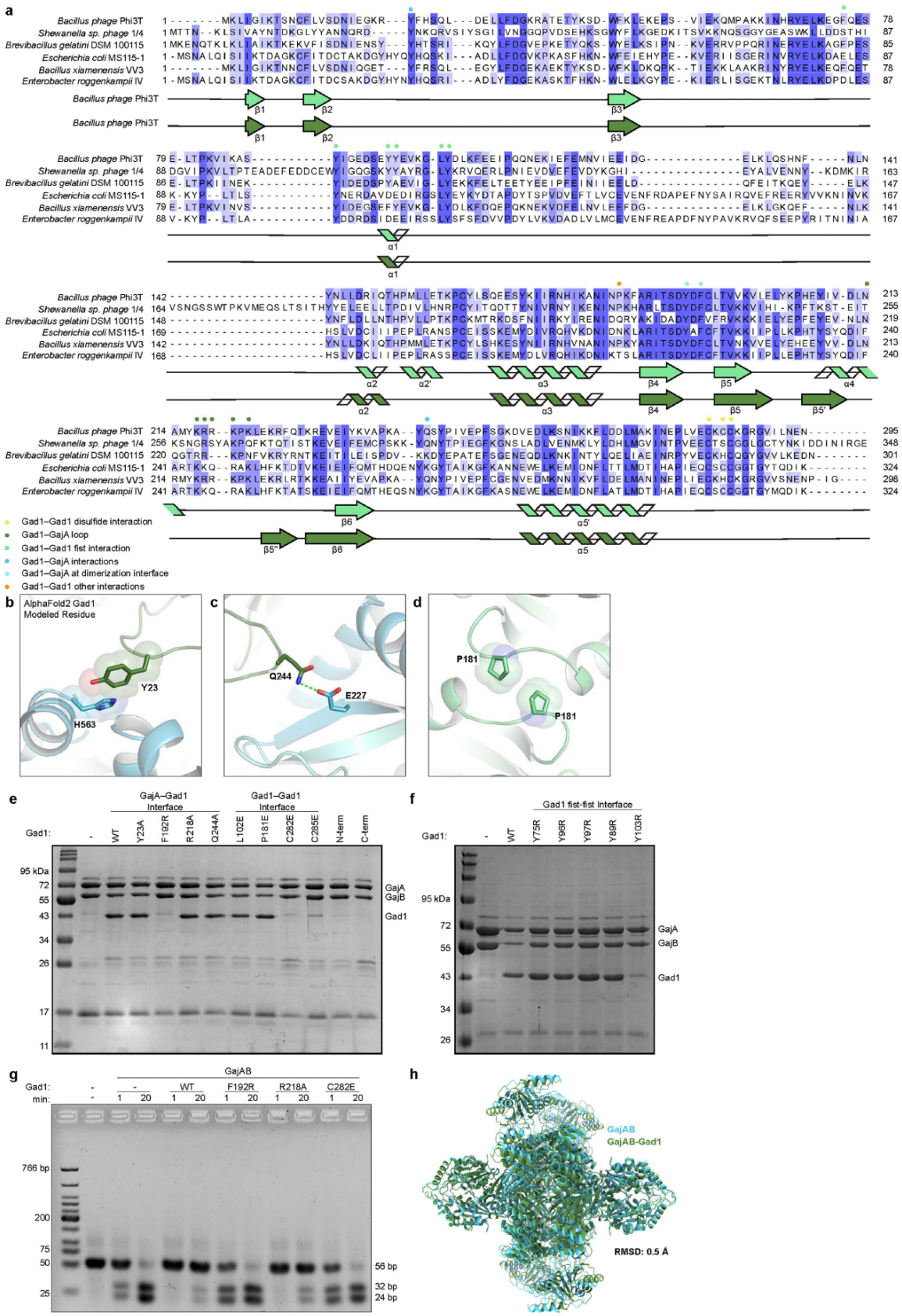
Table 1.2 | Cryo-EM data collection, refinement and validation statistics

	GajAB-Gad1 co-complex (EMD-**41983**) (PDB **8U7I**)
Data collection and processing	
Magnification	37,000
Voltage (kV)	300
Electron exposure (e–/Å^2)	41.1
Defocus range (μm)	−0.8 to −1.9
Pixel size (Å)	0.3115
Symmetry imposed	D2
Initial particle images (no.)	1,587,382
Final particle images (no.)	351,193
Map resolution (Å)	2.6
FSC threshold	0.143
Map resolution range (Å)	2.56–2.95
Refinement	
Initial model used (PDB code)	
Model resolution (Å)	2.57
FSC threshold	0.143
Model resolution range (Å)	2.56–2.58
Map sharpening *B* factor (Å^2)	−99.0
Model composition	
Non-hydrogen atoms	38,864
Protein residues	4,760
Ligands	0
B factors (Å^2)	
Proteins	162.67
R.m.s. deviations	
Bond lengths (Å)	0.004
Bond angles (°)	0.622
Validation	
MolProbity score	2.06
Clashscore	10.01
Poor rotamers (%)	3.07
Ramachandran plot	
Favored (%)	97.06
Allowed (%)	2.86
Disallowed (%)	0.09

Gad1 wraps around the GajAB complex using a network of homooligomeric interactions and notable conformational flexibility. On either side of the GajAB complex, four copies of Gad1 interlock into a tetrameric interface along the primary GajA-binding site (Figure1.7d). The Gad1 tetrameric interface is formed by hydrogen-bond interactions between the C-terminal 'shoulder' domain of each Gad1 monomer and a highly conserved set of three cysteine residues, C282, C284 and C285, which form disulfide interactions at an inter-subunit interface (Figure 1.7d,g and Figure 1.10a,d). The N terminus of each Gad1 monomer forms an 'arm' domain that extends out from the shoulder and reaches around the GajA nuclease active site to connect to a partnering Gad1 protomer from the opposite side of the complex. At the end of the Gad1 arm is an N-terminal 'fist' domain that allows two partnering Gad1 protomers to interact and complete the octameric web assembly (Figure 1.7c,h). Structural flexibility limits resolution in this portion of the cryo-EM map, but AlphaFold2 modelling[103,104] and rigid-body placement of the Gad1 N-terminal fist domain suggests that conserved hydrophobic residues around the Gad1 α1 helix mediate the fist–fist interactions (Figure 1.7h and Figure 1.10a). To fully encircle GajAB, Gad1 adopts two distinct structural conformations. Each pair of Gad1 proteins that wrap around and connect at the edge of the GajAB complex are formed by one Gad1 protomer reaching out from the shoulder with an arm domain extended straight down and one Gad1 protomer reaching out with an arm domain bent around 35 Å to the left (Figure 1.7i and Figure 1.10h). Sequence analysis of Gad1 proteins from phylogenetically diverse phages shows that the Gad1 N-terminal arm domain is highly variable in length (Figure 1.10a), providing further evidence that conformational flexibility in this region is crucial to inhibit host Gabija defence.

Figure 1.10 | Biochemical and structural characterization of the GajAB–Gad1 co-complex.

a, Structure-guided alignment of Gad1 proteins from indicated phage or prophage genomes colored according to amino acid conservation. The *Bacillus* phage Phi3T Gad1 secondary structure is displayed according to the two different conformations observed in the GajAB–Gad1 co-complex structure. Oligomerization interface residues are annotated according to the key below. **b,c,** Zoomed-in views of Gad1–GajA interface contacts including hydrophobic interactions in AlphaFold2 arm domain structure of Gad1 and the Toprim domain of GajA (b) and Gad1 shoulder domain residue Q244 interaction with GajA dimerization domain residue E277 (c). **d,** Zoomed-in view Gad1–Gad1 oligomerization interactions between shoulder domains of Gad1 protomers. **e,** SDS-PAGE analysis of the ability of *Shewanella* phage 1/4 Gad1 mutant proteins to interact with the GajAB complex. *Shewanella* phage 1/4 Gad1 mutant proteins were co-expressed with SUMO2-tagged GajAB (GajA-tagged) and co-purified by Ni-NTA pull-down. *Shewanella sp.* phage 1/4 Gad1 residues are numbered according to the Phi3T Gad1 structure. To measure high stringency of GajAB–Gad1 interactions, complexes were washed with a 1 M NaCl buffer prior to elution and co-purification. Notably, the Gad1 mutant C282E is no longer able to interact with GajAB *in vitro* under these stringent conditions, but retains the ability to disrupt Gabija defense *in vivo* suggesting that lower affinity interactions still occur. **f,** SDS-PAGE analysis of the ability of *Shewanella* phage 1/4 Gad1 fist–fist interface mutant proteins to interact with the GajAB complex. *Shewanella* phage 1/4 Gad1 mutant proteins were co-expressed with SUMO2-tagged GajAB (GajA-tagged), co-purified by Ni-NTA pull-down, and treated with SENP2 to cleave the SUMO2 tag prior to SDS-PAGE gel loading. *Shewanella sp.* phage 1/4 Gad1 residues are numbered according to the Phi3T Gad1 structure. **g,** Agarose gel analysis of the ability of GajAB–Gad1 mutant complexes to cleave target 56-bp dsDNA after a 1 min or 20 min incubation. **h,** Superposition of the GajAB crystal structure and GajAB from the GajAB–Gad1 cryo-EM structure demonstrates no significant conformational change upon Gad1 binding. Data in panels e, f, g are representative of 3 independent experiments.


a
Bacillus phage Phi3T
Shewanella sp. phage 1/4
Brevibacillus gelatini DSM 100115
Escherichia coli MS115-1
Bacillus xiamenensis VV3
Enterobacter roggenkampii IV
β1
β2
β3
α1
α2
α2'
α3
β4
β5
β5'
α4
β6
α5
α5'
β5''
Gad1–Gad1 disulfide interaction
Gad1–GajA loop
Gad1–Gad1 fist interaction
Gad1–GajA interactions
Gad1–GajA at dimerization interface
Gad1–Gad1 other interactions
b
AlphaFold2 Gad1
Modeled Residue
Y23
H563
c
Q244
E227
d
P181
P181
e
GajA–Gad1 Interface
Gad1–Gad1 Interface
Gad1:
WT
Y23A
F192R
R218A
Q244A
L102E
P181E
C282E
C285E
N-term
C-term
95 kDa
72
55
43
34
26
17
11
GajA
GajB
Gad1
f
Gad1 fist-fist Interface
Y75R
Y96R
Y97R
Y89R
Y103R
g
GajAB
min:
WT
F192R
R218A
C282E
766 bp
200
75
50
25
56 bp
32 bp
24 bp
h
GajAB
GajAB-Gad1
RMSD: 0.5 Å

To test the importance of individual GajAB–Gad1 interfaces, we next analysed a series of Gad1 substitution and truncation mutants for their ability to interact with GajAB and inhibit Gabija anti-phage defence. The Gad1 residue F192 is located between β4 and β5 and is part of a highly conserved region that forms the centre of the primary GajA–Gad1 interface (Figure 1.10a). The Gad1 substitution F192R blocked the ability of Gad1 to interact with GajAB *in vitro* and inhibit Gabjia anti-phage defence *in vivo* (Figure 1.7j and Figure 1.10a,e). However, individual mutations throughout the periphery were insufficient to disrupt Gad1 inhibition of Gabjia anti-phage defence. This shows that the large footprint of Gad1 is tolerant to small perturbations that might enable host resistance. Similarly, mutations to the conserved Gad1 cysteine residues in the tetrameric shoulder interface greatly reduced the stability of GajAB–Gad1 complex formation *in vitro* but only exhibited a threefold decrease and mostly still permitted Gad1 to block phage defence in *B. subtilis* cells (Figure 1.7j and Figure 1.10e). The formation of the GajAB–Gad1 complex was also disrupted *in vitro* by a Y103R mutation in the Gad1 fist–fist interface (Figure 1.7h and Figure 1.10f). Finally, in contrast to wild-type Gad1, expression of the Gad1 N-terminal fist–arm or C-terminal shoulder domains alone was unable to inhibit Gabija, providing evidence that full wrapping of Gad1 around the GajAB complex is necessary to enable phage evasion of anti-phage defence (Figure 1.7j and Figure 1.10e).

Superposition of the GajAB–Gad1 and GajAB complexes shows that Gad1 binding does not induce a notable conformational change in GajAB, and suggests that Gad1 instead functions through steric hindrance of Gabija anti-phage defence (Figure 1.10h). To define the mechanism of Gad1 inhibition of Gabija anti-phage defence, we modelled interactions between GajAB and target DNA. The GajA Toprim domain is structurally homologous to the *E. coli* protein MutS, which is involved in DNA repair[105]. Superimposing the MutS–DNA structure revealed positively charged patches lining a groove in the GajA Toprim domain that dips into the nuclease active site (Figure 1.11). Notably, the Gad1 arm domain directly occupies this putative DNA-binding surface, supporting a model in which the phage protein directly clashes with the path of target dsDNA (Figure 1.12a,b). To determine the effect of viral inhibition on GajAB catalytic function, we tested the role of Gad1 in individual steps of DNA binding and target DNA cleavage. Gad1 prevented GajAB from binding to target DNA and abolished all nuclease activity *in vitro* (Figure 1.12c,d). Gad1 proteins with F192R or C282E mutations were no longer able to inhibit DNA cleavage, in agreement with the inability of F192R-mutant proteins and the reduced ability of C282E-mutant proteins to block Gabija

defence *in vivo* and form stable GajAB–Gad1 complexes *in vitro* (Figure 1.10g). Together, these results show that phage Gad1 binds to and wraps around the GajAB complex to block target DNA degradation. Our findings reveal a complete mechanism by which phages evade the Gabija defence system of the host (Figure 1.12e).

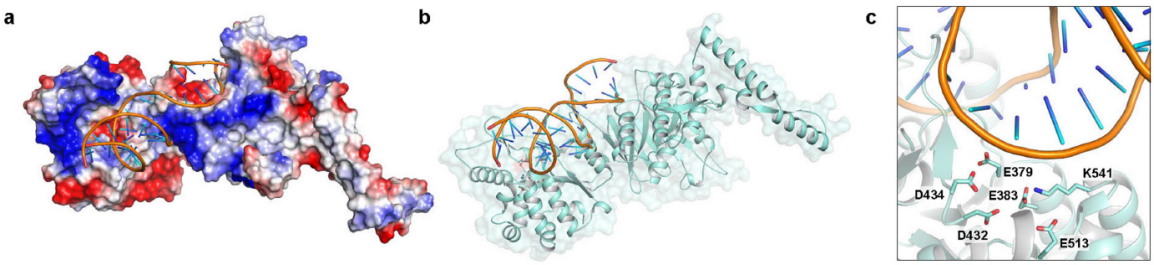


Figure 1.11 | Modeling DNA-bound GajA.

a,b, Isolated GajA protomer modeled with DNA bound to the Toprim domain shown with surface electrostatic potential (a) and in cartoon format (b). DNA modeling was performed using structural homology with the *E. coli* MutS–DNA complex (PDB ID 3K0S) (Lebbink et al. 2010). **c,** Zoomed-in view of the GajA Toprim active site with modeled DNA.

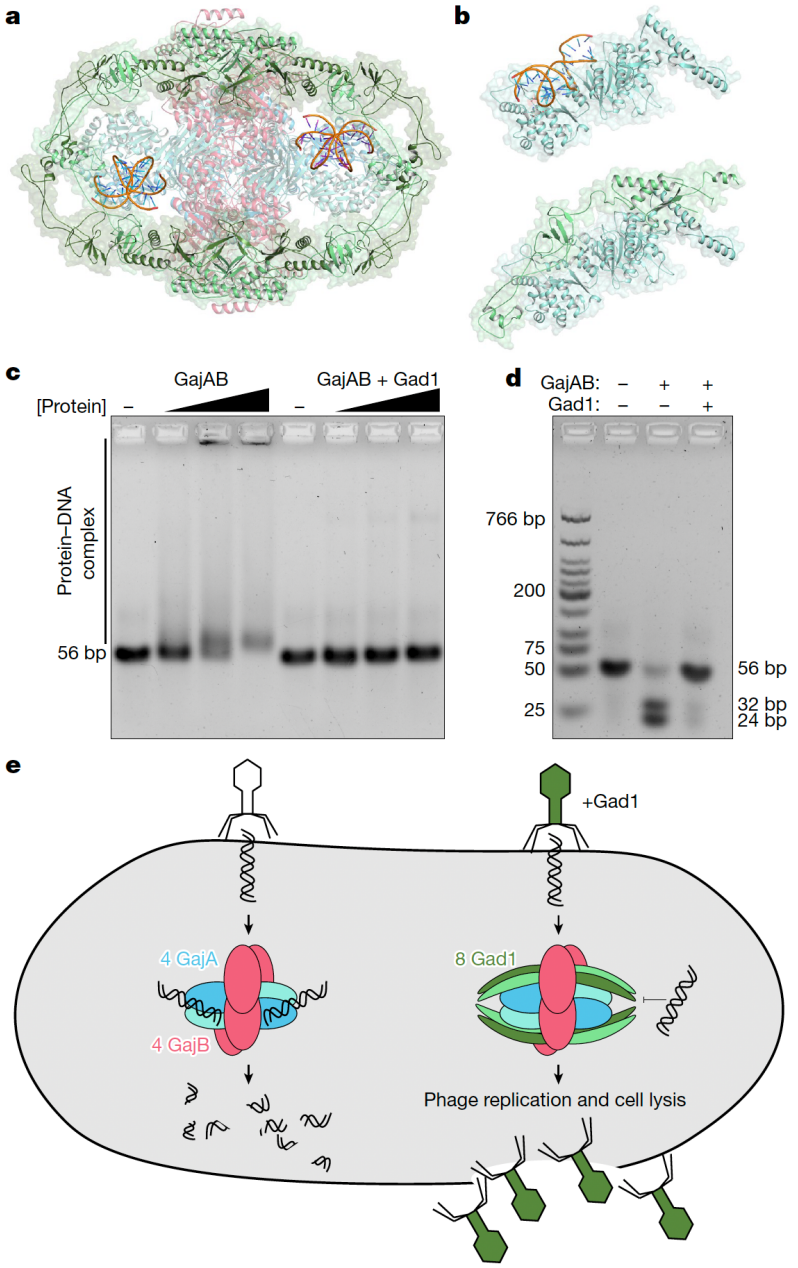


Figure 1.12 | Inhibition of Gabija DNA binding and cleavage enables viral evasion.

a, Cartoon representation of the GajAB–Gad1 co-complex structure with modeled DNA based on structural homology with *E. coli* MutS (PDB ID 3K0S). **b,** Isolated GajA protomer with modeled DNA bound to the Toprim domain (top) and same GajA promoter with Gad1 demonstrating significant steric clashes between Gad1 and the path of DNA (bottom). **c,d,** Biochemical analysis of GajAB 56-bp target DNA binding (c) and target cleavage (d) demonstrates that Gad1 potently inhibits GajAB activity. Data are representative of three independent experiments. **e,** Model of Gabija anti-phage defense and mechanism of Gad1 immune evasion.

Our study defines the structural basis of the formation of the Gabija supramolecular complex, and explains how phages block DNA cleavage to overcome this type of host immunity. The approximately 500-kDa GajAB complex expands an emerging theme in anti-phage defence, whereby protein subunits assemble into large machines to resist phage infection—similarly to the supramolecular complexes in CRISPR[106], CBASS[15,107] and RADAR[63,64] immunity. These results parallel human innate immunity, in which key effectors in inflammasome, Toll-like receptor, RIG-I-like receptor and cGAS–STING signalling pathways also oligomerize into large assemblies to block viral replication[2,108]. In contrast to the exceptionally large defence complexes of the host, phage evasion proteins are typically small, 5–20-kDa proteins that sterically occlude key protein binding and active-site motifs[100,101]. Breaking this rule, the 35-kDa anti-Gabija protein Gad1 is one of the largest described viral protein–protein inhibitors of host immune signalling (Figure 1.6). Whereas most viral evasion proteins that are larger than 20 kDa are enzymatic domains that catalytically modify target host factors or signalling molecules, the large size of Gad1 is necessary to bind to, oligomerize and encircle the entire host GajAB complex. Resistance to small phage proteins that simply block the GajA active site could explain why Gabija is a highly prevalent defence system in diverse bacterial phyla. A key question opened by our structures of the Gabija complex is how GajB helicase activity is linked to the activation of the GajA nuclease domain to control the cleavage of target DNA. Gad1 encasing the GajAB complex to trap it in an inactive state is a new mechanism by which phages evade host defences, and this finding provides a template to understand how viruses disrupt the complex mechanisms of activation of diverse anti-phage defence systems in bacteria.

Methods

Bacterial strains and phages

B. subtilis BEST7003 was grown in MMB (LB supplemented with 0.1 mM $MnCl_2$ and 5 mM $MgCl_2$) with or without 0.5% agar at 37°C or 30°C respectively. Whenever applicable, media were supplemented with ampicillin (100 μg mL^{-1}), chloramphenicol (34 μg mL^{-1}), or kanamycin (50 μg mL^{-1}) to ensure the maintenance of plasmids. *B. subtilis* phages phi3T (BGSCID 1L1) and SPβ (BGSCID 1L5) were obtained from the Bacillus Genetic Stock Center (BGSC). Prophages were induced using Mitomycin C (Sigma, M0503).

Phage titer was determined using the small drop plaque assay method[109] 400 μL of overnight culture of bacteria was mixed with 0.5% agar and 30 mL MMB and poured into a 10 cm^2 plate followed by incubation for 1 h at room temperature. In cases of bacteria expressing Gad1 homolog and Gad1 mutations, 0.1–1mM IPTG was added to the medium. 10-fold serial dilutions in MMB were performed for each of the tested phages and 10 μL drops were put on the bacterial layer. After the drops had dried up, the plates were inverted and incubated at room temperature overnight. Plaque forming units (PFUs) were determined by counting the derived plaques after overnight incubation and lysate titer was determined by calculating PFU mL^{-1}. When no individual plaques could not be identified, a faint lysis zone across the drop area was considered to be 10 plaques. Efficiency of plating (EOP) was measured by comparing plaque assay results on control bacteria and bacteria containing the defense system and/or a candidate anti-defense gene.

Plasmid Construction

For protein purification and biochemistry, *B. cereus* VD045 *GajA* (IMG ID 2519684552) and *GajB* (IMG ID 2519684553) genes were codon-optimized for expression in *E. coli* and synthesized as gBlocks (Integrated DNA Technologies) and cloned into custom pET vectors with an N-terminal 6×His-SUMO2 fusion tag (GajB alone) or a C-terminal 6×His tag (GajA alone). GajA and GajB proteins were co-expressed together using custom pET vector with an N-terminal 6×His-SUMO2 or N-terminal 6×His-SUMO2-5×GS tag on GajA and ribosome binding site between GajA and GajB. Phi3T and *Shewanella sp.* phage 1/4 Gad1 (IMG ID 2708680195) gBlocks were cloned into a custom pBAD vector containing a chloramphenicol resistance gene and IPTG-inducible promoter. For Gad1 pull-down assays, *Shewanella sp.* phage 1/4 Gad1 was

cloned with a ribosome binding site after the GajB gene in the N-terminal 6×His-SUMO2-5×GS GajAB plasmid.

For plaque assays, the DNA of Gad1 was amplified from phage phi3T genome using KAPA HiFi HotStart ReadyMix (Roche cat # KK2601). Since Gad1 was toxic in *B. subtilis* cells containing Gabija, *Shewanella sp.* phage 1/4 Gad1 was used and synthesized by Genscript. Gad1 and related homologs were cloned into the pSG-thrC-Phspank vector[18] and transformed to DH5α competent cells. The cloned vector and the vector containing Gad1 substitution and truncation mutants were subsequently transformed into *B. subtilis* BEST7003 cells containing Gabija integrated into the amyE locus[22], resulting in cultures expressing both Gabija and a Gad1 homolog. As a negative control, a transformant with an identical plasmid containing GFP instead of the anti-defense gene, was used. Transformation in *B. subtilis* was performed using MC medium as previously described[22]. Sanger sequencing was then applied to verify the integrity of the inserts and the mutations. The pSG1 plasmids containing point mutations in Gabija were constructed by restriction-enzyme subcloning Gabija sequence into pGEM9Z, site-directed mutagenesis as previously described[110], Gibson back into pSG1, and transformed into *B. subtilis* BEST7003 cells. Sanger sequencing of the mutations regions was applied to verify the mutations in Gabija.

Protein expression and purification

Recombinant GajAB and GajAB–Gad1 complexes were purified from *E. coli* as previously described[111]. Briefly, expression plasmids described above were transformed into BL21(DE3), BL21(DE3)-RIL (Agilent), or LOBSTR-BL21(DE3)-RIL cells (Kerafast), plated on MDG media plates (1.5% Bacto agar, 0.5% glucose, 25 mM Na_2HPO_4, 25 mM KH_2PO_4, 50 mM NH_4Cl, 5 mM Na_2SO_4, 0.25% aspartic acid, 2–50 μM trace metals, 100 μg mL^{-1} ampicillin, 34 μg mL^{-1} chloramphenicol) and grown overnight at 37°C. Five colonies were used to inoculate 30 mL of MDG starter overnight cultures (37°C 230 rpm). 10 mL of MDG starter cultures were then inoculated in 1 L M9ZB expression cultures (47.8 mM Na_2HPO_4, 22 mM KH_2PO_4, 18.7 mM NH_4Cl, 85.6 mM NaCl, 1% Cas-Amino acids, 0.5% glycerol, 2 mM $MgSO_4$, 2–50 μM trace metals, 100 μg mL^{-1} ampicillin, 34 μg mL^{-1} chloramphenicol) and induced with 0.5 mM IPTG after reaching an OD_{600} of ≥1.5 (overnight, 16°C, 230 rpm).

After overnight induction, cells were pelleted by centrifugation, resuspended, and lysed by sonication in 60 mL lysis buffer (20 mM HEPES pH 7.5, 400 mM NaCl, 10% glycerol, 20 mM Imidazole, 1

mM DTT). Lysate was clarified by centrifugation, and supernatant was poured over Ni-NTA resin (Qiagen). Resin was then washed with lysis buffer, lysis buffer supplemented to 1 M NaCl, lysis buffer again, and finally eluted with lysis buffer supplemented to 300 mM Imidazole. Samples were then dialyzed overnight in 14 kDa MWCO dialysis tubing (Ward's Science) with SUMO2-cleavage by hSENP2 as previously described[111]. hSENP2 did not efficiently cleave N-terminal 6×His-SUMO2-GajAB and the complex was therefore purified with an additional 5×GS linker. Proteins for crystallography and cryo-EM were dialyzed in dialysis buffer (20 mM HEPES-KOH pH 7.5, 250 mM KCl, and 1 mM DTT), purified by size exclusion chromatography using a 16/600 Superdex 200 column (Cytiva) and stored in gel filtration buffer (20 mM HEPES-KOH pH 7.5, 20 mM KCl, and 1 mM TCEP-KOH). Proteins for biochemical assays were dialyzed in dialysis buffer, purified by size exclusion chromatography using a 16/600 Superdex 200 column (Cytiva) or 16/600 Sephacryl 300 column (Cytiva) and stored in gel filtration buffer with 10% glycerol. Purified proteins were concentrated to >10 mg mL^{-1} using a 30 kDa MWCO centrifugal filter (Millipore Sigma), aliquoted, flash frozen in liquid nitrogen, and stored at −80°C.

Co-expression of Gabija with Phi3T Gad1 results in mild toxicity in *E. coli* grown on MDG media plates. No toxicity was observed using a closely related Gad1 homolog from the *Shewanella* phage 1/4. Biochemical analysis of Gabija–Gad1 interactions was therefore conducted with *Shewanella* phage 1/4 Gad1. Notably, all Gad1 residues analyzed are 100% conserved between Phi3T Gad1 and *Shewanella* phage 1/4 Gad1. For *Shewanella* phage 1/4 Gad1 pull-down assays, SUMO2-5×GS-GajA-GajB-Gad1 point mutant plasmids were transformed and expressed in BL21(DE3)-RIL or LOBSTR-BL21(DE3)-RIL cells and subject to Ni-NTA column chromatography and SUMO2 cleavage with SENP2. Gad1 pull-down was analyzed by SDS-PAGE and Coomassie Blue staining.

Crystallization and X-ray structure determination

Crystals were grown in hanging drop format using EasyXtal 15-well trays (NeXtal). Native GajAB crystals were grown at 18°C in 2 µL drops mixed 1:1 with purified protein (10 mg mL^{-1}, 20 mM HEPES 250 mM KCl, and 1 mM TCEP-KOH) and reservoir solution (100 mM HEPES-NaOH pH 7.5, 2.4% PEG-400, and 2.2 M ammonium sulfate). Crystals were grown for 7 days before cryo-protection with reservoir solution supplemented with 25% glycerol and harvested by plunging in liquid nitrogen. X-ray diffraction data were collected at the Advanced Photon Source (beamlines 24-ID-C and 24-ID-E). Data were processed using

the SSRL autoxds script (A. Gonzalez, Stanford SSRL). Experimental phase information was determined by molecular replacement using monomeric GajA and GajB AlphaFold2 predicted structures[103,104] in Phenix[112]. Model building was completed in Coot[113] and then refined in Phenix. The final structure was refined to stereochemistry statistics as reported in Extended Data Table 1. Structure images and figures were prepared in PyMOL.

Electrophoretic mobility shift assay

56-bp sequence-specific motif target dsDNA (5′ TTTTTTTTTTTTTTTTTAATAACCCGGTTATTTTTTTTTTTTTTTTTTTTTT 3′)[35] or scrambled dsDNA (5′ TTTTTTTTTTTTTTTTTGACATTACATTCAGTTTTTTTTTTTTTTTTTTTTTT 3′) was incubated with a final concentration of 2, 5, or 10 μM purified GajAB, GajA[E379A]–GajB, or GajAB–Gad1 complexes in 20 μL gel shift reactions containing 1 μM dsDNA, 5 mM $CaCl_2$, and 20 mM Tris-HCl pH 8.0 for 30 min at 4°C. 10 μL was then mixed with 2 μL of 50% glycerol and separated on a 2% TB (Tris-borate) agarose gel. The gel was then run at 250 V for 45 min, post-stained with TB containing 10 μg mL^{-1} ethidium bromide rocking at room temperature, de-stained in TB buffer for 40 min, and imaged on ChemiDoc MP Imaging System.

DNA cleavage assay

The same 56-bp dsDNA substrates as above were incubated with GajAB, GajA[E379A]–GajB, or GajAB–Gad1 complexes in a 20 μL DNA cleavage reaction buffer containing 1 μM dsDNA, 1 μM GajAB, GajA[E379A]–GajB, or GajAB–Gad1, 1 mM $MgCl_2$, 20 mM Tris-HCl pH 9.0 for 20 min at 37°C. Following incubation, reactions were stopped with DNA loading buffer containing 60 mM EDTA and 10 μL was analyzed on a 2% TB agarose gel, which was run at 250V for 45 min. The gel was then post-stained rocking at room temperature with TB buffer containing 10 μg mL^{-1} ethidium bromide, de-stained in TB buffer alone for 40 min, and imaged on a ChemiDoc MP Imaging System.

Cryo-EM sample preparation and data collection

For the SUMO2-GajAB–Gad1 co-complex sample, 3 μL of 1 mg mL^{-1} was vitrified using a Mark IV Vitrobot (Thermofisher). Prior to sample vitrification, 2/1 Carbon Quantfoil™ grids were glow discharged using an easiGlow™ (Pelco). Grids were then double-sided blotted for 9s, constant force of 0, 100% relative humidity chamber at 4°C, and a 10 s wait time prior to liquid ethane plunge and storage in liquid nitrogen.

GajAB–Gad1 co-complex cryo-EM grids were screened using a Talos Arctica microscope (Thermofisher) operating at 200 kV and the final map was collected on a Titan Krios microscope (ThermoFisher) operating at 300 kV. Both microscopes operated with a K3 direct electron detector (Gatan). SerialEM software version 3.8.6 was used for all data collection. For final data collection a total of 9,243 movies were taken at a pixel size of 0.3115 Å, a total dose of 41.1 e− /Å^2, dose per frame of 0.63 e− /Å^2 at a defocus range of −0.8 to −1.9 µm.

Cryo-EM data processing

SBGrid Consortium provided data-processing software[114]. Movies were imported into cryoSPARC[115] for patch-based motion correction, patch-based CTF estimation, 2D and 3D particle classification, and non-uniform refinement. cryoSPARC data processing is outlined in Figure 1.10. Briefly, after patch-based CTF estimation, five hundred micrographs were selected and autopicked using Blob Picker, which resulted in 625,295 particles after extracting from micrographs. 2D classifications were then used to generate 5 templates for Template Picker from which 110,654 particles were picked from 500 micrographs. After three more rounds of 2D classification 648,298 particles from all 9,243 micrographs were used in ab initios (K = 3), followed by heterogenous refinement. The best class with 573,410 particles was then used to go back and extract from all micrographs, which resulted in 570,485 particles used in a final 2D classification and Ab-initio. 351,193 particles from one Ab-initio class were used in non-uniform refinement along with Defocus and Global CTF refinement resulting in a 2.84 Å C1 symmetry and 2.57 Å D2 symmetry map, which was then used for model building.

Cryo-EM model building

Model building was performed in Coot[113] by manually docking AlphaFold2 predicted structures[103,104] as starting models and then manually completing refinement and model correction. To model the Gad1 fist domain, an AlphaFold2 model of the Gad1 arm–fist region was superimposed on the cryo-EM density of the manually built shoulder–arm region and then fit into density in Coot[113]. To complete the model for the sparse GajB density, the X-ray GajB structure was superimposed on the cryo-EM density. GajAB–Gad1 model was refined in Phenix[112], and the structure stereochemistry statistics are reported in Extended Data Table 2. Figures were prepared in PyMOL and UCSF ChimeraX[116].

Statistics and reproducibility

Experimental details regarding replicates are found within figure legends.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Coordinates and structure factors of the Gabija GajAB complex have been deposited in the PDB under the accession code 8SM3. Coordinates and density maps of the GajAB–Gad1 co-complex are deposited with the PDB and the Electron Microscopy Data Bank (EMDB) under the accession codes 8U7I and EMD-41983. All other data are available in the manuscript or Supplementary Fig. 1. Source data are provided with this paper.

