RESEARCH ARTICLE SUMMARY

MICROBIOLOGY

Prokaryotic innate immunity through pattern recognition of conserved viral proteins

Linyi Alex Gao*†, Max E. Wilkinson†, Jonathan Strecker†, Kira S. Makarova, Rhiannon K. Macrae, Eugene V. Koonin, Feng Zhang*

INTRODUCTION: Many organisms have evolved specialized immune pattern-recognition receptors, including nucleotide-binding oligomerization domain-like receptors (NLRs) of the STAND superfamily that are ubiquitous in plants, animals, and fungi. NLRs oligomerize upon recognition of pathogen-associated molecular patterns, leading to the activation of an effector domain that mediates an inflammatory or cell death response. Although the roles of NLRs in eukaryotic immunity are well established, it is unknown whether prokaryotes use similar defense mechanisms.

RATIONALE: We previously identified a set of bacterial and archaeal STAND nucleoside triphosphatases (NTPases), dubbed Avs (antiviral STAND), that protect bacteria from tailed phages through an unknown mechanism. Like eukaryotic NLRs, Avs proteins have a characteristic tripartite domain architecture consisting of a central NTPase, an extended C-terminal sensor, and an N-terminal effector. Here, we investigate the defense mechanism of these Avs proteins.

RESULTS: Using genetic screens in *Escherichia coli*, we characterized four Avs families (Avs1 to Avs4) and found that they detect hallmark viral proteins that are expressed during infection. In particular, Avs1 to Avs3 recognize the large terminase subunit, and Avs4 recognizes the portal. These two proteins together make up the conserved DNA packaging machinery of tailed phages. Coexpression of an Avs protein with its cognate target in *E. coli* resulted in cell death.

We assessed the specificity of Avs target recognition with a panel of terminases and portals from 24 tailed phages, spanning nine major families. Notably, a single Avs protein was capable of recognizing a large variety of targets (terminase or portal), with less than 5% sequence identity in some cases.

We next reconstituted Avs activity in vitro, focusing on representatives from *Salmonella enterica* (SeAvs3) and *E. coli* (EcAvs4), both of which contain N-terminal PD-DExK nuclease effectors. In the presence of their cognate target, SeAvs3 and EcAvs4 mediated degradation of double-stranded DNA. Nuclease activity required the presence of Mg^{2+} and adenosine triphosphate (ATP); however, the hydrolysis of ATP was not strictly required. Single-stranded DNA and RNA substrates were not cleaved.

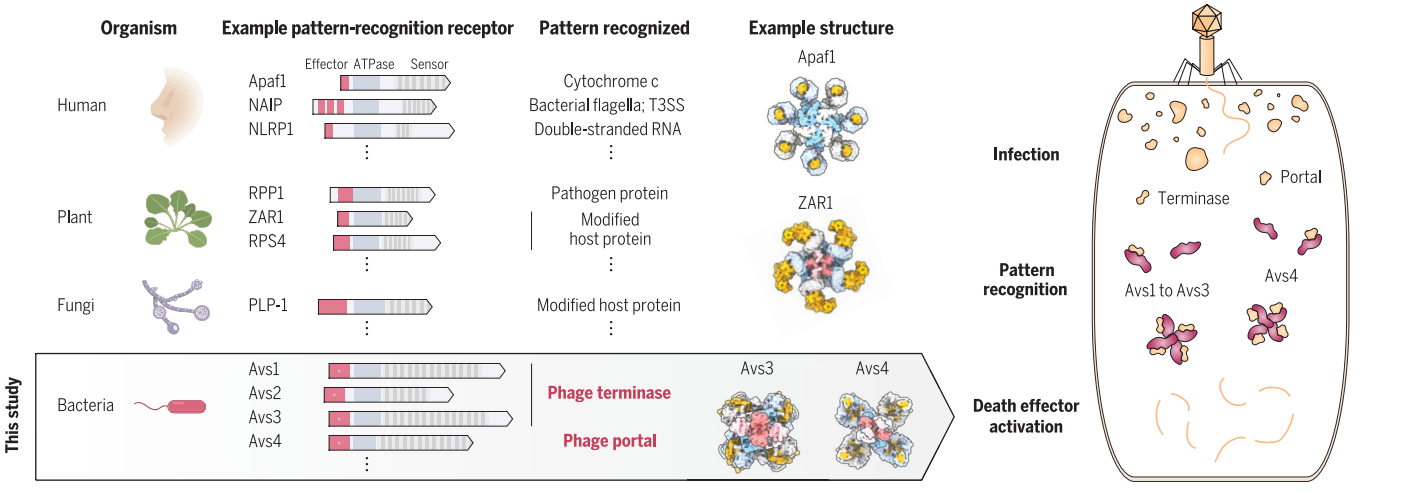
We determined the cryo–electron microscopy structures of the SeAvs3-terminase and EcAvs4-portal complexes, revealing that both form tetramers in which the C-terminal sensor domain of each Avs subunit binds to a single target protein. Binding is mediated by shape complementarity across an extended interface, consistent with fold recognition. Additionally, SeAvs3 directly recognizes terminase active-site residues and its ATP ligand. Tetramerization of both SeAvs3 and EcAvs4 is mediated by their STAND ATPase domains and allows the N-terminal nucleases to adopt active dimeric configurations.

Bioinformatic analysis of Avs proteins across prokaryotic lineages revealed at least 18 distinct types of N-terminal effectors that are modularly swapped between Avs homologs, as well as widespread distribution of *avs* genes across phyla with extensive horizontal gene transfer. Finally, we also discovered phage-encoded Avs inhibitors, highlighting an extensive arms race between prokaryotes and viruses.

CONCLUSION: NLR-like defense proteins in bacteria and archaea recognize the conserved folds of hallmark viral proteins and assemble into tetramers that activate diverse N-terminal effectors. The mechanism of these proteins highlights the similarity between the defense strategies of prokaryotes and eukaryotes and extends the paradigm of pattern recognition of pathogen-specific proteins across all three domains of life. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: zhang@broadinstitute.org (F.Z.); algao@stanford.edu (L.A.G.)
†These authors contributed equally to this work.
Cite this article as L. A. Gao *et al.*, *Science* **377**, eabm4096 (2022). DOI: 10.1126/science.abm4096

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abm4096



Mechanism and structures of NLR-like defense proteins in prokaryotes. (Left) Comparison of the domain architectures of 11 representative NLR-like pattern-recognition receptors across four kingdoms of life. Selected structures of activated complexes are shown as examples. T3SS, type 3 secretion system. (Right) Defense mechanism of Avs proteins in bacteria and archaea (this study). Target binding triggers the formation of Avs tetramers, which activates an N-terminal effector that disrupts the viral life cycle.

RESEARCH ARTICLE

MICROBIOLOGY

Prokaryotic innate immunity through pattern recognition of conserved viral proteins

Linyi Alex Gao[1,2,3,4,5,6]*†, Max E. Wilkinson[1,2,3,4,5]†, Jonathan Strecker[1,2,3,4,5]†, Kira S. Makarova[7], Rhiannon K. Macrae[1,2,3,4,5], Eugene V. Koonin[7], Feng Zhang[1,2,3,4,5]*

Many organisms have evolved specialized immune pattern-recognition receptors, including nucleotide-binding oligomerization domain–like receptors (NLRs) of the STAND superfamily that are ubiquitous in plants, animals, and fungi. Although the roles of NLRs in eukaryotic immunity are well established, it is unknown whether prokaryotes use similar defense mechanisms. Here, we show that antiviral STAND (Avs) homologs in bacteria and archaea detect hallmark viral proteins, triggering Avs tetramerization and the activation of diverse N-terminal effector domains, including DNA endonucleases, to abrogate infection. Cryo–electron microscopy reveals that Avs sensor domains recognize conserved folds, active-site residues, and enzyme ligands, allowing a single Avs receptor to detect a wide variety of viruses. These findings extend the paradigm of pattern recognition of pathogen-specific proteins across all three domains of life.

Bacteria and archaea have evolved numerous defense mechanisms against viral infections that involve a wide range of strategies and enzymatic activities (*1*–*5*). Defense systems are activated by viral nucleic acids, in the case of restriction-modification and CRISPR-Cas systems, or by different types of infection-induced cellular stress, including DNA double-strand breaks (*6*), inhibition of host transcription (*7*), cytosolic nucleotide depletion (*8*), and the disruption of translation elongation factor EF-Tu (*9*) or RecBCD repair nuclease (*10*). Alternatively, some systems constitutively synthesize small molecules that interfere with phage replication (*11*, *12*). However, for numerous defense systems, the mechanisms of activation remain uncharacterized, and it appears likely that distinct modes of activation exist within the diverse repertoire of recently discovered systems (*1*, *3*, *4*, *13*).

STAND nucleoside triphosphatases (NTPases), which include nucleotide-binding oligomerization domain–like receptors (NLRs) such as animal inflammasomes and plant resistosomes, are among the key players in immunity, cell signaling, and particularly programmed cell death in eukaryotes (*14*–*21*). STAND NTPases have a conserved tripartite domain architecture, consisting of a central NTPase domain, a C-terminal sensor that contains superstructure-forming repeats, and, in many cases, an N-terminal effector that mediates inflammation or cell death. In animal and plant innate immunity, STAND proteins function by recognizing diverse pathogen-associated molecular patterns (PAMPs), including peptidoglycan fragments from the bacterial cell wall (*18*), double-stranded RNA (*21*), bacterial flagellin and type 3 secretion systems (*16*, *17*), and endogenous host proteins that have been modified by pathogens (*19*, *20*). In all of these cases, recognition of the PAMP leads to oligomerization of the STAND NTPase and activation or recruitment of effector proteins.

Bacteria and archaea, especially those with complex signaling systems, also encode a diverse repertoire of STAND NTPases that are predicted to be involved in signal transduction and possibly in programmed cell death (*14*, *15*). However, the functions of these proteins are largely unknown, with the exception of several that have been characterized as transcriptional regulators (*22*–*24*). We recently identified a group of prokaryotic STAND NTPases, dubbed Avs (antiviral STAND) (*4*), that are often encoded next to restriction-modification and other defense systems (fig. S1) and protect bacteria from tailed phages (fig. S2). Here, we investigate the mechanism of Avs proteins.

Avs systems are activated by two conserved phage proteins

Although the domain architectures of Avs proteins resemble those of eukaryotic NLRs (Fig. 1, A and B), it is unclear whether they function through similar molecular mechanisms. We identified four distinct families of Avs proteins (Avs1 to Avs4) (fig. S3), each of which contains highly divergent tetratricopeptide repeat (TPR)–like sensor domains, and selected two representatives for further characterization: SeAvs3 from *Salmonella enterica* NCTC13175 and EcAvs4 from *Escherichia coli* NCTC11132, both of which provide robust protection against the T7-like coliphage PhiV-1 (fig. S2). We first asked how phage infection leads to Avs activation and whether a specific phage-encoded trigger exists for these defense systems. We cloned fragments covering the whole PhiV-1 phage genome into expression plasmids and transformed the resulting fragment library into *E. coli* containing either Avs proteins or empty-vector controls (Fig. 1C and data S1). We hypothesized that coexpression of Avs proteins with their putative triggers might lead to cell death and depletion of the respective phage genes from the pool, and we performed deep sequencing to detect enrichment or depletion of phage genes. Four phage genes were generally toxic to all cells; however, two genes were depleted only in the presence of Avs proteins, namely the large terminase subunit (gp19) when coexpressed with SeAvs3 and the portal protein (gp8) when expressed with EcAvs4 (Fig. 1D and fig. S4). By Southern blot, we observed that Avs3- and Avs4-mediated depletion of phage DNA during infection was abolished in gp19 and gp8 knockout phage strains, respectively (Fig. 1E and fig. S5), indicating that gp19 and gp8 are both necessary and sufficient for Avs activation.

To validate these findings, we transformed plasmids expressing gp19 or gp8 into *E. coli* harboring SeAvs3 or EcAvs4 and measured cell viability. Consistent with our previous results, we observed cell death after coexpression of SeAvs3 and gp19, as well as coexpression of EcAvs4 and gp8, but not with the reciprocal pairs (Fig. 1E). This toxicity depended on the predicted nuclease activity of both SeAvs3 and EcAvs4 and, importantly, was not due to any intrinsic features of the natural phage gene sequence, because recoded gene sequences also led to cell death (Fig. 1F). Furthermore, the enzymatic activity of the phage terminase, which contains adenosine triphosphatase (ATPase) and nuclease domains unrelated to those of Avs proteins, was not required for SeAvs3-mediated toxicity (Fig. 1F).

Avs proteins recognize a diverse range of terminase and portal proteins

To investigate the specificity of Avs activation, we cloned the portal and large terminase subunit genes from 24 tailed phages, spanning nine major phage families, and coexpressed these genes in *E. coli* with 15 Avs systems spanning all four Avs families (data S2 and S3).

[1]Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [2]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA. [3]McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [4]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [5]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [6]Society of Fellows, Harvard University, Cambridge, MA 02138, USA. [7]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA.
*Corresponding author. Email: zhang@broadinstitute.org (F.Z.); algao@stanford.edu (L.A.G.)
†These authors contributed equally to this work.

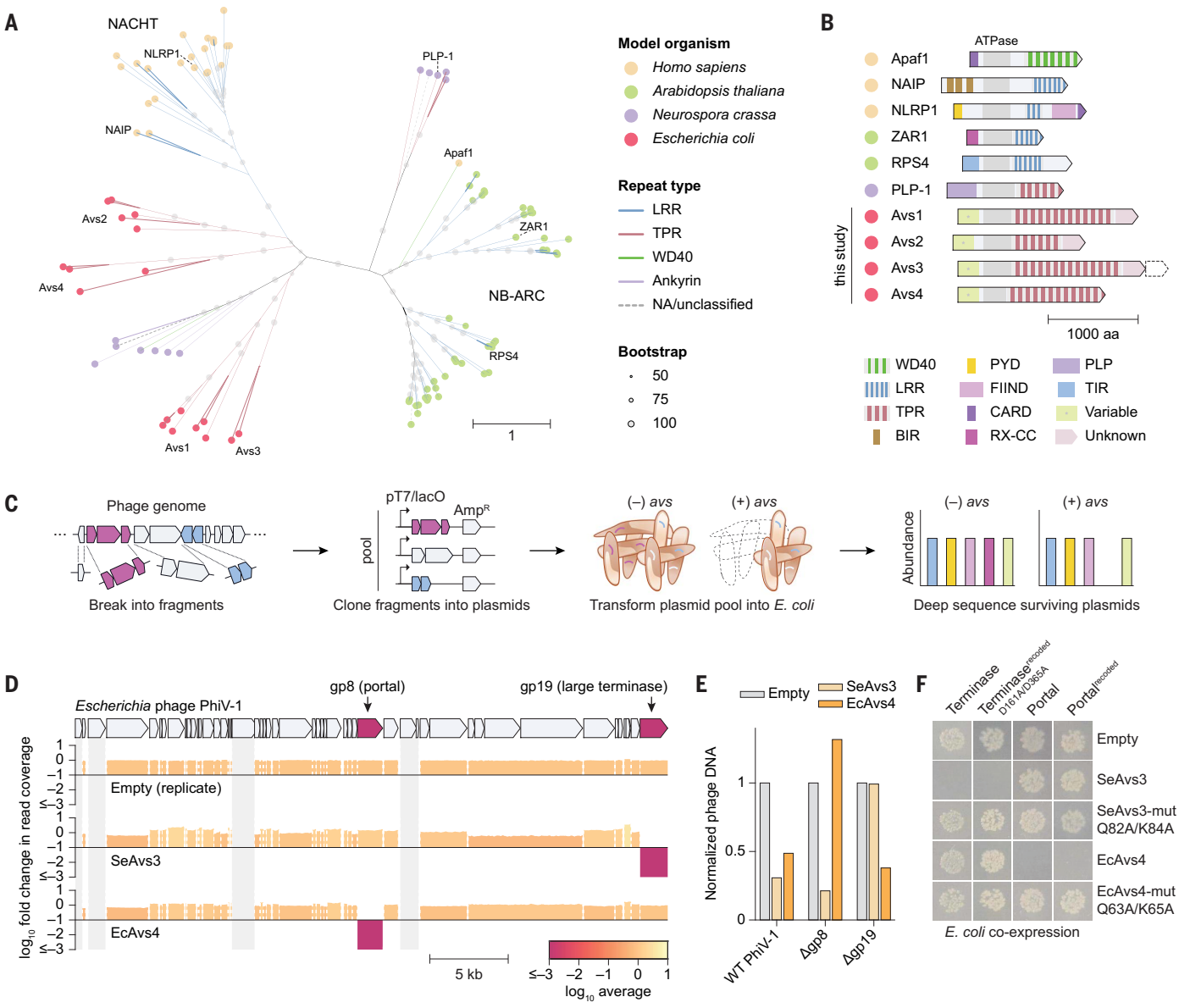


Fig. 1. Prokaryotic STAND NTPases recognize phage terminase and portal proteins. (**A**) Maximum likelihood tree of the ATPase domain of selected NLR-like STAND NTPases in four model organisms across kingdoms of life. (**B**) Domain architectures of representative NLR-like genes in (A). Ankyrin, ankyrin repeat; BIR, baculoviral inhibitor of apoptosis repeat; CARD, caspase activation and recruitment domain; FIIND, function to find domain; LRR, leucine-rich repeat; PLP, patatin-like phospholipase; PYD, pyrin domain; RX-CC, potato virus X resistance protein coiled-coil domain; TPR, tetratricopeptide repeat; WD40, WD40 repeat. (**C**) Schematic of the genetic screening approach used to identify phage-encoded activators of Avs proteins that induce cell death. Amp^R, ampicillin resistance gene. (**D**) Genetic screen results for phage-encoded activators. (**E**) Quantification of the phage DNA band intensity in a Southern blot of DNA isolated from phage-infected *E. coli*. WT, wild type. (**F**) Photographs of *E. coli* cotransformation assays with *avs* genes and phage activators identified in (D).

We quantified cellular toxicity and depletion of specific Avs-phage protein pairs for all 720 combinations by deep sequencing (Fig. 2A and data S2). These experiments revealed precise target specificity: Avs1 to Avs3 recognized only large terminase subunits, whereas Avs4 recognized only the portal protein (Fig. 2B and fig. S6). To assess the robustness of the assay, we repeated these experiments by varying the Avs promoter or the amount of terminase and portal induction and obtained similar results (fig. S7). Surprisingly, Avs1 and Avs2 also recognized terminases despite the lack of substantial sequence similarity among the C-terminal TPRs of Avs1 to Avs3, although we detected a structurally similar β sheet–rich domain at the end of the TPR arrays in all three proteins (fig. S8). These findings demonstrate conservation of target recognition across Avs families and suggest that the portal and large terminase subunit are key PAMPs that are recognized by prokaryotes. Moreover, Avs systems recognize PAMPs from diverse phages; for example, SeAvs3 and EcAvs2 were strongly activated by 20 of 24 and 19 of 24 tested terminases, respectively, and EcAvs4 was strongly activated by 15 of 24 tested portals (>100-fold depletion) (Fig. 1F). Because the portals and terminases from different phage families have limited sequence similarity, with less than 5% pairwise sequence identity in some cases (Fig. 2C), but share the same core fold (figs. S9 to S11), this broad range of activity implies that Avs proteins are triggered by conserved structural features rather than by specific peptide sequences. Consistent with

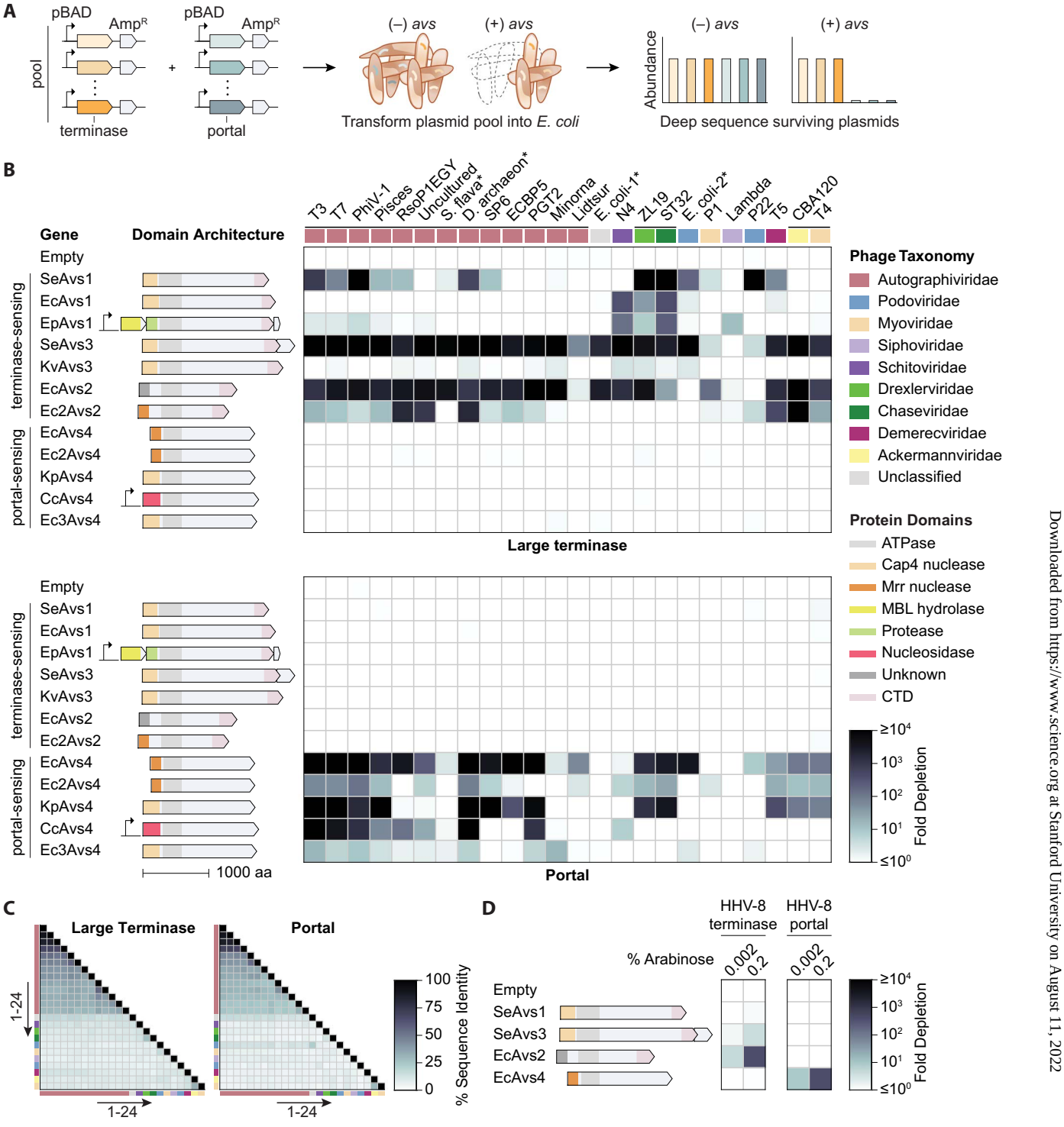


Fig. 2. Avs proteins are pattern-recognition receptors for the terminase and portal of diverse tailed phages. (**A**) Schematic of the plasmid depletion assay. (**B**) Heatmaps of plasmid depletion for the terminase and portal proteins of representative phages spanning nine major tailed phage families. The native *avs* promoter was retained for all homologs except for those outside of the Enterobacteriaceae family (EpAvs1 and CcAvs4). Terminases and portals were induced with 0.002% arabinose. Horizontal black lines at the top of the panel indicate groups of terminase proteins with at least 20% pairwise sequence identity. Asterisks indicate prophages. aa, amino acids; CTD, C-terminal domain; D. archaeon, *Desulfurococcales archaeon* ex4484_217_2; E. coli-1, *E. coli* NCTC9020; E. coli-2, *E. coli* M885; S. flava, *Sphingopyxis flava* R11H. (**C**) Pairwise amino-acid sequence identity between the core folds of the terminases and portals in (B), excluding nonconserved regions. (**D**) Activity of four Avs proteins against the human herpesvirus 8 (HHV-8) terminase and portal.

this possibility, EcAvs2 and EcAvs4 displayed weak but clear recognition of the terminase and portal, respectively, of human herpesvirus 8 (Fig. 2D), which is a highly diverged evolutionary derivative of tailed phages (*25*) and does not infect prokaryotes.

SeAvs3 and EcAvs4 are phage-activated DNA endonucleases

SeAvs3 and EcAvs4 contain predicted N-terminal PD-DExK–family nuclease domains (Fig. 3A), which we hypothesized degrade phage and cellular DNA upon target recognition. The nuclease domain of SeAvs3 is most similar to Cap4 effector nucleases of cyclic oligonucleotide–based defense systems (*26*, *27*), whereas EcAvs4 has an Mrr-like restriction endonuclease (*28*). Both Avs proteins contain conserved D-QxK or E-QxK catalytic motifs (Fig. 3B), and, in addition to the STAND NTPase, the SeAvs3 system contains a small open reading frame (ORF), the deletion of which reduced antiphage activity in *E. coli* (fig. S12A).

To biochemically reconstitute Avs activity in vitro, we purified recombinant SeAvs3, the protein encoded by the small ORF, EcAvs4, and the PhiV-1 terminase (gp19) and portal (gp8) proteins (fig. S13A). We incubated SeAvs3 and the small ORF product with linear double-stranded DNA (dsDNA) and observed progressive degradation of the substrate in the presence of gp19 but not gp8 (Fig. 3, C and D). This nuclease activity was dependent on the catalytic residues of SeAvs3 but did not require the small ORF product (Fig. 3C). We further investigated the substrate specificity of Avs systems and found that the nuclease activity was specific for dsDNA, whereas single-stranded DNA and RNA were not cleaved (fig. S13B). Moreover, SeAvs3 cleaved both linear and circular dsDNA, including *E. coli* genomic DNA (fig. S13C), indicative of endonuclease activity with no specificity for phage DNA, which is consistent with an abortive infection defense mechanism.

We next investigated cofactor requirements of SeAvs3 and found that in vitro activity depends on both Mg^{2+} and adenosine triphosphate

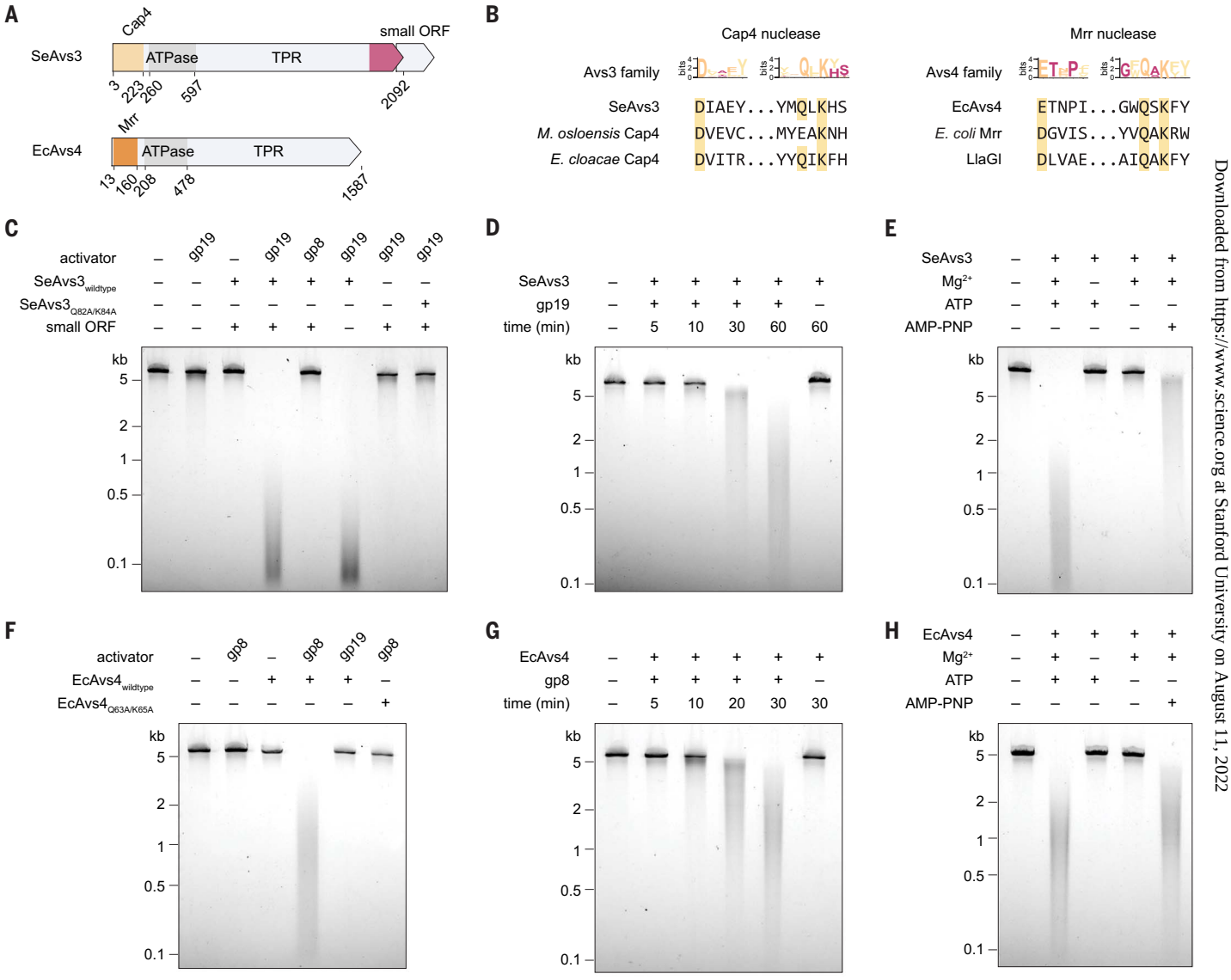


Fig. 3. SeAvs3 and EcAvs4 are phage-activated DNA endonucleases. (**A**) Domain architecture of SeAvs3 and EcAvs4. (**B**) Alignment of Avs D-QxK nuclease motifs with characterized Cap4 and Mrr representatives. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr. (**C** to **E**) Agarose gel analysis of SeAvs3 nuclease activity in vitro with a linear dsDNA substrate [(C) and (D)] and cofactor requirements (E). (**F** to **H**) Agarose gel analysis of EcAvs4 nuclease activity in vitro with a linear dsDNA substrate [(F) and (G)] and cofactor requirements (H).

(ATP); however, ATP hydrolysis is not strictly required because nuclease activity was observed at a reduced level in the presence of the nonhydrolyzable ATP analog adenylyl-imidodiphosphate (AMP-PNP) (Fig. 3E). We also found that the nuclease activity of EcAvs4 is activated by gp8, but not by gp19, and was abolished in an EcAvs4 Q63A/K65A (Gln^{63}→Ala/Lys^{65}→Ala) nuclease mutant (Fig. 3F). Similar to SeAvs3, nuclease activity of EcAvs4 required the presence, but not the hydrolysis, of ATP (Fig. 3H), consistent with phage plaque assays of SeAvs3 and EcAvs4 ATPase active-site mutants (fig. S12B). Together, these experiments indicate that SeAvs3 and EcAvs4 are promiscuous DNA endonucleases that are activated by distinct, highly conserved phage proteins in an ATP-dependent manner.

Structural basis of Avs binding and target recognition

To investigate how Avs systems recognize and bind their cognate phage proteins, we solved cryo–electron microscopy (cryo-EM) structures of the SeAvs3-terminase and EcAvs4-portal complexes in the presence of ATP and Mg^{2+} (figs. S14 to S17 and table S1). A reconstruction at 3.4-Å resolution revealed that SeAvs3 forms a tetramer, with each C-terminal TPR domain gripping the ATPase and nuclease domains of the gp19 terminase (Fig. 4, A and B). These TPR lobes are flexible and required symmetry expansion to improve their local resolution to 3.4 Å (Materials and methods and figs. S15 to S17). For EcAvs4 bound to the PhiV-1 gp8 portal (Fig. 4, C and D), image processing revealed equal abundances of a tetrameric complex and an octameric complex corresponding to tetramer head-to-head dimerization (fig. S14). At lower protein concentrations, however, we observed only the tetramer, indicating that it is most likely the functional complex (fig. S14). Negative-stain and cryo-EM imaging of SeAvs3 and EcAvs4 in the absence of phage proteins revealed only smaller monomeric particles (fig. S14), indicating that phage protein binding is required for the assembly of SeAvs3 and EcAvs4 into tetramers.

Tetramerization of both SeAvs3 and EcAvs4 is mediated through their STAND ATPase domains, which interact in a manner distinct from each other and from other characterized STAND ATPase oligomers like the Roq1 resistosome tetramer (*29*) or the Apaf1 apoptosome heptamer (*30*) (fig. S18). The SeAvs3 ATPase domain forms a C4-symmetric tetramer by interactions between the nucleotide-binding (NBD) and winged-helix (WHD) subdomains and the NBD subdomain of the adjacent protomer, whereas the EcAvs4 ATPase domain forms a C2-symmetric dimer of dimers with a tighter interface (1232 $Å^2$ of buried surface area, compared to 436 $Å^2$ for SeAvs3), with adjacent WHDs and NBDs both interacting (fig. S18). The smaller interface in SeAvs3 is compensated for by additional contacts between its C-terminal TPR domains (Fig. 4, A and B). SeAvs3 and EcAvs4 both maintain ATP in their active sites with an adjacent magnesium ion coordinated by the canonical Walker A and B motifs (Fig. 4, E and F). Notably, in both cases, tetramerization of the STAND ATPase domains brings adjacent N-terminal nuclease domains close together, forming two nuclease dimers with overall C2 symmetry (Fig. 4, G and L).

SeAvs3 and EcAvs4 contain nuclease effectors of the PD-DExK superfamily. Conventional PD-DExK nucleases (e.g., restriction endonucleases) use a pair of acidic residues to coordinate at least one metal ion and a conserved lysine residue to bind the scissile phosphate and stabilize the transition state for nucleolytic cleavage (Fig. 4J) (*31*, *32*). In the SeAvs3 Cap4 tetramer, this arrangement of residues is found in all four protomers; however, in the two "outward-facing" protomers, an extended β strand makes a steric block for DNA binding and/or metal coordination (Fig. 4I). In the two "inward-facing" protomers, a homotypic interface between the N-terminal α helices prevents the formation of this β strand, enabling the catalytic site to adopt the active configuration (Fig. 4H). Based on the crystal structure of the Hind III restriction endonuclease (*33*), the inward-facing protomers can be predicted to form a cavity for DNA binding, with each protomer likely cleaving opposite strands of the DNA (Fig. 4K). The EcAvs4 Mrr tetramer shows a similar principle, whereby the two inward-facing protomers contain active sites that resemble canonical PD-DExK nucleases, but in the outward-facing protomers, Glu^{49}, which is part of the conserved trio of active-site residues, is displaced (Fig. 4, M and N). Glu^{49} is found in the loop that spans residues 33 to 52, and interactions between this loop on an "inward" protomer and an adjacent "outward" protomer likely stabilize its position in the inward protomer. Like the SeAvs3 Cap4 tetramer, these two inward protomers form a cavity that accommodates DNA in a manner similar to Hind III (Fig. 4K).

SeAvs3 and EcAvs4 both contain extensive TPR domains for binding their cognate phage proteins, which we confirmed using a bacterial two-hybrid system and protein copurification (figs. S19 and S20). The SeAvs3 TPR domain forms a left-hand-like structure capped by a β sheet–rich C-terminal domain (Fig. 5A). This domain has two cavities in which the terminase ATPase and nuclease domains are nestled. Consistent with the ability of SeAvs3 to bind terminases with little sequence similarity (Fig. 2C), there are few specific residue-residue pair contacts between SeAvs3 and the PhiV-1 terminase. Instead, binding is determined by shape and charge complementarity between the two proteins, burying more than 3700 $Å^2$ of solvent-accessible surface area. This complementarity is maintained across a diverse range of experimental structures and AlphaFold models of phage terminases (Fig. 5B). Additionally, SeAvs3 directly recognizes residues within the two terminase active sites. In particular, Asp^{1710} in SeAvs3 forms a salt bridge with the highly conserved Arg^{61} within the Walker A motif of the terminase ATPase (Fig. 5, C and E). An arginine in this position is found in most terminase ATPases that activate SeAvs3 but not in nonactivating terminases (Fig. 5E and fig. S21). These observations suggest that Arg^{61} within the Walker A motif is a determinant of recognition specificity, and indeed, mutation of the cognate arginine in the T4 terminase ATPase domain substantially reduced SeAvs3 activation by the ATPase domain (Fig. 5F). Notably, an arginine is not typically found in this position in endogenous cellular ATPases (*34*), suggesting a possible mechanism for avoiding off-target activation. Furthermore, Arg^{1196} and Lys^{1198} in SeAvs3 form salt bridges to the four conserved aspartates that make up the active site of the terminase nuclease (Fig. 5, D and E), and mutation of Asp^{365} in the PhiV-1 terminase nuclease notably reduced SeAvs3 activation by the nuclease domain (Fig. 5F). Thus, SeAvs3 directly reads the active-site residues of both domains of the terminase. Furthermore, the ATP ligand bound by the terminase is detected by interactions between the gamma phosphate and His^{1770} and Tyr^{1714} of SeAvs3 (Fig. 5C). Targeting this ligand presumably helps avoid phage escape mutations, because ATP binding is required for the function of the terminase.

Because SeAvs3 detects both domains of the terminase, we hypothesized that there might be some functional redundancy in these interactions. Indeed, SeAvs3 was activated by the nuclease domain alone from some phages, including T7, but was also activated by the ATPase domain alone from T4 and ZL19, a T1 family phage (fig. S22). Likewise, SeAvs1 was activated by the nuclease domain from T7, but in the case of ST32, both the nuclease and ATPase domains were required. These results suggest that Avs1 and Avs3 recognize both the nuclease and ATPase domains but differ in the extent of activation by either domain, depending on the terminase. By contrast, deletion of the nuclease domain had no impact on Avs2 activity for any of the five tested terminases, suggesting that Avs2 recognizes the ATPase domain only (fig. S22). This pattern of recognition is consistent with the larger size of Avs1 and Avs3 compared with Avs2.

The TPR domain of EcAvs4 also binds the PhiV-1 portal with a large interface, burying 5800 $Å^2$ of solvent-accessible surface area, that includes notably few residue-residue contacts (Fig. 5G). The portal protein is recognized

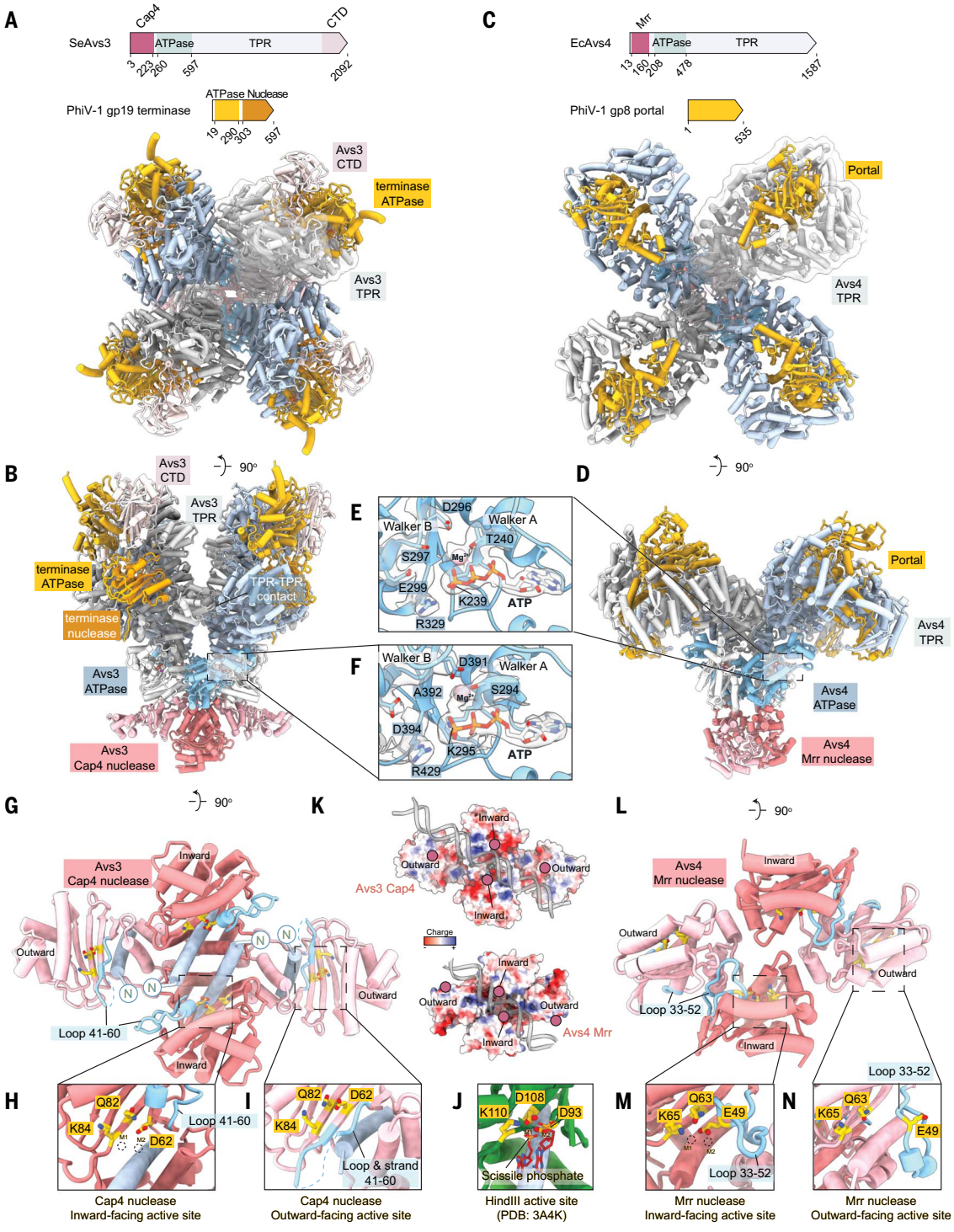


Fig. 4. Cryo-EM structures of SeAvs3 and EcAvs4 in complex with their cognate triggers. (**A** and **B**) Structure of the SeAvs3-terminase complex. (**C** and **D**) Structure of the EcAvs4-portal complex. (**E** and **F**) ATP molecule in the STAND ATPase active site of EcAvs4 and SeAvs3. The cryo-EM density is shown as a transparent surface. (**G**) SeAvs3 Cap4-like nuclease effector domain. (**H** and **I**) Active sites for the inward- and outward-facing protomers of the SeAvs3 Cap4-like nuclease. (**J**) Equivalent view of the active site of Hind III bound to target DNA with two divalent metal ions [Protein Data Bank (PDB) ID 3A4K]. (**K**) Electrostatic surface potential for the SeAvs3 Cap4-like nuclease and the EcAvs4 Mrr-like nuclease. Active sites are indicated by purple circles. Ideal B-form DNA is modeled on both surfaces based on the crystal structure of Hind III bound to its target (PDB ID 3A4K). (**L**) EcAvs4 Mrr-like nuclease effector domain. (**M** and **N**) Active sites for the inward- and outward-facing protomers of the EcAvs4 Mrr-like nuclease.

through its stem, clip, and part of its wing domain. In an assembled dodecameric portal complex, these regions are found toward the interior and are therefore more constrained in their fold requirements (Fig. 5I). Consistent with this observation, we performed random mutagenesis by polymerase chain reaction (PCR) to screen for portal mutations that abrogate Avs4 activation and found that all 29 identified mutations were nonconservative and located in the core wing or stem regions (fig. S23), possibly disrupting the core portal fold. The clip domain, which contains a conserved antiparallel β sheet with an intervening α helix, is recognized by β-sheet augmentation with a hairpin of EcAvs4 (Fig. 5H), a mode of fold recognition that does not depend on the amino acid sequence of the target. Because portal proteins are not enzymes, there are no active-site residues to target as in the SeAvs3-terminase complex. Finally, portal oligomerization is not compatible with the Avs4-bound state (Fig. 5H), suggesting that Avs4 recognizes portal monomers before they assemble into the procapsid.