Acknowledgements

We thank J. Grippen and members of the P.J.K. and R.S. laboratories for comments and discussion, and A. Lu for assistance with X-ray data collection. The work was funded by grants to P.J.K. from the Pew Biomedical Scholars program, the Burroughs Wellcome Fund PATH program, the Mathers Foundation, the Mark Foundation for Cancer Research, the Cancer Research Institute, the Parker Institute for Cancer Immunotherapy and the National Institutes of Health (1DP2GM146250-01), and by grants to R.S. from the European Research Council (ERC-AdG GA 101018520), the Israel Science Foundation (MAPATS grant 2720/22), the Ernest and Bonnie Beutler Research Program of Excellence in Genomic Medicine, the Deutsche Forschungsgemeinschaft (SPP 2330, grant 464312965) and the Knell Family Center for Microbiology. E.Y. is supported by the Clore Scholars Program and in part by the Israeli Council for Higher Education (CHE) via the Weizmann Data Science Research Center. A.G.J. is supported by a Life Science Research Foundation postdoctoral fellowship of the Open Philanthropy Project. X-ray data were collected at the Northeastern Collaborative Access Team beamlines 24-ID-C and 24-ID-E (P30 GM124165), and used a Pilatus detector (S10RR029205), an Eiger detector (S10OD021527) and the Argonne National Laboratory Advanced Photon Source (DE-AC02-06CH11357). Cryo-EM data were collected at the Harvard Cryo-EM Center for Structural Biology at Harvard Medical School. We thank T. Humphreys for help with cryo-EM data collection. Part of this research was supported by the NIH grant U24GM129547 and was performed at the Pacific Northwest Center for Cryo-EM at Oregon Health & Science University, with access through EMSL (grid.436923.9), a DOE Office of Science User Facility sponsored by the Office of Biological and Environmental Research.

Author Contributions

Sadie P. Antine: Designed and conceived study, performed protein purification and biochemical assays, performed X-ray crystallography, cryo-EM, and model building analysis, and designed figures and wrote manuscript.

Alex G. Johnson: Performed cryo-EM structural analysis.

Sarah E. Mooney: Performed protein purification and biochemical assays and prepared figures.

Azita Leavitt: Performed phage challenge assays, bioinformatics, and proteins sequence analysis.

Megan L. Mayer: Performed cryo-EM structural analysis.

Erez Yirmiya: Performed bioinformatics and protein sequence analysis.

Gil Amitai: Performed bioinformatics and protein sequence analysis.

Rotem Sorek: Performed bioinformatics and protein sequence analysis.

Philip J. Kranzusch: Designed and conceived the study, performed model building analysis, and wrote the manuscript.

Competing Interests

R.S. is a scientific cofounder and advisor of BiomX and Ecophage. The rest of the authors declare no competing interests.

Chapter 2: Phages overcome bacterial immunity via diverse anti-defense proteins

Erez Yirmiya[1,#], Azita Leavitt[1,#] , Allen Lu[2,3], Adelyn E. Ragucci[2,3], Carmel Avraham[1], Ilya Osterman[1], Jeremy Garb[1], Sadie P. Antine[2,3], Sarah E. Mooney[2,3], Samuel J. Hobbs[2,3], Philip J. Kranzusch[2,3,4], Gil Amitai[1,*], Rotem Sorek[1,*]

[1]Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 7610001, Israel

[2]Department of Microbiology, Harvard Medical School, Boston, MA, USA

[3]Department of Cancer Immunology and Virology, Dana-Farber Cancer Institute, Boston, MA, USA

[4]Parker Institute for Cancer Immunotherapy at Dana-Farber Cancer Institute, Boston, MA, USA

[#] These authors contributed equally to this work

The published manuscript including additional Supplementary Tables can be accessed as follows:

Nature. (2024) **625**, 352–359.
https://doi.org/10.1038/s41586-023-06869-w

Abstract

It was recently shown that bacteria use, apart from CRISPR–Cas and restriction systems, a considerable diversity of phage resistance systems[101,117–119], but it is largely unknown how phages cope with this multilayered bacterial immunity. Here we analysed groups of closely related *Bacillus* phages that showed differential sensitivity to bacterial defence systems, and discovered four distinct families of anti-defence proteins that inhibit the Gabija, Thoeris and Hachiman systems. We show that these proteins Gad1, Gad2, Tad2 and Had1 efficiently cancel the defensive activity when co-expressed with the respective defence system or introduced into phage genomes. Homologues of these anti-defence proteins are found in hundreds of phages that infect taxonomically diverse bacterial species. We show that the anti-Gabija protein Gad1 blocks the ability of the Gabija defence complex to cleave phage-derived DNA. Our data further reveal that the anti-Thoeris protein Tad2 is a ‘sponge’ that sequesters the immune signalling molecules produced by Thoeris TIR-domain proteins in response to phage infection. Our results demonstrate that phages encode an arsenal of anti-defence proteins that can disable a variety of bacterial defence mechanisms.

Introduction

The arms race between bacteria and their viruses has fuelled the evolution of defence systems that protect bacteria from phage infection[101,117–119]. Phages, in return, developed mechanisms that allow them to overcome bacterial defences[120]. Multiple phages were shown to encode anti-restriction proteins, which inhibit restriction-modification systems by direct binding to the restriction enzyme[121,122] or by masking restriction sites[123]. Phages are also known to encode many CRISPR–Cas inhibitors, which function through a variety of mechanisms including inhibition of CRISPR RNA loading[124], diversion of the CRISPR–Cas complex to bind non-specific DNA[125], prevention of target DNA binding or cleavage[126] and many additional mechanisms[100,127–129]. Phage proteins and non-coding RNAs that inhibit toxin–antitoxin-mediated defence have also been described[130].

Whereas early research focused on restriction modification and later on CRISPR–Cas as the main mechanisms of defence against phage, recent studies exposed dozens of previously unknown defence systems that are widespread among bacteria and archaea[9,22–24,118,131,132]. These systems mediate defence by using a plethora of molecular mechanisms, including small-molecule signalling[8,9,19,133–135], production of antiviral compounds[132,136] and reverse transcription of non-coding RNAs[131,137]. Recently, several phage proteins that inhibit the bacterial cyclic oligonucleotide-based antiphage signalling system (CBASS) and Thoeris defence systems have been described[18,76,81]. However, it is still mostly unknown whether and how phages can overcome the wide variety of newly reported defence systems. In the current study we use comparative genomics of closely related phages to discover four distinct families of phage proteins that inhibit the Gabija, Thoeris and Hachiman defence systems.

Results and Discussion

In a previous study, we demonstrated that analysis of genomically similar phages that exhibit differential sensitivity to bacterial immunity enabled the discovery of a phage protein, called Tad1, that inhibits the Thoeris defence system[18]. To examine whether this methodology could be used to systematically detect anti-defence proteins within phages, we isolated and analysed several groups of closely related phages, and tested their sensitivity to several defence systems that protect *Bacillus* species from phage infection[22] (Figure 2.1a). One group included eight newly isolated phages from the SPβ group, which also includes the previously isolated phages SPβ, phi3T and SPR[138–140]. These are temperate *Siphoviridae* phages with genomes of approximately 130 kilobases (kb) in length, with 43–96% alignable genomes when comparing phage pairs from this group (Figure 2.2a). A second group included six phages similar to the phage SPO1, a lytic *Myoviridae* phage with an approximately 130-kb-long genome[141]. More than 85% of the genome was alignable when comparing phage pairs from this group, with high (80–99%) sequence identity in alignable regions (Figure 2.1b). The third group of phages included eight previously isolated phages from the SBSphiJ group, which were reported in a recent study[18] (Figure 2.2b).

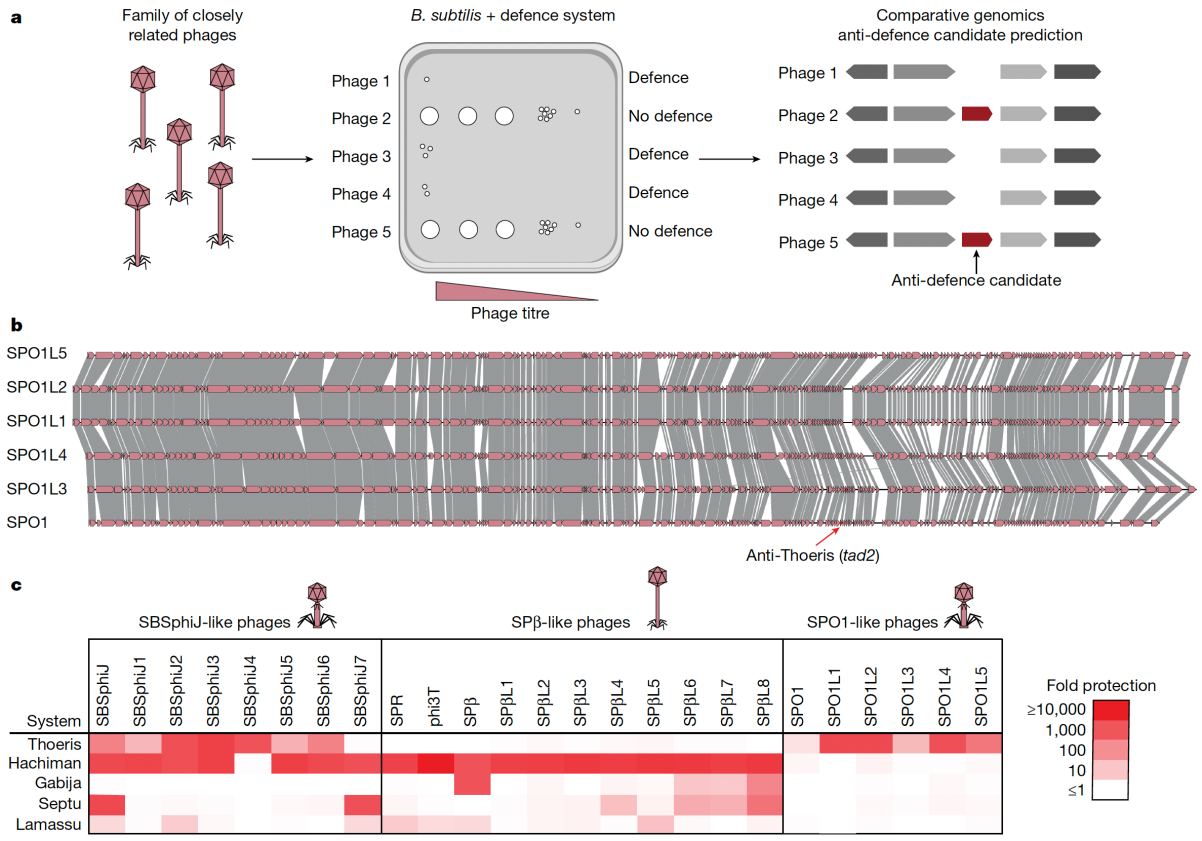


Figure 2.1 | Identification of anti-defense genes based on differential sensitivity to defense systems.

a, A flowchart of the experiments and analyses used in this study to detect candidate anti-defense genes. **b,** Genome comparison of six phages from the SPO1 group. Amino acid sequence similarity is marked by grey shading. Genome similarity was visualized using clinker (Gilchrist & Chooi 2021). **c,** Infection profile of SBSphiJ-like, SPbeta-like and SPO1-like phages infecting five *Bacillus subtilis* strains that express each of the defense systems Thoeris, Hachiman, Gabija, Septu and Lamassu. Fold defense was measured using serial dilution plaque assays, comparing the efficiency of plating (EOP) of phages on the system-containing strain to the EOP on a control strain that lacks the systems and contains an empty vector instead. Data represent an average of three replicates. Detailed data from individual plaque assays are found in Figure 2.3.

We next used this set of 25 phages to infect strains of *Bacillus subtilis* that expressed each of the defence systems described in Doron et al. 2018 as protecting against phages in *B. subtilis*. Five of these systems protected against at least one of the phages tested (Figure 2.1c). However, phages from the same group exhibited remarkably different properties when infecting defence-system-containing bacteria. For example, the Gabija defence system provided strong protection against the phages SPβ and SPβL8 but not against other phages from the SPβ group such as SPR and phi3T; and the Hachiman defence system provided defence against all phages of the SBSphiJ group except SBSphiJ4 (Figure 2.1c). To identify phage

genes that may explain the differential defence phenotype, we analysed the gene content in groups of phages that overcame each defence system and compared it to the gene content in phages that were blocked by the system. Genes common to phages that overcame the defence system, and not found in any of the phages that were blocked by the defence system, were considered as candidate anti-defence genes and were further examined experimentally.

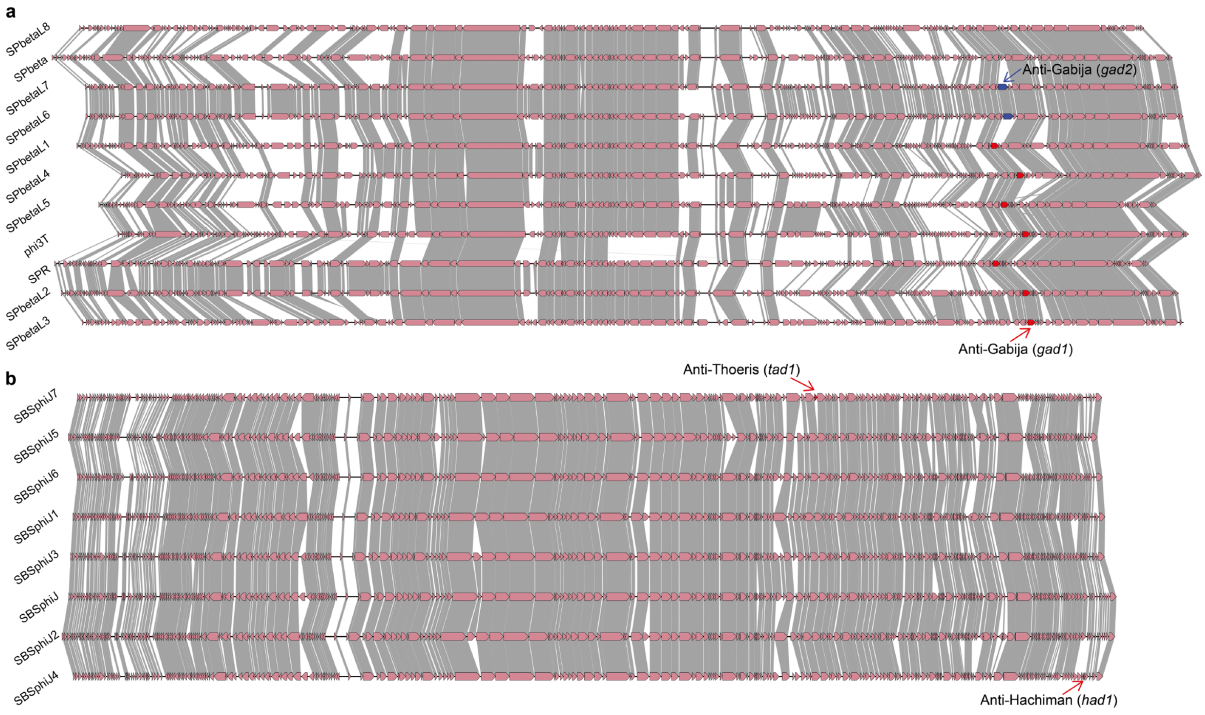


Figure 2.2 | Genome comparisons of phages from the SPbeta and SBSphiJ groups.

Genome comparison of (**a**) eleven phages from the SPbeta group and (**b**) eight phages from the SBSphiJ group. Amino acid sequence similarity is marked by grey shading. Genome similarity was visualized using clinker (Gilchrist & Choii 2021).

Gabija is a widespread bacterial defence system found in >15% of sequenced bacterial and archaeal genomes[20]. This system comprises two genes, *gajA* and *gajB*, which encode a DNA endonuclease and a UvrD-like helicase domain, respectively[22,35]. The Gabija system was shown to provide defence against a diverse set of phages[22,69]. The Gabija system from *Bacillus cereus* VD045, when cloned within *B. subtilis*, provided strong protection against some phages of the SPβ group including SPβ and SPβL8, and intermediate, weaker defence against the phages SPβL6 and SPβL7 (Figure 2.1c and Figure 2.3a). The remaining seven phages of the SPβ group were able to completely overcome Gabija-mediated defence.

We found two genes that were present in the seven Gabija-overcoming phages and missing from phages that were sensitive to Gabija defence (Figure 2.4a and Figure 2.2a). One of these genes (open reading frame 129 (ORF129)) did not show a Gabija-inhibiting phenotype when co-expressed with Gabija, and we were unable to clone the second gene (ORF128) into Gabija-expressing cells, presumably owing to toxicity. To examine the possible function of the non-cloned ORF128 gene, we knocked out that gene from the genome of the phage phi3T. Our results show that phi3T knocked out for ORF128 was no longer able to overcome Gabija defence, suggesting that this gene inhibits Gabija (Figure 2.4b). We denote the anti-Gabija gene *gad1* (Gabija anti-defence 1). Engineering gad1 from phi3T together with its native promoter into the genome of the phage SPβ, which naturally lacks this gene, rendered SPβ resistant to Gabija, confirming that *gad1* is both necessary and sufficient for the anti-Gabija phenotype (Figure 2.4b).

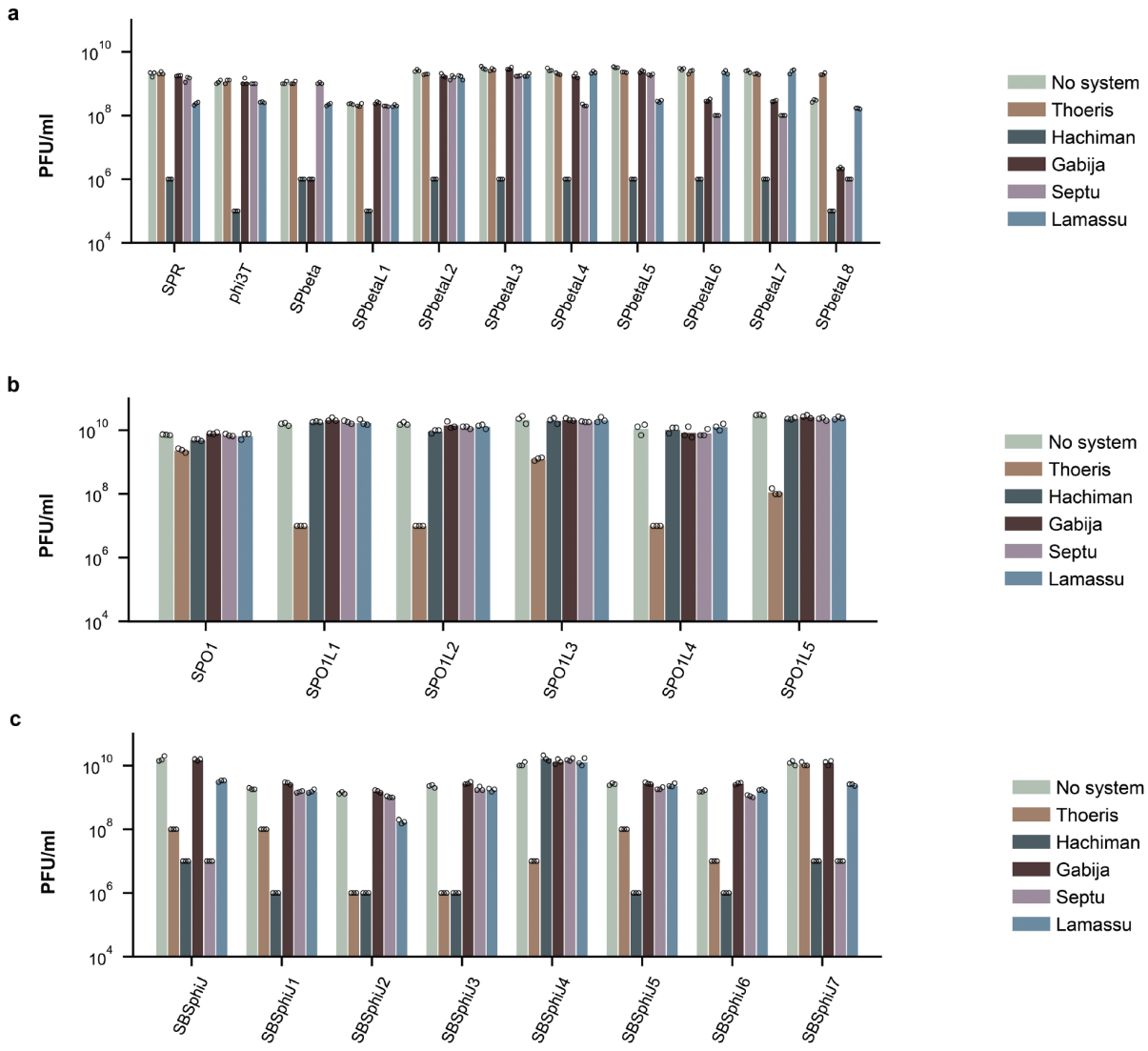


Figure 2.3 | Phages from the same family are differentially sensitive to bacterial defense systems.

Results of phage infection experiments with (**a**) eleven phages of the SPbeta group, (**b**) six phages of the SPO1 group, and (**c**) eight phages of the SBSphiJ group. Data represent plaque-forming units per ml (PFU/ml) of phages infecting control cells ("no system"), and cells expressing the respective defense systems. Shown is the average of three technical replicates, with individual data points overlaid. The Thoeris and Hachiman data presented here are the same as those presented in Figures 2.7b and 2.13b, respectively.

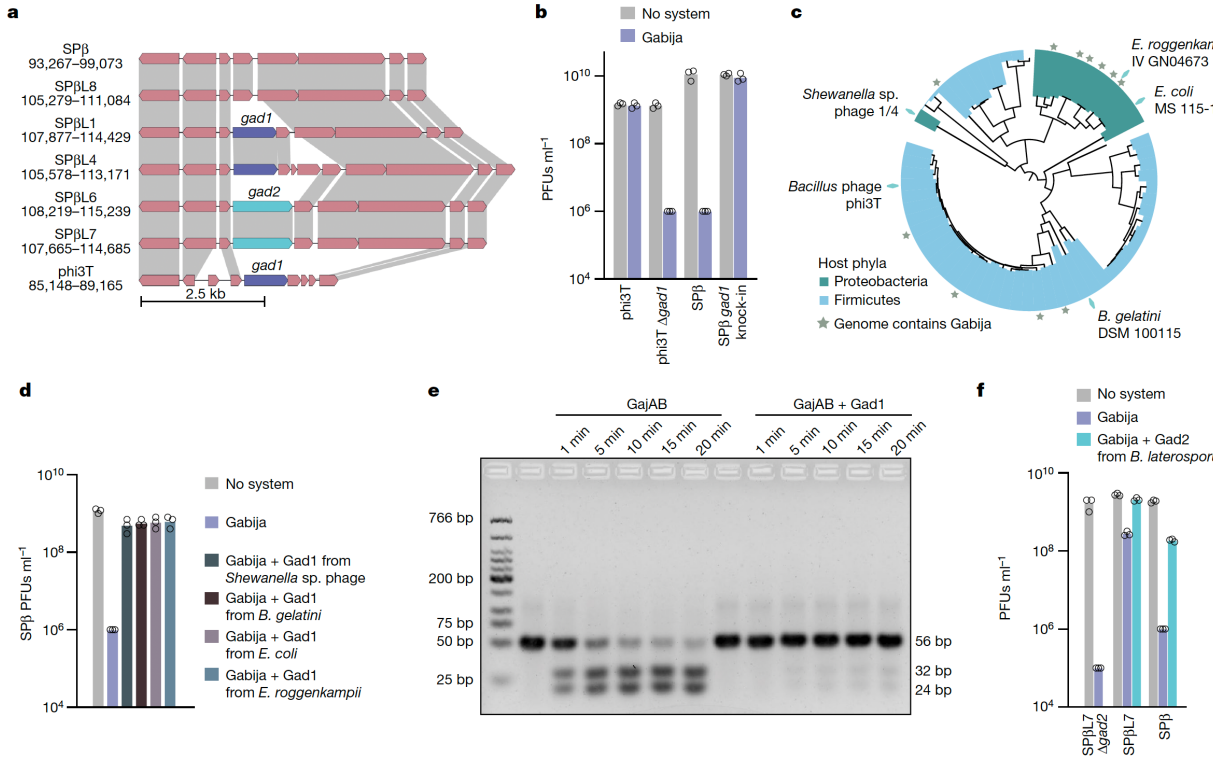


Figure 2. 4 | Gad1 and Gad2 proteins inhibit Gabija defense.

a, The anti-Gabija locus in phages of the SPbeta group. Amino acid sequence similarity is marked by grey shading. Genome similarity was visualized using Clinker (Gilchrist & Choii 2021). **b**, Deletion of *gad1* from phage phi3T eliminates the ability of the phage to cancel Gabija-mediated defense, while knock in of *gad1* into phage SPbeta renders the phage resistant to Gabija. Data represent plaque-forming units per ml (PFU/ml) of phages infecting control cells ("no system") and cells expressing the Gabija defense system. Shown is the average of three technical replicates, with individual data points overlaid. **c**, Phylogeny and distribution of Gad1 homologs. The names of bacteria in which Gad1 homologs were found in prophages and verified experimentally are indicated on the tree by cyan diamonds. **d**, Results of phage infection experiments. Data represent plaque-forming units per ml (PFU/ml) of SPbeta infecting control cells ("no system"), cells expressing the Gabija system ("Gabija"), and cells co-expressing the Gabija system and a Gad1 homolog. Shown is the average of three technical replicates, with individual data points overlaid. **e**, Gad1 blocks Gabija-mediated DNA cleavage. Incubation of purified Gabija (GajAB) complex, or Gabija co-purified with Gad1 from *Shewanella sp.* phage 1/4 (GajAB + Gad1) with a previously described DNA substrate from phage Lambda (Cheng et al. 2021). Shown is representative agarose gel from three independent experiments of proteins with 1, 5, 10, 15 or 20 min incubation with DNA. **f**, Knockout of *gad2* from phage SPbetaL7 renders the phage sensitive to Gabija defense, while expression of a Gad2 homolog from a *Brevibacillus laterosporus* prophage allows SPbeta to overcome Gabija-mediated defense. Phage infection experiments were conducted as in panel (**d**).

Gad1 is a 295-amino-acid (aa)-long protein, which does not exhibit sequence similarity to proteins of known function. We found 94 homologues of Gad1, distributed in genomes of various phages and prophages infecting host bacteria from the phyla Proteobacteria and Firmicutes (Figure 2.4c). Notably, many of the Gad1-containing prophages were integrated in bacterial genomes that also encoded the Gabija system, suggesting that Gad1 enabled these phages to overcome Gabija-mediated defence of their hosts (Figure 2.4c). Five Gad1 homologues from phages infecting *Shewanella sp.*, *Enterobacter roggenkampii*, *Escherichia coli*, *Brevibacillus gelatini* and *Bacillus xiamenensis* were selected for further experimental examination as representatives of the phylogenetic diversity of the Gad1 family (Figure 2.4c and Figure 2.5a). Unlike the Gad1 protein from the phage phi3T, four of the Gad1 homologues were not toxic when expressed in Gabija-containing cells (all except the homologue cloned from the *B. xiamenensis* prophage). Each of these four homologues efficiently inhibited Gabija defence when co-expressed in Gabija-containing cells (Figure 2.4d and Figure 2.5b).

It was recently shown that Gabija identifies and cleaves DNA having a specific nearly palindromic sequence motif derived from the phage lambda[35]. By purifying GajA and GajB and reconstituting the Gabija complex *in vitro*, we were able to confirm that Gabija cleaves DNA that contains the specific sequence motif (Figure 2.4e). Gabija purified in the presence of Gad1 was unable to cleave DNA (Figure 2.4e). In a companion paper, we show that Gad1 binds the Gabija complex as an octamer and inhibits its ability to bind and cleave DNA[37].

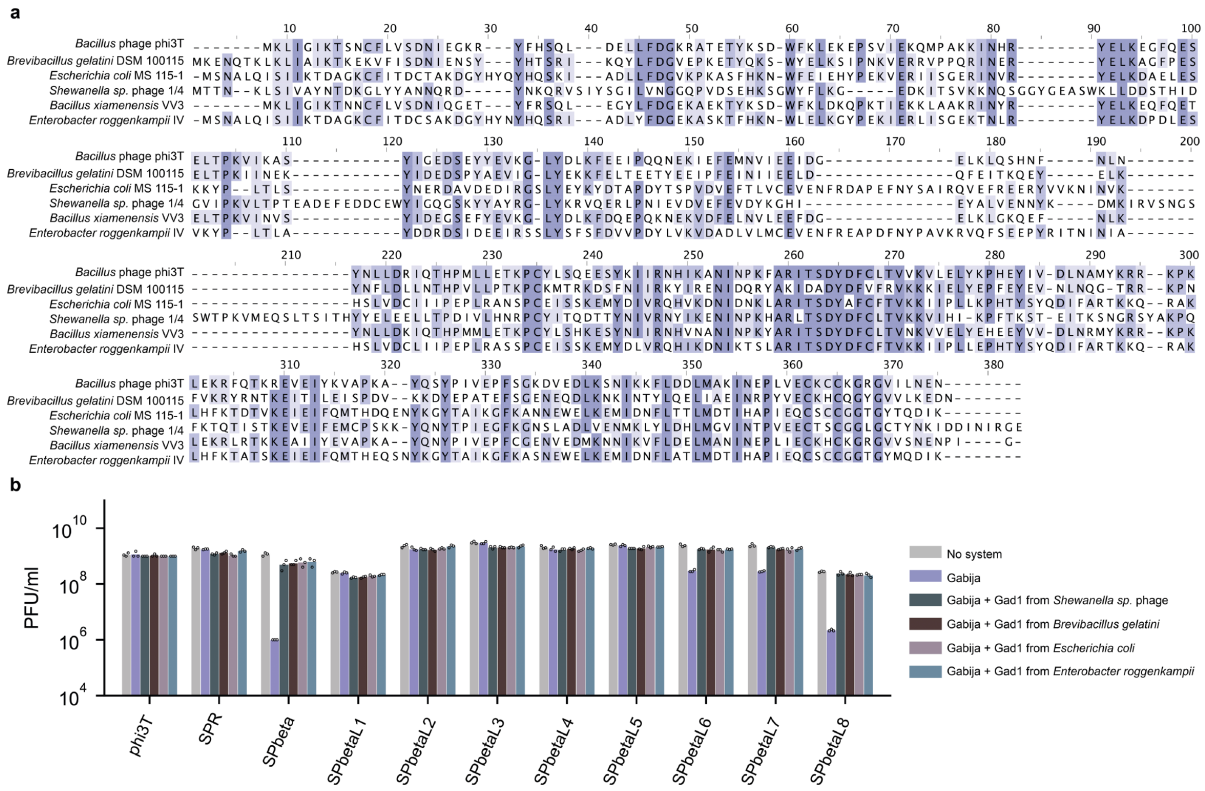


Figure 2.5 | Gad1 proteins inhibit Gabija mediated defense.

a, Multiple sequence alignment of the original Gad1 from phage phi3T and five Gad1 homologs that were chosen for experimental verification. Conserved residues are in purple. **b**, Results of phage infection experiments with eleven phages of the SPbeta group. Data represent plaque-forming units per ml (PFU/ml) of phages infecting control cells ("no system"), cells expressing the Gabija system ("Gabija"), and cells co-expressing the Gabija system and a Gad1 homolog. Shown is the average of three technical replicates, with individual data points overlaid. The SPbeta data presented here are the same as those presented in Figure 2.4d.

We next examined the phages SPβL6 and SPβL7, which lack gad1 but were still partially resistant to Gabija (Figure 2.1c). Notably, these phages encoded, at the same locus where *gad1* was encoded in other phages, another gene of unknown function. Knocking out this gene from the phage SPβL7 rendered this phage completely sensitive to Gabija (Figure 2.4f). The gene from SPβL7 was toxic when expressed in bacteria, but co-expression of a non-toxic homologue from a prophage of *Brevibacillus laterosporus* with Gabija completely inhibited Gabija defence, further verifying that it is an anti-Gabija gene, which we denote here as *gad2* (Figure 2.4a,f). Gad2 is a 400-aa-long protein that shows no sequence similarity to Gad1 or to other known proteins. Homology searches identified 170 homologues of Gad2 that almost always reside in genomes of phages and prophages infecting diverse host bacteria (Figure 2.6a). Notably, structural analysis using Alphafold2[103] predicted that Gad2 is an enzyme with a nucleotidyltransferase protein domain, suggesting that it inhibits Gabija through a mechanism of action different from that of Gad1 (Figure 2.6b). Point mutations in residues predicted to form the active site of the *B. laterosporus* Gad2 nucleotidyltransferase domain rendered Gad2 inactive, suggesting that nucleotidyltransferase activity is necessary for the anti-defence function (Figure 2.6c). Purified *B. laterosporus* Gad2 did not bind the Gabija complex *in vitro* and did not inhibit DNA cleavage, implying that Gad2 may operate upstream to Gabija to modify the phage molecules sensed by the GajA–GajB defence complex (Figure 2.6d–f). Combined together, our results suggest that SPβ-like phages encode anti-Gabija genes in a dedicated locus in their genomes, where multiple different Gabija-inhibiting genes can reside.