Avs proteins are widespread and possess diverse, modular effector domains

To assess the diversity of *avs* genes in prokaryotes, we collected all intact homologs

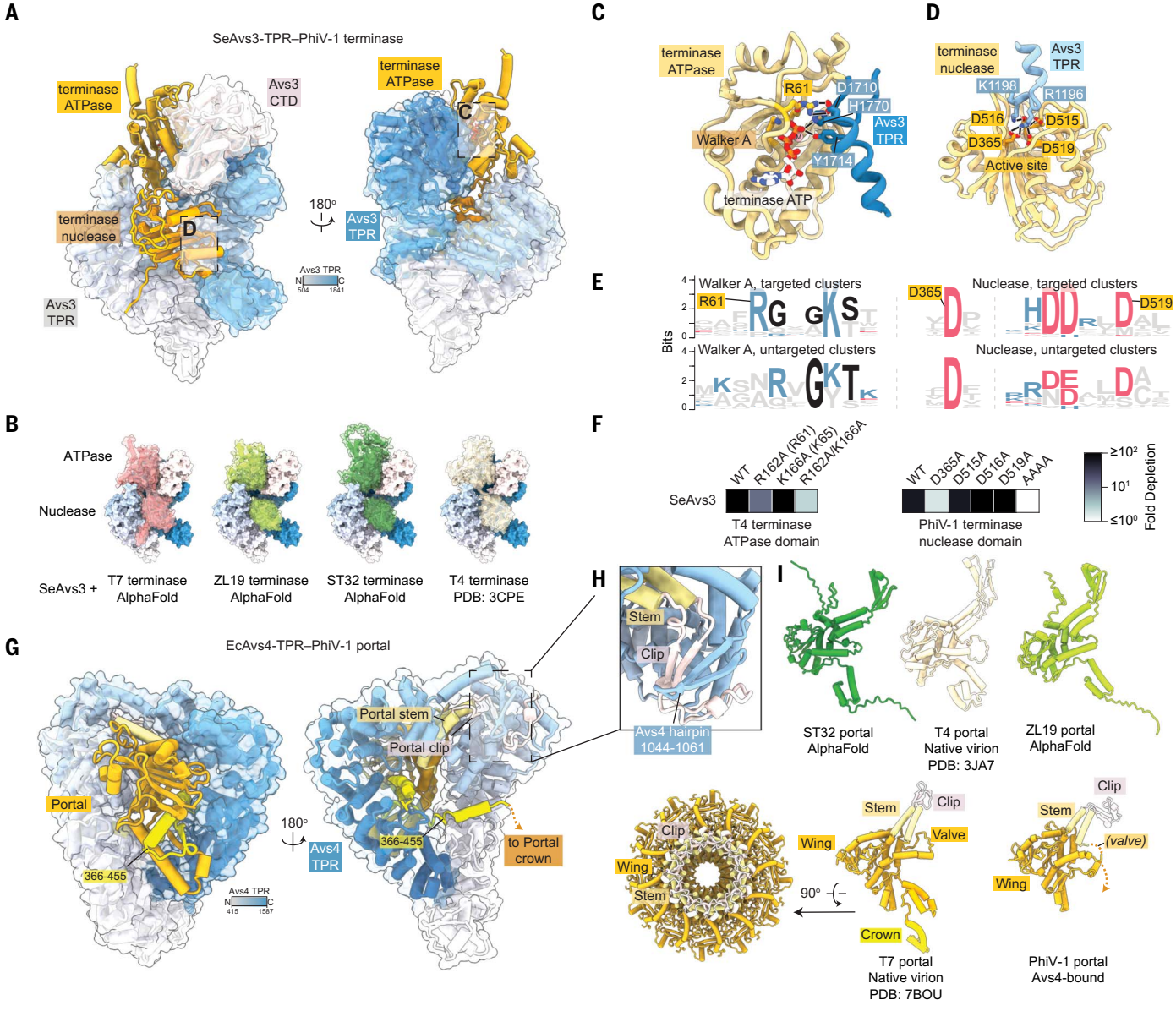


Fig. 5. Structural basis for viral-fold recognition by SeAvs3 and EcAvs4. (**A**) The interface between SeAvs3 and the PhiV-1 terminase. An SeAvs3 surface view is shown in transparency. SeAvs3 is colored from the N to C terminus according to the key. (**B**) AlphaFold or crystal structures of different terminases modeled into SeAvs3. The ATPase and nuclease domains were individually aligned to the PhiV-1 terminase domains. (**C** and **D**) Recognition of the PhiV-1 terminase ATPase and nuclease active sites by the SeAvs3 TPR domain. (**E**) Sequence logos for terminase ATPase Walker A motifs and terminase nuclease active sites. A total of 11,000 terminase sequences were clustered at 30% sequence identity, and motifs were extracted from clusters containing terminases targeted or not targeted by SeAvs3 according to Fig. 2B (see also fig. S21). (**F**) Plasmid depletion assay for SeAvs3 coexpressed in *E. coli* with a terminase ATPase or nuclease domain harboring active-site mutations. (**G**) The interface between EcAvs4 and the PhiV-1 portal. An EcAvs4 surface view is shown in transparency. EcAvs4 is colored from the N to C terminus according to the key. (**H**) β-sheet augmentation between EcAvs4 and the portal clip domain. (**I**) Comparison of the EcAvs4-bound state of the PhiV-1 portal, the cryo-EM structure of the highly homologous T7 portal in its native virion, and AlphaFold models of diverse portals. A top view of the assembled dodecamer of the T7 portal is also shown.

from each of the four families present in the National Center for Biotechnology Information (NCBI) nonredundant protein sequence database (data S4 to S8). The *avs* genes were identified in ~4 to 5% of sequenced prokaryotic genomes and are broadly distributed (Fig. 6A), with at least one *avs* gene detected in 27 of 29 and 3 of 10 well-represented bacterial and archaeal phyla, respectively (Fig. 6B and fig. S24A). Each Avs family has a characteristic protein size (Fig. 6C). We next constructed phylogenetic trees of each of the four families (Fig. 6, D and E; and fig. S24, B and C) and found that these trees did not follow bacterial and archaeal phylogenies, suggesting extensive horizontal gene transfer, particularly for *avs2* and *avs4*, in agreement with previous analyses of STAND NTPases (*14*, *15*). Furthermore, we detected at least 18 distinct types of N-terminal effector domains present in Avs proteins, including non-nuclease domains such as proteases, nucleosidases, sirtuins (SIR2), Toll/interleukin-1 receptor homology (TIR) domains, cytidine monophosphate (CMP) hydrolases, transmembrane

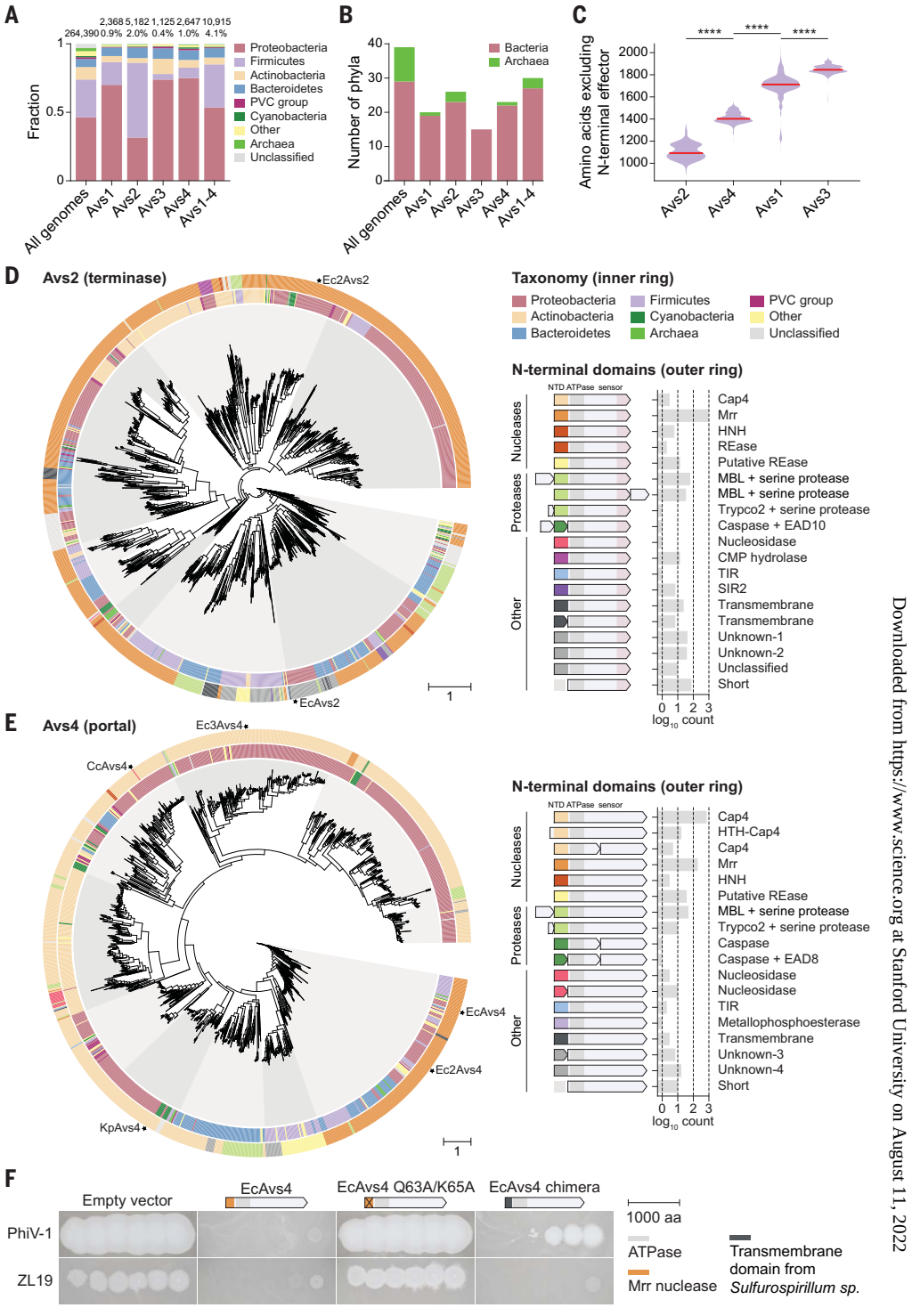


Fig. 6. Taxonomic distribution and domain architectures of Avs families. (**A**) Distribution of *avs* genes across phyla. The values above the bars indicate the number and percentage of genomes containing each gene. PVC, Planctomycetota, Verrucomicrobiota, and Chlamydiota. (**B**) Number of bacterial and archaeal phyla (minimum 100 sequenced isolates) with at least one detected instance of an *avs* gene. (**C**) Kernel density plots of the length distribution of Avs proteins, excluding the N-terminal domain. The red lines indicate medians. ****$p <$ 0.0001 (Mann-Whitney). (**D** and **E**) Maximum likelihood tree of representatives of the ATPase + C-terminal domain of Avs2 (n = 1255) (D) and Avs4 (n = 1089) (E) clustered at 95% sequence identity. See fig. S24 for the trees for Avs1 and Avs3. Stars on the outer ring indicate homologs investigated experimentally in this study. HTH, helix-turn-helix; MBL, metallo-β-lactamase; REase, restriction endonuclease. (**F**) Phage plaque assays showing antiphage defense activity of a chimeric Avs4 with transmembrane N-terminal helices from *Sulfurospirillum* sp. replacing the Mrr-like nuclease domain of EcAvs4. The X indicates a nuclease domain mutation.

helices, and domains with unknown functions (data S4 to S8). Some less common variants are predicted to participate in intracellular signaling networks through effector-associated domains (EADs) that may recruit a caspase-like protease by protein-protein interaction (*35*, *36*) (fig. S25), reminiscent of animal NLRs.

The apparent frequent exchange of N-terminal domains in the evolution of the Avs families emphasizes the modular organization characteristic of STAND NTPases (*15*) and implies that closely related ATPase and TPR domains can activate a wide range of effector functions beyond DNA cleavage. To test this hypothesis, we chose an Avs4 homolog from *Sulfurospirillum* sp. that contains ATPase and TPR domains that are highly similar to those of EcAvs4, with 44% overall amino acid identity, but encompasses an N-terminal region with predicted transmembrane helices instead of a nuclease (fig. S26A). We generated a chimeric Avs4 protein by transplanting the transmembrane domain to EcAvs4 and found that the chimera conferred protection against T7, PhiV-1, and ZL19 (Fig. 6F and fig. S26B) while retaining the ability to recognize the portal proteins from diverse phages (fig. S26C).

Phages inhibit Avs defense through diverse antidefense proteins

Bacterial and archaeal viruses have evolved diverse mechanisms to counteract defense systems (*37*), including numerous anti-restriction and anti-CRISPR proteins (*38*, *39*). We hypothesized that Avs inhibitors might exist among phage early genes, which are expressed before the portal and terminase genes during the phage life cycle. Focusing on the Autographiviridae family of T7-like coliphages, which have readily identifiable early genes, as well as portals and terminases that strongly activate Avs proteins, we identified a set of 122 representative early genes that typically encode small proteins (median length 77 amino acids), typical of antidefense genes (data S9). We performed a genetic screen for suppressors of Avs toxicity by coexpressing these genes with SeAvs3, EcAvs4, or KpAvs4 and their cognate trigger (Fig. 7A). We identified several early genes that rescued cell growth (Fig. 7B), most of which originate from a hypervariable region within a group of closely related phages isolated from wastewater (Fig. 7C) (*40*, *41*).

To validate these observations, we produced three of the Avs inhibitors by cell-free

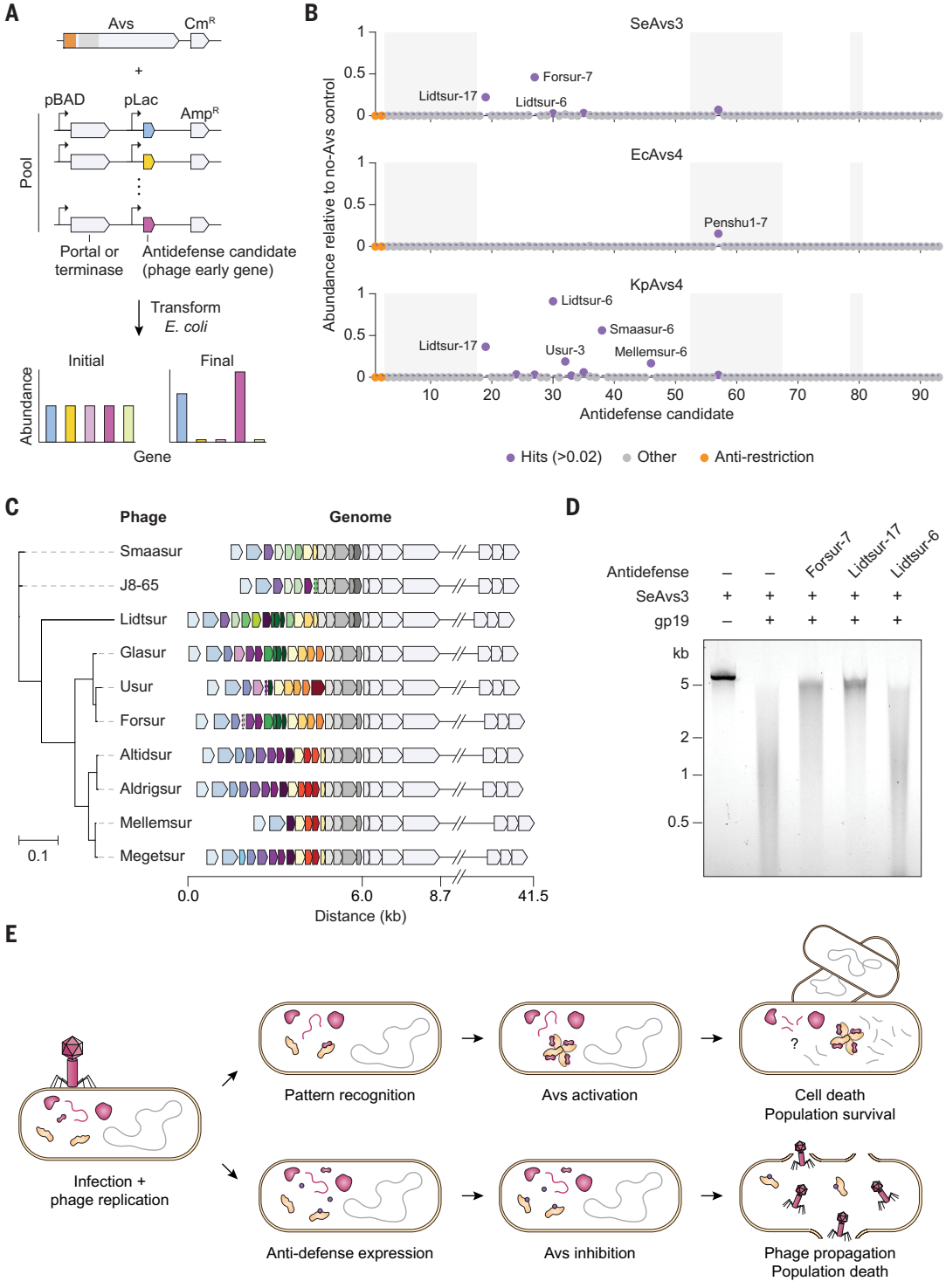


Fig. 7. Phage-encoded genes inhibit Avs activity. (**A**) Schematic of a pooled screen in *E. coli* for phage early genes that rescue Avs-mediated toxicity. Cm^R, chloramphenicol resistance gene. (**B**) Deep-sequencing readout of antidefense candidate genes coexpressed with SeAvs3, EcAvs4, or KpAvs4. (**C**) A hypervariable early-gene locus within a closely related set of Autographiviridae phages contains abundant antidefense genes. The tree was constructed from a concatenated alignment of conserved proteins present in all 10 phages. Colors represent groups of proteins clustered at 40% sequence identity and 70% coverage. (**D**) Agarose gel analysis showing in vitro reconstitution of anti-SeAvs3 activity by three antidefense candidates. (**E**) Schematic of the mechanism of Avs proteins as antiphage pattern-recognition receptors.

translation and observed inhibition of SeAvs3 nuclease activity in vitro by Lidtsur-17 and, to a lesser degree, by Forsur-7 (Fig. 7D). Lidtsur-17, Forsur-7, and Lidtsur-6 were also active in phage plaque assays and restored phage propagation on Avs-containing *E. coli* (fig. S27). Surprisingly, these inhibitors were active against different Avs families, including the chimeric Avs4, where the effector nuclease of EcAvs4 was replaced with a transmembrane domain. Furthermore, the lack of detectable sequence similarity between these inhibitors suggests distinct modes of action, which resembles the case of the highly diverse anti-CRISPRs (*38*, *39*). Further studies will be required to elucidate how these phage proteins block Avs activity.

Discussion

Here, we characterize four families of prokaryotic STAND NTPases and demonstrate that they are pattern-recognition receptors against two phage proteins: the large terminase subunit and the portal. These proteins, along with the major capsid protein, are the signature proteins of the virus realm Duplodnaviria, which unites tailed phages and tailed archaeal viruses with herpesviruses (*25*). Members of this realm, particularly tailed phages, are the most abundant among known viruses (*42*, *43*). The portal protein nucleates virion assembly, occupying the distinctive pentameric vertex of capsids and providing the attachment site for the phage tail, and serves as the channel for genome entry into and exit from the capsid (*44*). The terminase is the motor that packages the phage genome into the capsid at high density and pressure, using the energy of ATP hydrolysis, and cleaves DNA concatemers into genome-size units (*45*). The universal, complex molecular functions of these proteins engender strong selective constraints and hence evolutionary conservation. It is therefore logical that these particular proteins were selected as the targets for pattern recognition during the coevolution of prokaryotes with viruses. The three groups (Avs1 to Avs3) that recognize terminases do not form a clade in the phylogeny of the STAND domain (fig. S3), suggesting that defense based on terminase recognition evolved independently on multiple occasions.

The in vitro reconstitution and structural analysis of two Avs systems described here provide insight into their mechanism of defense, including promiscuous DNA endonuclease activity (Fig. 7E). Although many Avs proteins contain predicted nucleases, we identified diverse N-terminal effector domains throughout Avs families, indicating distinct mechanisms of defense that remain to be characterized. The demonstration that at least some of these effectors can be swapped without compromising Avs defense function highlights the modular functionality of these proteins, which appears important for the diversification of defense mechanisms. The effectors of SeAvs3 and EcAvs4 are both activated by tetramerization, suggesting that the diverse Avs effectors are unified by the requirement for oligomerization for activity, a common mechanism for signal transduction by STAND NTPases (*46*). Indeed, oligomerization is also involved in the activation of Cap4 nucleases in cyclic oligonucleotide-based antiphage signaling systems (*26*).

Notably, Avs proteins recognize conserved structural features of their cognate targets across an extreme variety of amino acid sequences, including those originating from both tailed phages and archaeal viruses, as well as eukaryotic herpesviruses, which are only distantly related and do not infect prokaryotes. Structural analysis of Avs3 revealed that it directly detects the active-site residues and ATP ligand of the terminase, thereby targeting the moieties that are the most difficult for phages to mutate without abrogating function.

Notable similarities, but also several differences, exist between eukaryotic NLRs and prokaryotic Avs proteins. Both are intracellular receptors of the STAND superfamily that detect PAMPs through C-terminal repetitive structures. Both exhibit triggered oligomerization, but with distinct interfaces between the central ATPase domains (fig. S18). Similar to plant NLRs like RPP1 and the Roq1 resistosome, both Avs3 and Avs4 form tetramers, with the effector domains activated by forming a twofold symmetric dimer of dimers (*29*, *47*). In the absence of their ligands, animal and plant NLRs have autoinhibited states that prevent oligomerization and effector activation (*19*). Whether such states exist for Avs proteins will require further investigation. In contrast to Avs proteins, NLRs often contain N-terminal adaptor domains that recruit other proteins to form macromolecular complexes known as inflammasomes or resistosomes (*19*). In these complexes, the effector (e.g., caspase-1) is a separate protein rather than a domain of the NLR. By contrast, Avs effectors are usually the N-terminal domain of the STAND NTPase. This simpler organization might be advantageous because counteracting phage replication requires a rapid, direct cellular response. This contrast parallels the distinction between the mechanisms of prokaryotic and eukaryotic STING proteins, whereby bacterial STING homologs directly activate TIR domain effectors rather than regulate transcription, as mammalian STINGs do (*19*, *48*).

Bacteria and archaea encode numerous diverse STAND NTPases beyond the four families characterized in this study (*15*). Although some are not involved in defense, such as the transcriptional regulators MalT (*22*), AfsR (*23*), and GutR (*24*), several are confirmed defense genes or are predicted to have a defense function based on their enrichment in genome regions adjacent to known defense systems (*4*, *49*). We investigated several of these other defense-associated systems (data S2) but observed no detectable toxicity when coexpressed with any of the 48 tested terminases or portals, suggesting that they are triggered by other pathogen-related patterns that remain to be identified. Further investigation will reveal whether these triggers are proteins and whether they are phage-specific or endogenous to the host. For instance, most characterized plant and fungal NLRs sense the state of host pathways rather than pathogen-specific proteins (*19*, *20*), and it remains a possibility that other groups of prokaryotic STAND NTPases function similarly.

Given the extensive sequence divergence among the STAND NTPases, it is unclear whether Avs proteins are direct evolutionary ancestors of eukaryotic NLRs, although this remains a possibility. Alternatively, or additionally, the characteristic tripartite domain architectures of diverse STAND NTPases could have evolved convergently, suggesting that this modular organization is a facile way to create allosterically activated enzymes that respond to PAMPs and could inspire the design of engineered molecular sensors. Overall, the results of this work advance our understanding of host-virus interactions in diverse microbes and extend the paradigm of pattern recognition of pathogen-specific proteins to all three domains of life.

Materials and methods

Phylogenetic analysis of STAND NTPases

For STAND phylogenetic analysis (fig. S3), PSI-BLAST searches (*50*) against the database of complete bacterial and archaeal genomes (extracted from GenBank, March 2019) were performed for three iterations using ATPase domains of seven Avs1 to Avs3 homologs (WP_126523998.1, WP_115407481.1, WP_084007836.1, WP_060615938.1, WP_139964370.1, WP_063118745.1, and WP_001017806.1) investigated experimentally. The 2000 best hits from each run were taken and combined with 949 Avs4 homologs found in the NCBI nonredundant protein sequence database (nr) in 2021. A nonredundant set of 4843 sequences was used for phylogenetic reconstruction using a hybrid UPGMA/FastTree approach as follows. At the first step, sequence clusters were obtained using MMseqs2 (*51*) with a sequence similarity threshold of 0.5, and the sequences within each cluster were aligned using MUSCLE (*52*). At the second step, cluster-to-cluster similarity scores were obtained using HHSEARCH (*53*) (including trivial clusters consisting of a single sequence each) and normalized by the minimum of the self-scores. Relative similarity scores (s) were converted to distances (d) defined as $d = -\ln s$, and a UPGMA (unweighted pair group method with arithmetic mean) dendrogram was constructed from the distance matrix (*54*). At the third step, sequence-based

trees were constructed from the cluster alignments using FastTree (*55*) (WAG evolutionary model, gamma-distributed site rates) and rooted by midpoint; these trees were grafted onto the tips of the profile similarity-based UPGMA dendrograms. FastTree was also used to calculate support values. Only the second step of the above procedure was applied to reconstruct the UPGMA dendrogram using multiple alignments of selected well-supported branches identified by the first procedure.

Construction of avs phylogenetic trees

Homologs of each of the four clades of *avs* genes were identified using PSI-BLAST searches against the NCBI nr database in June 2021 using position-specific scoring matrices for each clade derived from manually curated multiple sequence alignments (MSAs) of conserved regions. After a round of curation to remove false positives hits and partial proteins, referencing the corresponding genome assemblies to correct misannotated start codons, a list of 1584, 2342, 1018, and 1813 nonredundant full-length proteins were obtained for Avs1, Avs2, Avs3, and Avs4, respectively. To reduce sampling bias, sequences were then clustered at 95% sequence identity (minimum 80% coverage) using MMseqs2 with parameters–min-seq-id 0.95 -c 0.8–cov-mode 1. One representative from each cluster was selected for subsequent analyses, resulting in 843, 1255, 630, and 1089 sequences for Avs1, Avs2, Avs3, and Avs4, respectively.

MSAs of each Avs clade, excluding the variable N-terminal domains, were generated using MAFFT v7.450 (*56*) with global pairwise alignment (parameters–maxiterate 1000–globalpair). Alignments were trimmed using trimAl 1.2 with a gap threshold of 0.25 (-gt 0.25). Phylogenetic trees were built from the trimmed MSAs using IQ-TREE 1.6.12 (*57*) with the LG+G4 model and 2000 ultrafast bootstrap replicates (parameters -nstop 500 -bb 2000 -m LG+G4). To categorize the N-terminal domains, the N-terminal sequences were clustered using MMseqs2 with parameters–min-seq-id 0.4 -c 0.8, and a representative sequence from each cluster was analyzed using HHpred (*58*). Phyla classification was determined from the NCBI taxonomy database, and trees were rooted by midpoint and visualized using iTOL (*59*).

The phylogenetic tree comparing the ATPase domains of NLR-like genes across model organisms (Fig. 1A) was constructed in a similar manner, incorporating the set of 23 human NLRs and the best characterized NLRs from *Arabidopsis thaliana* (*60*) and *Neurospora crassa* (*20*, *61*).

Taxonomic distribution of avs genes

To determine the taxonomic distribution of *avs* genes, genome assemblies containing one or more full-length Avs homologs were identified through the NCBI Identical Protein Groups (IPG) database (data S8). Redundant assemblies were removed on the basis of their nine-digit accession numbers. To determine the percentage of genomes containing *avs* genes, the list of all available prokaryotic assemblies was downloaded from https://ftp.ncbi.nlm.nih.gov/genomes/GENOME_REPORTS/prokaryotes.txt.

Construction of terminase and portal alignments

Structures of all tested terminase and portal proteins were predicted using AlphaFold2 (*62*), and structures were aligned and visualized using PyMOL 2.3.4. Representatives of the predicted structures were used as input for MSA construction using PROMALS3D (*63*). Before computing pairwise sequence identities, MSAs were trimmed to retain only the regions corresponding to the core terminase or portal fold.

Cloning

Genes were chemically synthesized or amplified with Q5 [New England Biolabs (NEB)] or Phusion Flash (Thermo Scientific) polymerase. Plasmids were assembled using Gibson assembly. Plasmid sequences were verified by Tn5 tagmentation and high-throughput sequencing, as previously described (*4*, *64*).

Bacterial strains

E. coli NovaBlue and NovaBlue(DE3) were obtained from Millipore Sigma. *E. coli* K-12 (ATCC 25404) and strain C (ATCC 13706) were obtained from the American Type Culture Collection. All genetic assays were performed with *E. coli* NovaBlue(DE3) unless indicated otherwise.

Competent cell production

E. coli strains were cultured in ZymoBroth with 25 μg/ml chloramphenicol and made competent using Mix & Go buffers (Zymo) according to the manufacturer's recommended protocol.

PhiV-1 fragment screen

DNA fragments consisting of intact ORFs were amplified from phage PhiV-1 and cloned into expression plasmids after a LacI-repressed T7 promoter. Plasmids were pooled with an mNeonGreen-expressing control plasmid and transformed into *E. coli* NovaBlue(DE3) containing either SeAvs3, EcAvs4, or a pACYC184 empty vector. An additional sample consisting of the plasmid pool transformed into empty vector–containing *E. coli* NovaBlue, which does not express from T7 promoters, was also included to assess the basal toxicity of the phage genes.

After 1 hour of outgrowth in SOC (super optimal broth with catabolite repression) medium at 37°C, cells were plated on LB agar plates containing 25 μg/ml chloramphenicol and 100 μg/ml ampicillin in the absence of isopropyl-β-D-thiogalactopyranoside (IPTG). Plates were incubated for an additional 12 hours at 37°C, after which surviving plasmids were isolated by miniprep (Qiagen). A total of 200 ng of plasmid for each condition was tagmented with Tn5 to yield an average fragment size of ~500 base pairs. After addition of 0.5 volumes of 0.1% SDS and column purification, tagmented fragments were amplified over eight cycles by Q5 DNA polymerase (NEB) with sample-specific i5 and i7 index primers. Amplicons were gel extracted and sequenced on a NextSeq (Illumina) using 150 cycles for the forward read. Reads were mapped to reference sequences using Geneious Prime. The read coverage of each sample was then normalized to the read coverage of the mNeonGreen control within the same sample. Finally, for each sample, the read coverage per base was divided by the corresponding read coverage per base for the empty vector NovaBlue(DE3) control (Fig. 1D) or by that of the empty-vector NovaBlue control (fig. S4).

Terminase and portal depletion screens

Terminase and portal genes were amplified directly from phage samples or chemically synthesized (Twist Bioscience) with codon optimization for *E. coli*. Genes were expressed under the control of a pBAD promoter. Plasmids were pooled with an mCherry-expressing control plasmid and transformed into *E. coli* NovaBlue(DE3) containing an Avs system or a pACYC184 empty vector. After 1 hour of outgrowth in SOC at 37°C, cells were plated on LB agar plates containing 25 μg/ml chloramphenicol and 100 μg/ml ampicillin with 0.002% arabinose or, in some cases, with 0.2% arabinose, as detailed in the figures. After an additional 12 hours at 37°C, plasmids were isolated and sequenced, and depletion values (Fig. 2, B and D; and figs. S7 and S26C) were computed as described for the PhiV-1 depletion screen.

Portal and terminase mutant screens

Two synonymous versions of the T7 portal gene were randomly mutagenized by PCR using KAPA HiFi HotStart ReadyMix DNA polymerase (Roche) and cloned by Gibson assembly into a plasmid backbone containing a LacI-repressed T7 promoter. Plasmids were column purified and transformed into *E. coli* NovaBlue(DE3) containing EcAvs4, KpAvs4, or CcAvs4. Cells were plated on LB agar plates containing 25 μg/ml chloramphenicol and 100 μg/ml ampicillin in the absence of IPTG. After overnight growth, surviving colonies were sampled at random, cultured, and sequenced, and those containing single amino acid substitutions in the portal were retained for subsequent analysis. To reduce the number of stop codons and frameshift mutants sampled, a fluorescent protein (mNeonGreen) was included in

the plasmid backbone immediately after the portal ORF with a single-nucleotide overlap, such that mNeonGreen was translated only if the portal ORF remained intact (*65*) (fig. S23, A and B). Both portal and mNeonGreen were translated as separate polypeptides.

To quantitatively assess the effect of each mutant on Avs4-mediated toxicity (fig. S23C), mutant plasmids were pooled and retransformed into Avs4-containing *E. coli* as described for the PhiV-1 depletion screen. Fold depletion was also quantified as described, with the exception that only reads containing 20-mer sequences specific to one mutant (i.e., mapping to the mutation site) were counted in the analysis. A similar procedure was followed to quantify the effect of truncation of the terminase or portal (figs. S22 and S23E), as well as terminase-domain mutations (Fig. 5F).

Antidefense screen

Putative early genes from Autographiviridae coliphages were tabulated and clustered at 50% sequence identity and 50% coverage using MMseqs2 (–min-seq-id 0.5 -c 0.5), resulting in 120 clusters. One representative was selected from each cluster, along with two additional sequences, for a total of 122 initial candidates (data S9). Genes were synthesized by Twist Bioscience and cloned by Gibson assembly into expression vectors containing either the portal or terminase from phage PhiV-1 driven by a pBAD promoter. Antidefense candidates were expressed under the control of a *lac* promoter. Plasmids were pooled and transformed into *E. coli* containing SeAvs3, EcAvs4, KpAvs4, or an empty vector. Cells were grown at 37°C for 16 hours on LB agar plates containing 25 μg/ml chloramphenicol and 100 μg/ml ampicillin with no added arabinose. After plasmid isolation, antidefense candidates were amplified over two rounds of PCR to attach eight-nucleotide i7 and i5 index barcodes and sequenced with a 600-cycle MiSeq kit to ensure maximal coverage of each ORF. Reads containing mutations were discarded in the subsequent analysis.

Phage plaque assays

E. coli host strains were grown to saturation at 37°C in LB or Terrific Broth. To 10 ml of top agar (10 g/liter tryptone, 5 g/liter yeast extract, 10 g/liter NaCl, 7 g/liter agar), chloramphenicol was added to a final concentration 25 μg/ml and, if needed, ampicillin to a final concentration of 100 μg/ml. Ten-fold dilutions of phage in phosphate-buffered saline were spotted on the plates. After overnight incubation at 37°C, plates were photographed in a dark room with a white backlight.

Construction of mutant phages

PhiV-1 gp8 and gp19 knockout phages were constructed as previously described (*66*, *67*) using plasmid donors with homology arms to gp8 or gp19 in a *trxA*$^{-}$ *E. coli* strain [JW5856 from the Keio collection (*68*)]. The *trxA* gene was inserted via the donor into the PhiV-1 genome as a selection marker.

Protein purification

Avs genes were cloned into pCDF-Duet expression plasmids containing a C-terminal 6xHis tag. PhiV-1 gp8 and gp19 genes were cloned into TwinStrep-SUMO expression plasmids. Proteins were expressed in *E. coli* BL21(DE3) cells (NEB no. C2527H). Cells were grown in Terrific Broth to mid–log phase, and the temperature was lowered to 18°C. Expression was induced at OD_{600} 0.6 with 0.25 mM IPTG for 16 to 20 hours before harvesting and freezing cells at −80°C. Cell paste was resuspended in lysis buffer (50 mM Tris-HCl pH 7.5, 500 mM NaCl, 5% glycerol) supplemented with EDTA-free cOmplete protease inhibitor (Roche). Cells were lysed using a LM20 microfluidizer device (Microfluidics), and cleared lysate was bound to either Strep-Tactin Superflow Plus (Qiagen) or Ni-NTA Superflow resin (Qiagen). For TwinStrep-SUMO phage proteins, the resin was washed with lysis buffer, and proteins were eluted with lysis buffer supplemented with 5 mM desthiobiotin. The TwinStrep-SUMO tag was removed by overnight digest at 4°C with homemade SUMO protease Ulp1 at a 1:100 weight ratio of protease to target. Cleaved proteins were run on a Superose 6 Increase column (GE Healthcare Life Sciences) with a final storage buffer of 25 mM Tris-HCl pH 7.5, 500 mM NaCl, 10% glycerol, and 1 mM dithiothreitol (DTT).

Avs proteins containing 6xHis tags were bound to Ni-NTA resin in the presence of 25 mM imidazole, washed with lysis buffer containing 50 mM imidazole, and eluted with lysis buffer containing 300 mM imidazole. SeAvs3 was diluted to a final concentration of 100 mM NaCl and purified using a Resource Q column on an AKTA pure 25 L (GE Healthcare Life Sciences) with a 100 mM-1M NaCl gradient. EcAvs4 was further purified by diluting to a final concentration of 100 mM NaCl and absorbing contaminants by flowing the protein over a Resource Q and Heparin HP column. SeAvs3 and EcAvs4 were concentrated and loaded onto a Superose 6 Increase column with a final storage buffer of 25 mM Tris-HCl pH 7.5, 500 mM NaCl, 10% glycerol, and 1 mM DTT. SeAvs3 for cryo-EM analysis was purified in the same buffer without glycerol and only 300 mM NaCl, then concentrated to 1.4 mg/ml in a 500-μl 100,000 molecular weight cutoff (MWCO) Amicon spin concentrator.

Avs complex purification

Avs-TwinStrep constructs were cotransformed with plasmids expressing either gp8 or gp19 into electrocompetent *E. coli* BL21(DE3) (Sigma Aldrich CMC0016) and grown and induced as before. Avs pulldowns using Strep-Tactin Superflow Plus resin were run on SDS-PAGE (SDS–polyacrylamide gel electrophoresis) gels, and gp8 and gp19 bands were excised and confirmed by mass spectrometry (Taplin Biological Mass Spectrometry Facility, Harvard Medical School). For tandem affinity purification, plasmids containing SeAvs3-6xHis and gp19-StrepTag were cotransformed into electrocompetent *E. coli* BL21(DE3) and grown and induced as before. An SeAvs3-gp19 complex was purified using Ni-NTA followed by Strep-Tactin SuperFlow Plus resin. The final elution was run on a Superose 6 Increase column and yielded a peak elution at 13 ml containing a 1:1 ratio of SeAvs3 and gp19, as determined by SDS-PAGE band intensity analysis. A standard curve was generated using the Bio-Rad Gel Filtration Standard (1511901), and the gel-phase distribution coefficient (K_{ave}) was calculated as (elution volume – void column)/(column volume – void volume).

SeAvs3-terminase complex formation for cryo-EM

A total of 20 μg of SeAvs3 was mixed with 8.3 μg of PhiV-1 gp19 terminase in a total volume of 24 μl in the presence of 17 mM Tris-HCl pH 7.5, 280 mM NaCl, 0.8 mM DTT, 2% glycerol, 5 mM $MgCl_2$, and 1 mM ATP. The reaction was incubated at 37°C for 30 min and placed on ice for ~1 hour before cryo-EM grid preparation. Cryo-EM grids were prepared on a Thermo Scientific Vitrobot Mark IV at 4°C and 100% humidity. A total of 3 μl of reaction was applied to a freshly glow-discharged (12 s at 15 mA) Cu 300 R2/2 holey carbon grid with a 2-nm layer of amorphous carbon (Quantifoil). After 30 s, the grid was manually blotted with Whatman Grade 1 filter paper and plunged into liquid ethane.

EcAvs4-portal complex formation for cryo-EM

PhiV-1 gp8 was cloned into an MBP-bdSUMO expression plasmid, and EcAvs4 was cloned into a pCDF-Duet plasmid with an internal TwinStrep tag added between residues 114 and 115. The EcAvs4 Mrr-like nuclease active site was mutated (Q63A/K65A) to allow co-expression with the portal. These two plasmids were cotransformed into *E. coli* BL21(DE3). A total of 6 liters of culture was grown in Terrific Broth to mid–log phase, and the temperature was lowered to 18°C. IPTG (0.25 mM) was added to induce expression, and growth was continued overnight. Cell paste was resuspended in lysis buffer (50 mM Tris-HCl pH 7.4, 250 mM NaCl, 5% glycerol, 5 mM β-mercaptoethanol, 2 mM $MgCl_2$, 0.1 mM ATP) supplemented with EDTA-free cOmplete protease inhibitor (Roche). Cells were lysed using a LM20 microfluidizer (Microfluidics) and cleared lysate was bound to Amylose Resin

High Flow (NEB). After extensive washing with lysis buffer, the resin was eluted overnight at 4°C by addition of 10 μg of homemade bdSENP1 protease. Eluted protein was incubated with Strep-Tactin Superflow Plus resin, washed with lysis buffer, then eluted with lysis buffer supplemented with 5 mM desthiobiotin. The eluate was concentrated in a 6-ml Vivaspin spin concentrator (30,000 MWCO) and run on a Superose 6 Increase column using 20 mM Tris-HCl pH 7.4, 200 mM NaCl, 2 mM $MgCl_2$, and 0.1 mM ATP. Peak fractions containing EcAvs4 and gp8 were concentrated to 1.7 mg/ml using a 0.5 ml Amicon spin concentrator (100,000 MWCO) then immediately used for cryo-EM grid preparation. Cryo-EM grids were prepared on an Thermo Scientific Vitrobot Mark IV at 4°C and 100% humidity. A total of 3 μl of sample was applied to a freshly glow-discharged (60 s at 15 mA) Cu 300 R1.2/1.3 holey carbon grid (Quantifoil). The grid was blotted for 4 s with blot force +5 and drain time 1 s, then plunged into liquid ethane.

Cryo-EM data collection

All data were collected using the Thermo Scientific Titan Krios G3i cryo TEM at MIT.nano using a K3 direct detector (Gatan) operated in super-resolution mode with twofold binning and an energy filter with slit width of 20 eV.

For SeAvs3-gp19, 15,422 movies were collected at 105,000× magnification giving a real pixel size of 0.8697 Å, with defocus ranging from 1 to 3.5 μm with an exposure time of 1.15 s, fractionated into 30 frames and a flux of 19.7 e^-/pix/s giving a total fluence per micrograph of 30 $e^-/Å^2$. For EcAvs4-gp8, 22,902 movies were collected at 130,000× magnification giving a real pixel size of 0.6788 Å, with defocus ranging from 1 μm to 2.5 μm with an exposure time of 0.6 s, fractionated into 24 frames and a flux of 23.6 e^-/pix/s giving a total fluence per micrograph of 30.8 $e^-/Å^2$.

Cryo-EM data processing

All cryo-EM data were processed using RELION-4.0 (*69*). Movies were corrected for motion using the RELION implementation of MotionCor2, with 5-by-5 patches and dose-weighting. Contrast transfer function (CTF) parameters were estimated using CTFFIND-4.1. All reported resolutions use the gold-standard Fourier shell correlation with a cutoff of 0.143.

For the SeAvs3-gp19 dataset, particle picking was first carried out on 800 micrographs using the Topaz general model (*70*). A good subset of these particles, as determined by three-dimensional (3D) classification, was used to train Topaz, and this trained model was used to pick 128,500 particles from the entire dataset. Extracted particles, downscaled fourfold, were subjected to 3D classification without imposing symmetry using a reference derived from a preliminary dataset. A total of 44,489 particles, corresponding to 34.5% of picked particles, showed sharp features and apparent C4 symmetry and were reextracted without binning and refined with C4 symmetry imposed. After refining per-particle defocus and global magnification, beamtilt, trefoil, and performing Bayesian polishing, a reconstruction was yielded at 3.8-Å resolution with clear density for the SeAvs3 ATPase domain but blurred density for both the N-terminal nuclease and C-terminal TPR+terminase domains.

To improve density for the N-terminal nuclease domains, 3D classification without alignment was performed while imposing C2 symmetry. This revealed two equal populations of particles each with clear density for the nuclease domains, related by a 90° rotation in the *z* axis. In the refinement STAR file, the parameter *rlnAngleRot* was therefore incremented by 90° for one of these populations, before focused refinement starting at 1.8° local angular searches with a soft mask around the nuclease domains. This produced a reconstruction at 3.4 Å, measured using the same soft mask.

To improve density for the C-terminal TPR+terminase domains, C4 symmetry expansion was performed on the C4 refinement data.star file, followed by particle subtraction with recentering using a mask around one of the four asymmetric units. This generated four subparticles for each original particle. Refinement starting at 1.8° local angular searches with a soft mask, followed by CTF refinement and another round of refinement, produced a reconstruction at 3.4-Å resolution.

For the EcAvs4-gp8 dataset, 1825 particles from 80 micrographs were manually picked and used to train Topaz. The trained Topaz model then picked 444,626 particles from the entire dataset, which were extracted with fourfold binning and subjected with 3D classification without imposing symmetry using the octameric (pseudo-D4) EcAvs4-gp8 reference derived from a preliminary dataset. A total of 133,133 particles (29.9%) showing the same pseudo-D4 symmetry were reextracted at 1.034 Å/pix and refined with D4 symmetry imposed. After Bayesian polishing, this yielded a 3.7-Å resolution reconstruction. D1 symmetry expansion followed by particle subtraction was then used to convert these particles to 266,266 subparticles that correspond to the tetrameric complex. Like SeAvs3-gp19, these also had blurry density for the N-terminal nuclease and C-terminal TPR+terminase domains but additionally had poor density for the ATPase domains, suggesting a C2 reconstruction might be suitable for the whole tetramer.

To improve overall density, a 3D classification without alignment was first performed with C2 symmetry imposed. This produced two equally occupied classes, collectively representing a 169,977-particle subset (63.8%), that appeared identical but for a 90° rotation, but less clearly distinguished than the same analysis on SeAvs3-gp19. Therefore, they were refined together with local 1.8° angular searches and C2 symmetry but with "Relax symmetry: C4" to account for the pseudo-C4 symmetry. This produced a consensus C2 refinement but still with relatively blurred densities for the nuclease and C-terminal TPR domains. The nuclease domain density was improved by focused refinement with a soft mask, followed by refining anisotropic magnification, per-(sub)particle defocus, and beamtilt, trefoil and fourth-order aberrations, and second refinement, yielding a 2.9-Å resolution reconstruction. The C-terminal TPR domains were improved by C4 symmetry expansion, followed by C1 focused refinement with a soft mask and CTF refinement, but still had unclear density at the periphery at the site of an important EcAvs4-portal contact. Therefore, a final 3D classification was performed with a soft mask just around this contact and a regularization parameter (T) of 20. A total of 500,066 selected subparticles (73%) were then focus-refined with the same mask to yield a reconstruction at 3.0-Å resolution with better density for this region.

Model building

Initial models for SeAvs3, EcAvs4, PhiV-1 gp8, and PhiV-1 gp19 were generated using AlphaFold and fit into the cryo-EM maps using ISOLDE (*71*) with adaptive distance restraints, followed by manual rebuilding in Coot (*72*) and further refinement in ISOLDE. Coordinates were refined in real space using PHENIX (*73*), performing one macrocycle of global minimization and atomic displacement parameter (ADP) refinement and skipping local grid searches.

In vitro cleavage reactions

Purified Avs proteins were incubated with nucleic acid substrates in reaction buffer (20 mM HEPES pH 7.5, 100 mM NaCl, 1 mM DTT, 5% glycerol). Typical reactions contained ~100 ng of DNA substrate, 100 ng Avs protein, and 100 ng gp8 or gp19 in a 10-μl reaction volume. $MgCl_2$ was added at 5 mM where indicated, and ATP and AMP-PNP at 1 mM. Reactions were carried out at 37°C for the indicated time, and products were purified using a QIAquick PCR Purification column (Qiagen) before agarose gel analysis with a 1% E-Gel EX (Thermo Fisher Scientific).

Bacterial two-hybrid assays

Expression plasmids were cloned by fusing either the T18 or T25 fragments of CyaA from *Bordetella pertussis* to nuclease-deficient Avs proteins, as well as the PhiV-1 gp8 portal and

gp19 terminase. *E. coli* BTH101 cells (*F*$^{-}$, *cya-99*, *araD139*, *galE15*, *galK16*, *rpsL1 (Str*^{r}*)*, *hsdR2*, *mcrA1*, *mcrB1*) were cotransformed with pairs of T18 and T25 containing plasmids. Overnight cultures were diluted 1:20 and plated on indicator plates containing 50 mg/ml ampicillin, 25 mg/ml kanamycin, 500 μg/ml ammonium iron(III) citrate, 300 μg/ml S-gal, and 0.5 mM IPTG. Cells were grown at 30°C overnight before imaging.

Southern blot analysis

E. coli K-12 (ATCC 25404) cultures were grown to mid–log phase (OD_{600} 0.5), and for each sample, 6 ml of culture was infected with wild-type or mutant PhiV-1 at a multiplicity of infection of 1. After 20 min at 37°C, before cell lysis, infected cells were pelleted and resuspended in 200 μl of media. After further incubation at 37°C, for a total of 90 min, samples were frozen in liquid nitrogen. DNA was extracted from 200-μl cultures by adding 200 μl of lysis buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 100 mM NaCl, 1% SDS, 2% Triton X-100), 100 μl of glass beads, and 200 μl of phenol-chloroform (1:1) followed by brief vortexing. Samples were centrifuged at 4°C, and DNA from the upper layer was extracted with chloroform and precipitated with the addition of 1 ml of ice-cold 100% ethanol and centrifugation at 4°C. DNA pellets were resuspended in 200 μl of Tris-EDTA (TE) with 300 μg RNAse A (Sigma-Aldrich) and incubated at 37°C for 1 hour. DNA was once again precipitated with the addition of 1 ml of ice-cold 100% ethanol and 20 μl of 4M ammonium acetate, centrifuged, dried, and resuspended in TE.

DNA was digested with Eco47 III and run on a 1% agarose gel in 1× Tris-borate-EDTA (TBE) at 100 V. The gel was denatured with 0.5 M NaOH and 1.5 M NaCl for 30 min and neutralized with 1.5 M NaCl and 0.5 M Tris-HCl pH 7.5 for 30 min. DNA was transferred to a Hybond N+ membrane (GE Healthcare Life Sciences) using overnight capillary flow and 10X SSC buffer (1.5 M NaCl, 150 mM sodium citrate, pH 7). Membranes were UV-crosslinked (Stratalinker 1800, Agilent) and blocked at 61°C with Church hybridization buffer (250 mM NaH_2PO_4 pH 7.2, 1 mM EDTA, 7% SDS). Radiolabeled probes complementary to the gp13 gene were generated from purified PCR products using the Prime-It Random labeling kit (Agilent) and [α-32P]-dCTP. Membranes were probed overnight, washed three times with 61°C Church hybridization buffer, and exposed overnight with x-ray film (GE Healthcare Life Science) before developing. Quantification of phage DNA bands was performed in Fiji with background signal subtracted.