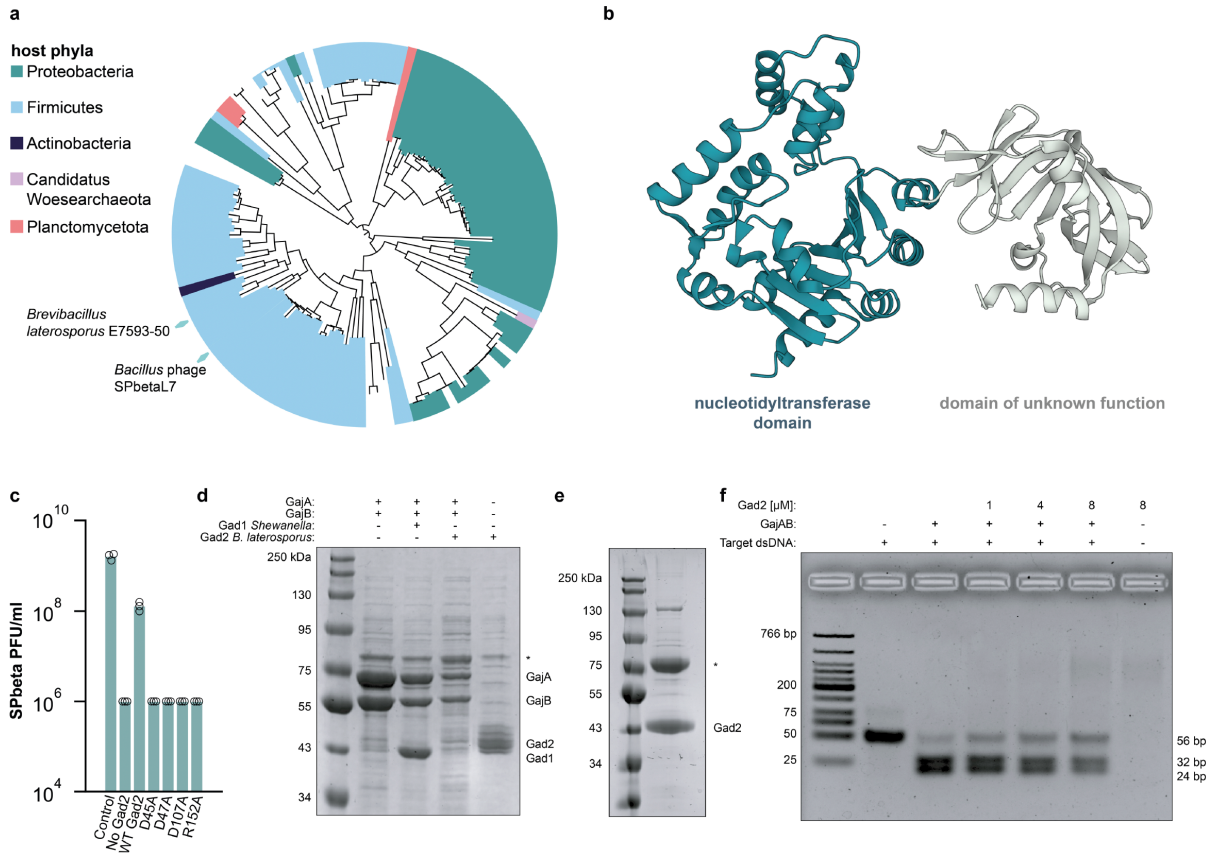


Figure 2.6 | Gad2 inhibits Gabija mediated defense.

a, Phylogeny and distribution of Gad2 homologs. Homologs that were tested experimentally are indicated on the tree by cyan diamonds. **b**, An Alphafold2 model for the structure of Gad2 from phage SPbetaL7. **c**, Mutations in the predicted nucleotidyltransferase active site in Gad2 result in loss of anti-defense activity. Data represent plaque-forming units per ml (PFU/ml) of phage SPbeta infecting cells co-expressing the Gabija system and WT or mutated Gad2 from *Brevibacillus laterosporus*, as well control cells expressing neither Gabija nor Gad2 ("Control") and cells expressing the Gabija system without Gad2 ("No Gad2"). Shown is the average of three technical replicates, with individual data points overlaid. **d**, SDS-PAGE analysis of Ni-NTA co-purified GajAB with *Shewanella* phage 1/4 Gad1 and *Brevibacillus laterosporus* Gad2 demonstrates that Gad2 does not stably interact with the GajAB complex. Asterisk indicates minor contamination with the *E. coli* protein ArnA. Data are representative of three independent experiments. **e**, SDS-PAGE analysis of purified *Brevibacillus laterosporus* Gad2. Asterisk indicates contamination with the *E. coli* protein ArnA. Data are representative of three independent experiments. **f**, Biochemical reconstitution of GajAB DNA degradation demonstrates that Gad2 does not directly inhibit GajAB cleavage of a 56-bp target DNA. Data are representative of three independent experiments.

The Thoeris defence system is present in approximately 4% of sequenced bacterial and archaeal genomes[22]. This system encodes ThsB, a protein with a Toll/interleukin-1 receptor (TIR) domain that serves as a sensor for phage infection, and ThsA, an NAD+-cleaving protein[142]. Following phage recognition, the Thoeris ThsB protein generates 1″–3′ gcADPR, a signalling molecule that activates ThsA, resulting in the depletion of NAD+ and the inhibition of phage replication[18]. Our recent discovery of the anti-Thoeris gene *tad1*, a 1″–3′ gcADPR sponge present in the phage SBSphiJ7 and absent from other phages in the SBSphiJ group, explains the observed insensitivity of SBSphiJ7 to Thoeris[18] (Figure 2.1c and Figure 2.3c). We reasoned that the phages SPO1 and SPO1L3 may also encode homologues of *tad1*, as these phages partially escaped Thoeris-mediated defence (Figure 2.1c and Figure 2.3b), but we were unable to identify *tad1* homologues in these phages. Instead, we identified a single gene present in SPO1 and SPO1L3 but absent from all other Thoeris-sensitive SPO1-like phages (Figure 2.7a). We expressed this gene, designated here as *tad2*, within *B. subtilis* cells that also express the Thoeris system from *B. cereus* MSX-D12. Tad2 robustly inhibited the activity of Thoeris, allowing Thoeris-sensitive phages to infect Thoeris-expressing cells (Figure 2.7b). Moreover, engineering t*ad2* into SBSphiJ, a phage that is normally blocked by Thoeris, resulted in a phage that overcomes Thoeris-mediated defence (Figure 2.7b). Silencing the expression of Tad2 in SPO1 using dead Cas9 (dCas9;[143]) further confirmed that Tad2 is responsible for the Thoeris-inhibiting phenotype of SPO1 (Figure 2.7c).

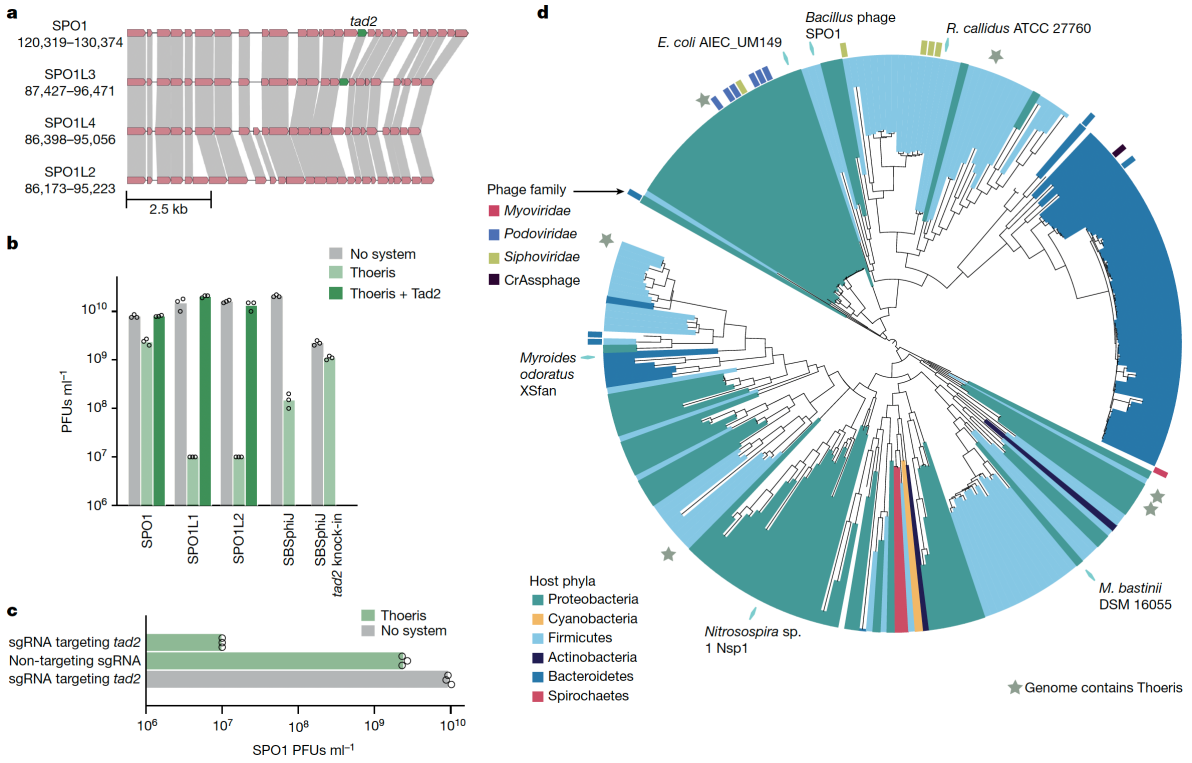


Figure 2.7 | Tad2 proteins inhibit Thoeris defense.

a, Genomic locus of the anti-Thoeris gene *tad2* in phages SPO1 and SPO1L3, and the respective locus in phages SPO1L4 and SPO1L2. Amino acid sequence similarity is marked by grey shading. Genome similarity was visualized using clinker (Gilchrist & Choii 2021). **b**, Anti-Thoeris activity of Tad2. Data represent plaque-forming units per milliliter (PFU/ml) of phages infecting control cells (no system), cells expressing the Thoeris system (Thoeris) and cells co-expressing the Thoeris system and the *tad2* gene from SPO1 (Thoeris + Tad2). Data for phage SBSphiJ as well as for SBSphiJ with a *tad2* knock in are also presented. Shown is the average of three technical replicates, with individual data points overlaid. **c**, Tad2 knockdown cancels anti-Thoeris activity. Data represent PFU/ml of SPO1 that infects cells expressing Thoeris and a dCas9 system targeting Tad2, as well as control cells. Shown is the average of three technical replicates, with individual data points overlaid. **d**, Phylogenetic analysis of Tad2 homologs in phage and prophage genomes. The names of bacteria in which Tad2 homologs were found in prophages and tested experimentally are indicated on the tree by cyan diamonds.

Tad2 is a short protein (89 aa) containing a DUF2829 protein domain. Sequence homology searches identified more than 650 homologues in the Integrated Microbial Genomes44 and Metagenomic Gut Virus45 databases. Phylogenetic analysis revealed that Tad2 is encoded by phages belonging to several phage morphology groups, including *Myoviridae*, *Podoviridae* and *Siphoviridae*, as well as by prophages integrated within more than 100 bacterial species from 6 different phyla (Figure 2.7d). We selected five Tad2 homologues representing the phylogenetic diversity of the family and cloned each one separately into *B. subtilis* cells expressing the Thoeris system (Figure 2.7d and Figure 2.8a). Four of these Tad2 homologues were able to inhibit Thoeris, including homologues derived from phages infecting distant organisms such as *Ruminococcus callidus* and *Maridesulfovibrio bastinii* (Figure 2.8b). These results demonstrate that Tad2 represents a large family of proteins used by phages to inhibit the Thoeris defence system.

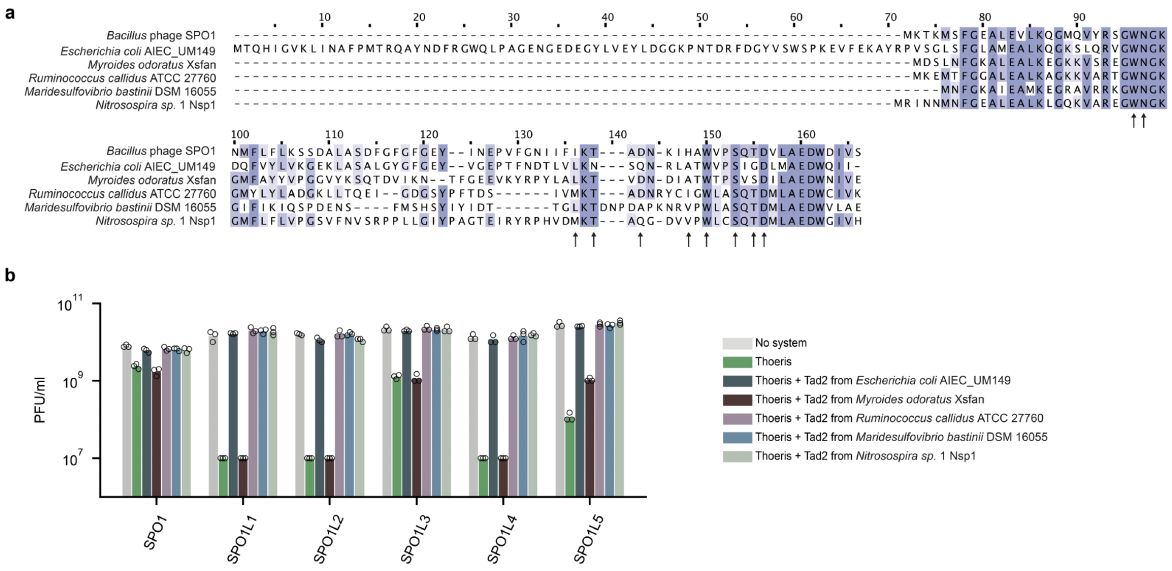


Figure 2.8 | Tad2 proteins inhibit Thoeris mediated defense.

a, Multiple sequence alignment of the original Tad2 from phage SPO1, and 5 Tad2 homologs that were chosen for experimental verification. Conserved residues are in purple. Black arrows indicate residues that are involved in 1″–3′ gcADPR binding. **b**, Results of phage infection experiments with six phages of the SPO1 group. Data represent plaque-forming units per ml (PFU/ml) of phages infecting control cells (“no system”), cells expressing the Thoeris system (“Thoeris”), and cells co-expressing the Thoeris system and a Tad2 homolog. Shown is the average of three technical replicates, with individual data points overlaid.

During phage infection, ThsB generates the 1″–3′ gcADPR signalling molecule to activate ThsA[18]. As expected, purified ThsA incubated with filtered cell lysates derived from SBSphiJ-infected, ThsB-expressing cells became a strong NADase, indicating that ThsB produced 1″–3′ gcADPR in response to SBSphiJ infection as previously shown[18] (Figure 2.9a). However, filtered lysates from similarly infected cells that co-expressed both ThsB and Tad2 failed to activate ThsA *in vitro*, suggesting that Tad2 functions upstream of ThsA (Figure 2.9a).

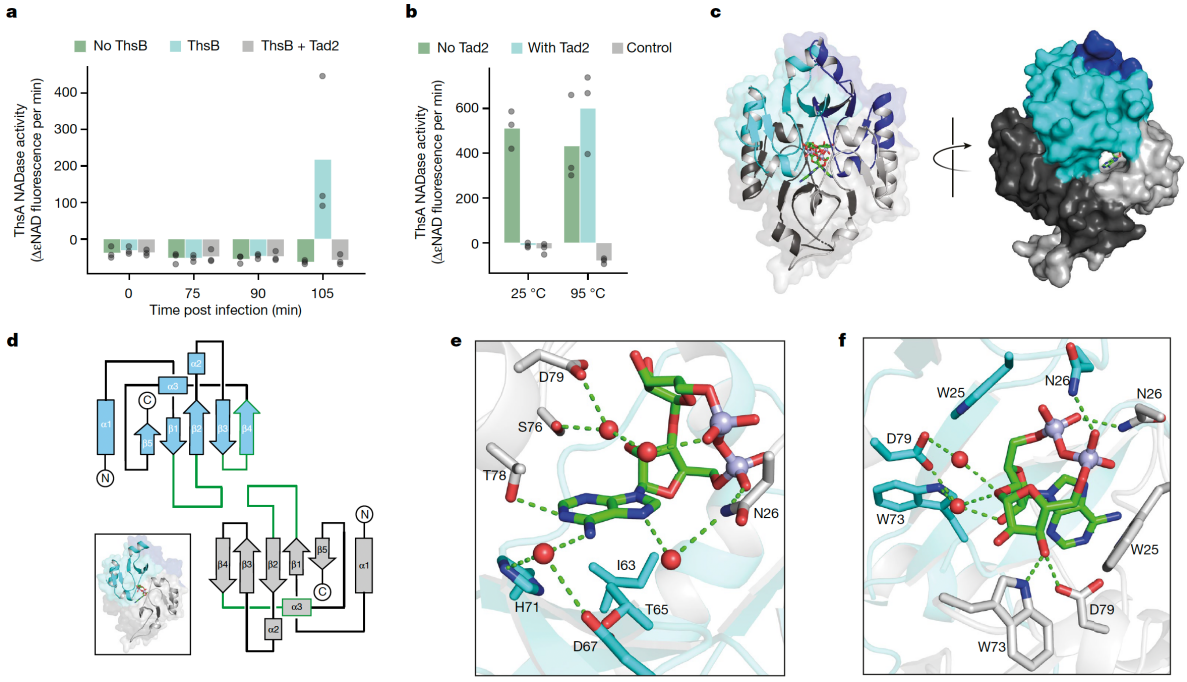


Figure 2.9 | Tad2 cancels Thoeris-mediated defense by sequestering 1″–3′ gcADPR.

a, Cells expressing ThsB, both ThsB and Tad2 or control cells that do not express ThsB were infected with phage SBSphiJ at a multiplicity of infection (MOI) of 10. NADase activity of purified ThsA incubated with filtered lysates was measured using a nicotinamide 1,N6-ethenoadenine dinucleotide (εNAD) cleavage fluorescence assay. Bars represent the mean of three experiments, with individual data points overlaid. **b**, Tad2 releases bound 1″–3′ gcADPR when denatured. Shown is the NADase activity of purified ThsA incubated with 600nM 1″–3′ gcADPR preincubated with 2.4 μM of purified Tad2 *in vitro* for 10 min, followed by an additional incubation of 10 min at either 25 °C or 95 °C. Control is buffer without 1″–3′ gcADPR. Bars represent the mean of three experiments, with individual data points overlaid. **c**, Overview of the crystal structure of Tad2 in complex with 1″–3′ gcADPR in cartoon (front) or surface (side) representation. Tad2 exists as a homotetramer formed by two dimer units (colored cyan/dark blue and grey/dark grey). Non-dimeric monomer pairs form two recessed ligand binding pockets that enclose 1″–3′ gcADPR. **d**, Topology map of two Tad2 monomers which come together to form the ligand binding pocket. Components that form the binding pocket are outlined in green. Each dimer subunit donates one monomer, as shown by the cartoon representation. **e**,**f** Detailed views centered around the adenine base (**e**) or ribose and phosphates (**f**) of residues that either directly interact with 1″–3′ gcADPR or coordinate key waters that reside within the binding pocket. Residues contributed by each of the two monomers that form the binding pocket are represented in cyan ($Tad2_a$) or grey ($Tad2_b$).

Recent studies have shown that phages can use dedicated proteins to degrade[81,144] or sequester[18,76] bacterial immune signalling molecules. Incubation of Tad2 with 1″–3′ gcADPR *in vitro* did not yield observable degradation products, suggesting that Tad2 is not an enzyme that cleaves 1″–3′ gcADPR (Figure 2.10a). To test whether Tad2 might act as a sponge that binds and sequesters the immune signal, we incubated purified Tad2 with 1″–3′ gcADPR, and then heated the reaction at 95 °C to denature Tad2. The denatured reaction readily activated ThsA, suggesting that Tad2 functions by binding and sequestering the Thoeris-derived signalling molecule, and that denaturation of Tad2 released the intact molecule back to the medium (Figure 2.9b). In support of this observation, our size-exclusion chromatography results showed that Tad2 that was pre-incubated with 1″–3′ gcADPR exhibited a substantial mobility shift (Figure 2.10b). In addition, Tad2 exhibited an increased absorption ratio of UV_{260nm}/UV_{280nm} following incubation with 1″–3′ gcADPR, further confirming that Tad2 binds this molecule as a ligand (Figure 2.10b). We found that Tad2 binds 1″–3′ gcADPR with an equilibrium dissociation constant (K_D) of 23.3 nM (Figure 2.10c). Notably, Tad2 could not bind the canonical cADPR, demonstrating high specificity of Tad2 to the ThsB-derived signalling molecule (Figure 2.10d).

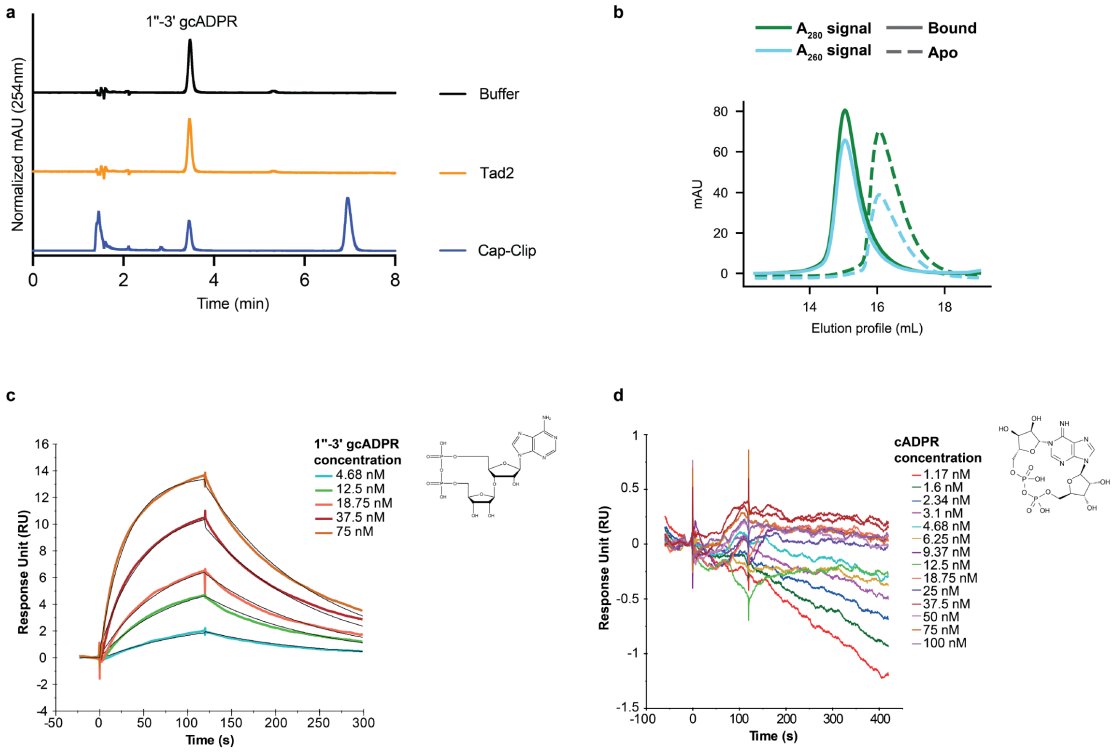


Figure 2.10 | Tad2 binds 1″–3′ gcADPR.

a, Incubation of Tad2 with 1″–3′ gcADPR *in vitro* does not yield observable degradation products. Representative HPLC traces of 1″–3′ gcADPR incubated with buffer, Tad2, or with the enzyme Cap-Clip known to cleave diphosphate linkages as a positive control. **b**, Size-exclusion chromatography of 1″–3′ gcADPR-bound or apo state Tad2. 1″–3′ gcADPR-bound Tad2 shows a substantial shift compared to Tad2 in the apo state. **c**, Surface plasmon resonance binding sensorgrams for Tad2 at five concentrations of 1″–3′ gcADPR. The black lines are the global fits using the instrument's evaluation software. k_a = 3.42E+05 ± 5.2E+02 (1/Ms), k_d = 0.00798 ± 1E-05 (1/s). **d**, Surface plasmon resonance binding sensorgrams for Tad2 at multiple concentrations of cADPR.

To define the mechanism of 1″–3′ gcADPR sequestration and Thoeris inhibition, we determined crystal structures of Tad2 from the phage SPO1 in the apo and 1″–3′ gcADPR ligand-bound states (Data Table 2.1). We also determined crystal structures of Tad2 bound to 1″–2′ gcADPR and Tad1 from a prophage of *Clostridium botulinum* bound to 1″–3′ gcADPR (Data Table 2.1). The structures of Tad2 reveal a homotetrameric assembly consistent with oligomerization observed during size-exclusion chromatography analysis (Figure 2.11). Two Tad2 protomers pack together at helix α2 and sheet β2 to form V-shaped homodimeric units, which then interlock perpendicularly along helix α1 to complete tetramerization (Figure 2.9c and Figure 2.12). The resulting assembly creates two identical ligand-binding pockets formed at the interface of two adjacent non-dimeric protomers, surrounded by loop β1–β2, loop β3–β4 and strand β4 of one protomer, and loop β1–β2, loop β4–α3 and helix α3 of the other (Figure 2.9c–f).

Table 2.1: Summary of crystallography data collection, phasing, and refinement statistics.

	cbTAD1–1″–3′ gcADPR **(8SMD)**	SPO1 TAD2–apo **(8SME)**	SPO1 TAD2–1″–3′ gcADPR **(8SMF)**	SPO1 TAD2–1″–2′ gcADPR **(8SMG)**	Had1 **(8TTO)**
Data collection					
Space group	P 2_1 3	P 6_3 2 2	C 1 2 1	P 6_3 2 2	P 1
Cell dimensions					
a, *b*, *c* (Å)	95.97, 95.97, 95.97	108.40, 108.40, 74.26	94.39, 82.47, 90.76	108.06, 108.06, 72.36	30.94, 40.18, 45.29
α, β, γ (°)	90.00, 90.00, 90.00	90.00, 90.00, 120.00	90.00, 107.88, 90.00	90.00, 90.00, 120.00	75.22, 89.42, 74.91
Resolution (Å)	42.92–2.10 (2.16–2.10)	46.94–2.36 (2.45–2.36)	46.06–1.75 (1.78–1.75)	39.29–2.10 (2.16–2.10)	37.45-2.00 (2.05-2.00)
R_{pim}	2.5 (69.8)	1.9 (81.5)	5.7 (63.0)	2.2 (71.9)	0.079 (0.684)
$I / \sigma(I)$	12.2 (1.0)	19.6 (1.5)	6.5 (1.1)	17.3 (1.5)	4.9 (1.0)
Completeness (%)	99.7 (96.3)	100.0 (100.0)	98.8 (97.1)	100.0 (99.8)	97.3 (95.2)
Redundancy	13.5 (13.3)	27.0 (26.7)	5.2 (4.7)	27.5 (26.8)	1.8 (1.8)
Refinement					
Resolution (Å)	42.92–2.10	46.94–2.36	46.06–1.75	39.29–2.10	37.45-2.00
No. reflections					
Total	236478	298974	344731	414196	24521
Unique	17522	11073	65939	15065	13372
Free	876	1103	1994	1503	1327
R_{work} / R_{free}	20.12 / 23.61	26.29/ 29.87	20.77/ 24.14	22.39/ 24.99	21.53 / 25.54
No. atoms					
Protein	1984 (2 copies)	1281	5054 (8 copies)	1284 (2 copies)	1533
Ligand / ion	70	–	142	35	–
Water	60	2	599	27	172
B-factors					
Protein	61.50	86.46	24.43	72.56	31.07
Ligand / ion	55.31	–	16.62	58.93	–
Water	55.46	59.19	34.43	59.70	36.04
R.m.s. deviations					
Bond lengths (Å)	0.001	0.002	0.004	0.002	0.001
Bond angles (°)	0.433	0.434	0.795	0.463	0.41

*All datasets were collected from individual crystals. *Values in parentheses are for the highest resolution shell.

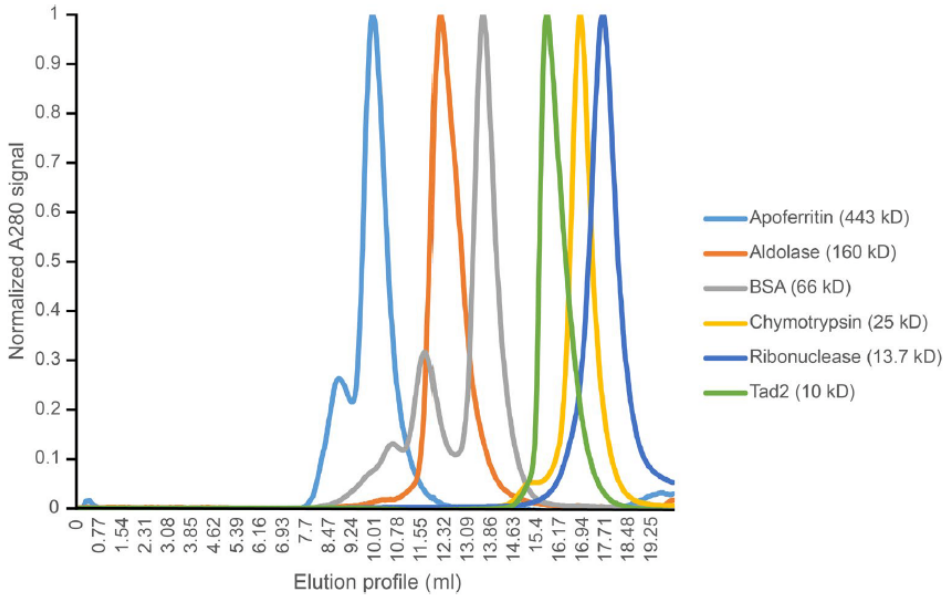


Figure 2.11 | Size-exclusion chromatography of Tad2 and various standards. Observed peak demonstrates that Tad2 forms a homomultimer.

A 1.75 Å structure of Tad2 in complex with 1″–3′-gcADPR explains the molecular basis of signal recognition. In the Tad2–1″–3′-gcADPR complex, ligand binding is mediated by extensive van der Waals interactions from $W25_{a,b}$ and $W73_{a,b}$, as well as polar interactions to phosphates and free hydroxyls from $N26_{a,b}$ and D79a,b (Figure 2.9e,f). Compared to the Tad1–1″–3′-gcADPR complex, Tad2 forms very few contacts to the adenine base of 1″–3′ gcADPR, with one hydrogen bond from $T78_b$ and nonpolar interactions from $I63_a$ and $T65_a$, whereas Tad1 forms base-stacking interactions via $F82_b$ and $R109_a$, as well as specific hydrogen bonding to the Hoogsteen edge via $N92_b$ (Figure 2.9e,f and Figure 2.12). Additionally, we observed several key water molecules in the Tad2 structure that contribute to 1″–3′ gcADPR binding and are coordinated by $N26_b$, $D67_a$, $H71_a$, $W73_b$, $S76_b$ and $D79_{a,b}$ (Figure 2.9e,f). Tad2 binds both 1″–3′ gcADPR and 1″–2′ gcADPR via the same binding residues, and both molecules are oriented nearly identically in the Tad2 binding pocket (Figure 2.12b,c). Although the molecular basis for ligand recognition is dissimilar between Tad2 and Tad1, we observe that Tad1, too, binds both 1″–3′ gcADPR and 1″–2′ gcADPR in the same pocket and orientation (Figure 2.12e,f), highlighting how both proteins function as sponges for gcADPR molecules. Taken together, our findings show that Tad2 inhibits Thoeris defence by binding and sequestering 1″–3′ gcADPR, preventing the activation of the Thoeris immune effector and mitigating Thoeris-mediated defence.

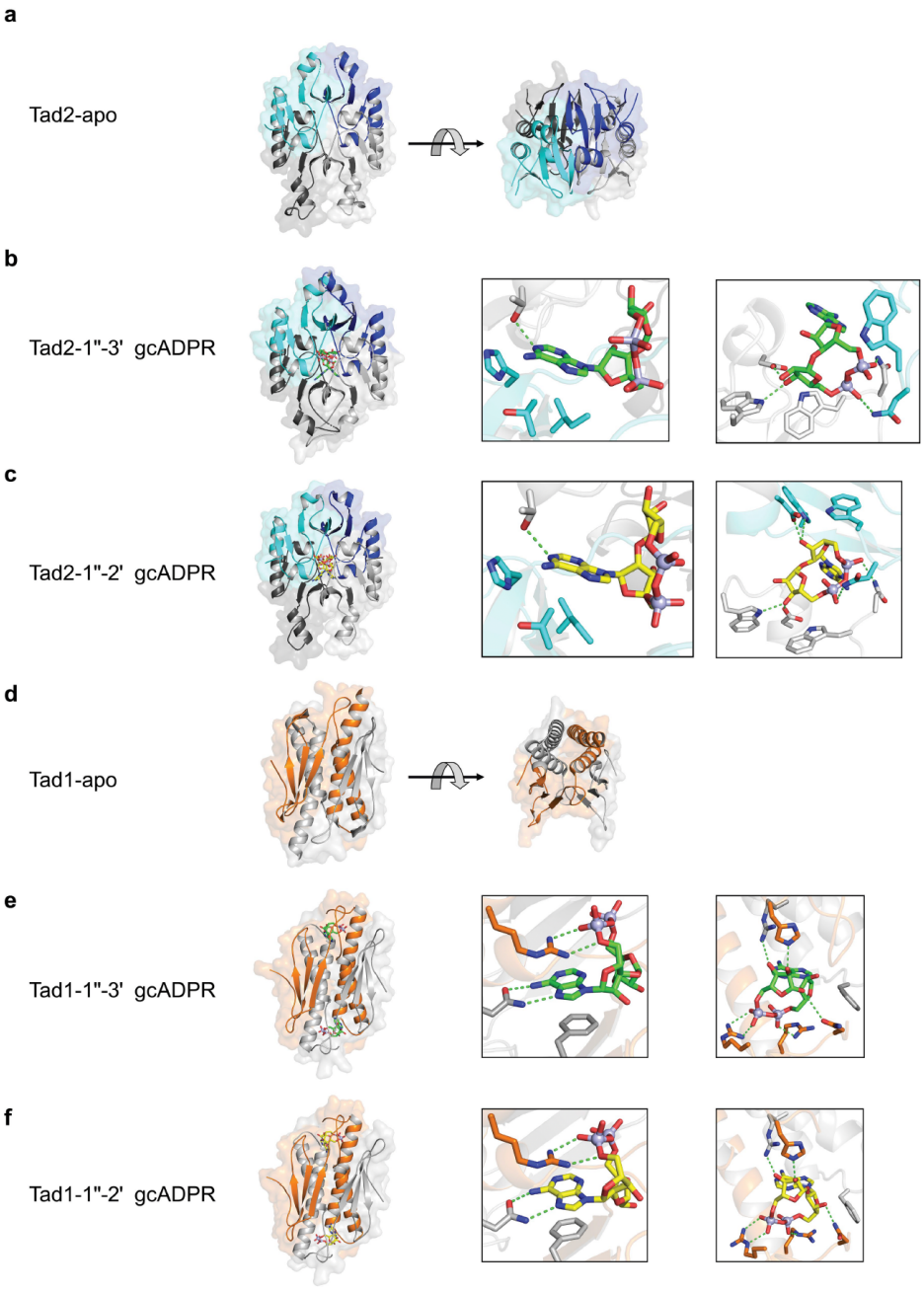


Figure 2.12 | Comparison of Tad2 and Tad1 in the apo and ligand-bound states.

a, Overview of the crystal structure of SPO1 Tad2 in the apo state in front and top view. **b**,**c**, Overview and detailed binding pocket views of adenine interactions (left) and ribose/phosphate interactions (right) of the crystal structures of SPO1 Tad2 in complex with 1″–3′ gcADPR **b**, or 1″–2′ gcADPR (**c**). **d**, Overview of the crystal structure (PDB: 7UAV) of cbTad1 in the apo state in front view and top view. **e**,**f** Overview and detailed binding pocket views of adenine interactions (left) and ribose/phosphate interactions (right) of the crystal structures of cbTad1 in complex with 1″–3′ gcADPR (**e**) or 1″–2′ gcADPR (**f**, PDB: 7UAW).

Hachiman is a defence system whose mechanism of action remains unsolved. It encodes a protein with a predicted helicase domain, as well as an additional protein with no known functional domains. Hachiman is found in more than 3% of sequenced bacterial and archaeal genomes, and was shown to provide strong protection against a broad range of phages[22].

We cloned three short genes that were unique to the phage SBSphiJ4, a phage that overcame Hachiman-mediated defence (Figure 2.1c), into a *B. subtilis* strain that expresses the Hachiman system from *B. cereus* B4087[22]. One of these genes completely abolished Hachiman-mediated defence, and we therefore named it *had1* (Hachiman anti-defence 1) (Figure 2.13a). Silencing of Had1 expression in SBSphiJ4 resulted in a phage that could infect control strains, but was blocked by Hachiman defence (Figure 2.13c). In addition, SBSphiJ that was engineered to include Had1 with its native promoter was able to overcome Hachiman defence, demonstrating that Had1 is responsible for the anti-Hachiman phenotype (Figure 2.13b).