REFERENCES AND NOTES

1. K. S. Makarova, Y. I. Wolf, S. Snir, E. V. Koonin, Defense islands in bacterial and archaeal genomes and prediction of novel defense systems. *J. Bacteriol.* **193**, 6039–6056 (2011). doi: 10.1128/JB.05535-11; pmid: 21908672
2. E. V. Koonin, K. S. Makarova, Y. I. Wolf, Evolutionary genomics of defense systems in archaea and bacteria. *Annu. Rev. Microbiol.* **71**, 233–261 (2017). doi: 10.1146/annurev-micro-090816-093830; pmid: 28657885
3. S. Doron *et al.*, Systematic discovery of antiphage defense systems in the microbial pangenome. *Science* **359**, eaar4120 (2018). doi: 10.1126/science.aar4120; pmid: 29371424
4. L. Gao *et al.*, Diverse enzymatic activities mediate antiviral immunity in prokaryotes. *Science* **369**, 1077–1084 (2020). doi: 10.1126/science.aba0372; pmid: 32855333
5. H. G. Hampton, B. N. J. Watson, P. C. Fineran, The arms race between bacteria and their phage foes. *Nature* **577**, 327–336 (2020). doi: 10.1038/s41586-019-1894-8; pmid: 31942051
6. D. Klaiman, E. Steinfels-Kohn, G. Kaufmann, A DNA break inducer activates the anticodon nuclease RloC and the adaptive immunity in *Acinetobacter baylyi* ADP1. *Nucleic Acids Res.* **42**, 328–339 (2014). doi: 10.1093/nar/gkt851; pmid: 24062157
7. C. K. Guegler, M. T. Laub, Shutoff of host transcription triggers a toxin-antitoxin system to cleave phage RNA and abort infection. *Mol. Cell* **81**, 2361–2373.e9 (2021). doi: 10.1016/j.molcel.2021.03.027; pmid: 33838104
8. R. Cheng *et al.*, A nucleotide-sensing endonuclease from the Gabija bacterial defense system. *Nucleic Acids Res.* **49**, 5216–5229 (2021). doi: 10.1093/nar/gkab277; pmid: 33885789
9. R. Bingham, S. I. Ekunwe, S. Falk, L. Snyder, C. Kleanthous, The major head protein of bacteriophage T4 binds specifically to elongation factor Tu. *J. Biol. Chem.* **275**, 23219–23226 (2000). doi: 10.1074/jbc.M002546200; pmid: 10801848
10. A. Millman *et al.*, Bacterial retrons function in anti-phage defense. *Cell* **183**, 1551–1561.e12 (2020). doi: 10.1016/j.cell.2020.09.065; pmid: 33157039
11. S. Kronheim *et al.*, A chemical defence against phage infection. *Nature* **564**, 283–286 (2018). doi: 10.1038/s41586-018-0767-x; pmid: 30518855
12. A. Bernheim *et al.*, Prokaryotic viperins produce diverse antiviral molecules. *Nature* **589**, 120–124 (2021). doi: 10.1038/s41586-020-2762-2; pmid: 32937646
13. K. S. Makarova, Y. I. Wolf, E. V. Koonin, Comparative genomics of defense systems in archaea and bacteria. *Nucleic Acids Res.* **41**, 4360–4377 (2013). doi: 10.1093/nar/gkt157; pmid: 23470997
14. E. V. Koonin, L. Aravind, Origin and evolution of eukaryotic apoptosis: The bacterial connection. *Cell Death Differ.* **9**, 394–404 (2002). doi: 10.1038/sj.cdd.4400991; pmid: 11965492
15. D. D. Leipe, E. V. Koonin, L. Aravind, STAND, a class of P-loop NTPases including animal and plant regulators of programmed cell death: Multiple, complex domain architectures, unusual phyletic patterns, and evolution by horizontal gene transfer. *J. Mol. Biol.* **343**, 1–28 (2004). doi: 10.1016/j.jmb.2004.08.023; pmid: 15381417
16. Y. Zhao *et al.*, The NLRC4 inflammasome receptors for bacterial flagellin and type III secretion apparatus. *Nature* **477**, 596–600 (2011). doi: 10.1038/nature10510; pmid: 21918512
17. E. M. Kofoed, R. E. Vance, Innate immune recognition of bacterial ligands by NAIPs determines inflammasome specificity. *Nature* **477**, 592–595 (2011). doi: 10.1038/nature10394; pmid: 21874021
18. R. Caruso, N. Warner, N. Inohara, G. Núñez, NOD1 and NOD2: Signaling, host defense, and inflammatory disease. *Immunity* **41**, 898–908 (2014). doi: 10.1016/j.immuni.2014.12.010; pmid: 25526305
19. J. D. G. Jones, R. E. Vance, J. L. Dangl, Intracellular innate immune surveillance devices in plants and animals. *Science* **354**, aaf6395 (2016). doi: 10.1126/science.aaf6395; pmid: 27934708
20. J. Heller, C. Clavé, P. Gladieux, S. J. Saupe, N. L. Glass, NLR surveillance of essential SEC-9 SNARE proteins induces programmed cell death upon allorecognition in filamentous fungi. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E2292–E2301 (2018). doi: 10.1073/pnas.1719705115; pmid: 29463729
21. S. Bauernfried, M. J. Scherr, A. Pichlmair, K. E. Duderstadt, V. Hornung, Human NLRP1 is a sensor for double-stranded RNA. *Science* **371**, eabd0811 (2021). doi: 10.1126/science.abd0811; pmid: 33243852
22. O. Danot, A complex signaling module governs the activity of MalT, the prototype of an emerging transactivator family. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 435–440 (2001). doi: 10.1073/pnas.98.2.435; pmid: 11209048
23. S. Horinouchi *et al.*, Primary structure of AfsR, a global regulatory protein for secondary metabolite formation in *Streptomyces coelicolor* A3(2). *Gene* **95**, 49–56 (1990). doi: 10.1016/0378-1119(90)90412-K; pmid: 2253887
24. R. Ye, S. N. Rehemtulla, S. L. Wong, Glucitol induction in *Bacillus subtilis* is mediated by a regulatory factor, GutR. *J. Bacteriol.* **176**, 3321–3327 (1994). doi: 10.1128/jb.176.11.3321-3327.1994; pmid: 8195087
25. E. V. Koonin *et al.*, Global organization and proposed megataxonomy of the virus world. *Microbiol. Mol. Biol. Rev.* **84**, e00061-19 (2020). doi: 10.1128/MMBR.00061-19; pmid: 32132243
26. B. Lowey *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). doi: 10.1016/j.cell.2020.05.019; pmid: 32544385
27. K. S. Makarova *et al.*, Evolutionary and functional classification of the CARF domain superfamily, key sensors in prokaryotic antivirus defense. *Nucleic Acids Res.* **48**, 8828–8847 (2020). doi: 10.1093/nar/gkaa635; pmid: 32735657
28. J. Heitman, P. Model, Site-specific methylases induce the SOS DNA repair response in *Escherichia coli*. *J. Bacteriol.* **169**, 3243–3250 (1987). doi: 10.1128/jb.169.7.3243-3250.1987; pmid: 3036779
29. R. Martin *et al.*, Structure of the activated ROQ1 resistosome directly recognizing the pathogen effector XopQ. *Science* **370**, eabd9993 (2020). doi: 10.1126/science.abd9993; pmid: 33273074
30. M. Zhou *et al.*, Atomic structure of the apoptosome: Mechanism of cytochrome *c*- and dATP-mediated activation of Apaf-1. *Genes Dev.* **29**, 2349–2361 (2015). doi: 10.1101/gad.272278.115; pmid: 26543158
31. K. Steczkiewicz, A. Muszewska, L. Knizewski, L. Rychlewski, K. Ginalski, Sequence, structure and functional diversity of PD-(D/E)XK phosphodiesterase superfamily. *Nucleic Acids Res.* **40**, 7016–7045 (2012). doi: 10.1093/nar/gks382; pmid: 22638584
32. A. Pingoud, M. Fuxreiter, V. Pingoud, W. Wende, Type II restriction endonucleases: Structure and mechanism. *Cell. Mol. Life Sci.* **62**, 685–707 (2005). doi: 10.1007/s00018-004-4513-1; pmid: 15770420
33. N. Watanabe, Y. Takasaki, C. Sato, S. Ando, I. Tanaka, Structures of restriction endonuclease HindIII in complex with its cognate DNA and divalent cations. *Acta Crystallogr. D Biol. Crystallogr.* **65**, 1326–1333 (2009). doi: 10.1107/S0907444909041134; pmid: 19966419
34. A. M. Burroughs, L. M. Iyer, L. Aravind, Comparative genomics and evolutionary trajectories of viral ATP dependent DNA-packaging systems. *Genome Dyn.* **3**, 48–65 (2007). doi: 10.1159/000107603; pmid: 18753784
35. G. Kaur, L. M. Iyer, A. M. Burroughs, L. Aravind, Bacterial death and TRADD-N domains help define novel apoptosis and immunity mechanisms shared by prokaryotes and metazoans. *eLife* **10**, e70394 (2021). doi: 10.7554/eLife.70394; pmid: 34061031
36. G. Kaur, A. M. Burroughs, L. M. Iyer, L. Aravind, Highly regulated, diversifying NTP-dependent biological conflict systems with implications for the emergence of multicellularity. *eLife* **9**, e52696 (2020). doi: 10.7554/eLife.52696; pmid: 32101166
37. J. E. Samson, A. H. Magadán, M. Sabri, S. Moineau, Revenge of the phages: Defeating bacterial defences. *Nat. Rev. Microbiol.* **11**, 675–687 (2013). doi: 10.1038/nrmicro3096; pmid: 23979432
38. A. Pawluk, A. R. Davidson, K. L. Maxwell, Anti-CRISPR: Discovery, mechanism and function. *Nat. Rev. Microbiol.* **16**, 12–17 (2018). doi: 10.1038/nrmicro.2017.120; pmid: 29062071
39. J. Bondy-Denomy, A. Pawluk, K. L. Maxwell, A. R. Davidson, Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature* **493**, 429–432 (2013). doi: 10.1038/nature11723; pmid: 23242138
40. N. S. Olsen, L. Forero-Junco, W. Kot, L. H. Hansen, Exploring the remarkable diversity of culturable *Escherichia coli* phages in the Danish wastewater environment. *Viruses* **12**, 986 (2020). doi: 10.3390/v12090986; pmid: 32899836
41. M. Schmerer, I. J. Molineux, J. J. Bull, Synergy as a rationale for phage therapy using phage cocktails. *PeerJ* **2**, e590 (2014). doi: 10.7717/peerj.590; pmid: 25279269
42. G. F. Hatfull, R. W. Hendrix, Bacteriophages and their genomes. *Curr. Opin. Virol.* **1**, 298–303 (2011). doi: 10.1016/j.coviro.2011.06.009; pmid: 22034588
43. H.-W. Ackermann, 5500 Phages examined in the electron microscope. *Arch. Virol.* **152**, 227–243 (2007). doi: 10.1007/s00705-006-0849-1; pmid: 17051420
44. C. L. Dedeo, G. Cingolani, C. M. Teschke, Portal protein: The orchestrator of capsid assembly for the dsDNA tailed

bacteriophages and herpesviruses. *Annu. Rev. Virol.* **6**, 141–160 (2019). doi: 10.1146/annurev-virology-092818-015819; pmid: 31337287
45. S. R. Casjens, The DNA-packaging nanomotor of tailed bacteriophages. *Nat. Rev. Microbiol.* **9**, 647–657 (2011). doi: 10.1038/nrmicro2632; pmid: 21836625
46. O. Danot, E. Marquenet, D. Vidal-Ingigliardi, E. Richet, Wheel of life, wheel of death: A mechanistic insight into signaling by STAND proteins. *Structure* **17**, 172–182 (2009). doi: 10.1016/j.str.2009.01.001; pmid: 19217388
47. S. Ma *et al*., Direct pathogen-induced assembly of an NLR immune receptor complex to form a holoenzyme. *Science* **370**, eabe3069 (2020). doi: 10.1126/science.abe3069; pmid: 33273071
48. B. R. Morehouse *et al*., STING cyclic dinucleotide sensing originated in bacteria. *Nature* **586**, 429–433 (2020). doi: 10.1038/s41586-020-2719-5; pmid: 32877915
49. S. A. Shmakov, K. S. Makarova, Y. I. Wolf, K. V. Severinov, E. V. Koonin, Systematic prediction of genes functionally linked to CRISPR-Cas systems by gene neighborhood analysis. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E5307–E5316 (2018). doi: 10.1073/pnas.1803440115; pmid: 29784811
50. S. F. Altschul *et al*., Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997). doi: 10.1093/nar/25.17.3389; pmid: 9254694
51. M. Steinegger, J. Söding, MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, 1026–1028 (2017). doi: 10.1038/nbt.3988; pmid: 29035372
52. R. C. Edgar, MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 113 (2004). doi: 10.1186/1471-2105-5-113; pmid: 15318951
53. M. Steinegger *et al*., HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinformatics* **20**, 473 (2019). doi: 10.1186/1471-2105-5-113; pmid: 15318951
54. K. S. Makarova *et al*., Evolutionary classification of CRISPR-Cas systems: A burst of class 2 and derived variants. *Nat. Rev. Microbiol.* **18**, 67–83 (2020). doi: 10.1038/s41579-019-0299-x; pmid: 31857715
55. M. N. Price, P. S. Dehal, A. P. Arkin, FastTree 2—Approximately maximum-likelihood trees for large alignments. *PLOS ONE* **5**, e9490 (2010). doi: 10.1371/journal.pone.0009490; pmid: 20224823
56. K. Katoh, K. Misawa, K. Kuma, T. Miyata, MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002). doi: 10.1093/nar/gkf436; pmid: 12136088
57. L.-T. Nguyen, H. A. Schmidt, A. von Haeseler, B. Q. Minh, IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015). doi: 10.1093/molbev/msu300; pmid: 25371430
58. L. Zimmermann *et al*., a completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J. Mol. Biol.* **430**, 2237–2243 (2018). doi: 10.1016/j.jmb.2017.12.007; pmid: 29258817
59. I. Letunic, P. Bork, Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021). doi: 10.1093/nar/gkab301; pmid: 33885785
60. A.-L. Van de Weyer *et al*., A species-wide inventory of NLR genes and alleles in *Arabidopsis thaliana*. *Cell* **178**, 1260–1272.e14 (2019). doi: 10.1016/j.cell.2019.07.038; pmid: 31442410
61. W. Dyrka *et al*., Diversity and variability of NOD-like receptors in fungi. *Genome Biol. Evol.* **6**, 3137–3158 (2014). doi: 10.1093/gbe/evu251; pmid: 25398782
62. J. Jumper *et al*., Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021). doi: 10.1038/s41586-021-03819-2; pmid: 34265844
63. J. Pei, B. H. Kim, N. V. Grishin, PROMALS3D: A tool for multiple protein sequence and structure alignments. *Nucleic Acids Res.* **36**, 2295–2300 (2008). doi: 10.1093/nar/gkn072; pmid: 18287115
64. S. Picelli *et al*., Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.* **24**, 2033–2040 (2014). doi: 10.1101/gr.177881.114; pmid: 25079858
65. M. Huber *et al*., Translational coupling via termination-reinitiation in archaea and bacteria. *Nat. Commun.* **10**, 4006 (2019). doi: 10.1038/s41467-019-11999-9; pmid: 31488843
66. U. Qimron, B. Marintcheva, S. Tabor, C. C. Richardson, Genomewide screens for Escherichia coli genes affecting growth of T7 bacteriophage. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 19039–19044 (2006). doi: 10.1073/pnas.0609428103; pmid: 17135349
67. A. M. Grigonyte *et al*., Comparison of CRISPR and marker-based methods for the engineering of phage T7. *Viruses* **12**, 193 (2020). doi: 10.3390/v12020193; pmid: 32050613
68. T. Baba *et al*., Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: The Keio collection. *Mol. Syst. Biol.* **2**, 0008 (2006). pmid: 16738554
69. D. Kimanius, L. Dong, G. Sharov, T. Nakane, S. H. W. Scheres, New tools for automated cryo-EM single-particle analysis in RELION-4.0. *Biochem. J.* **478**, 4169–4185 (2021). doi: 10.1042/BCJ20210708; pmid: 34783343
70. T. Bepler *et al*., Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nat. Methods* **16**, 1153–1160 (2019). doi: 10.1038/s41592-019-0575-8; pmid: 31591578
71. T. I. Croll, ISOLDE: A physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr. D Struct. Biol.* **74**, 519–530 (2018). doi: 10.1107/S2059798318002425; pmid: 29872003
72. A. Casañal, B. Lohkamp, P. Emsley, Current developments in Coot for macromolecular model building of electron cryo-microscopy and crystallographic data. *Protein Sci.* **29**, 1069–1078 (2020). doi: 10.1002/pro.3791; pmid: 31730249
73. D. Liebschner *et al*., Macromolecular structure determination using x-rays, neutrons and electrons: Recent developments in Phenix. *Acta Crystallogr. D Struct. Biol.* **75**, 861–877 (2019). doi: 10.1107/S2059798319011471; pmid: 31588918

ACKNOWLEDGMENTS

We thank D. Li, H. Altae-Tran, and G. Faure for valuable discussions; R. Tomaino for assistance with mass spectrometry; E. Brignole and C. Borsa for the smooth running of the MIT.nano cryo-EM facility; and the entire Zhang lab for support and advice. **Funding:** L.A.G. is supported by the Society of Fellows of Harvard University. K.S.M. and E.V.K. are supported by intramural funds of the US Department of Health and Human Services (to the National Library of Medicine). F.Z. is supported by NIH grants 1DP1-HL141201 and 2R01HG009761-05; the Howard Hughes Medical Institute; Open Philanthropy; the Edward Mallinckrodt, Jr. Foundation; the Poitras Center for Psychiatric Disorders Research at MIT; the Hock E. Tan and K. Lisa Yang Center for Autism Research at MIT; the Yang-Tan Molecular Therapeutics Center at McGovern; and the Phillips family and J. and P. Poitras. **Author contributions:** L.A.G. performed genetic assays and computational analysis. M.E.W. performed cryo-EM and determined the structures. J.S. performed in vitro biochemistry and bacterial two-hybrid assays. L.A.G. and J.S. carried out phage plaque assays. L.A.G. and K.S.M. performed the phylogenetic analysis. L.A.G., M.E.W., J.S., K.S.M., E.V.K., and F.Z. analyzed data. F.Z. supervised the research with support from R.K.M. L.A.G., M.E.W., and J.S. wrote the manuscript with input from all authors. **Competing interests:** F.Z. is a scientific advisor and cofounder of Editas Medicine, Beam Therapeutics, Pairwise Plants, Arbor Biotechnologies, and Sherlock Biosciences. The Broad Institute has filed a patent application that includes work described in this manuscript. **Data and materials availability:** Expression plasmids for all Avs systems and inhibitors are available from Addgene (#188786-188861). The genome sequence of phage ZL19 has been deposited in GenBank under accession number OM258170.1. The cryo-EM maps have been deposited in the Electron Microscopy Data Bank with the following codes: EMD-27421 (SeAvs3-gp19, C2 refinement), EMD-27424 (SeAvs3-gp19, C4-expanded C1 refinement), EMD-27425 (EcAvs4-gp8, C2 refinement), EMD-27422 (EcAvs4-gp8, C2 refinement focused on Mrr), and EMD-27426 (EcAvs4-gp8, C4-expanded C1 refinement). The coordinates for the composite atomic models have been deposited in the Protein Data Bank under accession codes 8DGC (SeAvs3-gp19) and 8DGF (EcAvs4-gp8).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abm4096
Figs. S1 to S27
Table S1
References (*74*–*82*)
MDAR Reproducibility Checklist
Data S1 to S19

View/request a protocol for this paper from *Bio-protocol*.

Submitted 15 September 2021; accepted 7 July 2022
10.1126/science.abm4096

Prokaryotic innate immunity through pattern recognition of conserved viral proteins

Linyi Alex GaoMax E. WilkinsonJonathan StreckerKira S. MakarovaRhiannon K. MacraeEugene V. KooninFeng Zhang

Science, 377 (6607), eabm4096.

STAND against viral invaders

The innate immune systems of animals, plants, and fungi universally use nucleotide binding oligomerization domain–like receptors (NLRs) of the STAND superfamily to detect molecular patterns common to pathogens. Gao *et al*. show that NLR-based immune pattern recognition is also prevalent in bacteria and archaea, something that was not known before. In particular, the authors characterized four families of NLR-like genes, finding that they are specific sensors for two highly conserved bacteriophage proteins. Upon binding to the target, these NLRs activate diverse effector domains, including nucleases, to prevent phage propagation. These findings demonstrate that pattern recognition of pathogen-specific proteins is a common mechanism of immunity across all domains of life. —DJ

Science
AAAS

Supplementary Materials for

Prokaryotic innate immunity through pattern recognition of conserved viral proteins

Linyi Alex Gao *et al*.

Corresponding authors: Linyi Alex Gao, algao@stanford.edu; Feng Zhang, zhang@broadinstitute.org

Science **377**, eabm4096 (2022)
DOI: 10.1126/science.abm4096

The PDF file includes:

Figs. S1 to S27
Table S1
References

Other Supplementary Material for this manuscript includes the following:

MDAR Reproducibility Checklist
Data S1 to S19

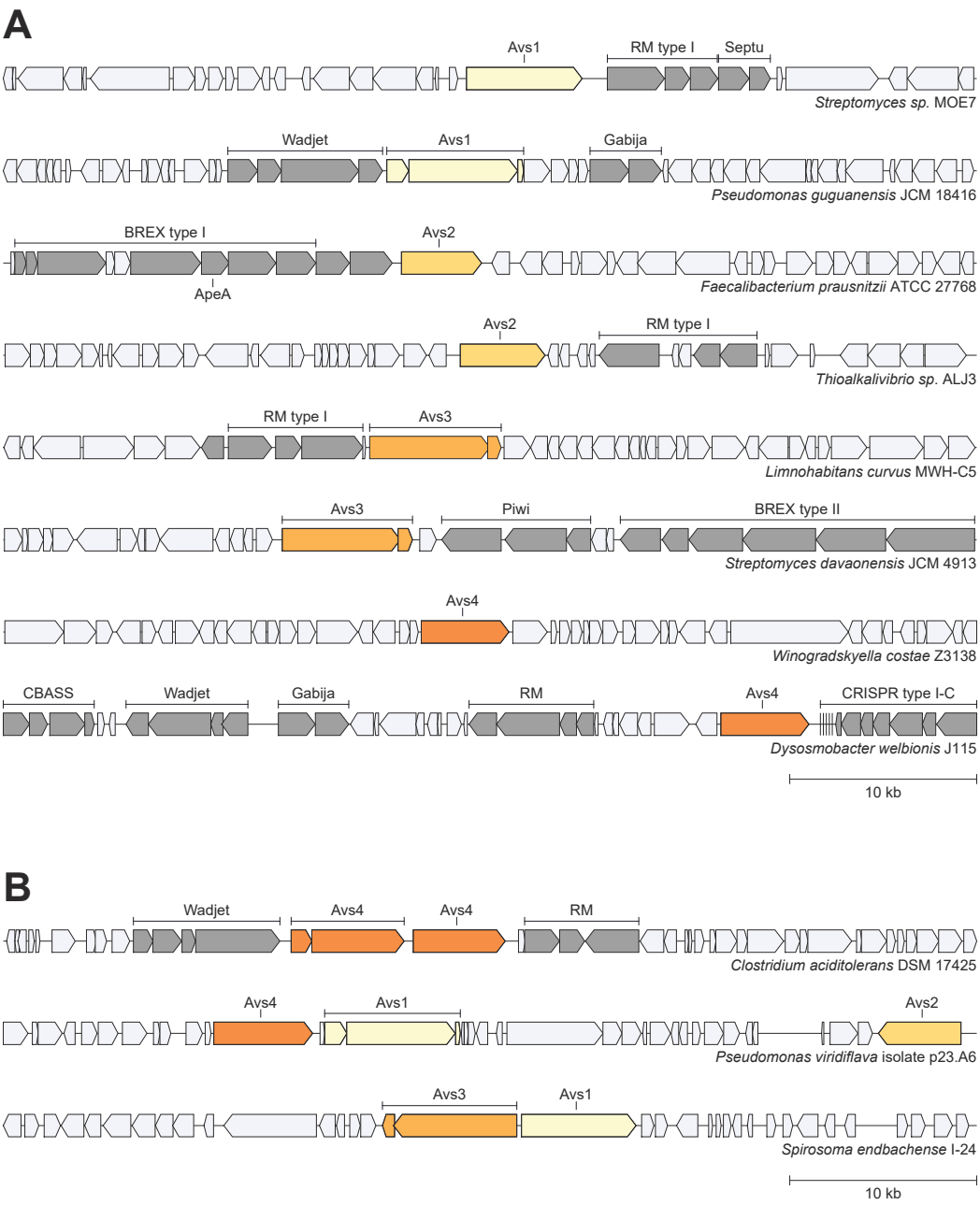


Figure S1: Example of *avs* genes (**A**) present in defense islands and (**B**) clustered with other *avs* genes. Other defense genes are highlighted in gray.

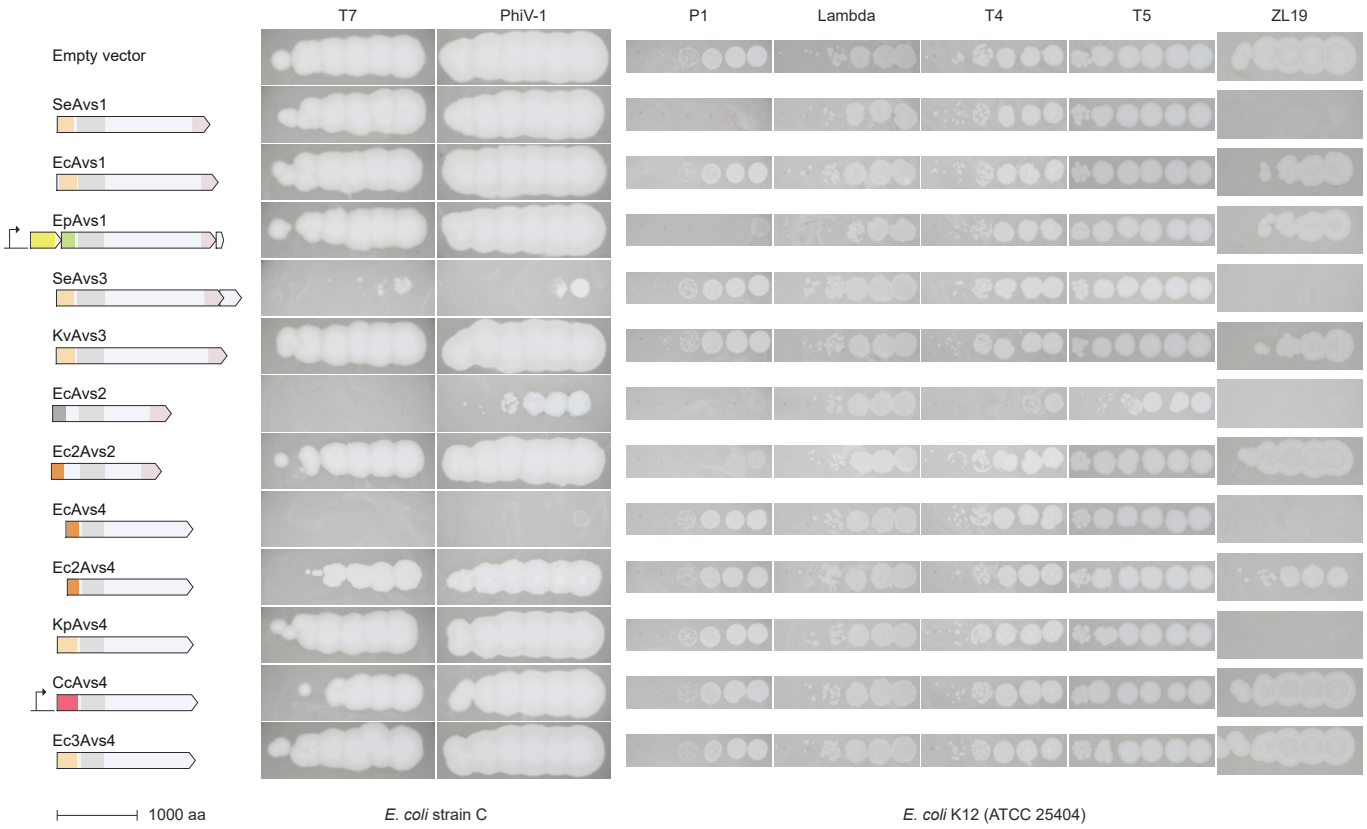


Figure S2: Heterologous reconstitution of Avs anti-phage activity in *E. coli*. Plaque assay spots correspond to 10-fold dilutions of phages T7, PhiV-1, P1, Lambda, T4, T5, and ZL19 on *E. coli* containing Avs-expressing plasmids.

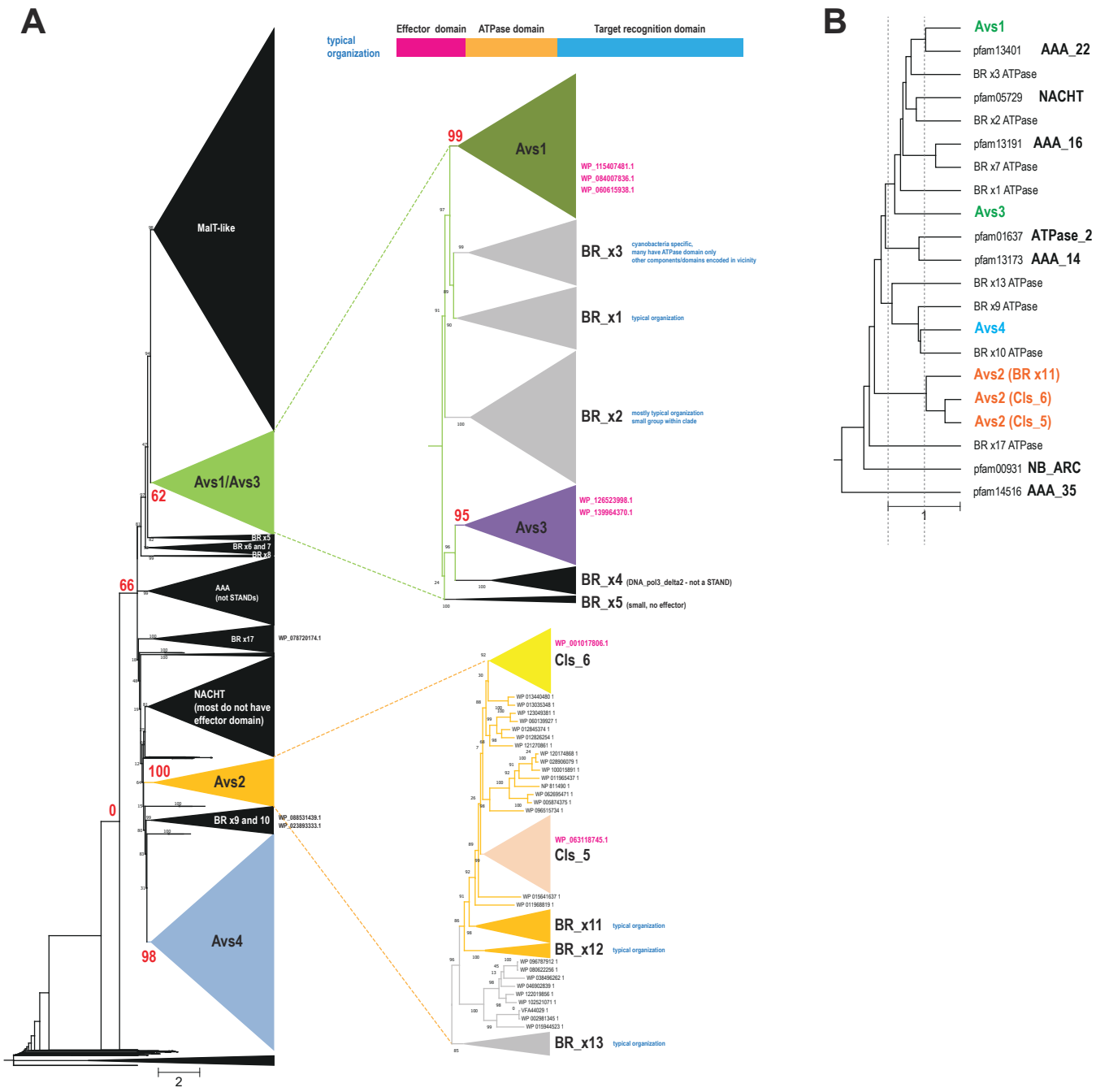


Figure S3: Four distinct clades of Avs proteins. (**A**) Phylogenetic tree of the ATPase domain of selected Avs proteins and other related ATPases identified by PSIBLAST. (**B**) UPGMA dendrogram including the ATPase domains in (A) and profiles of additional ATPases in pfam.

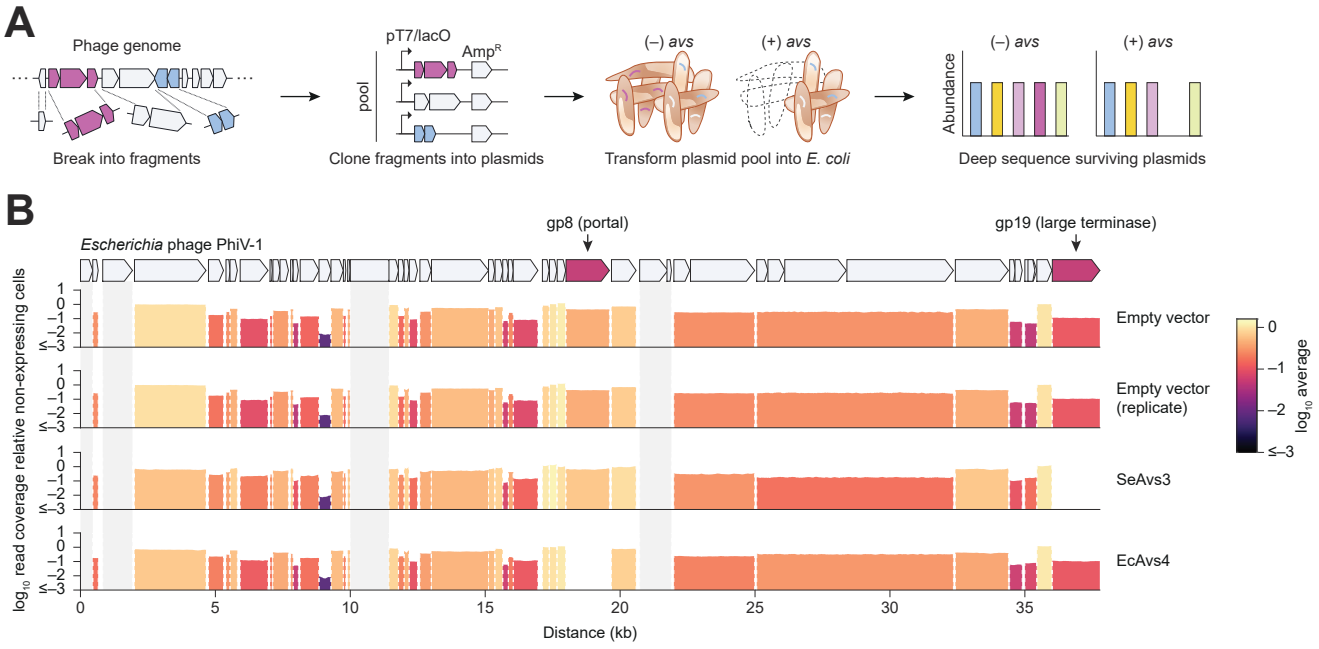


Figure S4: Related to Fig. 1. (**A**) Schematic of PhiV-1 fragment screen (Fig. 1C). (**B**) Read coverage of PhiV-1 fragment screen, without normalizing to the empty vector control.

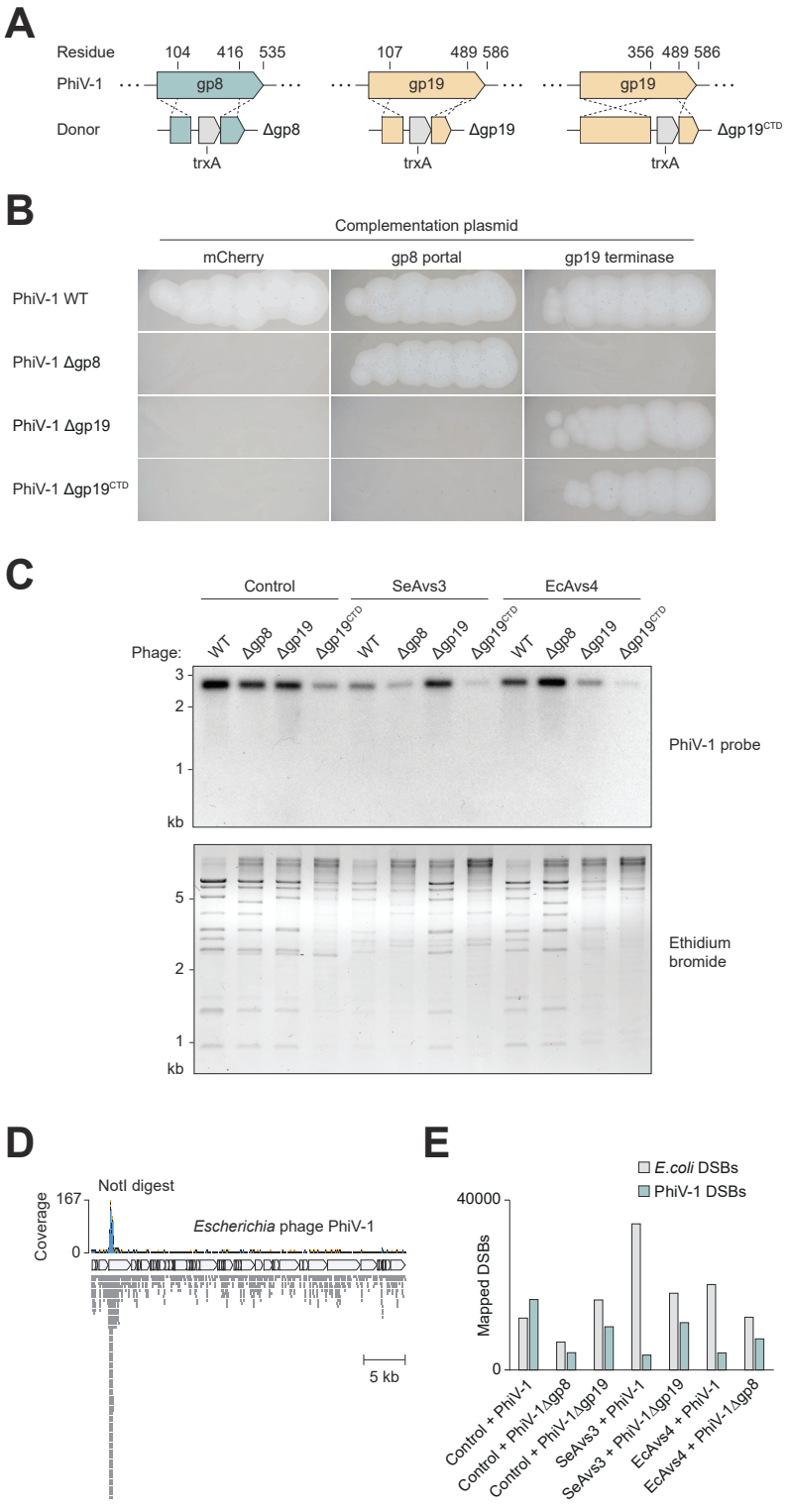


Figure S5: Related to Fig. 1. (**A**) Schematic for PhiV-1 mutant construction via plasmid homology donors in *E. coli*. A trans complementation plasmid encoding gp8 or gp19 was maintained in the cells to support phage growth. (**B**) Plaque assay validation of PhiV-1 knockout phages across different complementation plasmids. Spots correspond to 10-fold phage dilutions from right to left. (**C**) Southern blot analysis of phage-infected *E. coli* cell lysates using a PhiV-1 specific probe. WT, Δgp8, and Δgp19 were used at an MOI of 1, and Δgp19^{CTD} was used at an MOI of 0.25. END-seq analysis (*74*) of DNA double-strand breaks (DSBs) in (**D**) NotI-digested PhiV-1 DNA as a positive control and (**E**) *E. coli* cultures infected with wild type or mutant PhiV-1. Input DNA was normalized across samples.

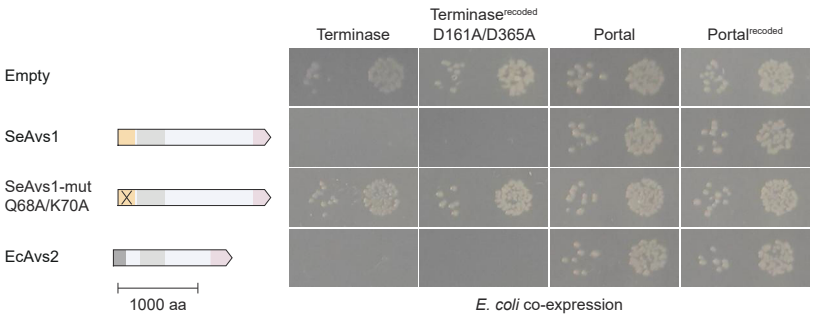


Figure S6: Related to Fig. 1. Photographs of *E. coli* co-transformation assays with Avs1-2 and activators from phage PhiV-1. The left spot on each image corresponds to a 10-fold dilution of the right spot.

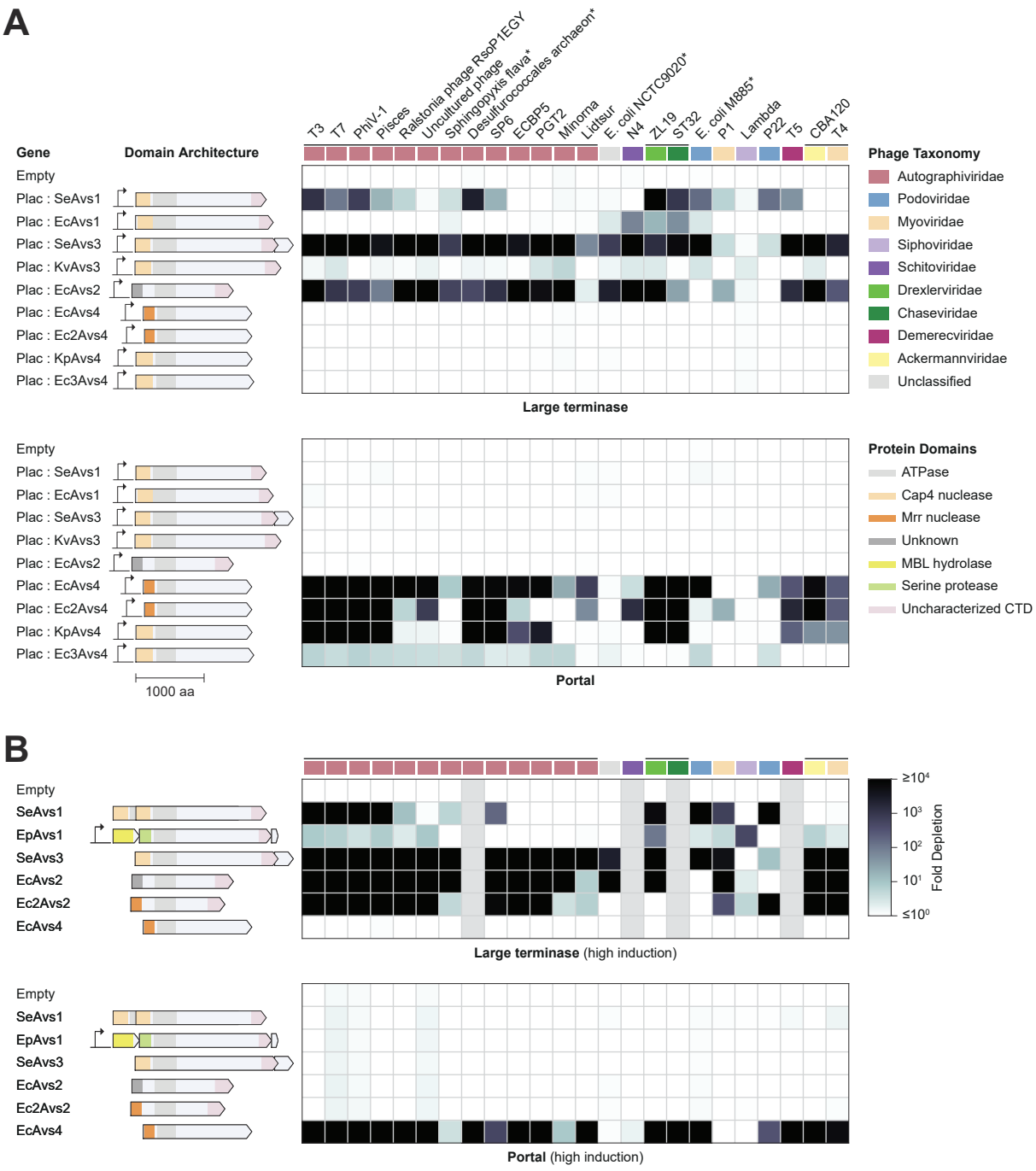


Figure S7: Robustness of the terminase and portal plasmid depletion assay in Fig. 2. (**A**) Specificity of Avs target recognition with *avs* genes expressed under the control of a *lac* promoter and weak induction of terminases and portals (0.002% arabinose). (**B**) Specificity of Avs target recognition with native *avs* promoters and strong induction of terminases and portals (0.2% arabinose). Terminases and portals were expressed under the control of a pBAD promoter. Gray boxes indicate pairwise combinations not assessed due to the toxicity of terminase overexpression.