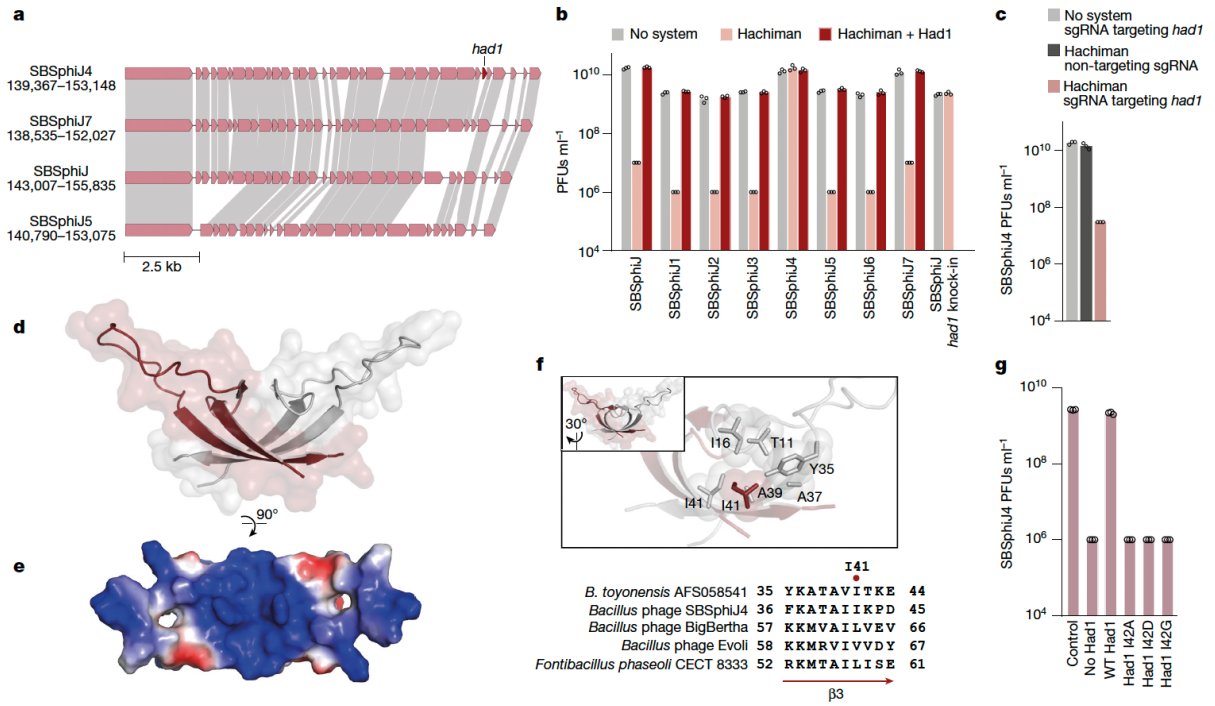


Figure 2.13 | Had1 proteins inhibit Hachiman defense.

a, Genomic locus of the anti-Hachiman gene *had1* in phages SBSphiJ4, as well as the relevant locus in phage SBSphiJ7, SBSphiJ and SBSphiJ5. Amino acid sequence similarity is marked by grey shading. Genome similarity was visualized using clinker (Gilchrist & Choii 2021). **b**, Differential defense of Hachiman against phages from the SBSphiJ group, and anti-Hachiman activity of Had1. Data represent plaque-forming units per milliliter (PFU/ml) of phages infecting control cells (no system), cells expressing the Hachiman system (Hachiman) and cells co-expressing the Hachiman system and the *had1* gene from SBSphiJ4 (Hachiman + Had1). Data for phage SBSphiJ with a *had1* knock-in are also presented. Shown is the average of three technical replicates, with individual data points overlaid. **c**, Had1 knockdown cancels anti-Hachiman activity. Results of phage SBSphiJ4 infection experiments. Data represent plaque-forming units per ml (PFU/ml) of SBSphiJ4 infecting cells expressing Hachiman and a dCas9 with an sgRNA targeting Had1, as well as control cells. Shown is the average of three technical replicates, with individual data points overlaid. **d**,**e**, *Bacillus toyonensis* Had1 structure overview and surface electrostatics. Had1 is a homodimer (colored red/light grey) with a central β-barrel core formed by strands β1–β3 and splayed loops that create a complex with an overall platform-like shape. **f**, Zoomed in cutaway and sequence alignment of *B. toyonensis* Had1 residue I41 that forms hydrophobic packing interactions with T11, I16, Y35, A37, and A39 from the opposing protomer. **g**, Had1 mutations at the center of the dimeric interface result in loss of anti-defense activity. Data represent plaque-forming units per ml (PFU/ml) of phage SBSphiJ infecting cells that co-express the Hachiman system and WT or mutated Had1 from phage SBSphiJ4, as well as control cells lacking Had1 and control cells lacking both Had1 and Hachiman. Shown is the average of three technical replicates, with individual data points overlaid.

Had1 is a short protein of 53 aa in size, which does not show sequence homology to any protein of known function. We found 23 homologues of Had1 in *Bacillus* phages as well as prophages integrated within *Bacillus* and *Paenibacillus* genomes, and selected five homologues that span the protein sequence diversity of Had1 for further experimental examination. Four of these proteins efficiently inhibited the activity of Hachiman, and we could not clone the fifth into Hachiman-expressing cells, possibly owing to toxicity (Figure 2.14a,b). These results confirm that Had1 is a Hachiman-inhibiting family of phage proteins.

We determined the crystal structure of Had1 from *Bacillus toyonensis*, revealing a homodimeric complex with two splayed loops that form an overall platform-like shape (Figure 2.13b and Figure 2.14c,d). The Had1 dimeric interface consists of four β-strands, two contributed by each protomer (Figure 2.13d). At the centre of the Had1 complex platform there is a positively charged patch formed by conserved residues in strand β2 including R17 and K18 (Figure 2.13e). A Had1 residue I41 at the centre of the dimeric interface in each protomer forms hydrophobic packing interactions with five residues on the other protomer, supporting the dimerization interface (Figure 2.13f). Substitution of the equivalent isoleucine residue in SBSphiJ4 Had1 rendered the protein unable to inhibit Hachiman, suggesting that the dimeric structure is necessary for Had1 evasion of host antiphage defence (Figure 2.13g).

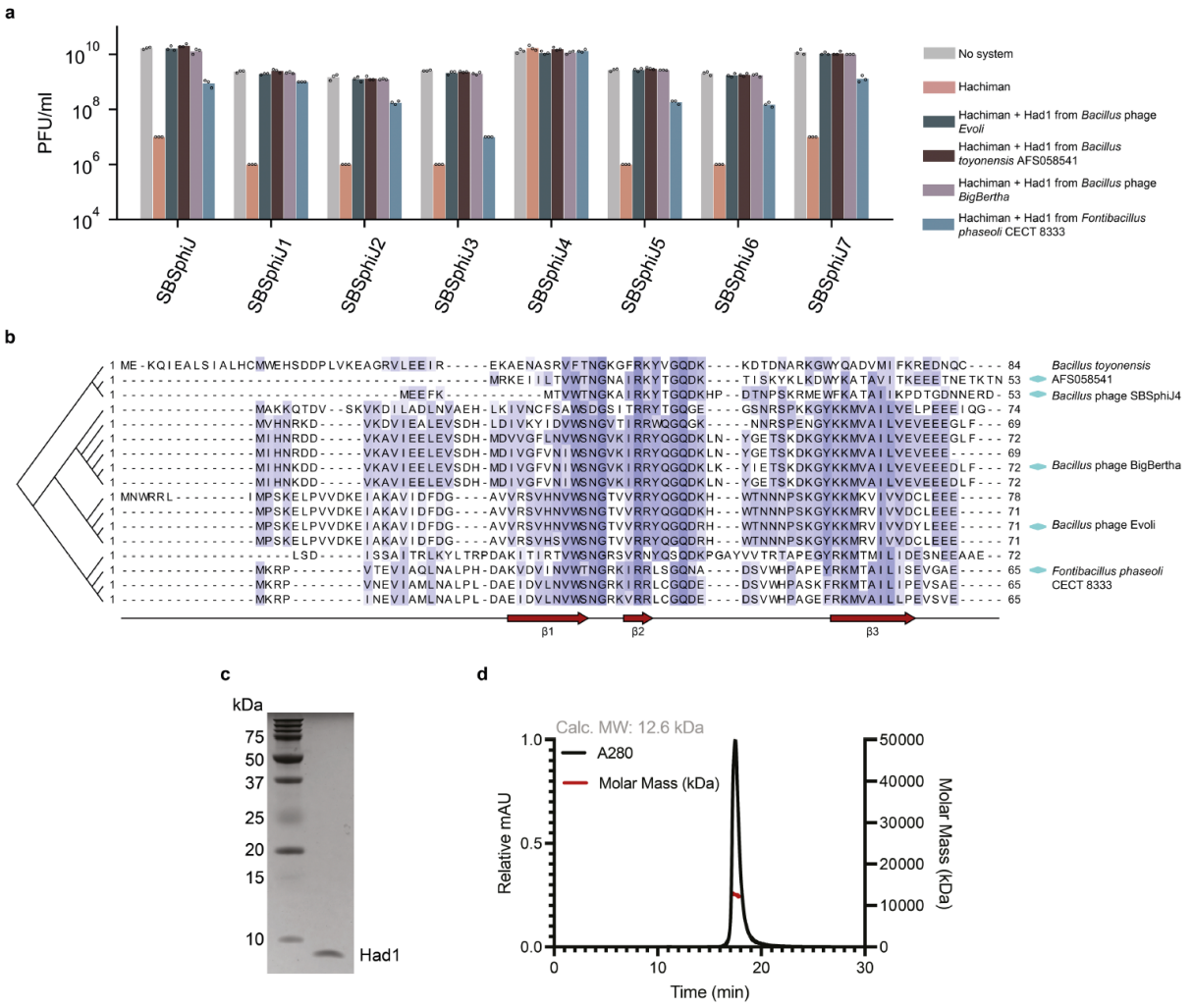


Figure 2.14 | Had1 proteins inhibit Hachiman-mediated defense.

a, Results of phage infection experiments with eight phages of the SBSphiJ group. Data represent plaque-forming units per ml (PFU/ml) of phages infecting control cells (“no system”), cells expressing the Hachiman system (“Hachiman”), and cells co-expressing the Hachiman system and a Had1 homolog. Shown is the average of three technical replicates, with individual data points overlaid. **b**, Structure-guided sequence alignment of Had1 homologs colored by BLOSUM62 score. **c**, SDS-PAGE and (**d**) SEC-MALS analysis of purified Had1. Full-length Had1 elutes as a single species that is consistent with a homodimeric complex (predicted homodimer 12.5 kDa, observed 12.6 kDa). Data are representative of three independent experiments.

It was previously shown that the SBSphiJ phage can escape Hachiman defence by mutating its single-stranded-DNA-binding protein, and it was suggested that Hachiman may recognize protein–DNA complexes produced as an intermediate of phage DNA replication or recombination[69]. The positively charged patch at the centre of the Had1 dimeric complex (Figure 2.13e) may imply that Had1 might shield phage replication intermediates from being recognized by the Hachiman system. As the mechanism of Hachiman defence is unknown at present, understanding how Had1 inhibits Hachiman in future studies may assist in solving the mechanism of Hachiman defence.

In this study, we identified multiple families of phage anti-defence proteins using comparative analysis of closely related phage genomes. Our findings demonstrate that phages have evolved diverse strategies to counter the complex, multilayered bacterial defence arsenal. Our data suggest that SPβ-like phages store anti-Gabija genes in a dedicated anti-Gabija locus in their genomes. This is reminiscent of anti-CRISPR loci identified in phages of Pseudomonas and other species, where different sets of anti-CRISPR genes are present in each individual phage[99]. Our discovery of a specific anti-Gabija locus in SPβ-like phages may point to a general rule for the organization of anti-defence genes in phage genomes, a genomic phenomenon that can assist in future searches for anti-defence genes.

The most widely distributed family of anti-defence proteins discovered here is that of Tad2, a protein that contains a domain of unknown function, DUF2829. A protein with DUF2829 was previously speculated to inhibit type II CRISPR–Cas systems and specifically Cas9[145], although binding to Cas9 was not demonstrated and the mechanism through which this protein antagonizes type II CRISPR–Cas was not determined[145]. Our findings that multiple DUF2829-containing proteins antagonize the Thoeris defence system by specifically binding and sequestering 1″–3′ gcADPR suggest that the DUF2829 family of proteins represent anti-Thoeris proteins rather than anti-CRISPR proteins.

Our results on Tad2 join recent discoveries of additional anti-defence proteins that function by sequestering immune signalling molecules[18,76]. These include Tad1, a completely different family of Thoeris-inhibiting proteins, which, similar to Tad2, also bind and sequester gcADPR molecules; and the phage-encoded Acb2 protein that inhibits CBASS defence by binding the cyclic oligonucleotide immune signalling molecules produced by CBASS[18,76]. It therefore seems that the production of protein ‘sponges’ that sequester immune signalling molecules is a highly efficient strategy that evolved within phages multiple

times in parallel to allow successful evasion of immune systems that utilize immune signalling molecules. The efficiency of this strategy may be perceived as counterintuitive, because it necessitates a 1:1 ratio between the number of phage sponge proteins and the immune signalling molecule (or even 2:1 in the case of Tad2). However, as Thoeris and CBASS become active and produce the immune signalling molecule relatively late in the infection cycle of the phage[9,19], phages have sufficient time to express a substantial amount of their sponge proteins at the early stages of infection. Thus, when the immune signalling molecule is produced by the defence system, there will already be enough copies of the sponge protein in the infected cell to efficiently block immune signalling. An alternative explanation for the efficiency of phage sponges could be that they have a higher affinity to the signalling molecule as compared to the bacterial effector. The affinity we measured for Tad2 binding to 1″–3′ gcADPR (K_D = 23.3 nM; Figure 2.10c) is indeed higher than the affinity measured for binding of the same molecule to Thoeris ThsA (K_D = 59.1 nM)[146], but both are within the same order of magnitude of nanomolar affinity.

We were not able to identify inhibitors of the Septu or Lamassu defence systems among the three groups of phages that we studied, although phages from these groups exhibited differential sensitivity to these defence systems. It is possible that escape from these two systems is mediated by mutations in existing genes rather than the presence of a dedicated anti-defence gene[69]. It is also possible that different phages in the same group utilize more than one protein to overcome defence, which would sometime hamper the analysis pipeline used in this study.

With the emergence of antibiotic-resistant bacteria[147], phage therapy, in which phages are used as an alternative to antibiotics, is being considered as a suitable therapeutic avenue for defeating bacterial pathogens[148–150]. One of the major obstacles for successful phage therapy is the recently discovered ability of bacteria to actively defend themselves by encoding a large variety of defence systems. Indeed, it was shown that the set of defence systems carried by a given bacterial strain is a strong determinant for the susceptibility of that strain to phages[151,152]. Engineering phages to carry a set of anti-defence proteins can enable them to overcome bacterial defences, resulting in phages with increased host ranges that will be more suitable for phage therapy. Thus, the anti-defence proteins we discovered here, as well as additional such proteins that were discovered and will be discovered in the future, could be used as tools for more efficient phage therapy applications.

Methods

Bacterial strains and growth conditions

Bacteria were grown in magnesium manganese broth (MMB) (LB + 0.1 mM MnCl2 + 5 mM MgCl2) at 37 °C shaking at 200 rpm, unless specified otherwise. To ensure the presence of an integrated antibiotics resistance cassette in the *B. subtilis* genome, the appropriate antibiotics were added at the following concentration: spectinomycin (100 µg ml^{-1}) or chloramphenicol (5 µg ml^{-1}). *Escherichia coli* strain *E. coli* NEB 5-alpha (NEB cat #C2987H) was grown in LB. Whenever applicable, media were supplemented with ampicillin (100 µg/ml).

Defense systems used in this study were cloned previously by Doron et al.[22] under their native promoter into the *amyE* locus of the *B. subtilis* BEST7003 genome. The source organisms of the specific systems used in this study are as follows: Thoeris (*Bacillus cereus* MSX-D12), Hachiman (*B. cereus* B4087), Gabija (*B. cereus* VD045), Septu (*Bacillus thuringiensis* HD12), and Lamassu (*Bacillus sp.* NIO-1130).

Phage strains, isolation, cultivation and sequencing

B. subtilis phages phi3T (BGSCID 1L1, GenBank accession KY030782.1), SPbeta (BGSCID 1L5, GenBank accession AF020713.1), SPR (BGSCID 1L56, GenBank accession OM236515.1) and SPO1 (BGSCID 1P4, GenBank accession NC_011421.1) were obtained from the *Bacillus* Genetic Stock Center (BGSC). Phages from the SBSphiJ group were isolated by us in previous studies[18,22]. Other phages used in this study were isolated by us from soil samples on *B. subtilis BEST7003* culture as described in Doron et al.[22]. For this, soil samples were added to *B. subtilis* BEST7003 culture and incubated overnight to enrich for B. subtilis phages. The enriched sample was centrifuged and filtered through 0.45 µm filters, and the filtered supernatant was used to perform double layer plaque assays as described in Kropinski et al.[153]. Single plaques that appeared after overnight incubation at RT were picked, re-isolated 3 times, and amplified as described below.

Phages were propagated by picking a single phage plaque into a liquid culture of *B. subtilis* BEST7003 grown at 37 °C to OD_{600} of 0.3 in MMB broth until culture collapse (or three hours in case of no lysis). The culture was then centrifuged for 10 min at 3200 g and the supernatant was filtered through a 0.45 µm filter to get rid of remaining bacteria and bacterial debris.

High titer phage lysates (>107 pfu ml^{-1}) were used for DNA extraction. 500 µl of the phage lysate was treated with DNase-I (Merck cat #11284932001) added to a final concentration of 20 mg ml^{-1} and incubated at 37 °C for 1 hour to remove bacterial DNA. DNA was extracted using the QIAGEN DNeasy blood and tissue kit (cat #69504) starting from the Proteinase-K treatment step.

Phages from the SBSphiJ and SPbeta groups were sequenced using a modified Nextera protocol as previously described[154]. Following Illumina sequencing, adapter sequences were removed from the reads using Cutadapt version 2.8[155] with the option -q 5. The trimmed reads from each phage genome were assembled into scaffolds using SPAdes genome assembler version 3.14.0[156], using the –careful flag.

The genomes of phages from the SPO1 group were sequenced via a long-read PacBio method, due to the high amount of modified bases in these phages. For library construction of phages from the SPO1 group, 1 µg of genomic DNA samples was fragmented using g-tubes (Covaris). Sheared DNA was purified with AMPure PB beads and was used to construct a SMRTbell library according to the PacBio library construction guidelines[157]. Samples were barcoded using Barcoded Overhang Adapters and pooled to one final library. Quantity and quality of the SMRTbell library were determined using the Qubit HS DNA kit and Agilent TapeStation Genomic DNA. No size selection was performed. The PacBio sequencing primer was then annealed to the SMRTbell library followed by binding of the polymerase to the primer–library complex. The library was loaded onto one SMRT cell in the PacBio Sequel system and sequenced in a Continuous Long Read (CLR) mode at a 10 h movie time. All phage genomes sequenced and assembled in this study were analyzed with Prodigal version 2.6.3[158,159] (default parameters) to predict ORFs.

Plaque assays

Phage titer was determined using the small drop plaque assay method[160]. 400 µl of overnight culture of bacteria were mixed with 0.5% agar and 30 ml MMB and poured into a 10 cm square plate followed by incubation for 1 h at RT. In cases of bacteria expressing anti-defense candidates and in the experiment with phage SBSphiJ with a Tad2 knock-in, 1 mM IPTG was added to the 30 ml MMB 0.5% agar. In cases of bacteria expressing dCas9-gRNA constructs, 0.002% xylose was added to the medium.10-fold serial dilutions in MMB were performed for each of the tested phages and 10 µl drops were put on the bacterial layer. After the drops had dried up, the plates were inverted and incubated at RT overnight. Plaque forming units (PFUs) were determined by counting the derived plaques after overnight incubation and lysate titer

was determined by calculating PFUs per ml. When no individual plaques could be identified, a faint lysis zone across the drop area was considered to be 10 plaques. Efficiency of plating (EOP) was measured by comparing plaque assay results on control bacteria and bacteria containing the defense system and/or a candidate anti-defense gene.

Prediction of candidate anti-defense genes

Predicted protein sequences from all phage genomes in each phage family were clustered into groups of homologs using the cluster module in MMSeqs2 release 12-113e3[161], with the parameters -e 10, -c 0.8, -s 8, --min-seq-id 0.3 and the flag --single-step-clustering. For each defense system, anti-defense candidates were defined as clusters that have a representation in all the phages that overcome the defense system and are absent from all the phages that are blocked by the defense system. One member was chosen from each cluster as a candidate anti-defense gene for further experimental testing. In the case of the Hachiman defense system, only predicted genes with no known function were tested experimentally. Gad2 was predicted based on its localization in the sane locus as *gad1* in phages SPbetaL6 and SPbetaL7.

Cloning of candidate anti-defense genes

The DNA of each anti-defense candidate was amplified from the source phage genome using KAPA HiFi HotStart ReadyMix (Roche cat # KK2601) and primers provided in Supplementary table 7, and cloned using the NEBuilder HiFI DNA Assembly cloning kit (NEB, no. E5520S) into the pSG-thrC-Phspank vector. Homologs of verified anti-defense genes were synthesized and cloned by Genscript Corp. The anti-defense candidates were cloned into the pSG-thrC-Phspank vector[18] and transformed into NEB 5-alpha competent cells. The cloned vector was subsequently transformed into *B. subtilis* BEST7003 cells containing the respective defense systems as applicable integrated into the *amyE* locus[22], resulting in cultures expressing both defense system and their corresponding anti-defense gene candidates, integrated into the *amyE* and *thrC* loci, respectively. As a negative control, a transformant with an identical plasmid containing sfGFP instead of the anti-defense gene, was used. Transformation to *B. subtilis* was performed using MC medium as previously described[22] and plated on LB agar plates supplemented with 5 μg ml^{-1} chloramphenicol incubated overnight at 30 °C. Whole-genome sequencing was then applied to all transformed *B. subtilis* strains, and Breseq (version 0.34.1) analysis[162] was used to verify the integrity of the inserts and lack of mutations.

Gad1 knockout in phi3T lysogenic strain

The upstream and downstream homologous arms of Gad1 were amplified from the phi3T phage genome using the PCR primer pairs Gad1_AF and Gad1_AR, and Gad1_BF and Gad1_BR, respectively. The spectinomycin resistance gene cassette was amplified from the vector pJmp3 (addgene plasmid #79875) using the PCR primers Spec_F and Spec_R. The pJmp3 backbone was amplified using the primers Vector_F and Vector_R.

These three parts were assembled together with the pJmp3 backbone using the NEBuilder HiFi DNA Assembly cloning kit (NEB cat # E5520S) and transformed to *E. coli* NEB 5-alpha competent cells. The cloned vector was then transformed into the phi3T lysogenic strain (BGSCID 1L1) using MC medium[22] and was plated on LB agar plates supplemented with 100 μg ml^{-1} spectinomycin and incubated overnight at 30 °C. The modified phi3T prophage was induced from the lysogenic bacterial strain grown to an OD600 of 0.3 by the addition of 0.5 μg ml^{-1} Mitomycin C (Sigma, M0503). After 3 h, the culture was centrifuged for 10 min at 3200 g and the supernatant containing the modified phages was collected and filtered through 0.2 μm filters. Whole-genome sequencing was performed to verify the sequence of the modified phage.

Gad2 knockout in SPbetaL7 using Cas13a

Cas13a was amplified from the pBA559 plasmid with the primers cas13a_fwd and cas13a_rev. The xylose promoter and homology arms for integration into the thrC site were amplified from plasmid pJG_thrC_dCAS9_gRNA[18] with primers dcas9xylProm_fwd and dcas9xylProm_rev. The gRNA complimentary to the beginning of Gad2 was amplified from the pBA559 plasmid with primers gRNAcas13_fwd and gRNAcas13L7159_rev. Plasmid assembly was conducted using NEBuilder HiFi DNA Assembly cloning kit (NEB, cat # E5520S) and transformed to NEB 5-alpha -competent cells. The cloned vector was subsequently transformed into the thrC site of *B. subtilis* BEST7003.

For the selection of Gad2 knockout SPbetaL7 phages, overnight culture of the Cas13a-gRNA containing bacteria was mixed was diluted 1:100 with MMB agar 0.5% with 0.2% xylose, and grown for 1 hour at RT. Then, 108 PFU of phage SPbetaL7 were spread on the plate. On the next day, several individual plaques were collected and propagated with 1 ml of the Cas13a-gRNA containing bacteria. Gad2 knockout was verified using PCR primers L7159Fchk and L7159Rchk. Sequencing of the product demonstrated

deletion of bases 109,505-110,753 from the SPbetaL7 genome, spanning the entire Gad2 gene and as well as the Gad2 promoter. The selected knockout phage was purified three times on *B. subtilis* BEST7003.

Gabija / Gabija + Gad1 complex assembly and *in vitro* nuclease activity

BcGajAB and BcGajAB + *Shewanella sp.* phage 1/4 Gad1 complexes were purified as described[37]. A 56-bp dsDNA with a sequence specific motif derived from phage Lambda (5′-TTTTTTTTTTTTTTTTTAATAACCCGGTTATTTTTTTTTTTTTTTTTTTTTT-3′)[35] was pre-incubated with purified GajAB or GajAB + Gad1 in 20 µL DNA cleavage reactions containing 1 µM dsDNA, 1 µM GajAB or GajAB + Gad1, 1mM MgCl2, 20 mM Tris-HCl pH 9.0 for 1, 5, 10, 15, and 20 min at 37 °C. Following incubation, reactions were stopped with DNA loading buffer containing EDTA and 10 µL was analyzed on a 2% TB (Tris-borate) agarose gel. Gels were run at 250V for 40 min at 4 °C, then stained by rocking at room-temperature in TB buffer with 10 µg ml−1 ethidium bromide for 30 min. Gels were de-stained in TB buffer for 40 min and imaged with a ChemiDoc MP Imaging System.

Gad2 pulldown assay

GajA, GajB, and Gad1 proteins were co-expressed together using custom pET vector with an N-terminal 6×His-SUMO2-5×GS tag on GajA and ribosome binding site between GajA and GajB, and GajB and Gad1. *Brevibacillus laterosporus* Gad2 was cloned from synthetic DNA (Integrated DNA Technologies) into a custom pBAD vector containing a chloramphenicol resistance gene and an IPTG-inducible promoter or a custom pET vector with an N-terminal 6×His-SUMO2 tag. Plasmids were transformed and expressed in BL21(DE3) or LOBSTR-BL21(DE3)-RIL (Kerafast) cells and subject to Ni-NTA column chromatography and SUMO2 cleavage with SENP2. Gad2 pulldown was analyzed by SDS-PAGE and Coomassie Blue staining.

GajAB DNA-Gad2 cleavage assay

GajAB complex was expressed and purified as previously described[37]. Briefly, Gad2 was expressed with an N-terminal 6×His-SUMO2 tag in LOBSTR-BL21(DE3)-RIL cells (Kerafast) and purified using Ni-NTA resin (Qiagen) followed by size exclusion chromatography using a 16/600 Superdex 200 column (Cytiva). The same 56-bp dsDNA substrate as used in GajAB DNA cleavage assays was incubated with 1 µM, 4 µM, or 8 µM Gad2 in a 20 µL DNA cleavage reaction containing 0.5 µM NAD+ (Sigma Aldrich), 0.5 µM NTPs, 1 µM dsDNA, 1 mM MgCl2, 20 mM Tris-HCl pH 9.0 for 10 min on ice. 1 µM GajAB was added to the reaction and incubated for 20 min at 37 °C. A control Gad2 only lane was run to demonstrate slight nucleic acid

contamination within the Gad2 protein prep. Following incubation, reactions were terminated with DNA loading buffer containing 60 mM EDTA and 10 µL was analyzed on a 2% TB agarose gel run at 250 V for 30 min. The gel was post-stained by rocking at room temperature with TB buffer containing 10 µg ml−1 ethidium bromide, de-stained in TB buffer alone for 30 min, and imaged on a ChemiDoc MP Imaging System.

Construction of dCas9 and gRNA cassettes for integration in *B. subtilis thrC* site

The plasmid pJG_thrC_dCAS9_gRNA was constructed as previously described[18]. To insert new spacers, two fragments were amplified from pJG_thrC_dCAS9_gRNA and the new spacer was introduced into the overlap of primers designed for NEBuilder HiFi DNA Assembly (NEB, no. E5520S). For the gRNA used to target tad2, the first fragment was amplified using primers JG528 and JG525, and the second using primers JG529 and JG524. The resulting assembled construct had the gRNA sequence “aagatgatgttcccaaacac”. For the gRNA used to target *had1*, the first fragment was amplified using primers JG389 and JG381, and the second using primers JG390 and JG382. The resulting assembled construct had the gRNA sequence “gcttgctaggattagtgtcc”. The gRNA sequence “ctatgattgatttttttagc” was used as a control. It was constructed as mentioned above, with primers JG389 and JG378, and JG390 and JG388. Shuttle vectors were propagated in *E. coli* NEB 5-alpha with 100 µg ml^{-1} ampicillin selection. Plasmids were isolated from *E. coli* NEB 5-alpha before transformation into the appropriate *B. subtilis* BEST7003 strains. The vectors containing the dCas9–gRNA sequences were cloned to *B. subtilis* strains containing the respective defense system, as well as to a control strain lacking the defense system.

Knock-in of Had1 and Tad2 into phage SBSphiJ and Gad1 into phage SPbeta

The DNA sequence of tad2, together with the Phspank promoter, was amplified from the Tad2-containing pSG-thrC-Phspank plasmid using KAPA HiFi HotStart ReadyMix (Roche, cat #KK2601) with the primer pair tad2KIF and tad2Kir. The DNA sequence of *had1*, together with its upstream intergenic region, was amplified from the genome of phage SBSphiJ4 with the primer pair had1KIF and had1KIR. The backbone fragment with the upstream and downstream genomic arms (±1.2 kbp) for the integration site of *tad2* and *had1* was amplified from the plasmid used previously for knock-in of the *tad1* gene[18], with the primer pair backboneKIF and backboneKIR. The DNA sequence of *gad1*, together with its upstream intergenic region, was amplified from the genome of phage phi3T with the primer pair gad1KIF and gad1KIR. The upstream

and downstream genomic arms (±1.2 kbp) for the integration site of the *gad1* gene within the SPbeta genome were amplified from the genome of phage SPbeta using the primer pair gad1LFF and gad1LFR, and the primer pair gad1RFF and gad1LRR. Cloning was conducted using the NEBuilder HiFi DNA Assembly cloning kit (NEB, no. E5520S) and transformed to NEB 5-alpha-competent cells. The cloned vector was subsequently transformed into the *thrC* site of *B. subtilis* BEST7003.

The *tad2*, *had1* and *gad1* containing *B. subtilis* BEST7003 strains were then infected with phages SBSphiJ (*tad2* and *had1*) or SPbeta (*gad1*) with an MOI of 0.1 and cell lysates were collected. Tad2 lysate was used to infect a Thoeris-containing B. subtilis culture in two consecutive rounds with an MOI of 2 in each round (30 °C, 1mM IPTG). Had1 lysate was used to infect a Hachiman-containing *B. subtilis* culture in two consecutive rounds with an MOI of 2 in each round (25 °C). Gad1 lysate was used to infect a Gabija-containing *B. subtilis* culture in two consecutive rounds with an MOI of 2 in each round (25 °C). Several plaques were collected and screened using PCR for the desired insertion within the phage genome. Phages with anti-defense genes were purified three times on *B. subtilis* BEST7003. Purified phages were verified again for the presence of *tad2*, *had1* and *gad1* using PCR amplifications.

Identification of anti-defense homologs and phylogenetic reconstruction

Homologs of anti-defense genes were searched in the metagenomic gut virus (MGV)[163] database using the "search" option of MMseqs release 12-113e3 with default parameters. Homologs of Gad1, Gad2 and Had1 were searched in the integrated microbial genomes (IMG) database[164] using the blast option in the IMG web server. Gad1 and Gad2 homologs were searched using the default parameters, while Had1 homologs were searched using an e-value of 10 due to their short size. For Gad1 and Had1 this process was repeated iteratively for homologs that were found, until convergence. For Tad2, due to the multitude of homologs, homology search was done using the "search" option of MMseqs release 12-113e3 with default parameters, against ~38000 prokaryotic genomes downloaded from the IMG database in October 2017.

For each family of anti-defense proteins, the unique (non-redundant) sequences were used for multiple sequence alignment with MAFFT version 7.402[165] using default parameters. Phylogenetic trees were constructed using IQ-TREE version 1.6.5[166] with the -m LG parameter. The online tool iTOL24 (v.5)[167] was used for tree visualization. Phage family annotations were based on the prediction in the MGV database. The host phyla annotations were either based on the prediction in the MGV database, or the IMG

taxonomy of the bacteria in which the prophage was found. Gabija and Thoeris defense systems were found in the bacterial genomes using DefenseFinder[20] version 1.0.9 and database release 1.2.3.

Preparation of filtered cell lysates

For generating filtered cell lysates, we used *B. subtilis* BEST7003 cells co-expressing Tad2 and the *B. cereus* MSX-D12 Thoeris system in which ThsA was inactivated (ThsB/ThsAN112A). Tad2 was integrated in the *thrC* locus as described above and expressed from an inducible Phspank promoter, and the Thoeris system was integrated in the amyE locus and expressed from its native promoter as described above. Controls included cells expressing only the ThsB/ThsAN112A Thoeris system without Tad2, as well as cells lacking both the Thoeris system and Tad2. These cultures were grown overnight and then diluted 1:100 in 250 ml MMB medium supplemented with 1 mM IPTG and grown at 37 °C, 200 rpm shaking for 120 min followed by incubation and shaking at 25 °C, 200 rpm until reaching an OD_{600} of 0.3. At this point, a sample of 40 ml was taken as the uninfected (time 0 min) sample, and SBSphiJ phage was added to the remaining 210 ml culture at an MOI of 10. Flasks were incubated at 25 °C with shaking (200 rpm), for the duration of the experiment. 40 ml samples were collected at time points 75, 90, 105 min post-infection. Immediately upon sample removal (including time point 0 min), the 40 ml sample tubes were centrifuged at 4 °C for 10 min at 3200 g to pellet the cells. The supernatant was discarded, and the pellet was flash frozen and stored at −80 °C.

To extract cell metabolites from frozen pellets, 600 µl of 100 mM Na phosphate buffer (pH 8.0) was added to each pellet. Samples were transferred to FastPrep Lysing Matrix B in a 2 ml tube (MP Biomedicals, no. 116911100) and lysed at 4 °C using a FastPrep bead beater for 2 × 40 s at 6 m s−1. Tubes were then centrifuged at 4 °C for 10 min at 15,000 g. Supernatant was then transferred to an Amicon Ultra-0.5 Centrifugal Filter Unit 3 kDa (Merck Millipore, no. UFC500396) and centrifuged for 45 min at 4 °C, 12,000 g. Filtered lysates were taken for *in vitro* ThsA-based NADase activity assay.

Expression and purification of ThsA

B. cereus MSX-D12 thsA fused to a C-terminal TwinStrep tag was cloned into a pACYC-Duet1 plasmid (addgene plasmid #71147). The protein was expressed under the control of the T7 promoter together with a C-terminal Twin-Strep tag for subsequent purification. Expression was performed in 5 L LB medium supplemented with chloramphenicol (34 mg ml^{-1}) in *E. coli* BL21(DE3) cells. Induction was performed with

0.2 mM IPTG at 15 °C overnight. The cultures were collected by centrifugation and lysed by a cooled cell disrupter (Constant Systems) in 100 ml lysis buffer composed of 20 mM HEPES pH 7.5, 0.3 M NaCl, 10% glycerol and 5 mM β-mercaptoethanol, 200 KU/100 ml lysozyme, 20 µg ml^{-1} DNase, 1 mM MgCl2, 1 mM phenylmethylsulfonyl fluoride (PMSF) and protease inhibitor cocktail (Millipore, 539134). Cell debris were sedimented by centrifugation, and the lysate supernatant was incubated with 2 ml washed StrepTactin XT beads (IBA, 2-5030–025) for 1 h at 4 °C. The sedimented beads were then packed into a column connected to an FPLC allowing the lysate to pass through the column at 1 ml min^{-1}. The column was washed with 20 ml lysis buffer. ThsA was eluted from the column using elution buffer containing 50 mM biotin, 100 mM Tris pH 8, 150 mM NaCl and 1 mM EDTA. Peaks containing the ThsA protein were injected to a size-exclusion chromatography (SEC) column (Superdex 200_16/60, GE Healthcare, 28-9893-35) equilibrated with SEC buffer (20 mM HEPES pH 7.5, 200 mM NaCl and 2 mM DTT). Peaks were collected from the SEC column, aliquoted and frozen at −80 °C to be used for subsequent experiments.

ThsA-based NADase activity assay

NADase reaction was performed in black 96-well half area plates (Corning, 3694). In each reaction microwell, purified ThsA protein was added to cell lysate, or to *in vitro* reactions of Tad2 with 1″–3′ gcADPR, or to 100 mM sodium phosphate buffer pH 8.0. 5 µl of 5 mM nicotinamide 1,N6-ethenoadenine dinucleotide (εNAD+, Sigma, N2630) solution was added to each well immediately before the beginning of measurements, resulting in a final concentration of 100 nM ThsA protein in a 50 µl final volume reaction. Plates were incubated inside a Tecan Infinite M200 plate reader at 25 °C, and measurements were taken at 300 nm excitation wavelength and 410 nm emission wavelength. Reaction rate was calculated from the linear part of the initial reaction.

Tad2 protein cloning, expression, and purification for biochemistry

The SPO1 *tad2* gene was cloned into the expression vector pET28-bdSumo as described previously described[18]. Tad2 was expressed in *E. coli* BL21(DE3) by induction with 200 µM IPTG at 15 °C overnight. A 2 L culture of bacteria expressing Tad2 was harvested and lysed by a cooled cell disrupter (Constant Systems) in lysis buffer (50mM Tris pH 8, 0.25M NaCl, 10% Glycerol) containing 200KU 100 ml^{-1} lysozyme, 20ug ml^{-1} DNase, 1mM MgCl2, 1 mM phenylmethylsulfonyl fluoride (PMSF) and protease inhibitor cocktail. After clarification of the supernatant by centrifugation, the lysate was incubated with 5 ml washed Ni beads

(Adar Biotech) for 1 h at 4 °C. After removing the supernatant, the beads were washed 4 times with 50 ml lysis buffer. Tad2 was eluted by incubation of the beads with 10 ml cleavage buffer (50 mM Tris pH 8, 0.25M NaCl, 10% Glycerol and 0.4 mg bdSumo protease) for 1 h at 23 °C. The supernatant, containing cleaved Tad2, was removed, and an additional 5 ml cleavage buffer was added to the beads and left overnight at 4 °C. The two elution solutions were combined, concentrated, and applied to a size exclusion (SEC) column (HiLoad_16/60_Superdex75 prep-grade, Cytiva) equilibrated with 50 mM Tris pH 8, 100 mM NaCl. Pure Tad2 migrating as a single peak was pooled and flash frozen in aliquots using liquid nitrogen and stored at −80 °C.

Incubation of purified 1″–3′ gcADPR with Tad2

2.4 μM purified Tad2 was incubated with 600 nM of 1″–3′ gcADPR in 100 mM Na Phosphate buffer, pH 8.0 for 10 min, at 25°C, followed by an additional 10 min incubation at either 95°C or 25°C. Following incubation, samples were left on ice for 1 min. The samples were then used for ThsA-based NADase activity assays as described above.

Analytical SEC analysis of apo and ligand bound Tad2

50 μl of Tad2 (158 μM) was incubated with 30 μl of 1 mM 1″–3′ gcADPR at 25 °C, in 0.1 M NaCl, 50 mM TrisHCl pH 8.0, for 20 min. The incubated mixture, and an apo protein incubated with an identical buffer without the ligand, were then loaded on a size exclusion Superdex_200_10/300 analytical column (PBS buffer) and monitored for absorption at both 260nm and 280nm. The oligomeric nature of Tad2 apo protein was evaluated by comparing the retention time of apo-Tad2 to that of five internal standard proteins (Ribonuclease 13.7 kD, Chymotrypsin 25 kD, BSA 66 kD, Aldolase 160 kD, and apoferritin 443 kD) using a size exclusion Superdex_200_10/300.

Surface Plasmon Resonance (SPR) measurements of Tad2 binding to 1″–3′ gcADPR and cADPR

Binding of 1″–3′ gcADPR to Tad2 was monitored by surface plasmon resonance with a BIAcore S200 apparatus (Cytiva, Sweden). Tad2 was immobilized on a CM5 S Series chip (Cytiva, Sweden) by amine coupling chemistry using the following protocol: chip activation was carried out with a freshly prepared mixture of N-hydroxysuccinimide (50 mM in water) and 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (195 mM in water) for 7.5 min in DPBS (Sartorius, SKU 02-023-5A) (flow rate of 10 μl/min). DPBS served as the running buffer along the experiment. Tad2 protein (5 μg ml−1 in 150 mM sodium acetate buffer, pH

3.8) was injected for 5 min (flow rate 10 µl/min) and the remaining activated carboxylic groups were blocked by injecting of 1 M ethanolamine hydrochloride, pH 8.0, for 5 min (flow rate 10 µl/min). Subsequently, a total of 1,800 RU (Response Units) of Tad2 were immobilized on to the chip. Before data collection, a normalization cycle followed by a priming cycle were run to stabilize the instrument. Binding of 1″–3′ gcADPR to Tad2 was monitored by injecting 1″–3′ gcADPR at multiple concentrations for 2 min at 25 °C and flow rate of 50 µl/min. Dissociation was carried out for 180 sec. No regeneration step was required after ligand binding as the dissociation reached baseline spontaneously. Sensograms were fit to a 1:1 binding model (S200 evaluation software 1.1) which yielded the kinetic constants (ka and kd) of the interaction. kd divided by ka yield the steady state KD. cADPR which served as a control did not bind, as expected.

Protein cloning, expression, and purification

Synthetic DNA fragments (Integrated DNA Technologies) of cmTad1, AbTIRTIR (Δ1–156), cbTad1, SPO1 Tad2 and *Bacillus toyonensis* Had1 genes were cloned into a custom pET expression vector containing an N-terminal 6×His-SUMO2 tag and an ampicillin resistance gene by Gibson assembly, as previously described[111]. Three colonies of BL21(DE3) RIL *E. coli* transformed with these plasmids and grown on MDG agar plates (1.5% agar, 2 mM MgSO4, 0.5% glucose, 25 mM Na2HPO4, 25 mM KH2PO4, 50 mM NH4Cl, 5 mM Na2SO4, 0.25% aspartic acid, 2–50 µM trace metals) were picked into a 30 ml MDG starter culture and shaken overnight at 230 rpm and 37 °C. A 1 L culture of M9ZB expression medium (2 mM MgSO4, 0.5% glycerol, 47.8 mM Na2HPO4, 22 mM KH2PO4, 18.7 mM NH4Cl, 85.6 mM NaCl, 1% Cas-amino acids, 2–50 µM trace metals, 100 µg ml^{-1} ampicillin and 34 µg ml^{-1} chloramphenicol) was seeded with 15 ml starter culture and grown at 230 rpm and 37 °C to an OD600 of 2 before induction of expression with 0.5 mM IPTG and incubation at 230 rpm and 16°C for 16 hours. For AbTIRTIR expression, nicotinamide-supplemented 2YT expression medium (1.6% w/v tryptone, 1% w/v yeast extract, 342 mM NaCl, 10 mM NAM, 100 µg ml−1 ampicillin, 34 µg ml−1 chloramphenicol) was used instead. Cells were harvested by centrifugation, resuspended in lysis buffer (20 mM HEPES-KOH pH 7.5, 400 mM NaCl, 30 mM imidazole, 10% glycerol, 1 mM DTT), lysed by sonication, and clarified by centrifugation at 25,000g for 20 min. Lysate was passed over 8 ml Ni-NTA resin (Qiagen), washed with 70 ml wash buffer (20 mM HEPES-KOH pH 7.5, 1 M NaCl, 30 mM imidazole, 10% glycerol, 1 mM DTT), eluted with lysis buffer supplemented to 300 mM

imidazole, and dialyzed in 14 kDa dialysis tubing in dialysis buffer (20 mM HEPES-KOH pH 7.5, 250 mM KCl, 1 mM DTT) overnight at 4°C with purified human SENP2 for 6× His-SUMO2 tag cleavage. For crystallography, proteins were further purified on a Superdex 75 16/600 size-exclusion chromatography column (Cytiva). Final samples were concentrated to >20 mg ml^{-1}, flash frozen, and stored at −80°C.

ThsB′ was cloned and purified similarly as described above, except that it was cloned into a custom pET vector containing a C-terminal 6× His tag and a chloramphenicol resistance gene, transformed into BL21(DE3) cells, grown in the presence of chloramphenicol only, expressed in 2YT medium with 10 mM NAM, and concentrated to 4 mg ml−1 after dialysis before flash-freezing and storage.

1′′–2′ gcADPR and 1′′–3′ gcADPR production and purification

gcADPR molecules were produced as described previously[18]. For 1′′–2′ gcADPR production, purified AbTIRTIR, a bacterial enzyme that efficiently converts NAD+ to 1′′–2′ gcADPR, was used to set up 300µl reactions (50 mM HEPES-KOH, 150 mM NaCl, 20 mM NAD+, 40 µM AbTIRTIR). For 1′′–3′ gcADPR production, Purified ThsB'[18] was used to set up 50 ml reactions (50 mM HEPES-KOH pH 7.5, 150 mM NaCl, 2 mM NAD+, 16 µM ThsB'). Reactions were carried out at RT for 24-48 hours before boiling at 95 °C for 10 min, pelleting at 13,500 g for 20 min, and filtering through a 10 kDa MWCO filter (Amicon). For the AbTIRTIR reaction, filtrate was diluted to 30 ml with PBS followed by addition of 200 µM purified cmTad1. For the ThsB' reaction, 10 µM purified cmTad1 was added directly to 50 ml of filtrate. Mixtures were then incubated for 1 hour at RT to allow cmTad1–gcADPR complex formation. The complexes were washed by successive concentration and dilution in a 10 kDa MWCO filter first with five 1:20 dilutions in PBS followed by five 1:20 washes in water. Complexes were concentrated to >3 mM before final release and extraction of gcADPR by boiling at 95 °C for 10 min, pelleting at 13,500 g for 20 min, filtering through a 3 kDa MWCO filter, and collecting the filtrate. Final purity and concentration were assessed by HPLC.

Protein crystallization and structural analysis

cbTad1, SPO1 Tad2, and Had1 crystals were grown by the hanging-drop method in EasyXtal 15-well trays (NeXtal) at 16°C. Hanging-drops were set using 1 µl of diluted protein solution (5–10 mg ml−1 protein, 20 mM HEPES-KOH pH 7.5, 70–80 mM KCl, 1 mM TCEP) and 1 µl reservoir solution over a 400µl well of reservoir solution. Proteins were crystallized and cryoprotected under the following conditions before being harvested by flash-freezing in liquid nitrogen: (1) cbTad1 complexed with 1′′–3′ gcADPR: Crystals were

grown for 3 weeks in drops supplemented with 500 μM ThsB′-derived 1″–3′ gcADPR using reservoir solution containing 0.2 M MgCl2, 0.1 M Tris-HCl pH 8.5, and 3.4 M 1,6-hexanediol. (2) SPO1 Tad2 in the apo state: Crystals were grown for 1 week using reservoir solution containing 0.1 M 2-(N-morpholino) ethanesulfonic acid pH 6.5, 10% (v/v) 1,4-dioxane, and 1.6 M ammonium sulfate before being cryoprotected with reservoir solution supplemented with 25% (v/v) glycerol. (3) SPO1 Tad2 complexed with 1″–3′ gcADPR: Crystals were grown for 1 week in drops supplemented with 500 μM ThsB′-derived 1″–3′ gcADPR using reservoir solution containing 0.2 M magnesium nitrate and 18% (w/v) PEG 3350 before being cryoprotected with reservoir solution supplemented with 20% (v/v) ethylene glycol and 1 mM 1″–3′ gcADPR. (4) SPO1 Tad2 complexed with 1″–2′ gcADPR: Crystals were grown for 1 week in drops supplemented with 500 μM AbTIRTIR-derived 1″–2′ gcADPR using reservoir solution containing 0.1 M Tris-HCl pH 8.5, 12% (v/v) glycerol, and 1.5 M ammonium sulfate before being cryoprotected with reservoir solution supplemented with 15% (v/v) ethylene glycol and 1 mM 1″–2′ gcADPR. (5) Had1 in the apo state: Crystals were grown for 1 week using reservoir solution containing 100 mM MES pH 6.2 and 30% PEG 4000 before being cryoprotected with reservoir solution supplemented with 10% (v/v) ethylene glycol.

CbTad1 and SPO1 Tad2 X-ray diffraction data were collected at the Advanced Photon Source (beamline 24-ID-C), and data were processed with XDS[168] and Aimless[169] using the SSRL autoxds script (A. Gonzalez, Stanford SSRL). Phases were determined by molecular replacement in Phenix using either previously determined cbTad1 structures[18] (PDB 7UAV, 7UAW) or sequence-predicted SPO1 Tad2 truncated structures from ColabFold v1.5.2[104]. Model building was performed in Coot[113], refinement was performed in Phenix[112]. Had1 X-ray diffraction data were collected at the Advanced Light Source (beamline 8.2.1) and data were processed with XDS[168] and Aimless[169] using the SSRL autoxds script (A. Gonzales, Stanford, SSRL). Experimental phase information was determined by molecular replacement using a dimeric Had1 AlphaFold2 predicted structure[104] in Phenix[112]. Model building was completed in Coot[113] and then refined in Phenix[112]. Statistics were analyzed as presented in Data Table 2.1[170–172], and structure figures were produced in PyMOL. Final structures were refined to stereochemistry statistics for Ramachandran plot (favored/allowed), rotamer outliers and MolProbity score as follows: cbTad1–1″–3′-gcADPR, 95.88/4.12%, 0.92%, 1.47; SPO1 Tad2 apo, 96.15/2.56%, 0.74%, 1.59; SPO1 Tad2–1″–3′-

gcADPR, 99.18/0.82%, 1.5%, 1.33; SPO1 Tad2–1″–2′-gcADPR, 99.36/0.64%, 2.19%, 1.27; Had1, 96.18/3.82%, 0.69%, 1.20.

HPLC analysis of Tad2 incubation with 1″–3′ gcADPR

Reactions to analyze cleavage of 1″–3′ gcADPR were performed in 120 µl reactions consisting of 50 mM Tris-HCl pH 7.5, 100 mM KCl, 5 mM MgCl2, 1 mM MnCl2, 50 µM gcADPR isomer, and either buffer or 1 µM Tad2. As a control, 1″–3′ gcADPR was also incubated with 1 µl Cap-Clip Acid Pyrophosphatase (known to cleave diphosphate linkages on mRNA caps, Fisher Scientific), using the manufacturer's recommended reaction conditions. Reactions were incubated at 37 °C for 1 h before filtration using a 3 kDa MWCO filter. Filtered reactions were analyzed using a C18 column (Agilent Zorbax Bonus-RP 4.6 x 150 mm) heated to 40 °C and run at 1 ml min−1 using a buffer of 50 mM NaH2PO4-NaOH pH 6.8 supplemented with 3% acetonitrile.

Data availability

Data that support the findings of this study are available within the article and its Supplementary Tables. IMG/MGV accessions, protein sequences and nucleotide sequences appear in Supplementary Tables 8–14. Coordinates and structure factors of cbTad1–1″–3′-gcADPR, SPO1 Tad2 apo, SPO1 Tad2–1″–3′-gcADPR, SPO1 Tad2–1″–2′-gcADPR and Had1 have been deposited in the PDB under the accession codes 8SMD, 8SME, 8SMF, 8SMG and 8TTO respectively. The genome sequences of phages SPO1L1-SPO1L5 and SPbetaL1-SPbetaL8 have been deposited with GenBank under accession codes OQ921336-OQ921348, respectively. Source data are available for all the main figures and for extended data figures.

Acknowledgements

We thank the Sorek laboratory members for comments on the manuscript and fruitful discussion. We also thank Y. Peleg and S. Albeck from the Center for Structural Proteomics within the Weizmann Institute of Science for assistance with protein expression, Y. Fridmann-Sirkis from the Life Sciences Core Facilities of the Weizmann Institute for help with SPR analysis, and H. Keren-Shaul and D. Pilzer from the Life Sciences Core Facilities of the Weizmann Institute for help with PacBio sequencing. R.S. was supported, in part, by the European Research Council (grant no. ERC-AdG GA 101018520), Israel Science Foundation MAPATS Grant 2720/22), the Deutsche Forschungsgemeinschaft (SPP 2330, Grant 464312965), the Ernest and Bonnie Beutler Research Program of Excellence in Genomic Medicine, and the Knell Family Center for Microbiology. E.Y. is supported by the Clore Scholars Program, and, in part, by the Israeli Council for Higher Education (CHE) via the Weizmann Data Science Research Center. P.J.K. was supported, in part, by the Pew Biomedical Scholars programme and The Mathers Foundation. S.J.H. is supported through a Cancer Research Institute Irvington Postdoctoral Fellowship (no. CRI3996). X-ray data were collected at the Northeastern Collaborative Access Team beamlines 24-ID-C and 24-ID-E (P30 GM124165), and used a Pilatus detector (S10RR029205), an Eiger detector (S10OD021527) and the Argonne National Laboratory Advanced Photon Source (DE-AC02-06CH11357), and at beamline 8.2.1 of the Advanced Light Source, a US DOE Office of Science User Facility under contract no. DE-AC02-05CH11231 and supported in part by the Howard Hughes Medical Institute, the ALS-ENABLE program, and the NIGMS grant P30 GM124169-01.

Author Contributions

Erez Yirmiya: Conceptualized and designed the study, built and executed the computational pipeline and analyzed the data. Wrote the manuscript.

Azita Leavitt: Conceptualized and designed the study, isolated the phages and conducted all the *in vivo* experiments unless stated otherwise.

Allen Lu: Performed the structural analysis of Tad1 and Tad2.

Adelyn E. Ragucci: Determined and analyzed the structure of Had1.

Carmel Avraham: Performed the biochemical experiments with cell lysates and led the mechanistic characterization of the Tad2 activity.

Ilya Osterman: Designed and conducted all the phage knock in experiments and the knockout of *gad2* from phage SPbetaL7.

Jeremy Garb: Designed and generated the knock out clones.

Sadie P. Antine: Performed protein purification for Gabija, Gad1, and Gad2, and performed DNA cleavage assays.

Sarah E. Mooney: Performed protein purification for Gabija, Gad1, and performed DNA cleavage assays.

Samuel J. Hobbs: Performed structural analysis, characterization of the Tad2 activity, and analysis of Had1 oligomerization.

Philip J. Kranzusch: Performed the structural analysis of Tad1 and Tad2 and performed DNA cleavage assays.

Gil Amitai: Performed the biochemical experiments with cell lysates and led the mechanistic characterization of the Tad2 activity and supervised the study.

Rotem Sorek: Conceptualized, designed and supervised the study. Wrote the manuscript.

Competing Interests

R.S. is a scientific cofounder and advisor of BiomX and Ecophage. The rest of the authors declare no competing interests.

Chapter 3: Comparative analyses define distinct classes of nuclease-helicase systems in bacterial anti-phage defense

Adelyn E. Ragucci[1,2#], Sadie P. Antine[1,2#], Ethan M. Leviss[1,2], Sarah E. Mooney[1,2] and Philip J. Kranzusch[1,2,3*]

[1] Department of Microbiology, Harvard Medical School, Boston, MA 02115, USA

[2] Department of Cancer Immunology and Virology, Dana-Farber Cancer Institute, Boston, MA 02115, USA

[3] Parker Institute for Cancer Immunotherapy at Dana-Farber Cancer Institute, Boston, MA 02115, USA

[#] These authors contributed equally to this work

Abstract

All phages must inject their nucleic acid genome to infect and replicate in bacterial cells. Therefore, a hallmark of phage infection is the presence of foreign nucleic acid. To combat phage nucleic acid, bacteria have evolved many nuclease-helicase anti-phage defense systems, however, the mechanisms for how they function remain mostly unknown. Recent work on Gabija and Hachiman demonstrates that they form remarkably different protein complexes and have contrasting phage nucleic acid-targeting specificities[37,92]. Therefore, from the genome level, how nuclease-helicase systems function cannot be discerned. Here, we present a biochemical screen of 14 nuclease-helicase systems. We found that all nuclease-helicase systems form complexes with operon-encoded proteins at varying oligomeric states, and either target all nucleic acid substrates given *in vitro*, have specific phage DNA cleavage, or do not cleave any nucleic acid *in vitro*. Overall, we present a biochemical characterization of how nuclease-helicase systems target phage nucleic acid during infection.

Introduction

Over 150 anti-phage defense systems have now been identified, yet mechanistically how they function remains somewhat unknown. A commonality that has emerged is the presence of nuclease and helicase domains in defense systems. Nuclease-helicase systems such as Gabija and Hachiman are distinguished as individual systems because the type of nuclease and helicase present in the operons are divergent and overall have very low sequence and structural homology. The Gabija operon contains two proteins, an OLD nuclease and an SF1A DNA helicase, which form a 4:4 nuclease-helicase complex[37]. Hachiman consists of a DUF4297 nuclease and DEAD-box/RNA helicase that come together in a 1:1 ratio[92]. *Bacillus cereus* Gabija cleaves a palindromic 15-bp sequence dsDNA derived from lambda phage[35]. *Escherichia coli* ECOR31 Hachiman degrades most DNA substrates given *in vitro*[92]. Unlike restriction-modification systems that exert specificity through methylation of host DNA or CRISPR systems that use RNA guides to target phage DNA, nuclease-helicase systems do not appear to have cleavage specificity that can be gleaned from predicted protein activity. Therefore, how nuclease-helicase systems are activated and inhibit phage infection remains an open question in the field.

To understand the biochemical activities of nuclease-helicase anti-phage defense systems we surveyed the literature and selected 16 different systems for analysis[22–25,70,173,174]. AbpAB, Azaca, DUF4297-HerA, Gabija, Hachiman, Hna, Menshen, Mokosh Type I, Mokosh Type II, Nhi-like, Olokun, PD-T4-1, PD-T4-4, PD-T4-8, Toprim-TA, and Upx operon architectures all contain a nuclease, helicase and/or other/unknown proteins (Figure 3.1a). All systems have distinct nuclease types such as DUF4297 (AbpAB, DUF4297-HerA), DUF1837 (AbpAB, Hachiman), (P)D-(D/E)-XK (Hna, Olokun, PD-T4-1, PD-T4-4, PD-T4-9, Upx), PLD (Mokosh Type I and II) and OLD/Toprim (Gabija, Menshen, Toprim-TA) (Figure 3.2). A bioinformatic screen of all toxin-antitoxin (TA) -like operons in bacteria revealed an ABC ATPase and Toprim TA-like operon that mirrors the domain architecture of the Gabija GajA protein[174]. We therefore named this system Toprim-TA.

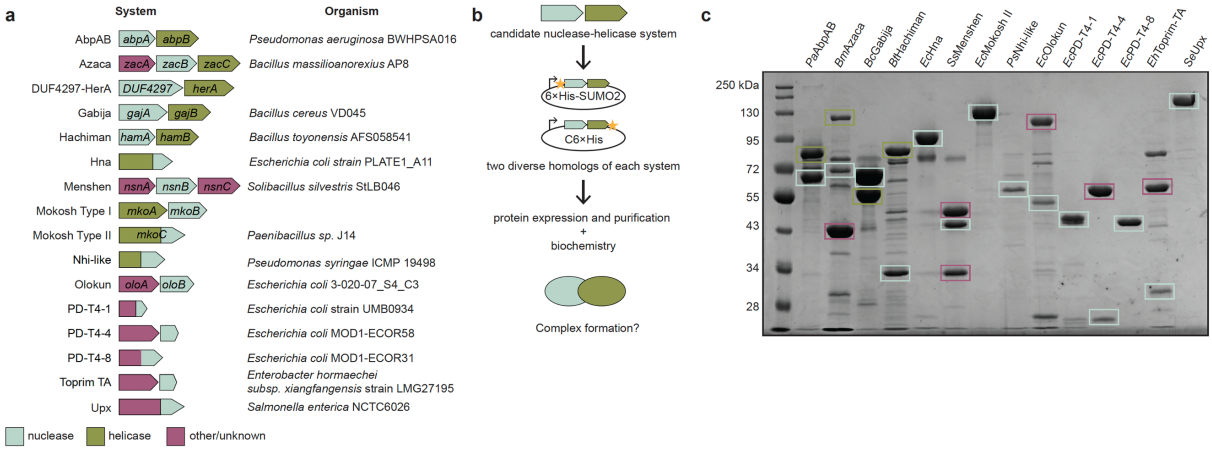


Figure 3.1 | Diverse nuclease-helicase systems form protein complexes.

a, Schematic of operon architecture of 16 selected nuclease-helicase anti-phage defense systems. If genes have annotated names, they are included inside the gene with color legend below. **b,** Nuclease-helicase defense system expression and purification strategy. Systems were cloned into custom pET vectors with an N-terminal 6×His-SUMO2 fusion tag on the first protein in the operon or a C-terminal 6×His tag on the last gene in the operon. **c,** SDS-PAGE**,** analysis of purified nuclease-helicase systems. Colored boxes indicate the identity of the nuclease, helicase, or other/unknown protein. Data are representative of at least 3 independent experiments.

We included Gabija and Hachiman as benchmarks for understanding the 14 other systems in our screen. Complex formation is required for Gabija and Hachiman defense, which led us to ask whether other nuclease-helicase operons form complexes. We tagged the native operon of each system with either an N-terminal 6×His-SUMO2 (first protein) or a C-terminal 6×His tag (last protein) and expressed and purified each system from *E. coli* (Figure 3.1b). We successfully purified 14 out of the 16 systems, excluding DUF4297-HerA and Mokosh Type I, and found that all multi-protein systems formed hetero-oligomeric complexes (Figure 3.1c). Complex formation is therefore a widespread rule for nuclease-helicase systems and likely allows for nuclease and helicase activities to be linked to target nucleic acid during phage infection.

Gabija and Hachiman form protein complexes with vastly different oligomeric states to exert antiviral defense. To assess the oligomeric states of the other 12 nuclease-helicase systems, we used mass photometry. Mass photometry uses light scattering of individual molecules in solution to determine the

mass, oligomeric state, and heterogeneity of a sample[175] (Figure 3.3a–i, Figure 3.4a–c). As a control, we performed mass photometry on Gabija, which accurately determined the mass of the 4:4 *B. cereus* GajAB complex[37] (Figure 3.3a). We then used the mass of each nuclease-helicase system to estimate stoichiometry and determined predicted structures using AlphaFold3[176] (Figure 3.3b-i). *Bacillus toyonensis* Hachiman forms a 1:1 complex with a similar structure to E. *coli* Hachiman[92] (Figure 3.3c). Intriguingly, *Pseudomonas aeruginosa* AbpAB forms a 1:1 nuclease-helicase structure that mirrors Hachiman with the addition of an N-terminal DUF4297 nuclease domain in AbpA (Figure 3.2a; Figure 3.3b,c).

E. coli Hna exists as both a monomer and dimer in solution with helicase domains interacting at the core of the dimer and (P)D-(D/E)-XK nuclease domains on the outside edges (Figure 3.3d). *E. coli* Mokosh Type II and *Salmonella enterica* Upx proteins are the only systems to exist as solely monomers in solution (Figure 3.3e, i). *E. coli* Mokosh Type II has a coiled-coiled extension similar to human Rad50, which binds to the dsDNA-break repair complex Mre11[177] (Figure 3.3e). *Pseudomonas fluorescens* Olokun makes a butterfly-shaped 2:2 complex of the adaptin-like protein OloA and (P)D-(D/E)-XK nuclease OloB. Adaptin-like proteins are found in diverse scaffolding and regulatory proteins[178]. OloA interacts with the N-terminal helical domain of OloB that extends to a C-terminal (P)D-(D/E)-XK nuclease domain. Ec-PD-T4-1 forms a dimer with a C-terminal nuclease core surrounded by an N-terminal unstructured region and helical bundles (Figure 3.3g). *E. coli* PD-T4-8 forms a tetramer with predicted DNA binding domain dimers on either side of the complex bridged by a C-terminal (P)D-(D/E)-XK nuclease core (Figure 3.3h). Predicting the structure of EcPD-T4-8 led us to realize that it is the Shedu anti-phage defense system, which nicks linear dsDNA 14bp from the 5′ end of the substrate to confer defense[179].

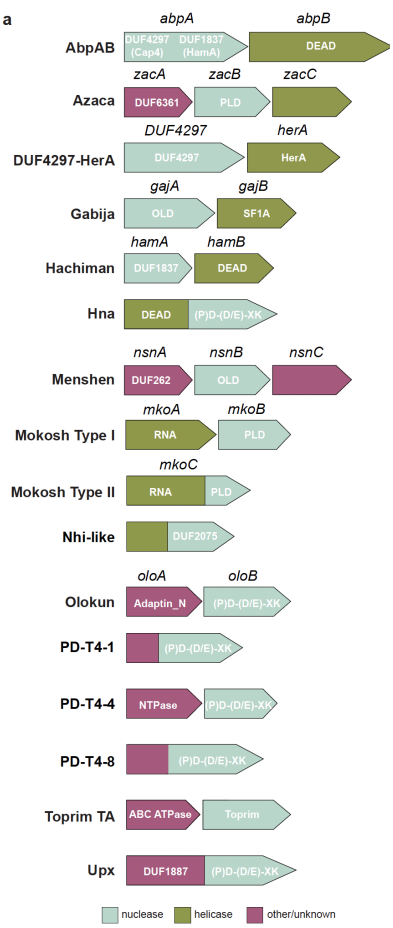
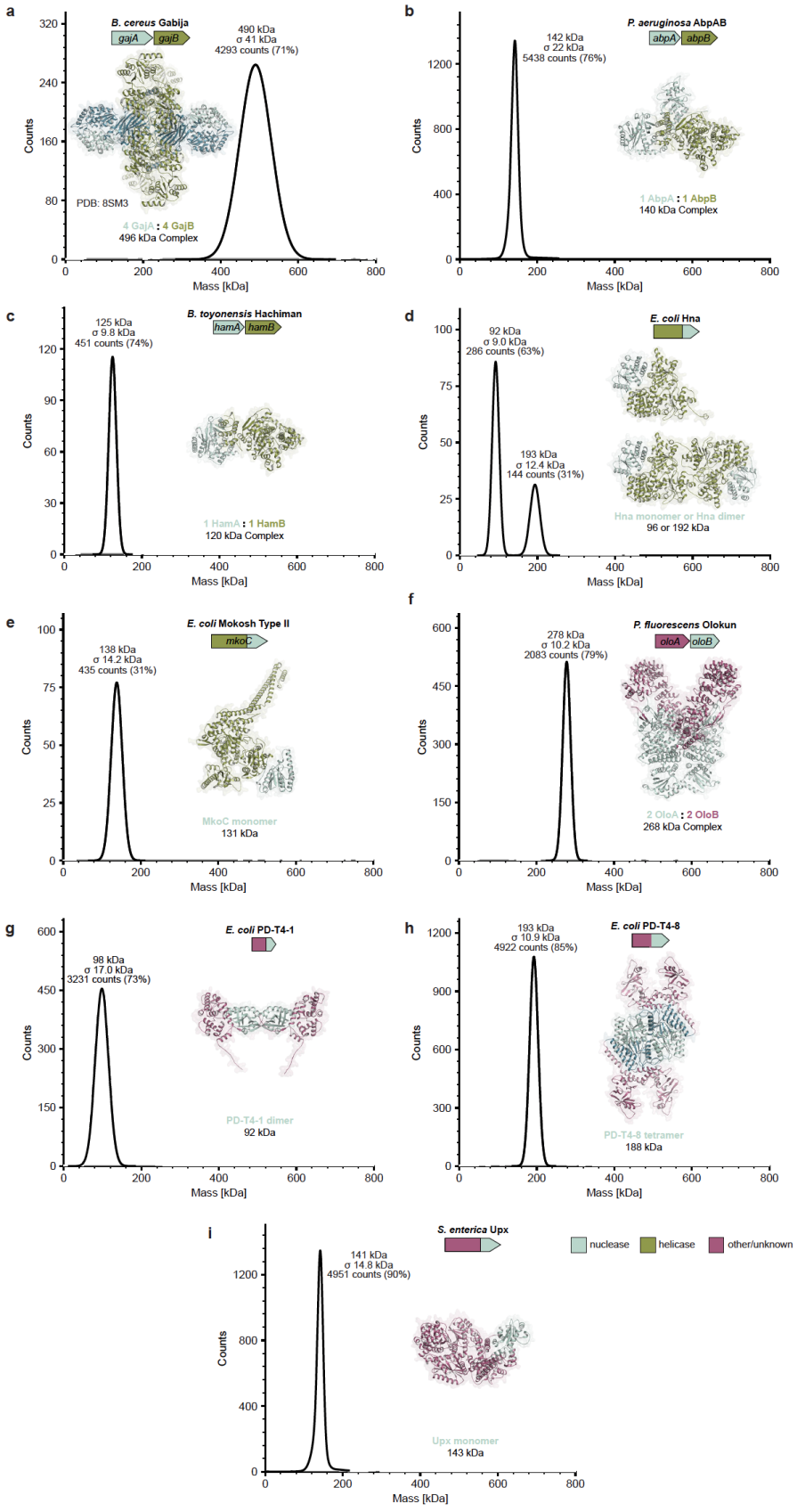


Figure 3.2 | Predicted domain architecture of nuclease-helicase systems.

a, Operon architecture of 16 selected nuclease-helicase anti-phage defense systems. Annotated domain architecture of each gene is shown in white, with the gene names above.