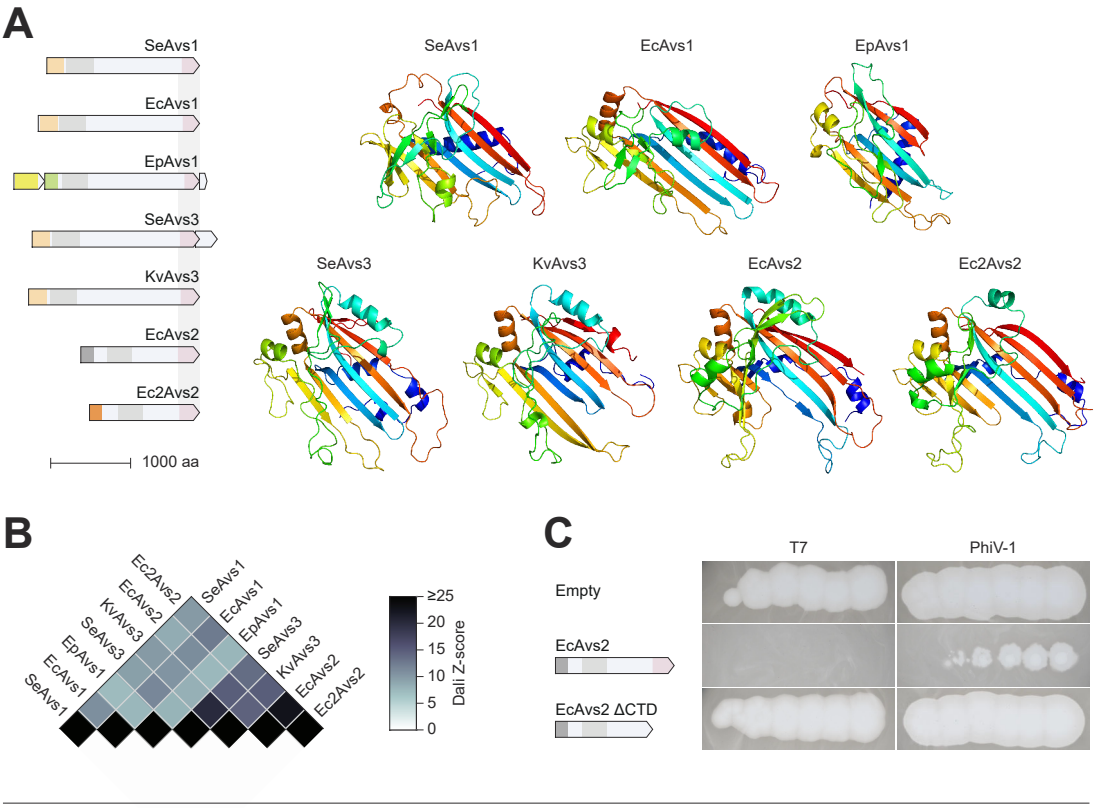


Figure S8: Avs1, Avs2, and Avs3 contain a structurally conserved C-terminal domain essential for defense activity. (**A**) Structures predicted by AlphaFold2 of the C-terminal domains (CTDs) of the seven Avs1-3 homologs investigated in this study. The N- and C-termini of the CTDs are colored from blue to red, respectively. (**B**) Heatmap of Dali Z-scores (*75*) of pairwise comparisons between the Avs1-3 CTDs in (A) (smallest Z-score = 7). Z-scores above 2 indicate significant structural similarity. (**C**) Effect of CTD deletion on EcAvs2 defense activity against T7 and PhiV-1. Spots correspond to 10-fold dilutions from right to left.

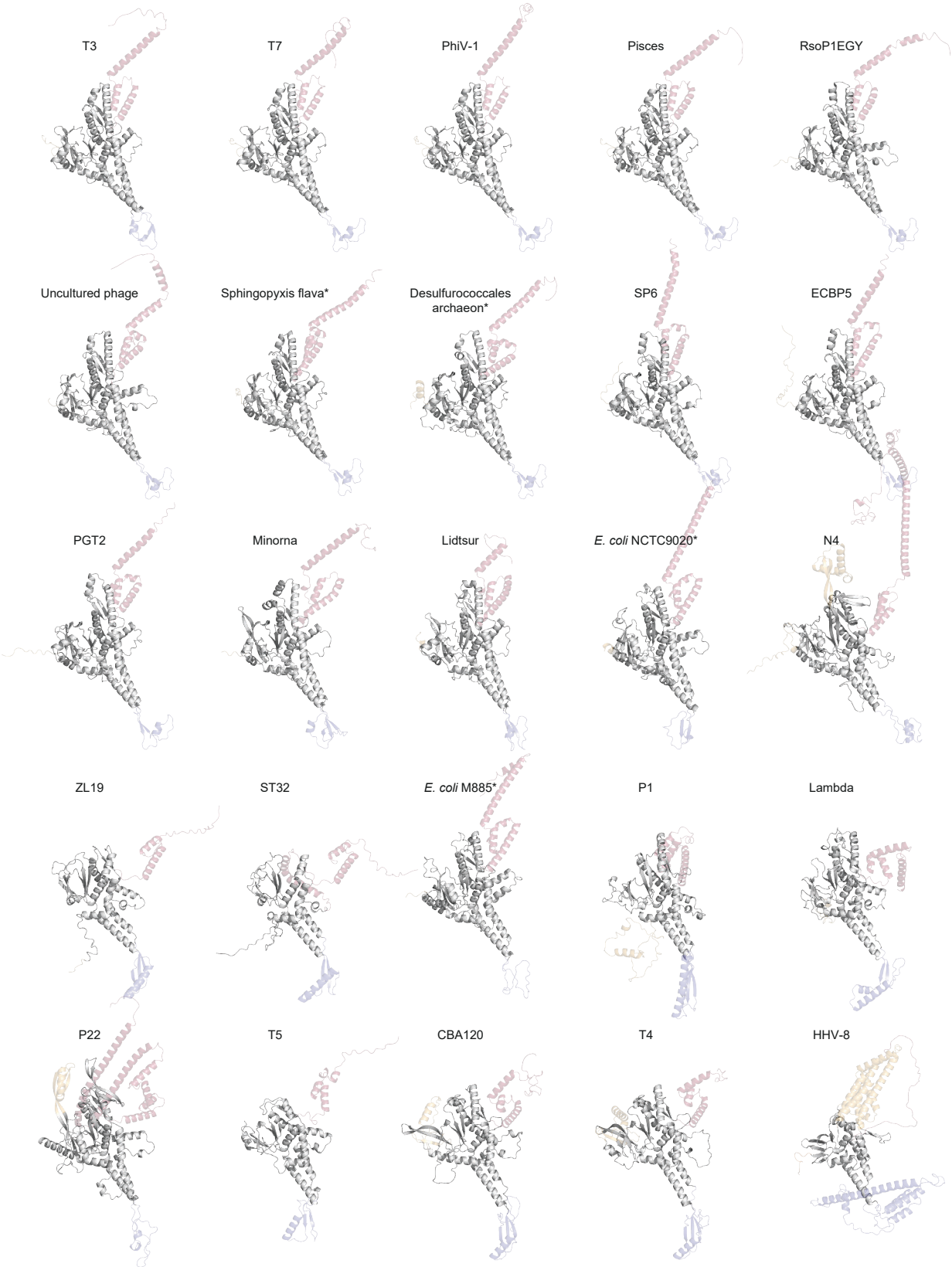


Figure S9: Structures of portal proteins predicted by AlphaFold2. The core portal fold is shown in gray. The clip, crown, and other insertions are colored blue, red, and orange, respectively. Asterisks indicate prophages.

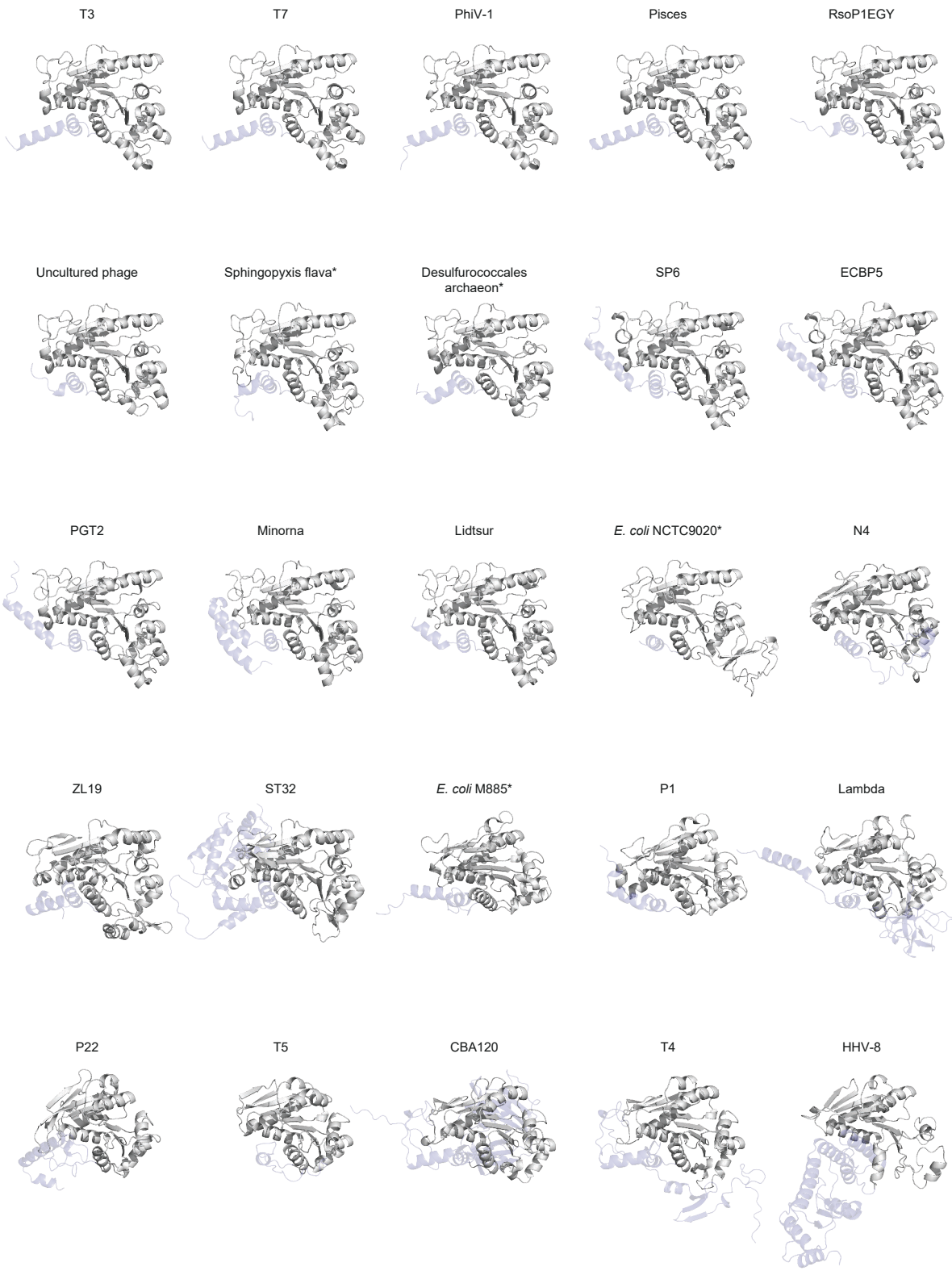


Figure S10: Structures of the N-terminal ATPase domains of large terminases predicted by AlphaFold2. The core ATPase fold is shown in gray.

Figure S11: Structures of the C-terminal nuclease domains of large terminases predicted by AlphaFold2. The core nuclease fold is shown in gray.

Figure S12: Requirements for Avs3 and Avs4 defense activity. Effects of (**A**) Avs3 small ORF deletion and (**B**) Avs3-4 nuclease and ATPase Walker A/B mutations on activity against T7 and PhiV-1.

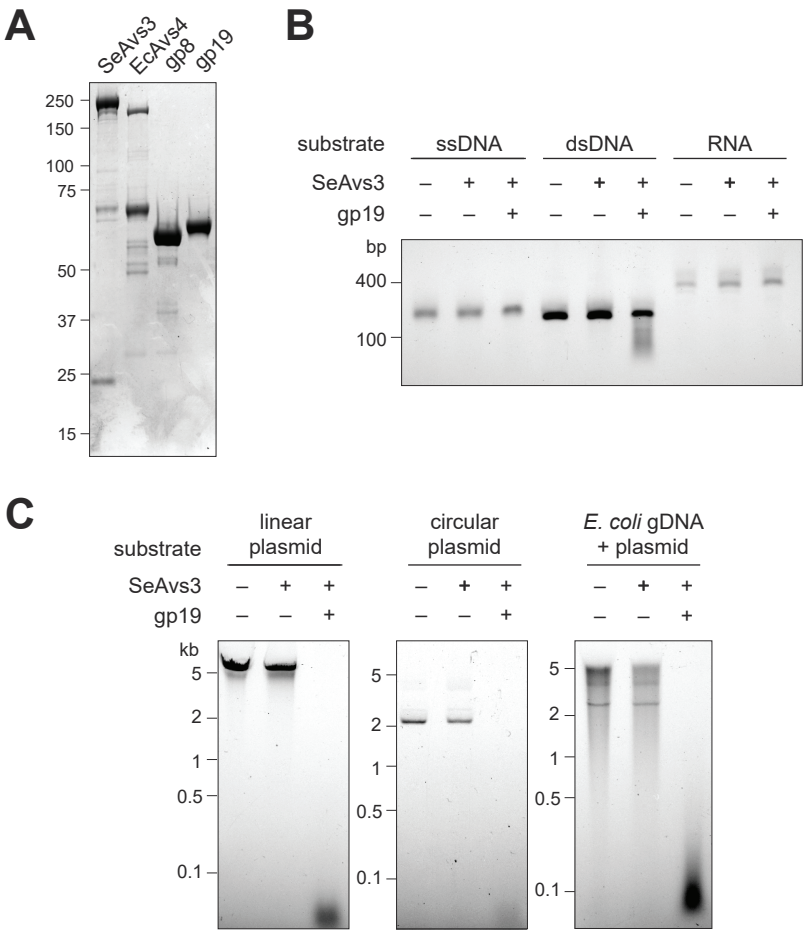


Figure S13: Related to Fig. 3. In vitro reconstitution of Avs activity. (**A**) Coomassie-stained SDS-PAGE gel of purified Avs proteins and phage triggers. (**B**, **C**) Agarose gel analysis of SeAvs3 nucleic acid substrate specificity.

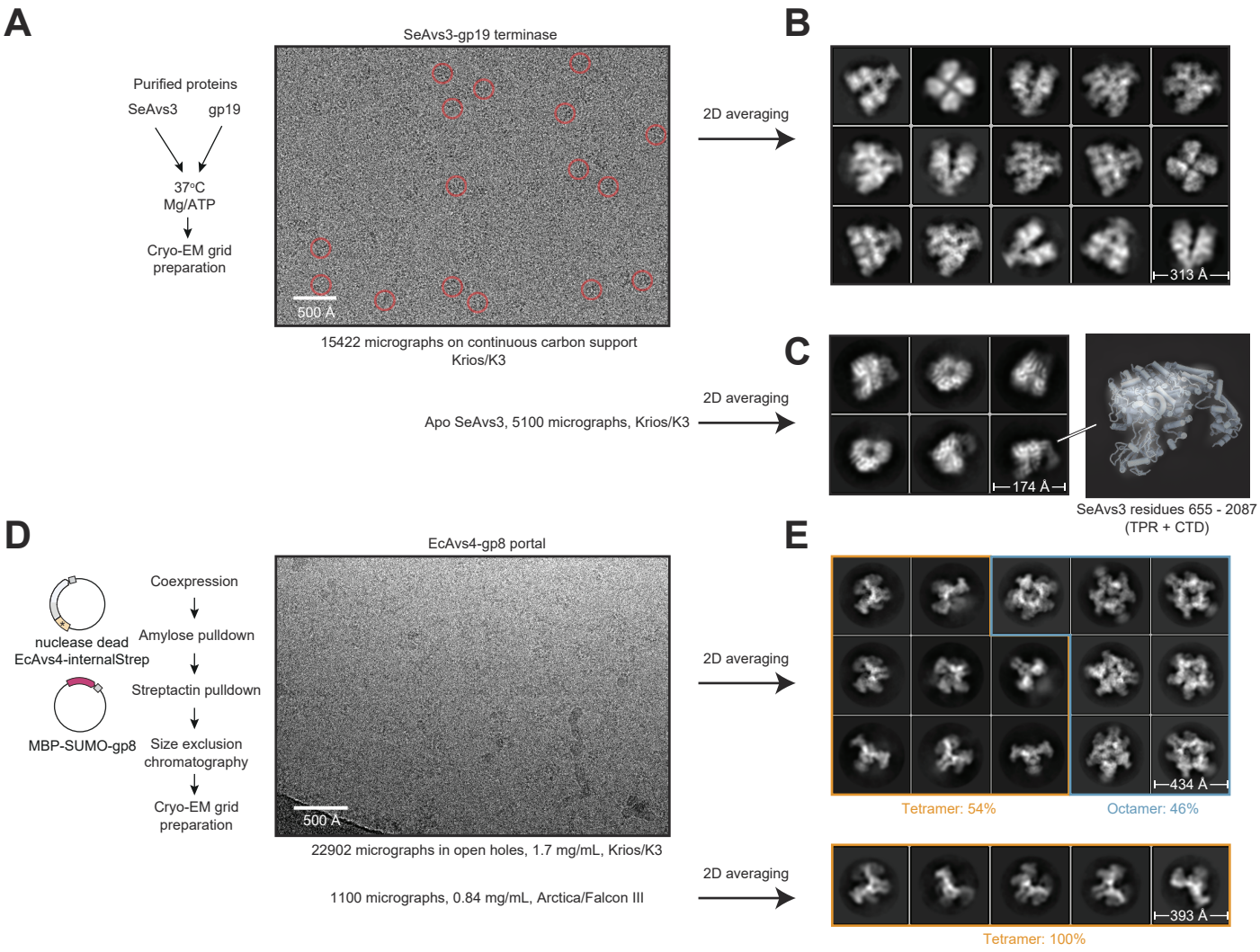


Figure S14: Imaging Avs proteins by electron microscopy. (**A**) Example cryo-EM micrograph of assembled SeAvs3-gp19 complex. (**B**) Representative 2D class averages of SeAvs3-gp19 from 128,500 automatically picked particles. (**C**) 2D averages from cryo-EM imaging of SeAvs3 alone. One class is shown magnified with the structure of SeAvs3 residues 655 – 2087 superimposed, based on the structure of the SeAvs3-gp19 complex. This dataset did not allow high resolution structure determination, potentially due to inherent flexibility in apo-SeAvs3. (**D**) Example cryo-EM micrograph of purified EcAvs4-gp8 complex. (**E**) Representative 2D class averages of the tetrameric and octameric species of EcAvs4-gp8 from 444,626 automatically picked particles. Also shown are 2D averages from a small screening cryo-EM dataset from the same sample diluted 2-fold, showing only the tetrameric species. (**F**) Avs samples imaged by negative-stain electron microscopy using an FEI Tecnai 12 microscope operated at 120 keV. Samples were applied to continuous carbon and stained using 2% uranyl formate. Avs3 and Avs4 do not assemble into tetramers in the absence of their cognate ligands.

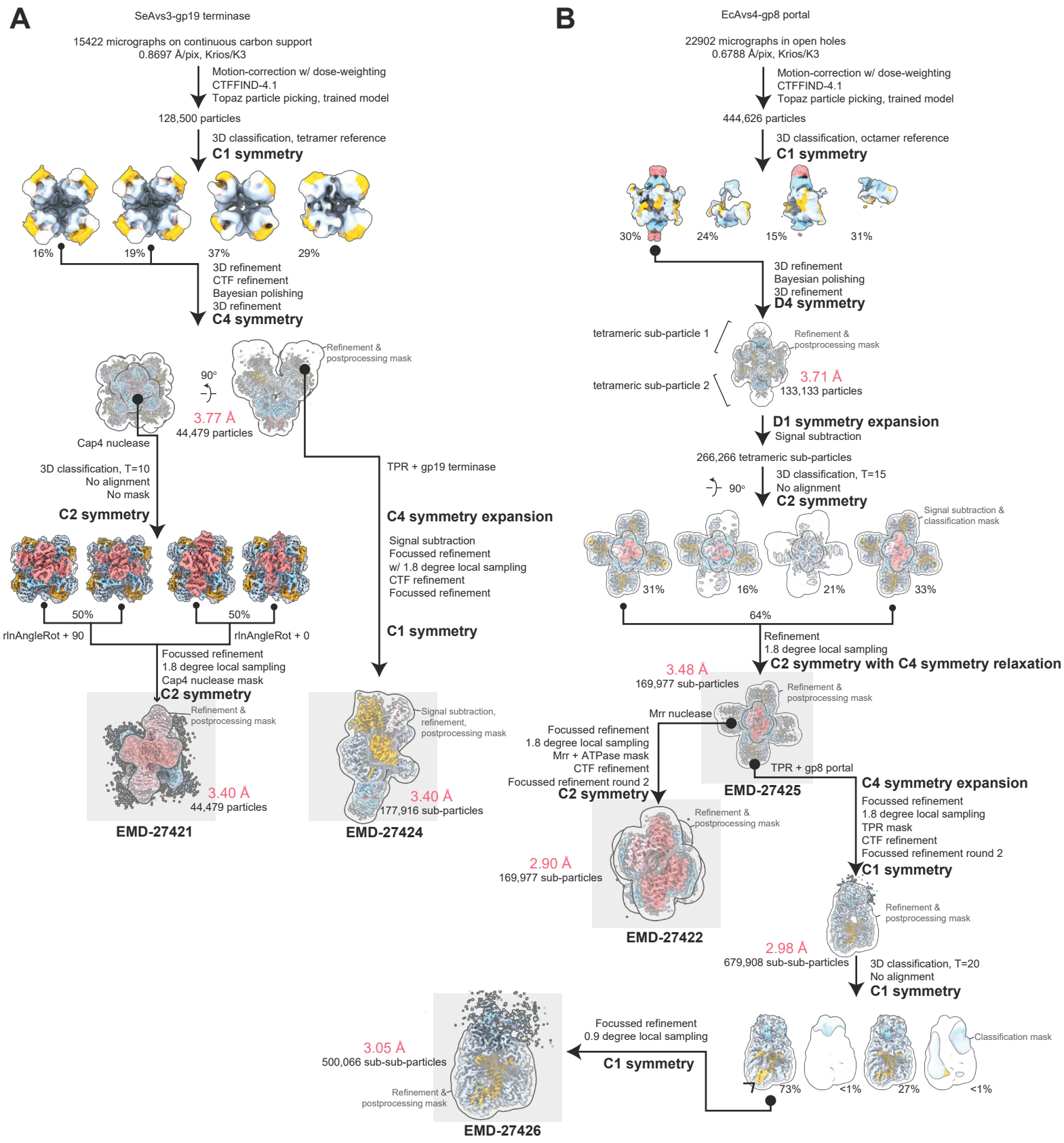


Figure S15: Cryo-EM data processing scheme. Flowchart outlining the data processing for (**A**) the SeAvs3-gp19 complex and (**B**) the EcAvs4-gp8 complex. Final maps deposited to the EMDB are highlighted.