Figure 3.3 | Nuclease-helicase systems form oligomeric complexes of varying stoichiometries.

a-i, Mass photometry analysis of nuclease-helicase proteins with indicated molecular weight, standard deviation of the Gaussian fit, and particle counts above each peak. Data are representative of at least 2 independent experiments. **a,** Cartoon model of Gabija GajAB complex (PDB ID 8SM3). **b-i,** Cartoon models of AlphaFold3 (Abramson et al. 2024) predicted nuclease-helicase complexes based on mass photometry oligomeric state.


a
B. cereus Gabija
gajA
gajB
490 kDa
σ 41 kDa
4293 counts (71%)
PDB: 8SM3
4 GajA : 4 GajB
496 kDa Complex
b
P. aeruginosa AbpAB
abpA
abpB
142 kDa
σ 22 kDa
5438 counts (76%)
1 AbpA : 1 AbpB
140 kDa Complex
c
B. toyonensis Hachiman
hamA
hamB
125 kDa
σ 9.8 kDa
451 counts (74%)
1 HamA : 1 HamB
120 kDa Complex
d
E. coli Hna
92 kDa
σ 9.0 kDa
286 counts (63%)
193 kDa
σ 12.4 kDa
144 counts (31%)
Hna monomer or Hna dimer
96 or 192 kDa
e
E. coli Mokosh Type II
mkoC
138 kDa
σ 14.2 kDa
435 counts (31%)
MkoC monomer
131 kDa
f
P. fluorescens Olokun
oloA
oloB
278 kDa
σ 10.2 kDa
2083 counts (79%)
2 OloA : 2 OloB
268 kDa Complex
g
E. coli PD-T4-1
98 kDa
σ 17.0 kDa
3231 counts (73%)
PD-T4-1 dimer
92 kDa
h
E. coli PD-T4-8
193 kDa
σ 10.9 kDa
4922 counts (85%)
PD-T4-8 tetramer
188 kDa
i
S. enterica Upx
nuclease
helicase
other/unknown
141 kDa
σ 14.8 kDa
4951 counts (90%)
Upx monomer
143 kDa
Counts
Mass [kDa]

E. coli PD-T4-4, *E. hormaechei* Toprim-TA, and *Solibacillus silvestris* Menshen all exist as supramolecular complexes (Figure 3.4a–c). However, each complex has substantial standard deviations for the predicted mass, making it hard to decipher the exact stoichiometries of each subunit. This could be due to the dynamic nature of these complexes, with subunits coming on and off, or nonglobular structures as the standards for mass photometry are globular proteins. Structural analysis of these complexes is essential to determine the oligomeric state. *Ec* PD-T4-4 exists as two species in solution, an NTPase dimer and a 228 kDa +/− 22 kDa complex of unknown stoichiometry of NTPase to (P)D-(D/E)-XK nuclease (Figure 3.4a). *Eh*Toprim-TA exists as a single complex in solution at 197 kDa +/− 25 kDa (Figure 3.4b). *Ss*Menshen exists as two species at 177 kDa +/− 28 kDa and 351 +/− 40 kDa (Figure 3.4c). Like Gabija, *Ec*PD-T4-4, *Eh*Toprim-TA, and *Ss*Menshen form supramolecular complexes, which mirror human immunity where innate immune proteins, such as inflammasomes, form large protein complexes to exert antiviral function[180]. Overall, nuclease-helicase antiviral systems form complexes of varying oligomeric states indicating they may target nucleic acids in different ways.

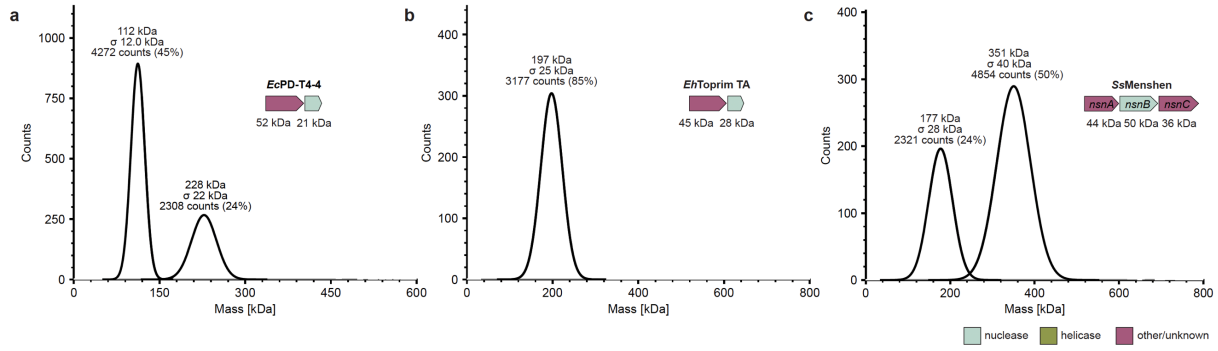


Figure 3.4 | *Ec*PD-T4-4, *Eh*Toprim TA, *Ss*Menshen form complexes of unclear stoichiometry.

a-c, Mass photometry analysis of nuclease-helicase proteins with indicated molecular weight, standard deviation of the Gaussian fit, and particle counts above each peak. Given the large standard deviation and molecular weight of each protein, the oligomeric state by mass photometry is unknown. Data are representative of at least 2 independent experiments

We next wanted to understand what type of nucleic acid nuclease-helicase systems target during phage infection. We reasoned that phage genomic DNA may be the most abundant substrate that these systems encounter during phage infection. To test this hypothesis we incubated the 14 nuclease-helicase proteins with a panel of mostly *E. coli* phage genomic DNAs (Figure 3.5a,b). We then looked for DNA cleavage/degradation by agarose gel electrophoresis. Testing each system against 11 DNA substrates led to an emerging pattern where systems either degrade all DNA tested, no DNA, or a specific type(s) of DNA (Figure 3.5b).

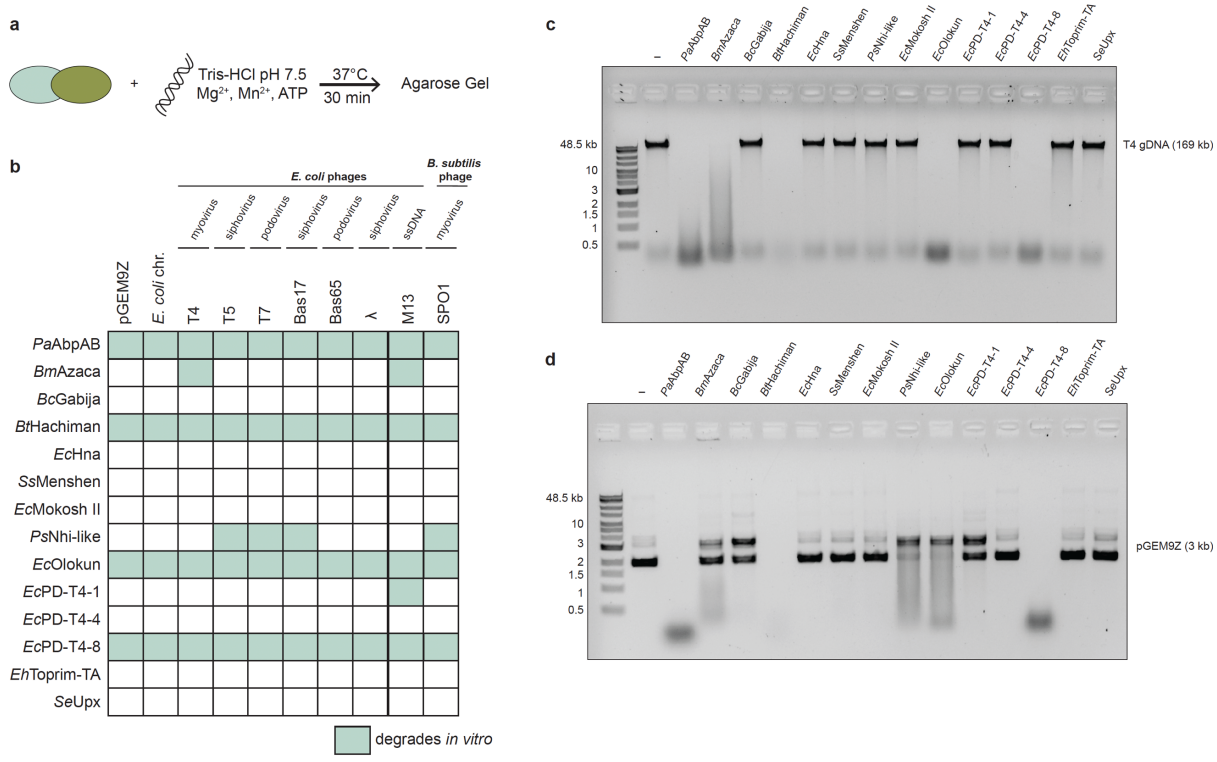


Figure 3.5 | Nuclease-helicase systems have different phage DNA cleave specificity.

a, Schematic of phage DNA cleavage assays. Nuclease-helicase systems were incubated with DNAs in b, for 30 min at 37°C followed by cleavage analysis by agarose gel electrophoresis. **b,** Phage DNA by nuclease-helicase system cleavage map. Systems that cleave DNA are in blue and white boxes indicate no cleavage. pGEM9z plasmid and BL21 *E. coli* chromosome were used as non-phage DNA controls. **c,** Agarose gel analysis of the ability nuclease-helicase systems to cleave phage T4 genomic DNA and **d,** pGEM9z plasmid DNA. Data are representative of at least 3 independent experiments.

*Pa*AbpAB, *Bt*Hachiman, *Ec*Olokun, and *Ec*PD-T4-8 cleave all *E. coli* phage DNAs in addition to a plasmid (pGEM9z), *E. coli* chromosomal DNA, and *Bacillus* phage SPO1 phage DNA (Figure 3.5b–d; Figure 3.6a–h). *E. coli* phage M13 genome is single-stranded indicating that these four systems can target both double-stranded and single-stranded DNA *in vitro* (Figure 3.6g,h). There is likely a mechanism by which bacteria inhibit these systems until a phage infection cue triggers nuclease activity. We can express and purify these systems from *E. coli*, albeit with expression tags, indicating they are not degrading host DNA *in vivo*. Supporting this hypothesis, following phage infection, *E. coli* Hachiman degrades both phage and host DNA leaving "phantom" cells devoid of nucleic acid in the cytoplasm[92]. Previous work shows that *Ec*Olokun defends against lambda-vir and SECphi27[24] and EcPD-T4-8 defends against T-even phages and SECphi18[25], which is indicative of a phage-specific cue that triggers DNA degradation in cells.

We previously showed that the *Bc*Gabija GajAB complex can cleave a palindromic dsDNA sequence derived from phage lambda DNA at pH 9 with no ATP in the reaction buffer[37]. Gabija does not cleave lambda DNA or any other substrate in these experiments due to more physiological buffer conditions (pH 7.5 and 1 mM ATP) (Figure 3.5b–d; Figure 3.6a–i).

*Ec*Hna, *Ss*Menshen, *Ec*PD-T4-4, *Eh*Toprim-TA, and *Se*Upx do not cleave any DNA substrate tested *in vitro* (Figure 3.5b–d; Figure 3.6a-i). Sensing a phage-specific cue other than phage genomic DNA may activate nuclease activity. For example, the PARIS defense system is activated when phage proteins, including phage T7 Ocr, bind to AriA, an ABC ATPase protein, which releases AriB, a Toprim domain protein, to cleave host lysine transfer RNA (tRNA)[50,51]. Future work will determine if nuclease-helicase defense systems are activated by specific phage proteins or detect specific phage DNA replication intermediates to halt phage infection. Previous work supports this hypothesis as phages escape AbpAB, Hachiman, and Hna by mutating phage single-stranded DNA binding proteins (SSBs), which bind to single-stranded DNAs produced during phage DNA replication and help with DNA polymerase processivity[69–71].

*Bm*Azaca, *Ps*Nhi-like, and *Ec*PD-T4-1 cleave specific phage DNAs. *Bm*Azaca is specific to T4 and M13 (Figure 3.5b,c; Figure 3.6g). The T4 genome contains glucosyl-hydroxymethylcytosines[181], and M13 is a single-stranded DNA phage. *Bm*Azaca defends against T-even phages, T2, T4, and T6[24], indicating that *Bm*Azaca likely cleaves phage DNA *in vivo* inhibiting T-even phage replication. *Ps*Nhi-like cleaves T5, T7, Bas17 and SPO1 DNA (Figure 3.5b; Figure 3.6b,c,d,h). SPO1 contains hydroxymethyluracil (hmU)

instead of thymine in its genome. T5, T7, and Bas17 do not have modified bases within their genomes. Therefore, *Ps*Nhi-like may target a specific DNA structure or sequence found within all five phage genomes. *Ec*PD-T4-1 cleaves M13 phage DNA only (Figure 3.5b; Figure 3.6g). PD-T4-1 was discovered in an *E. coli* screen for new defense systems and protects against T-even phages[25]. M13 phage DNA *in vitro* may mirror a T-even phage DNA replication intermediate that *Ec*PD-T4-1 targets *in vivo*. Further research will determine the patterns of phage genomic DNA cleavage vs. phage defense *in vivo* for all nuclease-helicase systems. Phage DNA cleavage *in vitro*, but no defense *in vivo* does not rule out the physiological relevance of DNA cleavage as phages can also encode proteins that directly block defense systems. For example, *Bacillus* phage Phi3T encodes Gabija anti-defense 1 (Gad1), which binds in an octameric web around the *Bc*Gabija GajAB complex preventing DNA targeting[37].

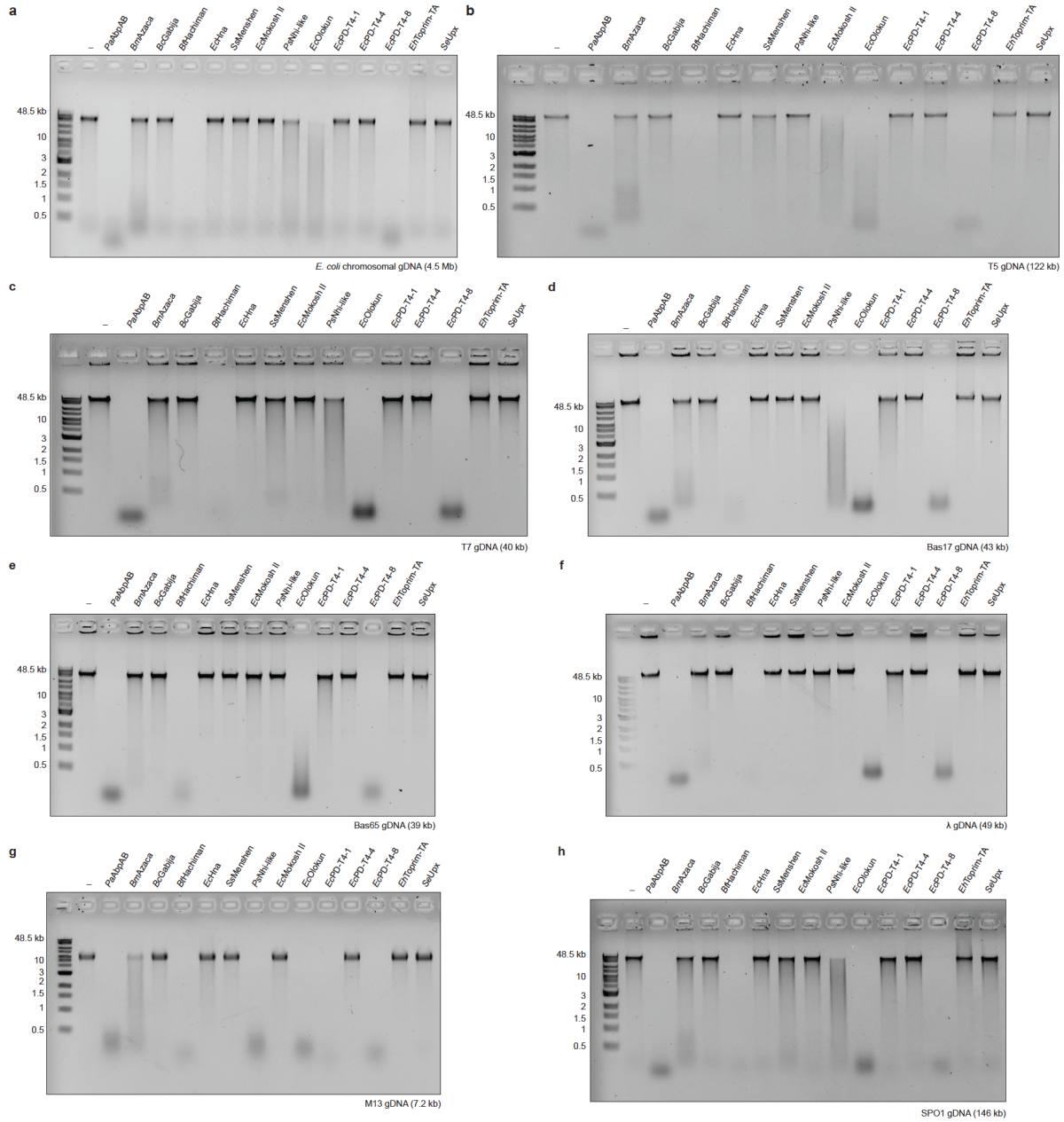


Figure 3.6 | Nuclease-helicase systems phage DNA cleavage agarose gel analysis.

a-h, Agarose gel analysis of nuclease-helicase system DNA cleavage for all substrates shown in **Figure 3.5b** map. Data are representative of at least 3 independent experiments.

Next, we looked at the active site/conserved residue requirements for phage DNA cleavage. We made double point mutations in previously annotated[22–25,70] or highly conserved residues of the nuclease, helicase, and other/unknown proteins and expressed and purified each system. For AbpAB, mutation of the DUF4297 nuclease active site residues in ApbA resulted in the loss of complete degradation of pGEM9z plasmid DNA, but still resulted in plasmid nicking resulting in an upshift of the nicked and relaxed plasmid (Figure 3.7a). Interestingly, mutation of the key residues in the DEAD-box helicase motif of AbpB still allowed for complete plasmid degradation indicating that *in vitro* AbpB helicase activity in complex with AbpA is not required to target DNA. Olokun also does not require conserved residues in the accessory OloA protein (Figure 3.7b), as mutation of a DLL motif[182] does not inhibit T4 phage DNA cleavage. Mutation of nuclease active site residues in OloB abolishes DNA cleavage (Figure 3.7b). OloA likely functions more as a scaffold protein to position phage DNA for cleavage by OloB. For Azaca, mutations in the DUF6361 domain of ZacA, PLD-like nuclease residues in ZacB, and Walker A helicase motif in ZacC, all led to the inhibition of T4 DNA cleavage (Figure 3.7c). ZacA mutations still allowed for partial DNA cleavage, but not to WT Azaca ZacABC complex level. Therefore, all catalytic active sites of the Azaca complex are required and work together for phage DNA targeting. Mutation of both the DUF1837 nuclease active site in Hachiman HamA and the helicase motifs of HamB resulted in almost complete loss of cleavage activity except for nicking of the pGEM9z plasmid and slight degradation for the HamB helicase mutant (Figure 3.7d). This recapitulates data from *E. coli* Hachiman where HamB unwinds DNA that is then fed into the HamA nuclease active site[92]. Mutation of the (P)D-(D/E)-XK nuclease active site abolishes DNA cleavage activity for both *Ec*PD-T4-1 and *Ec*PD-T4-8 (Figure 3.7e,f).

These results distinguish two groups of nuclease-helicase systems, one group that requires all active sites for DNA cleavage: Azaca, Hachiman, PD-T4-1, PD-T4-8, and another group that only requires that active site of the respective nuclease: AbpAB and Olokun. For AbpAB and Olokun, AbpB and OloA might more generally serve as scaffolds that orient DNA to be cleaved during phage infection. Azaca, Hachiman, PD-T4-1, and PD-T4-8 likely scan DNA and identify phage-derived DNA, which requires all catalytic activities of the individual proteins (Figure 3.7g).

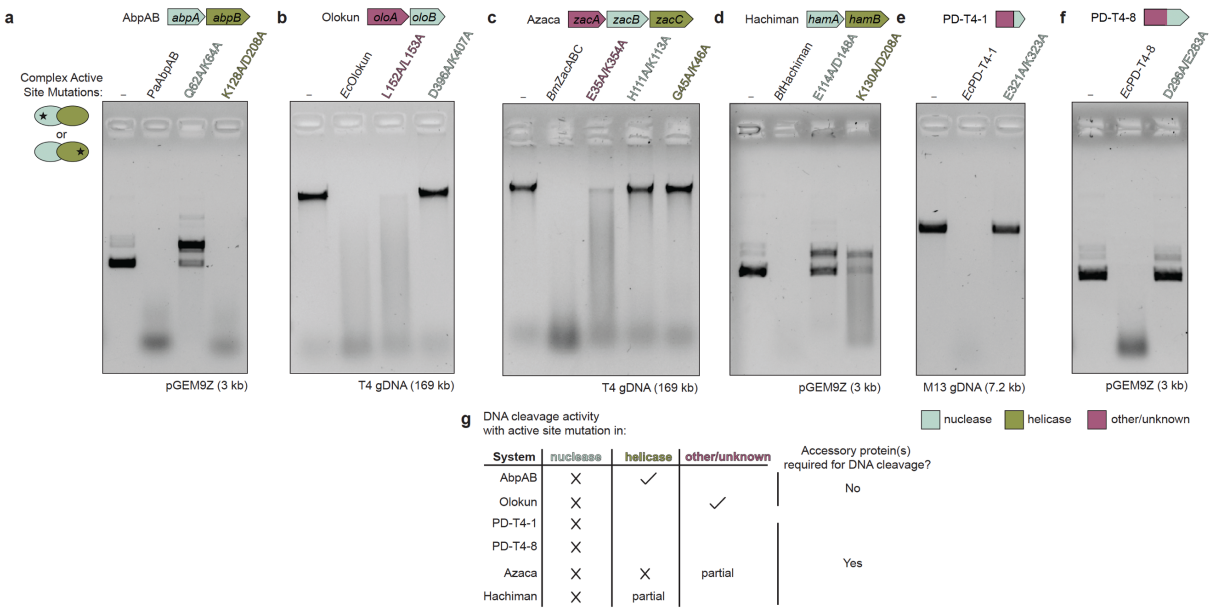
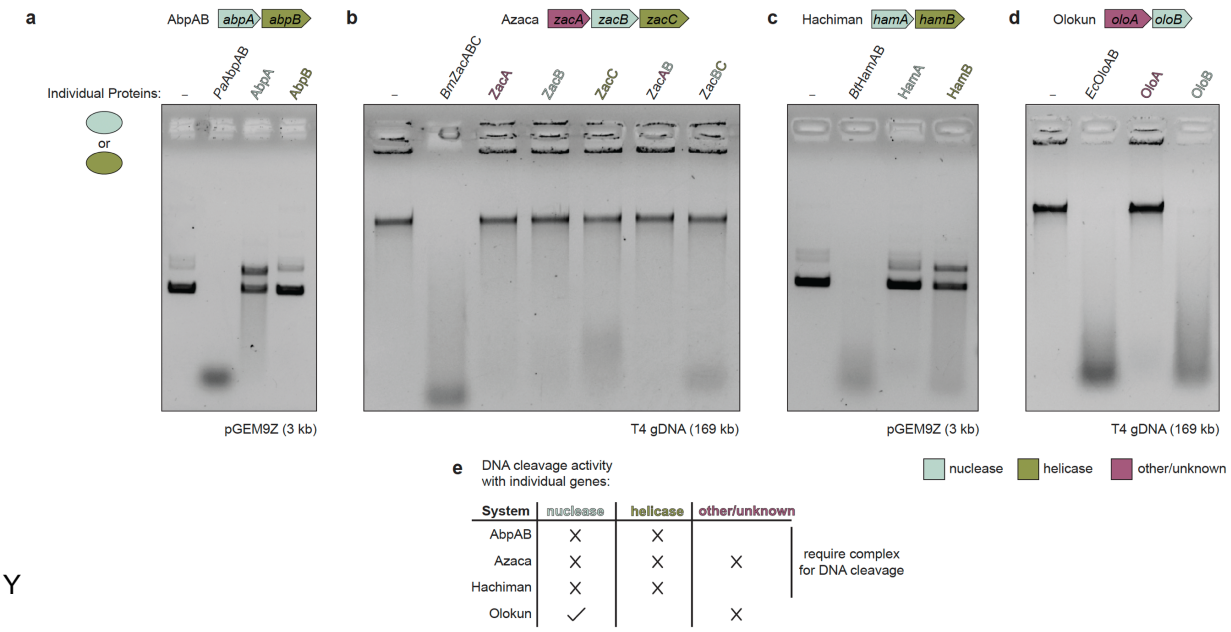


Figure 3.7 | Nuclease-helicase catalytic active site requirements for phage DNA cleavage.

a-f, Agarose gel analysis of t wild-type (WT), nuclease-, helicase-, other/unknown- catalytic active site mutants on a representative plasmid (pGEM9z) or phage DNA cleavage. Active site mutations are indicated by color above each well for the nuclease, helicase, or other/unknown proteins. Data are representative of at least 3 independent experiments. **g,** Summary of phage DNA cleavage results by system and whether mutating the predicted active site of the nuclease, helicase, or other/unknown stops (X) or still allows (check mark) for cleavage.

To understand if nuclease-helicase systems are activated by releasing the nuclease protein from the complex, we expressed and purified each of the proteins individually and tested phage DNA cleavage activity. The accessory protein(s) could inhibit nuclease cleavage activity until a phage infection cue is triggered leading to a conformational change and release of the nuclease protein. For AbpAB, AbpA nuclease alone does not completely degrade pGEM9z plasmid DNA but nicks it, while as expected AbpB helicase protein alone does not cleave DNA (Figure 3.8a). This again points to a more structural role of AbpB where it needs to bind to AbpA for nuclease activity but does not require helicase active site residues to cleave *in vitro* (Figure 3.7a). For *Bm*Azaca, ZacA (DUF/unknown), ZacB (nuclease), ZacC (helicase) alone, or a combination of the nuclease ZacB with ZacA (ZacAB) or ZacC (ZacBC), do not cleave T4 DNA (Figure 3.8b). Intriguingly there is a smaller shadow of DNA present in all lanes containing the helicase ZacC indicating ZacC may be bound by or generate a small DNA product (Figure 3.8b). For Hachiman, the HamAB complex is required for DNA degradation as there is no DNA cleavage for each of the individual nuclease and helicase proteins (Figure 3.8c). Finally, for Olokun, OloB alone cleaves phage DNA (Figure 3.8d). This is the only example of a mechanism where OloA may restrain OloB nuclease activity until a phage-generated cue is triggered. Overall, three out of the four nuclease-helicase systems tested required all proteins in a complex to cleave phage DNA: AbpAB, Azaca, and Hachiman (Figure 3.8e).



Y

Figure 3.8 | Nuclease-helicase complex formation requirements for phage DNA cleavage.

a-d, Agarose gel analysis of the individual nuclease, helicase, other/unknown proteins to cleave either pGEM9z or phage DNA. **b,** For Azaca, ZacB the nuclease, and ZacAB, and ZacBC complexes were tested for T4 phage DNA cleavage. Data are representative of at least 3 independent experiments. **e,** Summary of phage DNA cleavage results by system and with the nuclease, helicase, or other/unknown protein stops (X) or still allows (check mark) for cleavage.

All DNA cleavage reactions were performed at 1 μM protein concentration, which may not be a physiologically relevant concentration or might represent a state of DNA cleavage where nuclease-helicase systems need to undergo molecular crowding. To test this hypothesis, we performed a protein concentration titration for each active phage DNA degrading nuclease-helicase system. Using concentrations of 10 nM to 1 μM we found that *Pa*AbpAB, *Bm*Azaca, and *Ec*PD-T4-1 all start to cleave DNA at low concentrations (Figure 3.9). Therefore, these systems can potentially exert cleavage as individual complexes. *Bt*Hachiman, *Ps*Nhi-like, *Ec*Olokun, and *Ec*PD-T4-8 completely degrade phage or plasmid DNA at 1 μM (Figure 3.10). These systems may be activated by a phage-derived DNA that allows for many protein complexes to load onto, leading to activation of the system and complete DNA degradation.

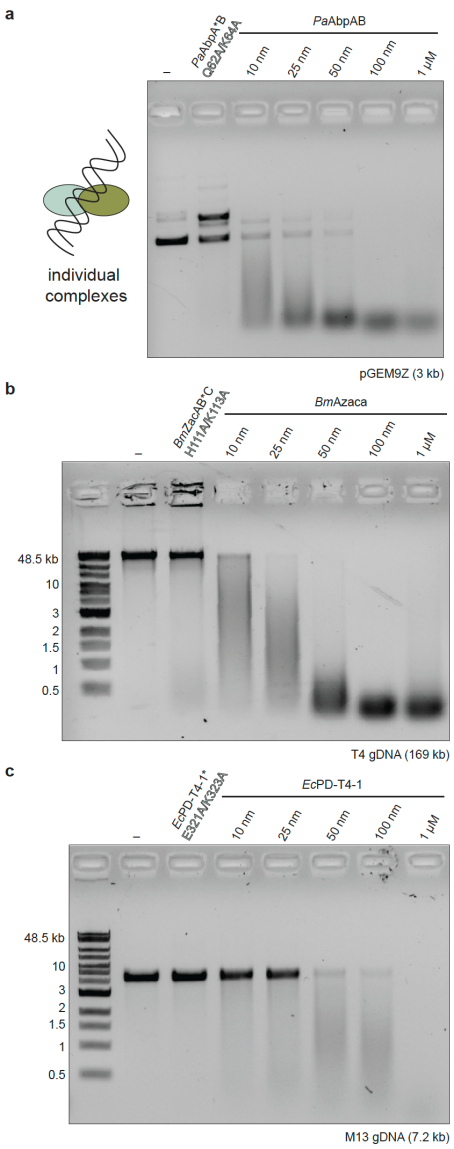


Figure 3.9 | AbpAB, Azaca, and PD-T4-1 cleave DNA at low concentrations.

a-c, Agarose gel analysis of a, *Pa*AbpAB, b, *Bm*Azaca, and c, *Ec*PD-T4-1 at 10 nM – 1 μM concentrations of proteins. Catalytic dead nuclease complexes were used as a negative control. Data are representative of at least 3 independent experiments.

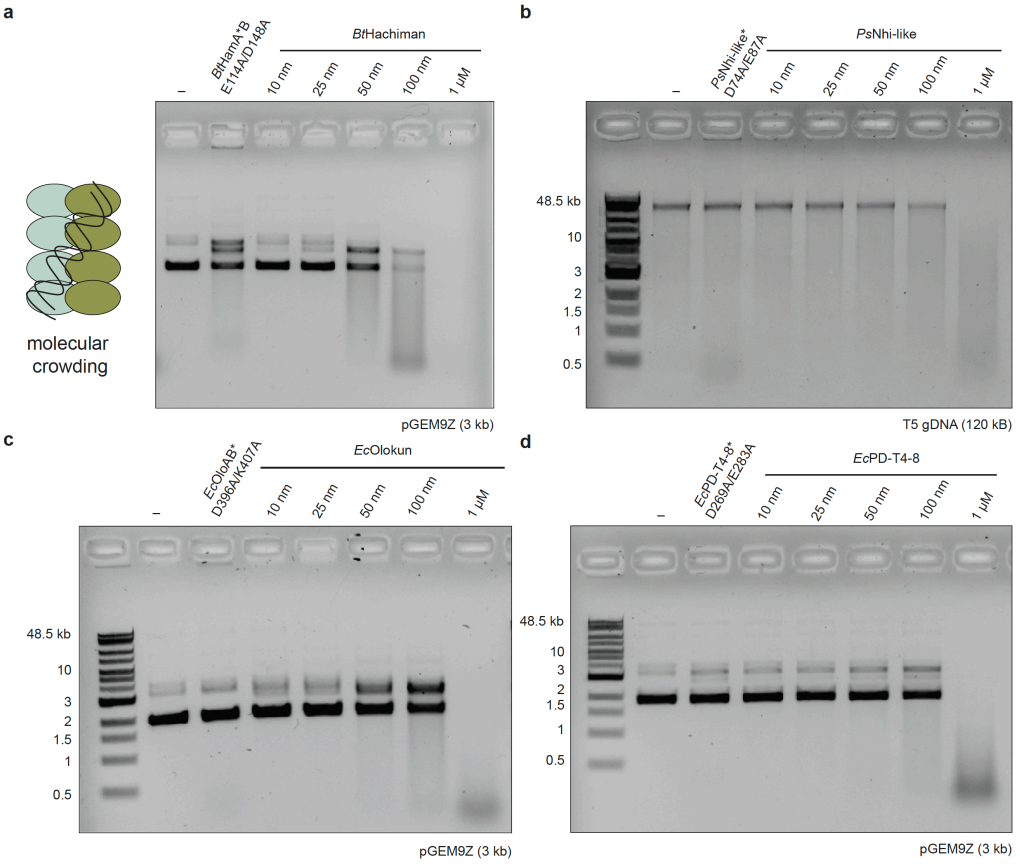


Figure 3.10 | Hachiman, Nhi-like, Olokun, and PD-T4-8 cleave DNA at high concentrations.

a-d, Agarose gel analysis of a, *Bt*Hachiman, b, *Ps*Nhi-like, c, *Ec*Olokun, *and d, Ec*PD-T4-8 at 10 nM – 1 µM concentrations of proteins. Catalytic dead nuclease complexes were used as a negative control. Data are representative of at least 3 independent experiments.

Discussion

This work highlights that the mechanisms by which nuclease-helicase systems defend against phage infection cannot be understood only by operon architecture alone. Nuclease-helicase systems can range from 1:1 structures to supramolecular complexes (Figure 3.11). The type of nuclease encoded by the system does not predict the oligomeric state of the system (Figure 3.3). Predicting the structures of nuclease-helicase systems reveals a great diversity in both complex composition and overall 3D structure of the complexes. While the core shape of the globular helicase proteins/domains and the beta-strand core of nucleases are conserved, the secondary structures of the accessory domains that decorate the active sites are drastically different. These accessory domains mostly form the interaction interfaces with other protomers in the complex forming structures that range from 1:1 to butterfly complexes. Evolutionarily this makes sense as interaction interfaces near the nuclease and helicase active sites would need to maintain a conserved structure for catalytic activity. However, when interaction interfaces are in accessory domains that do not directly influence catalytic activity, they can mutate to maintain or diversify complex formation. Our work highlights the importance of complex formation as most active nuclease-helicase systems require a complete protein complex to cleave DNA.

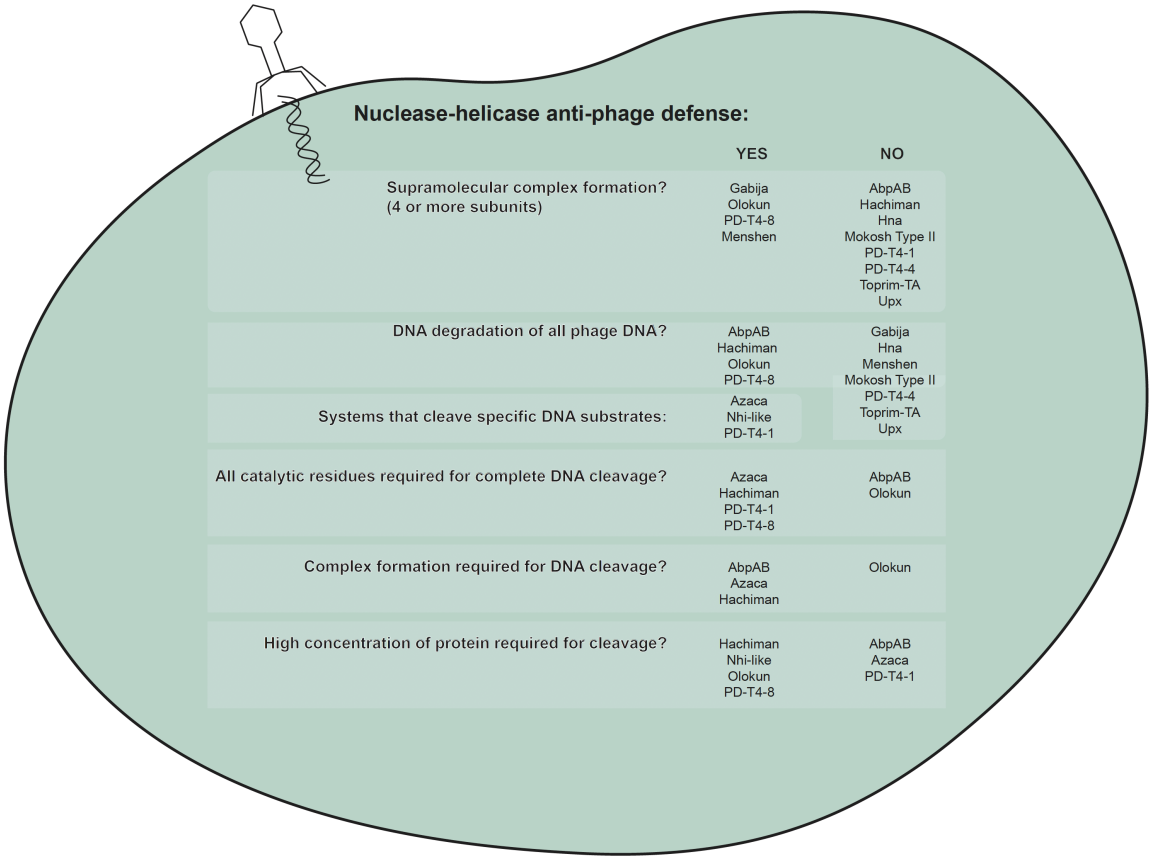


Figure 3.11 | A compendium of nuclease-helicase biochemical activity.

Summary of all biochemical results for nuclease-helicase systems.

Some systems, AbpAB, Hachiman, Olokun, and PD-T4-8 cleave all DNA *in vitro*, which leads to the question of whether these systems cause abortive infection. AbpAB, Hachiman, PD-T4-8 were previously shown to be abortive infection systems[25,71,92], supporting this hypothesis. When phage infection is sensed, these systems likely cleave both host and phage-derived DNA to save the rest of the bacterial colony. Unlike restriction-modification systems that cleave specific DNA sequences, these systems may indiscriminately cleave nucleic acid once activated. Sensing of DNA damage, which includes dsDNA breaks, could be a common trigger to unleash DNA cleavage by nuclease-helicase systems.

For systems that cleave only specific phage DNA including Azaca and PD-T4-1, phage defense may occur by targeting a specific DNA sequence or structure. Azaca defends against T-even phages and cleaves phage T4 DNA, which all contain hydroxymethylcytosines bases[24,181]. Future work will determine if genome modification is involved in Azaca DNA cleavage by sequencing DNA degradation products.

There are nuclease-helicase systems in our screen that cleave no phage DNA substrates including Gabija, Hna, Menshen, Mokosh Type II, PD-T4-4, Toprim-TA, and Upx. These systems could potentially cleave RNA and/or protect against RNA phages or need a specific phage infection-induced cue *in vivo* to cleave nucleic acid. Nuclease activity could be activated by binding a specific phage protein or the activity of phage proteins on DNA and therefore *in vitro* phage genomic DNA alone would not lead to activation. From a more biochemical standpoint, these systems may function enzymatically under different buffer conditions. In this study, we used, pH 7.5 150 mM Salt and 1 mM ATP, which is a representative condition for nuclease activity *in vitro*, but enzyme buffer requirements can vary.

Molecular crowding or phase separation may be an activation signal for nuclease-helicase systems that are only active at high concentrations (1 μM): Hachiman, Nhi-like, Olokun, and PD-T4-8. For example, the BtsA defense system is diffuse within cells before infection, but then forms puncta on replicating phage DNA 20 min after infection, indicating that local concentration of the protein on phage DNA may activate defense[183].

When expressing individual proteins of the Olokun system, OloA (scaffold) and OloB (nuclease), we realized that OloB runs at both its expected and double the expected size on SDS-PAGE gels, which is around the size of OloA. Therefore, in our OloAB complex preps, we determined we mostly have OloB protein alone and that OloA comes off in another peak from size exclusion chromatography. All the data we present here is OloB nuclease with only a minimal concentration of OloA. Future work will determine if OloA restricts DNA cleavage by OloB and why OloB runs at double the expected size by SDS-PAGE.

Overall, this work sets the foundation to understand the molecular mechanisms by which nuclease-helicase systems defend against phage infection. Future work to understand how these systems directly target phage DNA by structure or sequence will reveal more commonalities and reasons for evolving so many nuclease-helicase systems.

Methods

Nuclease-helicase system selection and plasmid construction

Designated nuclease-helicase systems were chosen from the literature with the one requirement of having a nuclease-domain encoded in the operon. If applicable, the operon used in the discovery paper and a diverse homolog from a different bacterial species were cloned into custom pET vectors with an N-terminal 6×His-SUMO2 fusion tag on the first protein in the operon or a C-terminal 6×His tag on the last gene in the operon. Operons were not codon optimized leaving the sequence of the operon intact from the start codon of the first gene to the stop codon of the last gene. Individual genes were cloned into either pET vectors with an N-terminal 6×His-SUMO2 fusion tag on the first protein in the operon or a C-terminal 6×His tag. Full operon expression vectors and active site mutants used for protein expression and purification are as follows: N-terminal 6×His-SUMO2 *Pa*AbpAB, N-terminal 6×His-SUMO2 *Bm*Azaca, 6×His-SUMO2-5×GS *Bc*Gabija, N-terminal 6×His-SUMO2 *Bt*Hachiman, *Ec*Hna C-terminal 6×His tag, *Ss*Menshen C-terminal 6×His tag, *Ec*Mokosh II C-terminal 6×His tag, 6×His-SUMO2-5×GS *Ps*Nhi-like, *Ec*Olokun C-terminal 6×His tag, *Ec*PD-T4-1 C-terminal 6×His tag, *Ec*PD-T4-4 C-terminal 6×His tag, 6×His-SUMO2-5×GS *Ec*PD-T4-8, *Eh*Toprim-TA C-terminal 6×His tag, and N-terminal 6×His-SUMO2 *Se*Upx. Individual protein expression vectors are as follows: N-terminal 6×His-SUMO2 *Pa*AbpA, N-terminal 6×His-SUMO2 *Pa*AbpB, N-terminal 6×His-SUMO2 *Bm*ZacA, N-terminal 6×His-SUMO2 *Bm*ZacB, N-terminal 6×His-SUMO2 *Bm*ZacC, N-terminal 6×His-SUMO2 *Bm*ZacAB, N-terminal 6×His-SUMO2 *Bm*ZacBC, N-terminal 6×His-SUMO2 *Bt*HamA, N-terminal 6×His-SUMO2 *Bt*HamB, N-terminal 6×His-SUMO2 *Ec*OloA, and N-terminal 6×His-SUMO2 *Ec*OloB.

Protein expression and purification

Nuclease-helicase complexes were expressed in *E. coli* as previously described[111]. Briefly, expression vectors were transformed into BL21(DE3)-RIL (Agilent), or LOBSTR-BL21(DE3)-RIL cells (Kerafast), plated on MDG media plates (1.5% Bacto agar, 0.5% glucose, 25 mM Na_2HPO_4, 25 mM KH_2PO_4, 50 mM NH_4Cl, 5 mM Na_2SO_4, 0.25% aspartic acid, 2–50 µM trace metals, 100 µg mL^{-1} ampicillin, 34 µg mL^{-1} chloramphenicol) and grown overnight at 37°C. Five colonies were picked to inoculate 30 mL overnight MDG starter cultures (37°C 230 rpm). 1 L M9ZB expression cultures (47.8 mM Na_2HPO_4, 22 mM KH_2PO_4, 18.7 mM NH_4Cl, 85.6 mM NaCl, 1% Cas-Amino acids, 0.5% glycerol, 2 mM $MgSO_4$, 2–50 µM trace metals, 100 µg mL^{-1} ampicillin, 34 µg mL^{-1} chloramphenicol) were then inoculated with 10 mL or starting OD_{600} of

0.0475 of MDG starter cultures and then induced with 0.5 M IPTG after reaching an OD_{600} of ≥1.5 followed by overnight induction (16°C, 230 rpm).

After overnight inductions, cells were centrifuged, resuspended, and pellets were lysed by sonication in 60 mL lysis buffer (20 mM HEPES pH 7.5, 400 mM NaCl, 10% glycerol, 20 mM Imidazole, 1 mM DTT). Lysate was clarified by centrifugation, and resulting supernatants was added to Ni-NTA resin (Qiagen). Resin was washed with lysis buffer, lysis buffer supplemented to 1 M NaCl and then lysis buffer again for *Pa*AbpAB, *Bc*Gabija, *Bt*Hachiman, *Ec*Hna, *Ec*Mokosh II, *Ps*Nhi-like, *Ec*PD-T4-1, and *Se*Upx. The resin was washed with lysis buffer three times for *Bm*Azaca, *Ss*Menshen, *Ec*Olokun, *Ec*PD-T4-4, and *Eh*Toprim-TA. Samples were eluted with lysis buffer supplemented to 300 mM Imidazole. Samples were then dialyzed overnight in 14 kDa MWCO dialysis tubing (Ward's Science) with SUMO2-cleavage by hSENP2 as previously described[111]. *Bm*Azaca was dialyzed and stored in gel filtration buffer (20 mM HEPES-KOH pH 7.5, 20 mM KCl, and 1 mM TCEP-KOH) supplemented with 10% glycerol. *Ec*Mokosh II was next purified by ion-exchange chromatography using HiTrap Heparin HP columns (GE Healthcare) and eluted by a 150 mM – 1 M NaCl gradient. Fractions were pooled and further purified by size exclusion chromatography using a 16/600 Sephacryl 300 column (Cytiva) and stored in gel filtration buffer. *Bc*Gabija and *Ps*Nhi-like were purified by size exclusion chromatography using a 16/600 Sephacryl 300 column (Cytiva) and stored in gel filtration buffer. *Pa*AbpAB, *Bt*Hachiman, *Ec*Hna, *Ss*Menshen, *Ec*Olokun, *Ec*PD-T4-1, *Ec*PD-T4-4, *Ec*PD-T4-8, *Eh*Toprim-TA, and *Se*Upx were purified by size exclusion chromatography using a 16/600 Superdex 200 column (Cytiva) and stored in gel filtration buffer. Final proteins were concentrated to >10 mg mL^{-1} using a 30 kDa MWCO centrifugal filter (Millipore Sigma), aliquoted, flash frozen in liquid nitrogen, and stored at −80°C.

All catalytic active site mutants were purified by Ni-NTA, dialyzed and stored in gel filtration buffer (20 mM HEPES-KOH pH 7.5, 20 mM KCl, and 1 mM TCEP-KOH) supplemented with 10% glycerol with comparison to the WT protein/complex. Individual proteins from each system were purified by Ni-NTA dialyzed and stored in gel filtration buffer (20 mM HEPES-KOH pH 7.5, 20 mM KCl, and 1 mM TCEP-KOH) supplemented with 10% glycerol with comparison to the full protein/complex.

Mass Photometry

Refeyn MP was used at Center for Macromolecular Interactions at Harvard Medical School, RRID:SCR_018270. The oligomeric state of purified proteins was determined by mass photometry (TwoMP, Refeyn). Purified proteins were diluted to 200-400 nM in gel filtration buffer prior to analysis. Isopropanol and MilliQ-H_2O cleaned coverslips were loaded with gaskets and placed on the objective lens with a drop of Immersol immersion oil on it (Zeiss). AcquireMP software (Refeyn) was used to find focus on gel filtration buffer only, and then beta-amylase and thyroglobulin control or samples were added to the buffer 1:10 followed by immediate recording for 60 seconds. Ratiometric contrast measurements were converted to molecular weight measurements using DiscoverMP software (Refeyn). Automatic Gaussian curve fitting was performed on all oligomeric states. *Pseudomonas fluorescens* NZ007 Olokun was used instead of *Ec*Olokun for mass photometry due to better complex stability.

AlphaFold3 Modeling

AlphaFold3 server[176] was used to predict the structure of nuclease-helicase systems based on oligomeric states by mass photometry. All 5 models were compared in ChimeraX[116] using MatchMaker. All models for each system looked relatively similar and model 0 is presented in the figures.

Phage gDNA extraction

5 mL LB cultures supplemented with 5 mM $MgCl_2$, 5 mM $CaCl_2$, 1 mM $MnCl_2$ were inoculated with a single BW25113 bacterial colony (overnight, 37°C, 230 rpm). Using the overnight culture, 20 mL fresh LB supplemented with 5 mM $MgCl_2$, 5 mM $CaCl_2$, and 1 mM $MnCl_2$ were inoculated at an OD_{600} of 0.1 and grown at 37°C 230RPM until an OD_{600} of 0.3 was reached. Cultures were then infected at MOI of 3 with phage and incubated until cultures collapsed (~ 3 hours). Phage-infected cultures were centrifuged, and supernatants were passed through 0.22 µm filter. Phage supernatants were treated with DNase I and RNase A for 30 minutes at 37°C. 10 mL of phage precipitant solution (30% PEG 8000, 3M NaCl) was added to phage supernatants and incubated at 4°C overnight. Phage particles were pelleted by spinning at 10,000 x g for 10 minutes at 4°C. Resulting supernatants were decanted, and phage pellets were resuspended in 500 µL 10 mM EDTA and 5 mM $MgSO_4$ and incubated with 100 µg/mL proteinase K (50°C 30 min). Phage genomic DNA was then purified using Wizard DNA Clean-Up Kit (Promega). Briefly, phage DNA was mixed with 1 mL Wizard DNA Clean-Up Resin and added to a column/resin. DNA was washed with 2 mL of 80% isopropanol, dried by centrifugation at max speed 2 min, and eluted with 100 µL pre-warmed 80ºC distilled

water. For *Bacillus* Phage SPO1, phage was propagated using PY79 cells. pGEM9z was prepped by maxi-prep (Qiagen). BL21 gDNA was isolated using a DNeasy blood and tissue kit (Qiagen). M13mp18 Single-stranded DNA and Lambda DNA were purchased from NEB.

DNA Cleavage Assays

Phage gDNA or plasmid DNA substrates were incubated with indicated proteins in a 20 µL cleavage reaction with 200 ng DNA, 10 nM – 1 µM protein, 20 mM Tris-HCl pH 7.5, 150 mM KCl, 5 mM $MgCl_2$, 1 mM $MnCl_2$, 1 mM $MnCl_2$, 1 mM DTT. Proteins were diluted in gel filtration buffer supplemented with 10% glycerol. Reactions were incubated at 37°C for 30 min followed by 100 µg/mL Proteinase K treatment at 50°C for 30 min. DNA loading buffer containing 60 mM EDTA was added to the reactions and 10 µL was loaded onto a 0.8-1% TAE agarose gel run at 150 V for 30 min at room temperature. Gels were post-stained with TAE supplemented with 10 µg mL^{-1} ethidium bromide for 30 min, de-stained in water for 30 min, and imaged on a ChemiDoC MP Imaging System.

Acknowledgements

We thank the members of the Kranzusch Lab for helpful discussions and the Sorek lab for providing plasmids for Millman et al. 2022 nuclease-helicase systems. We thank Artem Egorov, Olena Shyrokova, Vasili Hauryliuk, and Gemma Atkinson for sharing the sequences of *E.coli*-like Toprim_TA systems. We thank the Center for Macromolecular Interactions at Harvard Medical School, RRID:SCR_018270 for the use of Refeyen TwoMP. The work was funded by grants to P.J.K. from the Pew Biomedical Scholars program, the Burroughs Wellcome Fund PATH program, the Mathers Foundation, the Mark Foundation for Cancer Research, the Cancer Research Institute, the Parker Institute for Cancer Immunotherapy and the National Institutes of Health (1DP2GM146250-01).

Author Contributions

Adelyn E. Ragucci: Designed and conceived the study, performed protein purification and DNA cleavage assays, and made figures.

Sadie P. Antine: Designed and conceived the study, performed protein purification, DNA cleavage assays, and mass photometry. Made figures and wrote the thesis chapter.

Ethan M. Leviss: Performed protein purification and DNA cleavage assays.

Sarah E. Mooney: Performed protein purification and DNA cleavage assays.

Philip J. Kranzusch: Designed and conceived the study.

Discussion

By studying anti-phage defense, new biology reveals itself in intricate mechanisms that bacteria use to fight the most abundant biological entity on earth, bacteriophages. Bacteria are constantly being infected by phages leading to a never-ending battle that drives biological innovation. The discovery of restriction-modification systems led to a revolution in molecular biology, which I believe most researchers don't realize that bacteria evolved these strategies to target foreign DNA. This sentiment is also true for CRISPR systems, which we still don't the diversity of or true potential for biotechnology. With restriction-modification and CRISPR in mind, these two anti-phage defense systems cannot be the only systems that will lead to new biological tools. Therefore, by studying the basic biology of the bacteria-phage arms race we can glean new molecular mechanisms of antiviral signaling, but also hopefully use these strategies for new biotechnology.

In addition to biotechnology, studying the bacteria-phage arms race can help the fight against anti-microbial resistant bacteria (AMR)[184]. Cases of AMR continue to rise, while the discovery of new antibiotics diminishes. Instead of using small molecular antibiotics, the potential use of phages to kill AMR bacteria has reinvigorated interest[185]. However, anti-phage defense systems that block phages impede the development of phage therapy. Understanding the mechanisms by which phages overcome defense systems will help "design" phages that can overcome the immune system of a specific AMR bacteria. This starts by understanding the molecular mechanisms by which phages defeat anti-phage defense systems to identify phage anti-defense proteins that have minimal toxicity and broad anti-defense targeting. The best phages for phage therapy will likely have anti-defense genes that target more than one defense system at a time.

The fight between bacteria and phage serves as a reservoir of untapped biology to understand the evolution of antiviral immunity. There is now clear evidence that the human immune system likely evolved from bacterial anti-phage defense[7,186]. While many mechanisms in anti-phage defense mirror the mechanisms of mammalian immunity, we can now uncover new molecular mechanisms humans use for antiviral defense by looking at conservation between newly identified anti-phage defense systems and proteins of unknown function in humans. A Silent Information Regulator 2 (SIR2) domain protein was

recently discovered to function in the human TLR4 innate immune signaling pathway because of its homology to SIR2 domains found in Thoeris and CBASS anti-phage defense systems[187].

Supramolecular complex formation in Gabija anti-phage defense

Just looking at the operon structure of Gabija would never have led to the finding that it forms a 4:4 nuclease-helicase complex (Chapter 1). Structural biology is necessary for understanding the full molecular mechanisms of anti-phage defense systems. Supramolecular complex formation is now a widespread theme in anti-phage defense[188]. For Gabija, a GajA nuclease tetramer forms two complete Toprim active site dimers on either side of the complex. Helicase dimers then bind to either side of the complex along the ABC ATPase interfaces of GajA. Forming such a large complex likely allows GajB helicase activity to regulate GajA ABC ATPase activity, which could lead to a conformation change required for GajA to bind target DNA sequences[39]. Gabija may be activated by sensing dsDNA damage through GajB helicase activity. Rep and UvrD helicase act as dimers and unwind dsDNA breaks to initiate a DNA damage response. Instead of initiating a DNA damage response, GajB could activate GajA nuclease activity to degrade phage, host, or both DNA. However, GajB has been shown to have minimal helicase activity *in vitro*[36,56]. Therefore, there is likely something *in vivo* that cannot yet be recapitulated with purified components that leads to Gabija activation. DNA is rarely free from DNA-binding proteins, and transcription and replication complexes in cells. Gabija may sense the activities of proteins on nucleic acid that are specific to phage infection. Forming a supramolecular complex therefore allows activation to be tightly regulated and likely leads to conformational changes that distinctly occur during phage infection to prevent autoactivation in the absence of phage infection. Our work shows that key interface mutations between GajA–GajA and GajA–GajB protomers results in loss of phage defense against SpBeta (Chapter 1). While many of the mutations disrupted GajAB complex formation, mutations in GajA–GajA still form a complex indicating that conformational changes that are governed by these interaction interfaces allow for phage DNA targeting.

GajA is an OLD nuclease, which are implicated in sensing RecBCD inhibition during phage P2 infection. P2 OLD protein alone can exert this activity, however, there are slightly differences in the protein architecture of P2 OLD and GajA, including alpha-helical extensions in the Toprim domain of GajA[37,49].

RecBCD binds to linear dsDNA that are generated from dsDNA breaks or phage replication. GajB could potentially sense DNA damage accumulated from RecBCD inhibition, which could then activate GajA to cleave this DNA. Future genetic work will determine if RecBCD inhibition activates Gabija anti-phage defense.

Fully uncovering the mechanism by which Gabija is activated could potentially lead to its use in biotechnology. Just like CRISPR systems, there is likely a lot of biological diversity among Gabija systems from different species, so assessing Gabija activity across this diversity will aid in understanding key differences and similarities in activation and antiviral defense.

Gabija anti-defense 1 wraps around Gabija for immune evasion

The Gabija 4:4 nuclease-helicase supramolecular structure is favorable genomically. A large complex forms only from two genes and likely makes it harder for phages to evade defense. Phages are limited in coding capacity by the size of the capsid head. Therefore, phages were thought to mostly encode genes that are essential for infection. While not all bacteria that phages encounter will encode the same anti-phage defense systems, phages likely evolve with their host and therefore encode dedicated anti-defense genes. To be evolutionarily favorable, but still function in anti-defense, these proteins are usually small. Anti-CRISPR and Anti-RM proteins are small negatively charged proteins that bind and block the active sites of Cas and restriction enzymes, respectively[82,83]. Gabija anti-defense breaks this rule as a 295 amino acid, 35 kDa protein (Chapters 1 & 2). As a monomer alone, Gad1 is distinct from other discovered anti-defense genes, but its molecular mechanism is even more unique as it forms an octameric web around the GajAB complex (Chapter 1). Gad1 only contacts GajA the nuclease in the GajAB complex but requires GajB the helicase to be pulled down with co-expressed in *E. coli*. When expressing GajA and Gad1, Gad1 is not pulled down with tagged GajA. This indicates that GajB likely locks GajA in an inactive conformation that Gad1 recognizes. GajA has structural flexibility in the dimerization domain, which is highlighted by changes in resolution from apo-GajAB to DNA-bound[39] and high B-factor in the X-ray crystal structure[37] (PDB 8SM3). Gad1 abuts the GajA dimerization domain on either side with a positively charged loop. While there are some distinct residue contacts, Gad1 more generally recognizes the shape of the GajA dimerization domain and locks it in place. This occurs in the C-terminal shoulder region of Gad1, which has

the most secondary structure. The N-terminal domain of Gad1 which was fitted into the low resolution cryo-EM density map with AlphaFold2 is mostly unstructured. Additionally, the N-terminal domain has varying lengths depending on the species. Gad1 likely recognizes diverse Gabija systems from many different bacterial species by recognizing the shape of the GajA dimerization interface and then using the N-terminal unstructured domain to drape over the GajA Toprim domain connecting protomers on the other side of the complex. The Gad1 homolog used for biochemistry in this study is from a gram-negative *Shewanella* phage that binds and blocks a gram-positive *Bacillus cereus* Gabija system. Therefore, Gad1 isn't a species-specific anti-defense protein so to speak, but it is likely transferred by mobile genetic elements across many different species.

Gad1 is 2.5 times as likely to be present on a prophage in a bacterium that chromosomally encodes Gabija[189]. Gabija may target a process that lysogenic phages require to integrate into bacterial genomes or Gad1 could be expressed from the integrated prophage when other phages try to infect. *Gad1* is in an early gene expression cassette in SpBeta-like phages and is likely expressed early on and binds Gabija before the phage-induced cue occurs[190]. Like Ocr, Tad1, and Acb2, it remains to be shown if Gad1 can block more than one anti-phage defense system. If Gad1 were to do this, the system would have to have a dimerization-like domain that interlocks protomers of a large complex. Gad1 could potentially block OLD proteins in PARIS, Menshen, and Retron systems if they mirror the tetrameric assembly of GajA in Gabija. This is not the case for PARIS as it forms a 6:3 ABC ATPase-Toprim domain complex, so even systems with the same genomic architecture form different stoichiometric complexes and overall structures.

An intriguing and potentially unanswerable question that remains is how Gad1 evolved? Some anti-CRISPR proteins act as decoy proteins or RNAs that likely came from the defense system itself but mutated to act as molecular traps[82,191]. Phages evade a toxin-antitoxin system called DarTG1 by encoding an antitoxin-like protein that reverses the toxin-mediated chemical modification on phage DNA[192]. Gad1 could have potentially evolved from a protein that binds to dimerization-like domains in bacteria including the Gabija GajA protein and then evolved to extend the N-terminus of the protein to drape around and connect to another Gad1 promoter. Overall, the Gad1 immune evasion mechanism highlights the wild strategies that phages use to overcome bacterial immunity.

There are likely small phage anti-defense DNA mimic proteins that bind and block the active site of the GajA Toprim domain, however, the extravagant Gad1 mechanism likely hints that they are not efficient inhibitors. GajA threads the target DNA sequence through the Toprim active with a 110° angle bend in the DNA[39]. Mimicking this type of DNA binding is likely harder than the B-form or slightly bent mimics of Ocr and anti-CRISPRs. Additionally, some phages encode their topoisomerases for DNA replication[193]. If these DNA mimics were to bind phage topoisomerases, phage replication would be inhibited. Future studies will determine the full extent to which phages evade Gabija defense.

Gad2 immune evasion

The molecular mechanism by which Gabija anti-defense 2 (Gad2) blocks Gabija anti-defense remains to be solved. However, Gad2 functions differently from Gad1 and does not bind the complex or inhibit DNA cleavage *in vitro* (Chapter 2). Gad2 has an N-terminal nucleotidyltransferase domain and a C-terminal domain of unknown function. Gad2 could potentially generate a nucleotide product that blocks phage DNA sensing by Gabija. For example, the phage protein Dam methylates phage DNA to avoid restriction-modification systems[67]. Anti-CRISPR, AcrIF11 acts as mono-ADP-ribosyltransferase to modify Cas8f proteins at residue N250, which inhibits PAM residue binding[194]. A genomic language model trained on amino acid sequence, genomic neighborhoods, and structure predicts that Gad2 is a Poly(A)-polymerase enzyme with additional homology to CCA-adding enzymes[195]. Gad2 could potentially modify the nucleic acid that is the cue for Gabija activation. Future work will determine what substrates Gad2 modifies during infection and if Gabija could potentially target RNA. Early work demonstrated that P2 OLD causes a defect in tRNAs during the lysogenization of *recBCD* mutant cells[196]. While we initially studied Gad1 to understand the mechanism of activation of the Gabija anti-phage defense, the immune evasion mechanism only showed complete wrapping of the inactive Gabija GajAB complex. Understanding the mechanism of how Gad2 evades Gabija immunity may shed light on the actual nucleic acid substrate targeted by Gabija *in vivo*.

Anti-defense strategies against Thoeris and Hachiman

The discovery of anti-defense proteins by comparing the phage susceptibility and genomic content of related phages reveals how intricately bacterial anti-phage defense and anti-defense phage evolution occurs (Chapter 2). The use of sponges to sequester immune signals produced by Thoeris indicates that sponges are produced in excess during infection to be able to sequester all the immune signals produced. When Thoeris is activated, it may only produce a low concentration of immune signal that sponges can sequester to stop the signaling cascade. While there are likely yet to be nucleases that cleave Thoeris signals, molecular protein sponges have emerged as a widespread and efficient way to stop immune activation in CBASS and Thoeris systems.

Had1 is a unique small protein with a positively charged patch that inhibits Hachiman in an unknown mechanism (Chapter 2). Phages evade the Hachiman by mutating phage single-stranded DNA binding proteins (SSBs)[69]. This indicates that Hachiman could be activated to cleave all DNA in the cell during phage infection by sensing a protein-DNA complex during phage DNA replication or damaged DNA that occurs during replication. Had1 may function to hide exposed single-stranded DNA that SSBs do not bind to. Hachiman can also be activated by DNA damage without phage infection[92], which hints at a more general mechanism of sensing DNA damage that occurs during phage infection. Sensing DNA damage mirrors human immunity where viruses manipulate and sometimes activate host DNA damage responses to evade or induce an immune response[197].

Diverse nucleases in anti-phage defense

We have shown that nuclease-helicase anti-phage defense systems all have different biochemical activities on phage nucleic acid. The main question that remains is how are these systems activated *in vivo*? Future research will determine if nuclease-helicase systems have target recognition sequences in phage genomes or if they are activated by a phage cue, such as a phage protein to cleave all DNA within cells. The 14 nuclease-helicase systems we surveyed have varying nuclease protein folds including DUF4297, DUF1837, (P)D-(D/E)-XK, and OLD/Toprim (Chapter 3). These nuclease folds likely respond to cleave phage nucleic acid after sensing phage infection by the helicase or other protein occurs. Diverse nuclease folds likely evolved in anti-phage defense to circumvent phage anti-defense. Phage-encoded DNA-mimic proteins bind and block (P)D-(D/E)-XK restriction enzymes, so being able to use a different

nuclease fold to circumvent anti-defense is important to inhibit phage infection. These types of nucleases could also be functionally redundant and be swapped in and out between systems allowing continued phage targeting. Determining how phages escape nuclease-helicase systems will lead to a deeper understanding of the evolutionary battle between phage DNA cleavage and bacterial immunity.

Complex formation by nuclease-helicase systems is now true for all systems we have tested biochemically. However, the oligomeric state of the system cannot be predicted only by the protein domains of the operon. Nuclease-helicase systems potentially form diverse oligomeric structures to be able to target diverse nucleic acids during infection. Determining nucleic acid bound states for nuclease-helicase systems will determine shared or divergent mechanisms for targeting phage nucleic acid. While determining structures using purified proteins will determine direct targeting, understanding how nuclease-helicase proteins assemble inside cells and target nucleic acid will give context to phage infection. Looking at where nuclease-helicase systems localize in cells, whether they form puncta, and if they directly degrade injecting phage DNA or replicating phage DNA will be an exciting new approach to take. *In vivo* experiments with fluorescence microscopy or cryo-electron tomography (cryo-ET) would lead to a deeper understanding of phage-host interactions.

Phage nucleic-acid targeting drives unique bacterial and phage immune mechanisms

Overall, our work highlights the unique mechanisms by which bacteria defend against phage infection by targeting nucleic acid and how phages fight back using proteins that specifically inhibit the activity of a defense system. Gabija and other nuclease-helicase systems form supramolecular complexes, which link human and bacterial immunity, where both species use large innate immune protein complexes to inhibit viral infection. Gad1 is an example of a completely novel way phages can evade Gabija defense by forming an octameric web around the GajA nuclease. In the future, new phage proteins will likely be discovered that utilize the same strategy for other defense systems. Gad2 stands out as an enzyme that has no homology to Gad1 and likely modifies the phage target that Gabija senses during infection. Tad2 and Had1 are other examples of molecular innovation in which phages evolved to stop Thoeris and Hachiman defense. While studying Gabija and Gad1 in detail led to an almost complete molecular mechanism for Gabija defense, there are many other nuclease-helicase systems. Surveying 14 additional

nuclease-helicase operons led us to group these defense systems based on oligomeric state and nucleic acid specificity. Future research following up on the biochemical activities of nuclease-helicase defense systems will uncover even more unique molecular mechanisms. By studying phage-host interactions untapped biology can uncover itself in the understanding of amazing molecular mechanisms.

Supplementary Materials

Supplementary information associated with Chapter 1 can be downloaded here:

https://www.nature.com/articles/s41586-023-06855-2#Sec19

Supplementary information associated with Chapter 2 can be downloaded here:

https://www.nature.com/articles/s41586-023-06869-w#Sec34

References

1. Liu, G., Park, H.-S., Pyo, H.-M., Liu, Q. & Zhou, Y. Influenza A Virus Panhandle Structure Is Directly Involved in RIG-I Activation and Interferon Induction. *J Virol* **89**, 6067–6079 (2015).

2. Ablasser, A. & Chen, Z. J. cGAS in action: Expanding roles in immunity and inflammation. *Science* **363**, eaat8657 (2019).

3. Kanneganti, T.-D. Central roles of NLRs and inflammasomes in viral infection. *Nat Rev Immunol* **10**, 688–698 (2010).

4. Fitzgerald, K. A. & Kagan, J. C. Toll-like Receptors and the Control of Immunity. *Cell* **180**, 1044–1066 (2020).

5. Iwasaki, A. & Medzhitov, R. Control of adaptive immunity by the innate immune system. *Nat Immunol* **16**, 343–353 (2015).

6. Ledvina, H. E. & Whiteley, A. T. Conservation and similarity of bacterial and eukaryotic innate immunity. *Nat Rev Microbiol* **22**, 420–434 (2024).

7. Wein, T. & Sorek, R. Bacterial origins of human cell-autonomous innate immune mechanisms. *Nat Rev Immunol* **22**, 629–638 (2022).

8. Whiteley, A. T. *et al.* Bacterial cGAS-like enzymes synthesize diverse nucleotide signals. *Nature* **567**, 194–199 (2019).

9. Cohen, D. *et al.* Cyclic GMP–AMP signalling protects bacteria against viral infection. *Nature* **574**, 691–695 (2019).

10. Banh, D. V. *et al.* Bacterial cGAS senses a viral RNA to initiate immunity. *Nature* **623**, 1001–1008 (2023).

11. Huiting, E. & Bondy-Denomy, J. Defining the expanding mechanisms of phage-mediated activation of bacterial immunity. *Current Opinion in Microbiology* **74**, 102325 (2023).

12. Slavik, K. M. & Kranzusch, P. J. CBASS to cGAS-STING: The Origins and Mechanisms of Nucleotide Second Messenger Immune Signaling. *Annual Review of Virology* **10**, 423–453 (2023).

13. Duncan-Lowey, B., McNamara-Bordewick, N. K., Tal, N., Sorek, R. & Kranzusch, P. J. Effector-mediated membrane disruption controls cell death in CBASS antiphage defense. *Molecular Cell* **81**, 5039-5051.e5 (2021).

14. Lowey, B. *et al.* CBASS Immunity Uses CARF-Related Effectors to Sense 3′–5′- and 2′–5′-Linked Cyclic Oligonucleotide Signals and Protect Bacteria from Phage Infection. *Cell* **182**, 38-49.e17 (2020).

15. Morehouse, B. R. *et al.* Cryo-EM structure of an active bacterial TIR-STING filament complex. *Nature* **608**, 803–807 (2022).

16. Kibby, E. M. *et al.* Bacterial NLR-related proteins protect against phage. *Cell* **186**, 2410-2424.e18 (2023).

17. Kibby, E. M. *et al.* A bacterial NLR-related protein recognizes multiple unrelated phage triggers to sense infection. 2024.12.17.629029 Preprint at https://doi.org/10.1101/2024.12.17.629029 (2024).

18. Leavitt, A. *et al.* Viruses inhibit TIR gcADPR signalling to overcome bacterial defence. *Nature* **611**, 326–331 (2022).

19. Ofir, G. *et al.* Antiviral activity of bacterial TIR domains via immune signalling molecules. *Nature* **600**, 116–120 (2021).

20. Tesson, F. *et al.* Systematic and quantitative view of the antiviral arsenal of prokaryotes. *Nat Commun* **13**, 2561 (2022).

21. Makarova, K. S., Wolf, Y. I., Snir, S. & Koonin, E. V. Defense Islands in Bacterial and Archaeal Genomes and Prediction of Novel Defense Systems. *Journal of Bacteriology* **193**, 6039–6056 (2011).

22. Doron, S. *et al.* Systematic discovery of antiphage defense systems in the microbial pangenome. *Science* **359**, eaar4120 (2018).

23. Gao, L. *et al.* Diverse enzymatic activities mediate antiviral immunity in prokaryotes. *Science* **369**, 1077–1084 (2020).

24. Millman, A. *et al.* An expanded arsenal of immune systems that protect bacteria from phages. *Cell Host Microbe* **30**, 1556-1569.e5 (2022).

25. Vassallo, C. N., Doering, C. R., Littlehale, M. L., Teodoro, G. I. C. & Laub, M. T. A functional selection reveals previously undetected anti-phage defence systems in the E. coli pangenome. *Nat Microbiol* **7**, 1568–1579 (2022).

26. DeWeirdt, P. C., Mahoney, E. M. & Laub, M. T. DefensePredictor: A Machine Learning Model to Discover Novel Prokaryotic Immune Systems. 2025.01.08.631726 Preprint at https://doi.org/10.1101/2025.01.08.631726 (2025).

27. Mordret, E. *et al.* Protein and genomic language models chart a vast landscape of antiphage defenses. 2025.01.08.631966 Preprint at https://doi.org/10.1101/2025.01.08.631966 (2025).

28. Luria, S. E. & Human, M. L. A nonhereditary, host-induced variation of bacterial viruses. *Journal of Bacteriology* **64**, 557–569 (1952).

29. Bertani, G. & Weigle, J. J. Host controlled variation in bacterial viruses. *Journal of Bacteriology* **65**, 113–121 (1953).

30. Linn, S. & Arber, W. Host specificity of DNA produced by Escherichia coli, X. In vitro restriction of phage fd replicative form. *Proceedings of the National Academy of Sciences* **59**, 1300–1306 (1968).

31. Roberts, R. J. *et al.* A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Research* **31**, 1805–1812 (2003).

32. Hille, F. *et al.* The Biology of CRISPR-Cas: Backward and Forward. *Cell* **172**, 1239–1259 (2018).

33. Makarova, K. S. *et al.* Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nat Rev Microbiol* **18**, 67–83 (2020).

34. Pinilla-Redondo, R. *et al.* Type IV CRISPR–Cas systems are highly diverse and involved in competition between plasmids. *Nucleic Acids Res* **48**, 2000–2012 (2020).

35. Cheng, R. *et al.* A nucleotide-sensing endonuclease from the Gabija bacterial defense system. *Nucleic Acids Research* **49**, 5216–5229 (2021).

36. Cheng, R. *et al.* Prokaryotic Gabija complex senses and executes nucleotide depletion and DNA cleavage for antiviral defense. *Cell Host & Microbe* (2023) doi:10.1016/j.chom.2023.06.014.

37. Antine, S. P. *et al.* Structural basis of Gabija anti-phage defence and viral immune evasion. *Nature* **625**, 360–365 (2024).

38. Yang, X.-Y. *et al.* Molecular basis of Gabija anti-phage supramolecular assemblies. *Nat Struct Mol Biol* **31**, 1243–1250 (2024).

39. Li, J. *et al.* Structures and activation mechanism of the Gabija anti-phage system. *Nature* **629**, 467–473 (2024).

40. Huo, Y. *et al.* Structural and biochemical insights into the mechanism of the Gabija bacterial immunity system. *Nat Commun* **15**, 836 (2024).

41. Christie, G. E. & Calendar, R. Bacteriophage P2. *Bacteriophage* **6**, e1145782 (2016).

42. Sironi, G. Mutants of Escherichia coli unable to be lysogenized by the temperate bacteriophage P2. *Virology* **37**, 163–176 (1969).

43. Sironi, G., Bialy, H., Lozeron, H. A. & Calendar, R. Bacteriophage P2: interaction with phage lambda and with recombination-deficient bacteria. *Virology* **46**, 387–396 (1971).

44. Dillingham, M. S. & Kowalczykowski, S. C. RecBCD Enzyme and the Repair of Double-Stranded DNA Breaks. *Microbiology and Molecular Biology Reviews* **72**, 642–671 (2008).

45. Cheng, K., Wilkinson, M., Chaban, Y. & Wigley, D. B. A conformational switch in response to Chi converts RecBCD from phage destruction to DNA repair. *Nat Struct Mol Biol* **27**, 71–77 (2020).

46. Lindahl, G. On the control of transcription in bacteriophage P2. *Virology* **46**, 620–633 (1971).

47. Schiltz, C. J., Lee, A., Partlow, E. A., Hosford, C. J. & Chappie, J. S. Structural characterization of Class 2 OLD family nucleases supports a two-metal catalysis mechanism for cleavage. *Nucleic Acids Research* **47**, 9448–9463 (2019).

48. Schiltz, C. J., Adams, M. C. & Chappie, J. S. The full-length structure of Thermus scotoductus OLD defines the ATP hydrolysis properties and catalytic mechanism of Class 1 OLD family nucleases. *Nucleic Acids Research* **48**, 2762–2776 (2020).

49. Dot, E. W., Thomason, L. C. & Chappie, J. S. Everything OLD is new again: How structural, functional, and bioinformatic advances have redefined a neglected nuclease family. *Molecular Microbiology* **n/a**,.

50. Burman, N. *et al.* A virally encoded tRNA neutralizes the PARIS antiviral defence system. *Nature* **634**, 424–431 (2024).

51. Deep, A., Liang, Q., Enustun, E., Pogliano, J. & Corbett, K. D. Architecture and activation mechanism of the bacterial PARIS defence system. *Nature* **634**, 432–439 (2024).

52. Rousset, F. *et al.* Phages and their satellites encode hotspots of antiviral systems. *Cell Host & Microbe* **30**, 740-753.e5 (2022).

53. Belukhina, S. *et al.* Specificity and Mechanism of tRNA cleavage by the AriB Toprim nuclease of the PARIS bacterial immune system. 2025.02.04.636281 Preprint at https://doi.org/10.1101/2025.02.04.636281 (2025).

54. Aravind, L., Leipe, D. D. & Koonin, E. V. Toprim--a conserved catalytic domain in type IA and II topoisomerases, DnaG-type primases, OLD family nucleases and RecR proteins. *Nucleic Acids Res* **26**, 4205–4213 (1998).

55. Krishnan, A., Burroughs, A. M., Iyer, L. M. & Aravind, L. Comprehensive classification of ABC ATPases and their functional radiation in nucleoprotein dynamics and biological conflict systems. *Nucleic Acids Res* **48**, 10045–10075 (2020).

56. Oh, H. *et al.* Structural and functional investigation of GajB protein in Gabija anti-phage defense. *Nucleic Acids Research* gkad951 (2023) doi:10.1093/nar/gkad951.

57. Raney, K. D., Byrd, A. K. & Aarattuthodiyil, S. Structure and Mechanisms of SF1 DNA Helicases. in *DNA Helicases and DNA Motor Proteins* (ed. Spies, M.) 17–46 (Springer, New York, NY, 2013). doi:10.1007/978-1-4614-5037-5_2.

58. Lee, J. Y. & Yang, W. UvrD Helicase Unwinds DNA One Base Pair At A Time By A Two-Part Power Stroke. *Cell* **127**, 1349–1360 (2006).

59. Ordabayev, Y. A., Nguyen, B., Kozlov, A. G., Jia, H. & Lohman, T. M. UvrD helicase activation by MutL involves rotation of its 2B subdomain. *Proceedings of the National Academy of Sciences* **116**, 16320–16325 (2019).

60. Fu, J., Schroder, K. & Wu, H. Mechanistic insights from inflammasome structures. *Nat Rev Immunol* **24**, 518–535 (2024).

61. Johnson, A. G. *et al.* Structure and assembly of a bacterial gasdermin pore. *Nature* **628**, 657–663 (2024).

62. Tamulaitiene, G. *et al.* Activation of Thoeris antiviral system via SIR2 effector filament assembly. *Nature* 1–6 (2024) doi:10.1038/s41586-024-07092-x.

63. Duncan-Lowey, B. *et al.* Cryo-EM structure of the RADAR supramolecular anti-phage defense complex. *Cell* **186**, 987-998.e15 (2023).

64. Gao, Y. *et al.* Molecular basis of RADAR anti-phage supramolecular assemblies. *Cell* **186**, 999-1012.e20 (2023).

65. Pleška, M. *et al.* Bacterial Autoimmunity Due to a Restriction-Modification System. *Curr Biol* **26**, 404–409 (2016).

66. Malone, L. M., Birkholz, N. & Fineran, P. C. Conquering CRISPR: how phages overcome bacterial adaptive immunity. *Current Opinion in Biotechnology* **68**, 30–36 (2021).

67. Kossykh, V. G., Schlagman, S. L. & Hattman, S. Phage T4 DNA [*N*]-*adenine*6*Methyltransferase*. OVEREXPRESSION, PURIFICATION, AND CHARACTERIZATION ∗. *Journal of Biological Chemistry* **270**, 14389–14393 (1995).

68. Bryson, A. L. *et al.* Covalent Modification of Bacteriophage T4 DNA Inhibits CRISPR-Cas9. *mBio* **6**, 10.1128/mbio.00648-15 (2015).

69. Stokar-Avihail, A. *et al.* Discovery of phage determinants that confer sensitivity to bacterial immune systems. *Cell* **186**, 1863-1876.e16 (2023).

70. Sather, L. M. *et al.* A broadly distributed predicted helicase/nuclease confers phage resistance via abortive infection. *Cell Host & Microbe* **31**, 343-355.e5 (2023).

71. Sasaki, T. *et al.* Phage single-stranded DNA-binding protein or host DNA damage triggers the activation of the AbpAB phage defense system. *mSphere* **8**, e00372-23 (2023).

72. Bendtsen, J. D., Nilsson, A. S. & Lehnherr, H. Phylogenetic and Functional Analysis of the Bacteriophage P1 Single-Stranded DNA-Binding Protein. *J Virol* **76**, 9695–9701 (2002).

73. Zhang, T. *et al.* A bacterial immunity protein directly senses two disparate phage proteins. *Nature* 1–8 (2024) doi:10.1038/s41586-024-08039-y.

74. Zhang, T. *et al.* Direct activation of a bacterial innate immune system by a viral capsid protein. *Nature* **612**, 132–140 (2022).

75. Roberts, C., Fishman, C., Banh, D. & Marraffini, L. A. A bacterial TIR-based immune system senses viral capsids to initiate defense. 2024.07.29.605636 Preprint at https://doi.org/10.1101/2024.07.29.605636 (2024).

76. Huiting, E. *et al.* Bacteriophages inhibit and evade cGAS-like immune function in bacteria. *Cell* **186**, 864-876.e21 (2023).

77. Jenson, J. M., Li, T., Du, F., Ea, C.-K. & Chen, Z. J. Ubiquitin-like conjugation by bacterial cGAS enhances anti-phage defence. *Nature* **616**, 326–331 (2023).

78. Cao, X. *et al.* Phage anti-CBASS protein simultaneously sequesters cyclic trinucleotides and dinucleotides. *Molecular Cell* **84**, 375-385.e7 (2024).

79. Chang, R. B. *et al.* A widespread family of viral sponge proteins reveals specific inhibition of nucleotide signals in anti-phage defense. 2024.12.30.630793 Preprint at https://doi.org/10.1101/2024.12.30.630793 (2024).

80. Li, D. *et al.* Single phage proteins sequester signals from TIR and cGAS-like enzymes. *Nature* **635**, 719–727 (2024).

81. Hobbs, S. J. *et al.* Phage anti-CBASS and anti-Pycsar nucleases subvert bacterial immunity. *Nature* **605**, 522–526 (2022).

82. Wiegand, T., Karambelkar, S., Bondy-Denomy, J. & Wiedenheft, B. Structures and Strategies of Anti-CRISPR-Mediated Immune Suppression. *Annual Review of Microbiology* **74**, 21–37 (2020).

83. Tock, M. R. & Dryden, D. T. The biology of restriction and anti-restriction. *Current Opinion in Microbiology* **8**, 466–472 (2005).

84. Jiang, F. *et al.* Temperature-Responsive Competitive Inhibition of CRISPR-Cas9. *Molecular Cell* **73**, 601-610.e5 (2019).

85. Kennaway, C. K. *et al.* The structure of M.EcoKI Type I DNA methyltransferase with a DNA mimic antirestriction protein. *Nucleic Acids Research* **37**, 762–770 (2009).

86. Wilkinson, M. *et al.* Structural basis for the inhibition of RecBCD by Gam and its synergistic antibacterial effect with quinolones. *eLife* **5**, e22963 (2016).

87. Yirmiya, E. *et al.* Structure-guided discovery of viral proteins that inhibit host immunity. *Cell* **0**, (2025).

88. Patel, K. & Seed, K. A Class 1 OLD family nuclease encoded by Vibrio cholerae is countered by a vibriophage-encoded direct inhibitor. 2025.01.06.631583 Preprint at https://doi.org/10.1101/2025.01.06.631583 (2025).

89. Niault, T., Houte, S. van, Westra, E. & Swarts, D. C. Evolution and ecology of anti-defence systems in phages and plasmids. *Current Biology* **35**, R32–R44 (2025).

90. Gottlin, E. B., Rudolph, A. E., Zhao, Y., Matthews, H. R. & Dixon, J. E. Catalytic mechanism of the phospholipase D superfamily proceeds via a covalent phosphohistidine intermediate. *Proc Natl Acad Sci U S A* **95**, 9202–9207 (1998).

91. Steczkiewicz, K., Muszewska, A., Knizewski, L., Rychlewski, L. & Ginalski, K. Sequence, structure and functional diversity of PD-(D/E)XK phosphodiesterase superfamily. *Nucleic Acids Research* **40**, 7016–7045 (2012).

92. Tuck, O. T. *et al.* Genome integrity sensing by the broad-spectrum Hachiman antiphage defense complex. *Cell* **187**, 6914-6928.e20 (2024).

93. Payne, L. J. *et al.* PADLOC: a web server for the identification of antiviral defence systems in microbial genomes. *Nucleic Acids Res* **50**, W541–W550 (2022).

94. Hopfner, K.-P. Invited review: Architectures and mechanisms of ATP binding cassette proteins. *Biopolymers* **105**, 492–504 (2016).

95. Raney, K. D., Byrd, A. K. & Aarattuthodiyil, S. Structure and Mechanisms of SF1 DNA Helicases. *Adv Exp Med Biol* **767**, 17–46 (2013).

96. Gorbalenya, A. E. & Koonin, E. V. Helicases: amino acid sequence comparisons and structure-function relationships. *Current Opinion in Structural Biology* **3**, 419–429 (1993).

97. Tanner, N. K. & Linder, P. DExD/H box RNA helicases: from generic motors to specific dissociation functions. *Mol Cell* **8**, 251–262 (2001).

98. Korolev, S., Hsieh, J., Gauss, G. H., Lohman, T. M. & Waksman, G. Major domain swiveling revealed by the crystal structures of complexes of E. coli Rep helicase bound to single-stranded DNA and ADP. *Cell* **90**, 635–647 (1997).

99. 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).

100. Stanley, S. Y. & Maxwell, K. L. Phage-Encoded Anti-CRISPR Defenses. *Annu Rev Genet* **52**, 445–464 (2018).

101. Hampton, H. G., Watson, B. N. J. & Fineran, P. C. The arms race between bacteria and their phage foes. *Nature* **577**, 327–336 (2020).

102. Yirmiya, E. *et al.* Phages overcome bacterial immunity via diverse anti-defence proteins. *Nature* **625**, 352–359 (2024).

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

104. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat Methods* **19**, 679–682 (2022).

105. Lebbink, J. H. G. *et al.* Magnesium coordination controls the molecular switch function of DNA mismatch repair protein MutS. *J Biol Chem* **285**, 13131–13141 (2010).

106. Wang, J. Y., Pausch, P. & Doudna, J. A. Structural biology of CRISPR–Cas immunity and genome editing enzymes. *Nat Rev Microbiol* **20**, 641–656 (2022).

107. Hogrel, G. *et al.* Cyclic nucleotide-induced helical structure activates a TIR immune effector. *Nature* **608**, 808–812 (2022).

108. Kagan, J. C., Magupalli, V. G. & Wu, H. SMOCs: supramolecular organizing centres that control innate immunity. *Nat Rev Immunol* **14**, 821–826 (2014).

109. Mazzocco, A., Waddell, T. E., Lingohr, E. & Johnson, R. P. Enumeration of bacteriophages using the small drop plaque assay system. *Methods Mol Biol* **501**, 81–85 (2009).

110. Liu, H. & Naismith, J. H. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol* **8**, 91 (2008).

111. Zhou, W. *et al.* Structure of the Human cGAS-DNA Complex Reveals Enhanced Control of Immune Surveillance. *Cell* **174**, 300-311.e11 (2018).

112. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol* **75**, 861–877 (2019).

113. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126–2132 (2004).

114. Morin, A. *et al.* Collaboration gets the most out of software. *eLife* **2**, e01456 (2013).

115. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290–296 (2017).

116. Pettersen, E. F. *et al.* UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci* **30**, 70–82 (2021).

117. Dy, R. L., Richter, C., Salmond, G. P. C. & Fineran, P. C. Remarkable Mechanisms in Microbes to Resist Phage Infections. *Annual Review of Virology* **1**, 307–331 (2014).

118. Bernheim, A. & Sorek, R. The pan-immune system of bacteria: antiviral defence as a community resource. *Nat Rev Microbiol* **18**, 113–119 (2020).

119. Tal, N. & Sorek, R. SnapShot: Bacterial immunity. *Cell* **185**, 578-578.e1 (2022).

120. Samson, J. E., Magadán, A. H., Sabri, M. & Moineau, S. Revenge of the phages: defeating bacterial defences. *Nat Rev Microbiol* **11**, 675–687 (2013).

121. Atanasiu, C., Su, T. -J., Sturrock, S. S. & Dryden, D. T. F. Interaction of the ocr gene 0.3 protein of bacteriophage T7 with EcoKI restriction/modification enzyme. *Nucleic Acids Research* **30**, 3936–3944 (2002).

122. Walkinshaw, M. D. *et al.* Structure of Ocr from Bacteriophage T7, a Protein that Mimics B-Form DNA. *Molecular Cell* **9**, 187–194 (2002).

123. Drozdz, M., Piekarowicz, A., Bujnicki, J. M. & Radlinska, M. Novel non-specific DNA adenine methyltransferases. *Nucleic Acids Research* **40**, 2119–2130 (2012).

124. Thavalingam, A. *et al.* Inhibition of CRISPR-Cas9 ribonucleoprotein complex assembly by anti-CRISPR AcrIIC2. *Nat Commun* **10**, 2806 (2019).

125. Lu, W.-T., Trost, C. N., Müller-Esparza, H., Randau, L. & Davidson, A. R. Anti-CRISPR AcrIF9 functions by inducing the CRISPR–Cas complex to bind DNA non-specifically. *Nucleic Acids Research* **49**, 3381–3393 (2021).

126. Bondy-Denomy, J. *et al.* Multiple mechanisms for CRISPR–Cas inhibition by anti-CRISPR proteins. *Nature* **526**, 136–139 (2015).

127. Li, Y. & Bondy-Denomy, J. Anti-CRISPRs go viral: The infection biology of CRISPR-Cas inhibitors. *Cell Host & Microbe* **29**, 704–714 (2021).

128. Jia, N. & Patel, D. J. Structure-based functional mechanisms and biotechnology applications of anti-CRISPR proteins. *Nat Rev Mol Cell Biol* **22**, 563–579 (2021).

129. Davidson, A. R. *et al.* Anti-CRISPRs: Protein Inhibitors of CRISPR-Cas Systems. *Annual Review of Biochemistry* **89**, 309–332 (2020).

130. Otsuka, Y. & Yonesaki, T. Dmd of bacteriophage T4 functions as an antitoxin against Escherichia coli LsoA and RnlA toxins. *Molecular Microbiology* **83**, 669–681 (2012).

131. Millman, A. *et al.* Bacterial Retrons Function In Anti-Phage Defense. *Cell* **183**, 1551-1561.e12 (2020).

132. Bernheim, A. *et al.* Prokaryotic viperins produce diverse antiviral molecules. *Nature* **589**, 120–124 (2021).

133. Lau, R. K. *et al.* Structure and Mechanism of a Cyclic Trinucleotide-Activated Bacterial Endonuclease Mediating Bacteriophage Immunity. *Molecular Cell* **77**, 723-733.e6 (2020).

134. Millman, A., Melamed, S., Amitai, G. & Sorek, R. Diversity and classification of cyclic-oligonucleotide-based anti-phage signalling systems. *Nat Microbiol* **5**, 1608–1615 (2020).

135. Tal, N. *et al.* Cyclic CMP and cyclic UMP mediate bacterial immunity against phages. *Cell* **184**, 5728-5739.e16 (2021).

136. Kronheim, S. *et al.* A chemical defence against phage infection. *Nature* **564**, 283–286 (2018).

137. Bobonis, J. *et al.* Bacterial retrons encode phage-defending tripartite toxin–antitoxin systems. *Nature* **609**, 144–150 (2022).

138. Rosenthal, R., Toye, P. A., Korman, R. Z. & Zahler, S. A. THE PROPHAGE OF SPβc2dcitK1, A DEFECTIVE SPECIALIZED TRANSDUCING PHAGE OF BACILLUS SUBTILIS. *Genetics* **92**, 721–739 (1979).

139. Tucker, R. G. Acquisition of Thymidylate Synthetase Activity by a Thymine-requiring Mutant of Bacillus subtilis following Infection by the Temperate Phage φ3. *Journal of General Virology* **4**, 489–504 (1969).

140. Noyer-Weidner, M., Jentsch, S., Pawlek, B., Günthert, U. & Trautner, T. A. Restriction and modification in Bacillus subtilis: DNA methylation potential of the related bacteriophages Z, SPR, SP beta, phi 3T, and rho 11. *Journal of Virology* **46**, 446–453 (1983).

141. Stewart, C. R. *et al.* The Genome of *Bacillus subtilis* Bacteriophage SPO1. *Journal of Molecular Biology* **388**, 48–70 (2009).

142. Ofir, G. *et al.* DISARM is a widespread bacterial defence system with broad anti-phage activities. *Nat Microbiol* **3**, 90–98 (2018).

143. Peters, J. M. *et al.* A Comprehensive, CRISPR-based Functional Analysis of Essential Genes in Bacteria. *Cell* **165**, 1493–1506 (2016).

144. Athukoralage, J. S. *et al.* An anti-CRISPR viral ring nuclease subverts type III CRISPR immunity. *Nature* **577**, 572–575 (2020).

145. Uribe, R. V. *et al.* Discovery and Characterization of Cas9 Inhibitors Disseminated across Seven Bacterial Phyla. *Cell Host & Microbe* **25**, 233-241.e5 (2019).

146. Manik, M. K. *et al.* Cyclic ADP ribose isomers: Production, chemical structures, and immune signaling. *Science* **377**, eadc8969 (2022).

147. MacLean, R. C. & San Millan, A. The evolution of antibiotic resistance. *Science* **365**, 1082–1083 (2019).

148. Kortright, K. E., Chan, B. K., Koff, J. L. & Turner, P. E. Phage Therapy: A Renewed Approach to Combat Antibiotic-Resistant Bacteria. *Cell Host & Microbe* **25**, 219–232 (2019).

149. Nobrega, F. L., Costa, A. R., Kluskens, L. D. & Azeredo, J. Revisiting phage therapy: new applications for old resources. *Trends in Microbiology* **23**, 185–191 (2015).

150. Federici, S. *et al.* Targeted suppression of human IBD-associated gut microbiota commensals by phage consortia for treatment of intestinal inflammation. *Cell* **185**, 2879-2898.e24 (2022).

151. Hussain, F. A. *et al.* Rapid evolutionary turnover of mobile genetic elements drives bacterial resistance to phages. *Science* **374**, 488–492 (2021).

152. LeGault, K. N. *et al.* Temporal shifts in antibiotic resistance elements govern phage-pathogen conflicts. *Science* **373**, eabg2166 (2021).

153. Kropinski, A. M., Mazzocco, A., Waddell, T. E., Lingohr, E. & Johnson, R. P. Enumeration of Bacteriophages by Double Agar Overlay Plaque Assay. in *Bacteriophages: Methods and Protocols, Volume 1: Isolation, Characterization, and Interactions* (eds. Clokie, M. R. J. & Kropinski, A. M.) 69–76 (Humana Press, Totowa, NJ, 2009). doi:10.1007/978-1-60327-164-6_7.

154. Baym, M. *et al.* Inexpensive Multiplexed Library Preparation for Megabase-Sized Genomes. *PLOS ONE* **10**, e0128036 (2015).

155. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).

156. Nurk, S. *et al.* Assembling Genomes and Mini-metagenomes from Highly Chimeric Reads. in *Research in Computational Molecular Biology* (eds. Deng, M., Jiang, R., Sun, F. & Zhang, X.) 158–170 (Springer, Berlin, Heidelberg, 2013). doi:10.1007/978-3-642-37195-0_13.

157. Procedure & Checklist – Preparing Multiplexed Microbial Libraries Using SMRTbell Express Template Prep Kit 2.0. (2021).

158. Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).

159. Gilchrist, C. L. M. & Chooi, Y.-H. clinker & clustermap.js: automatic generation of gene cluster comparison figures. *Bioinformatics* **37**, 2473–2475 (2021).

160. Mazzocco, A., Waddell, T. E., Lingohr, E. & Johnson, R. P. Enumeration of Bacteriophages Using the Small Drop Plaque Assay System. in *Bacteriophages: Methods and Protocols, Volume 1: Isolation, Characterization, and Interactions* (eds. Clokie, M. R. J. & Kropinski, A. M.) 81–85 (Humana Press, Totowa, NJ, 2009). doi:10.1007/978-1-60327-164-6_9.

161. Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol* **35**, 1026–1028 (2017).

162. Deatherage, D. E. & Barrick, J. E. Identification of Mutations in Laboratory-Evolved Microbes from Next-Generation Sequencing Data Using breseq. in *Engineering and Analyzing Multicellular Systems: Methods and Protocols* (eds. Sun, L. & Shou, W.) 165–188 (Springer, New York, NY, 2014). doi:10.1007/978-1-4939-0554-6_12.

163. Nayfach, S. *et al.* Metagenomic compendium of 189,680 DNA viruses from the human gut microbiome. *Nat Microbiol* **6**, 960–970 (2021).

164. Chen, I.-M. A. *et al.* IMG/M v.5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. *Nucleic Acids Research* **47**, D666–D677 (2019).

165. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* **30**, 772–780 (2013).

166. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution* **32**, 268–274 (2015).

167. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Research* **49**, W293–W296 (2021).

168. Kabsch, W. XDS. *Acta Cryst D* **66**, 125–132 (2010).

169. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Cryst D* **69**, 1204–1214 (2013).

170. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Cryst D* **66**, 12–21 (2010).

171. Karplus, P. A. & Diederichs, K. Linking Crystallographic Model and Data Quality. *Science* **336**, 1030–1033 (2012).

172. Weiss, M. S. Global indicators of X-ray data quality. *J Appl Cryst* **34**, 130–135 (2001).

173. Yasui, R., Washizaki, A., Furihata, Y., Yonesaki, T. & Otsuka, Y. AbpA and AbpB provide anti-phage activity in Escherichia coli. *Genes Genet Syst* **89**, 51–60 (2014).

174. Ernits, K. *et al.* The structural basis of hyperpromiscuity in a core combinatorial network of type II toxin–antitoxin and related phage defense systems. *Proceedings of the National Academy of Sciences* **120**, e2305393120 (2023).

175. Young, G. *et al.* Quantitative mass imaging of single biological macromolecules. *Science* **360**, 423–427 (2018).

176. Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).

177. Park, Y. B. *et al.* Eukaryotic Rad50 functions as a rod-shaped dimer. *Nat Struct Mol Biol* **24**, 248–257 (2017).

178. Shin, J., Nile, A. & Oh, J.-W. Role of adaptin protein complexes in intracellular trafficking and their impact on diseases. *Bioengineered* **12**, 8259–8278.

179. Loeff, L., Walter, A., Rosalen, G. T. & Jinek, M. DNA end sensing and cleavage by the Shedu anti-phage defense system. *Cell* **188**, 721-733.e17 (2025).

180. Shi, M., Zhang, P., Vora, S. M. & Wu, H. Higher-order assemblies in innate immune and inflammatory signaling: A general principle in cell biology. *Current Opinion in Cell Biology* **63**, 194–203 (2020).

181. Miller, E. S. *et al.* Bacteriophage T4 Genome. *Microbiology and Molecular Biology Reviews* **67**, 86–156 (2003).

182. Morgan, J. R., Prasad, K., Hao, W., Augustine, G. J. & Lafer, E. M. A Conserved Clathrin Assembly Motif Essential for Synaptic Vesicle Endocytosis. *J Neurosci* **20**, 8667–8676 (2000).

183. Owen, S. V. *et al.* Prophages encode phage-defense systems with cognate self-immunity. *Cell Host & Microbe* **29**, 1620-1633.e8 (2021).

184. Ho, C. S. *et al.* Antimicrobial resistance: a concise update. *The Lancet Microbe* **6**, (2025).

185. Strathdee, S. A., Hatfull, G. F., Mutalik, V. K. & Schooley, R. T. Phage therapy: From biological mechanisms to future directions. *Cell* **186**, 17–31 (2023).

186. Cury, J. *et al.* Conservation of antiviral systems across domains of life reveals immune genes in humans. *Cell Host & Microbe* **32**, 1594-1607.e5 (2024).

187. Bonhomme, D. *et al.* A human homolog of SIR2 antiphage proteins mediates immunity via the TLR pathway. 2024.09.18.613514 Preprint at https://doi.org/10.1101/2024.09.18.613514 (2024).

188. Payne, L., Jackson, S. & Pinilla-Redondo, R. Supramolecular assemblies in bacterial immunity: an emerging paradigm. *Trends in Microbiology* (2024) doi:10.1016/j.tim.2024.06.003.

189. Tesson, F. *et al.* Exploring the diversity of anti-defense systems across prokaryotes, phages and mobile genetic elements. *Nucleic Acids Research* **53**, gkae1171 (2025).

190. Kohm, K. & Hertel, R. The life cycle of SPβ and related phages. *Arch Virol* **166**, 2119–2130 (2021).

191. Camara-Wilpert, S. *et al.* Bacteriophages suppress CRISPR–Cas immunity using RNA-based anti-CRISPRs. *Nature* **623**, 601–607 (2023).

192. Johannesman, A., Awasthi, L. C., Carlson, N. & LeRoux, M. Phages carry orphan antitoxin-like enzymes to neutralize the DarTG1 toxin-antitoxin defense system. *Nat Commun* **16**, 1598 (2025).

193. Grose, J. H. *et al.* The Genomes, Proteomes, and Structures of Three Novel Phages That Infect the Bacillus cereus Group and Carry Putative Virulence Factors. *J Virol* **88**, 11846–11860 (2014).

194. Niu, Y. *et al.* A Type I-F Anti-CRISPR Protein Inhibits the CRISPR-Cas Surveillance Complex by ADP-Ribosylation. *Molecular Cell* **80**, 512-524.e5 (2020).

195. Jha, N. *et al.* Gaia: A Context-Aware Sequence Search and Discovery Tool for Microbial Proteins. 2024.11.19.624387 Preprint at https://doi.org/10.1101/2024.11.19.624387 (2024).

196. Bregegere, F. Bacteriophage P2-λ interference: Inhibition of protein synthesis involves transfer RNA inactivation. *Journal of Molecular Biology* **90**, 459–467 (1974).

197. Lopez, A. *et al.* Viral Modulation of the DNA Damage Response and Innate Immunity: Two Sides of the Same Coin. *J Mol Biol* **434**, 167327 (2022).