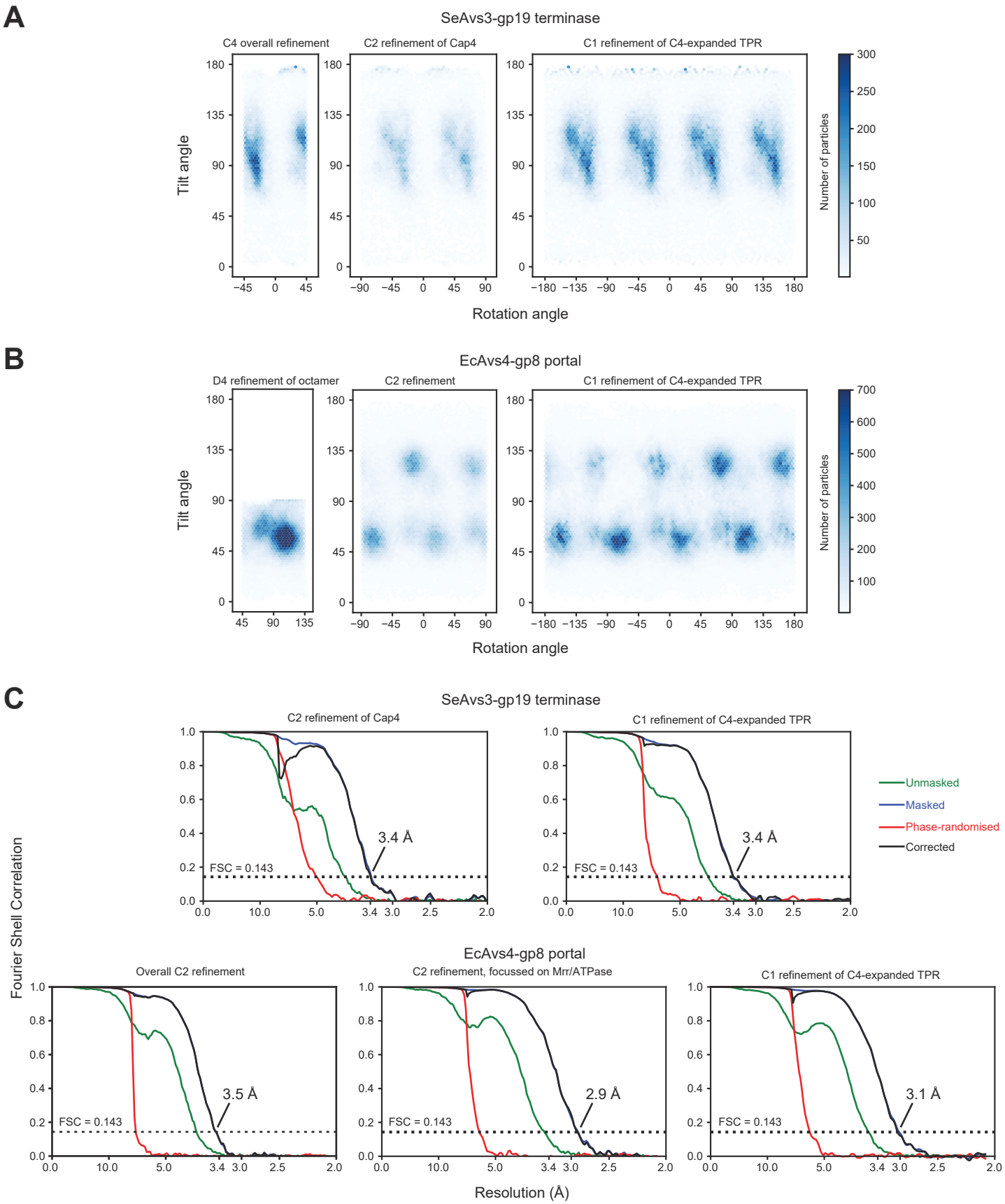


Figure S16: Cryo-EM data statistics. (**A**,**B**) Orientation distributions for reconstructions of the SeAvs3-gp19 terminase complex and EcAvs4-gp8 complex. The range of the x-axis, corresponding to the RELION metadata parameter 'rlnAngleRot,' is set according to the symmetry of the reconstruction. (**C**) Gold-standard Fourier-Shell Correlation curves.

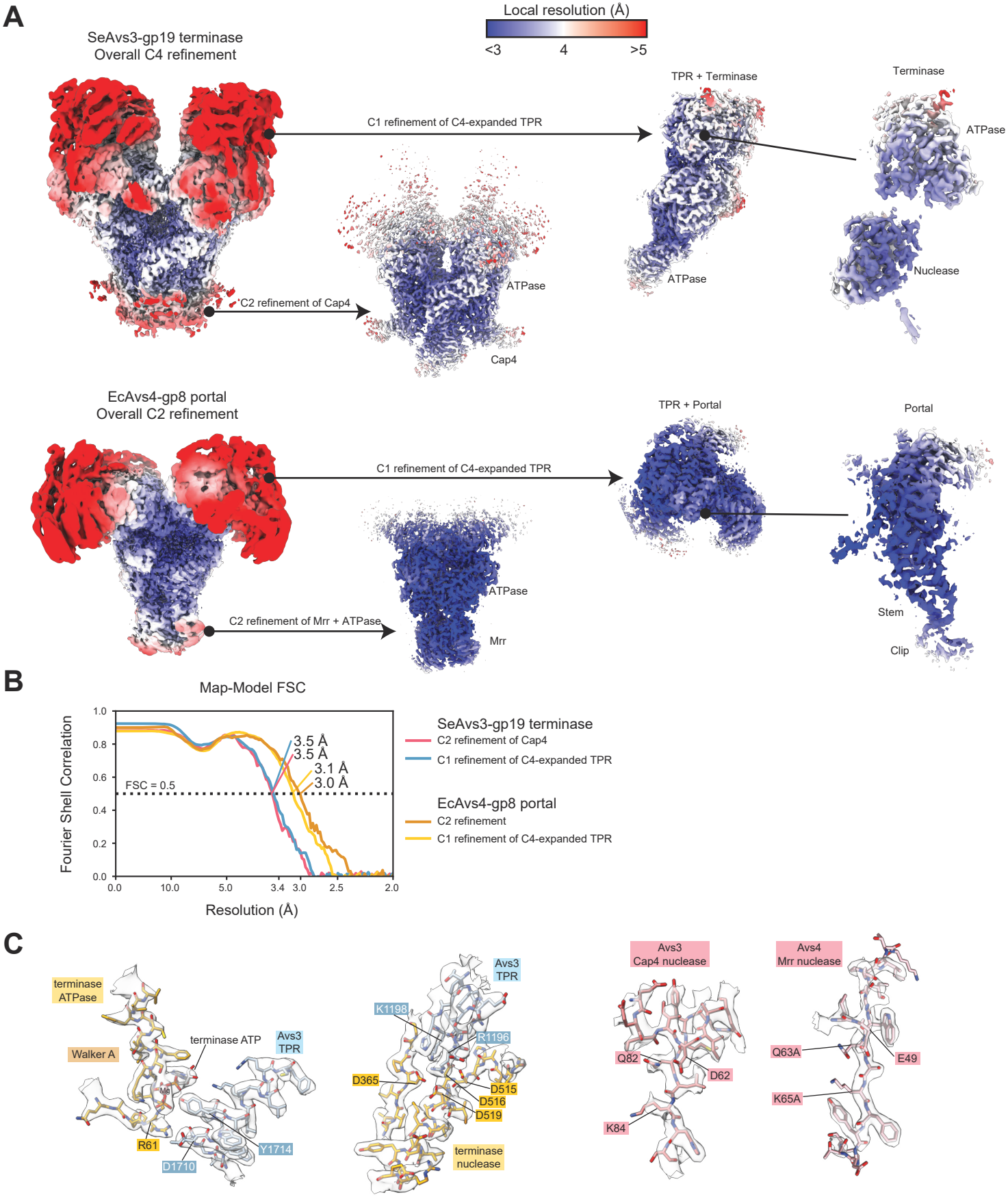


Figure S17: Cryo-EM map quality and map-to-model fitting. (**A**) Cryo-EM densities colored by local resolution as calculated within RELION. The overall maps are shown filtered by local resolution, while the focus-refined maps are shown auto-sharpened. Sharpened maps are also shown just around the phage ligands. (**B**) Map-to-model Fourier-Shell Correlation as calculated in PHENIX, softly masking each map around the fitted model. (**C**) Example cryo-EM densities for different parts of the structures.

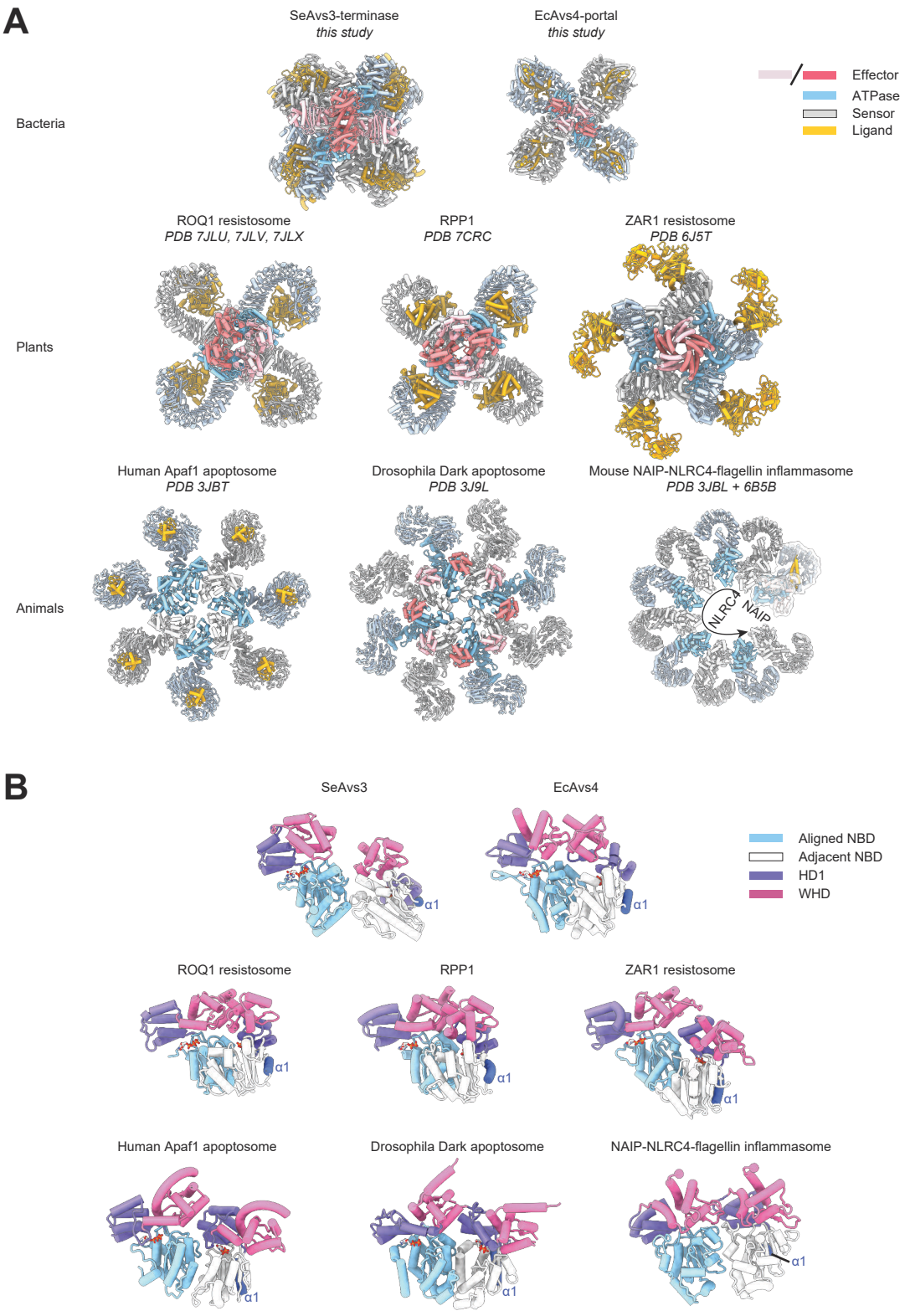


Figure S18: Comparison of activated STAND structures. (**A**) STAND oligomers from different domains of life (*29*, *30*, *47*, *76*–*79*), colored by function. The ROQ1 resistosome structure is a composite by imposing C4 symmetry on PDB 7JLU and merging it with PDB 7JLV and 7JLX (*29*). The NAIP inflammasome structure is an alignment of the C11 symmetric NLRC4 oligomer (with four subunits hidden) (*78*) with the NAIP-NLRC4-flagellin filament structure (*79*). (**B**) Two adjacent STAND ATPase domains from these structures, aligned on the nucleotide-binding domain of one ATPase (blue), showing different relative positions of the adjacent ATPase. NBD; nucleotide-binding domain. HD1; helical domain 1. WHD; winged-helix domain.

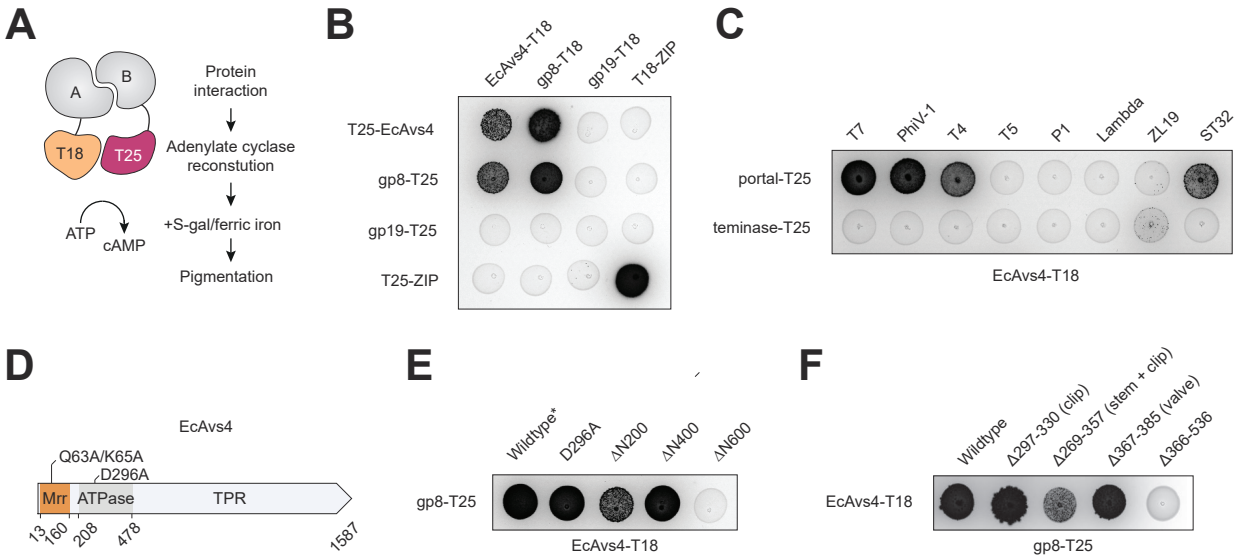


Figure S19: Bacterial two-hybrid analysis of EcAvs4-portal interactions. (**A**) Schematic of a bacterial two-hybrid system for detecting protein-protein interactions. (**B**) Two-hybrid analysis of pairwise interactions between EcAvs4 and PhiV-1 proteins grown on S-gal indicator plates. (**C**) Interactions between EcAvs4 and the portal and terminase genes from eight phages. (**D**) Schematic of EcAvs4 protein domains. (**E**) Two-hybrid analysis of EcAvs4 mutations and truncations. (**F**) Two-hybrid analysis of PhiV-1 portal deletions.

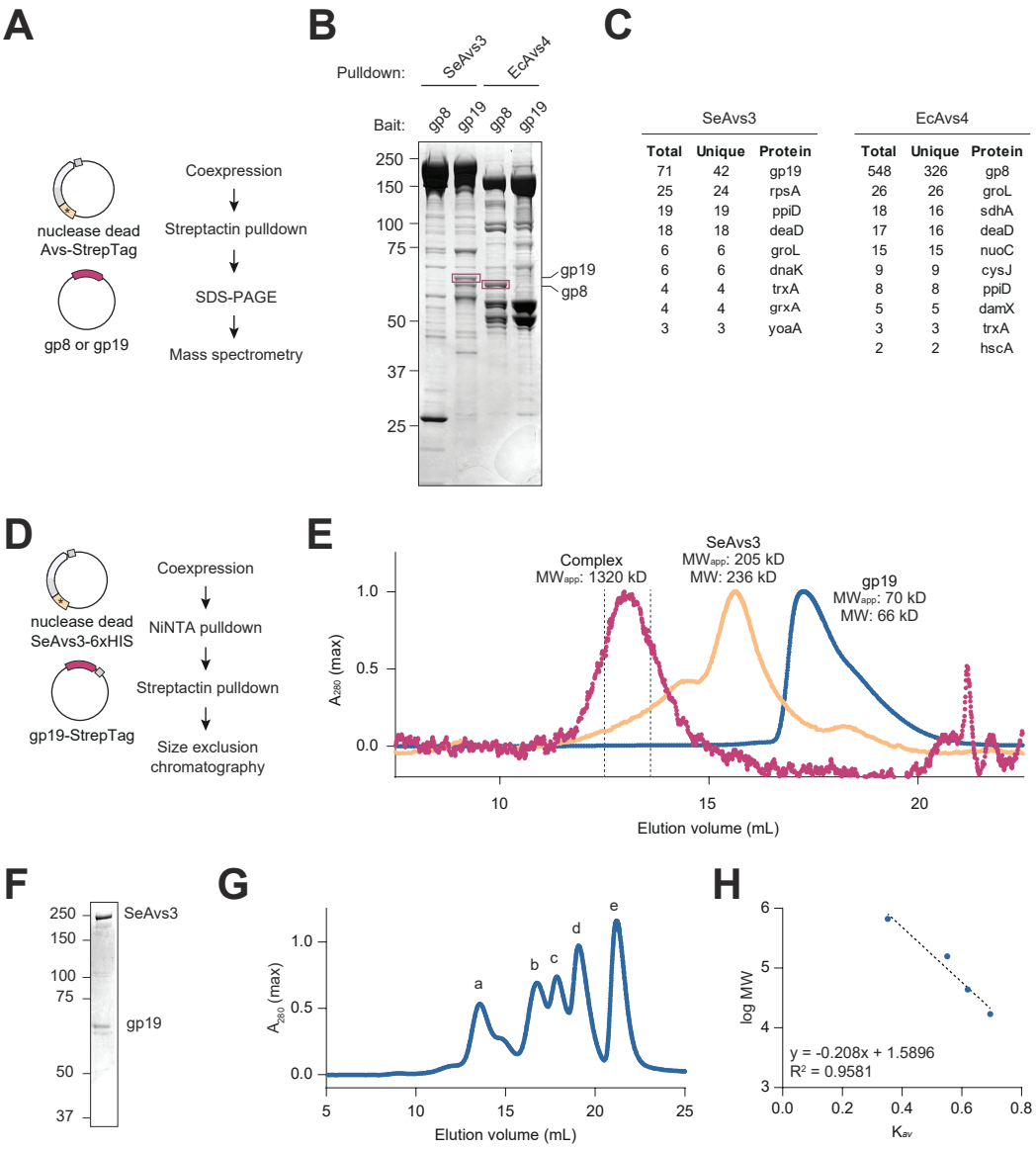


Figure S20: Analysis of Avs proteins co-purified with their cognate triggers. (**A**) Co-purification strategy. (**B**) SDS-PAGE analysis of SeAvs3 and EcAvs4 purified in the presence of gp19 terminase or gp8 portal. Highlighted bands were excised and analyzed by mass spectrometry. (**C**) Total and unique mapped peptides from mass spectrometry analysis of gp19 and gp8 gel bands. (**D**) Schematic of tandem affinity purification of the SeAvs3-terminase complex. (**E**) Size exclusion chromatography of SeAvs3, PhiV-1 terminase, and the SeAvs3-terminase complex. (**F**) Coomassie-stained SDS-PAGE protein gel of the SeAvs3-terminase complex. (**G**) Size exclusion chromatography of protein standards (a: thyroglubulin, 670 kDa, b: γ-globulin, 158 kDa, c: ovalbumin, 44 Kda, d: myoglobin, 17 kDa, e: vitamin B12, 1.35 kDa). (**H**) Calibration curve of the Superose 6 Increase column.

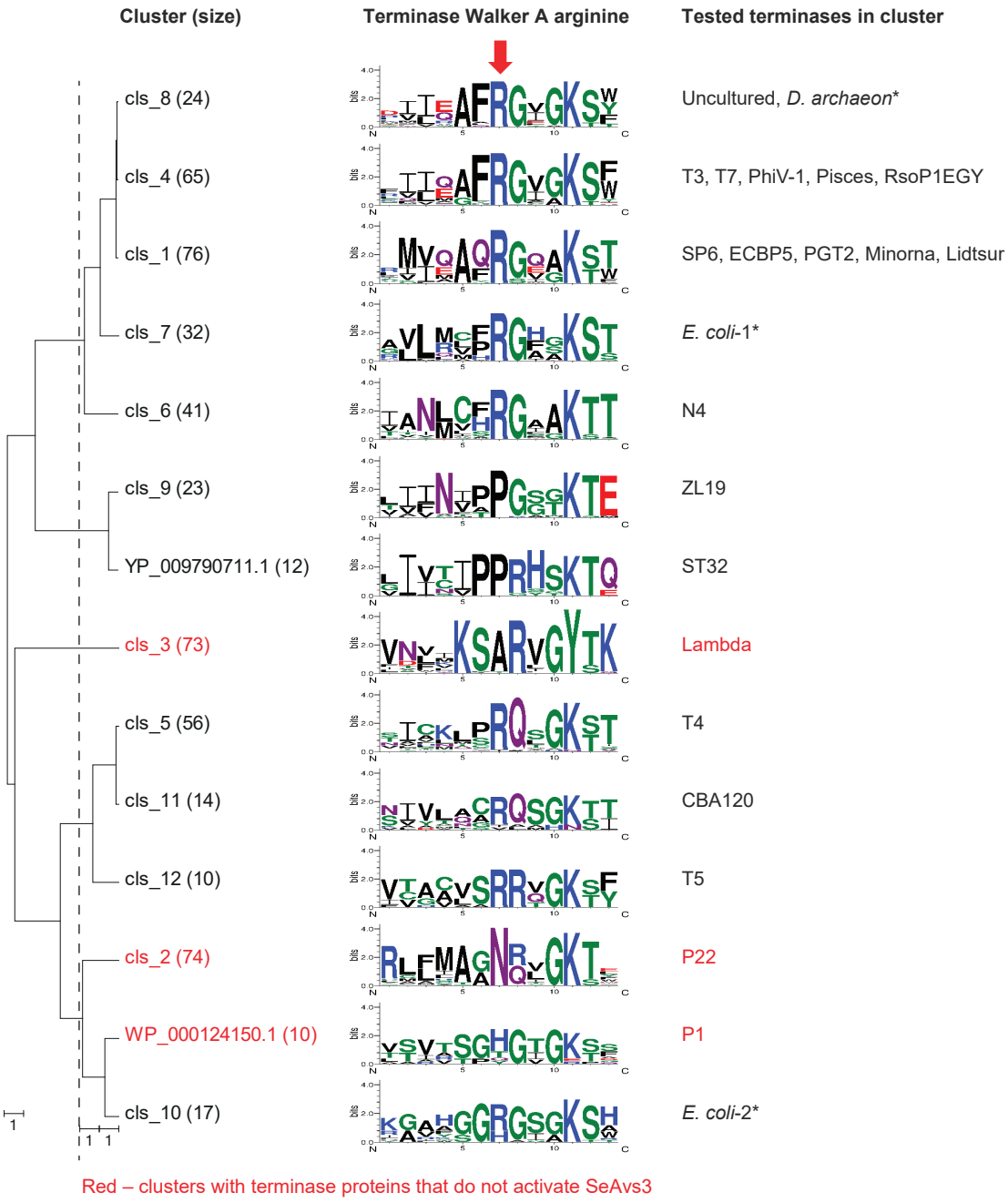


Figure S21: Related to Fig. 5. Weblogos of the Walker A motifs of phage terminases. Each motif represents a cluster of terminases that contain at least one representative that was tested experimentally in this study. Terminase sequences (*80*) were supplemented with the 24 terminases in this study and clustered at 30% sequence identity. Clusters containing terminases that do not activate SeAvs3 are shown in red. The UPGMA tree was built using a procedure described previously (*54*).

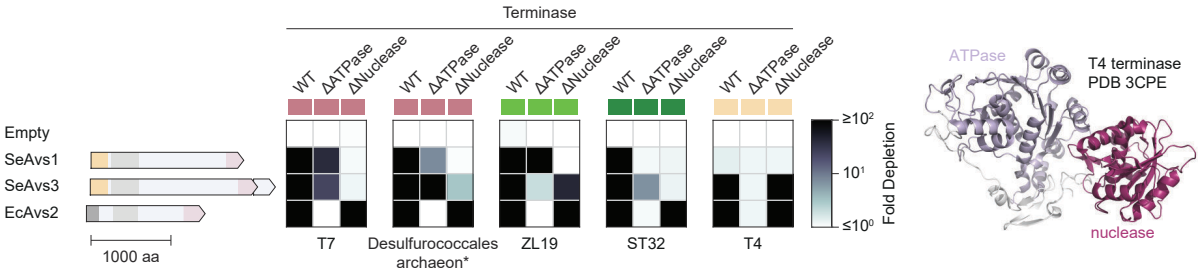


Figure S22: Effect of terminase domain deletions on the activation of Avs1, Avs2, and Avs3. The structure of the T4 terminase (*81*) is shown as an example.

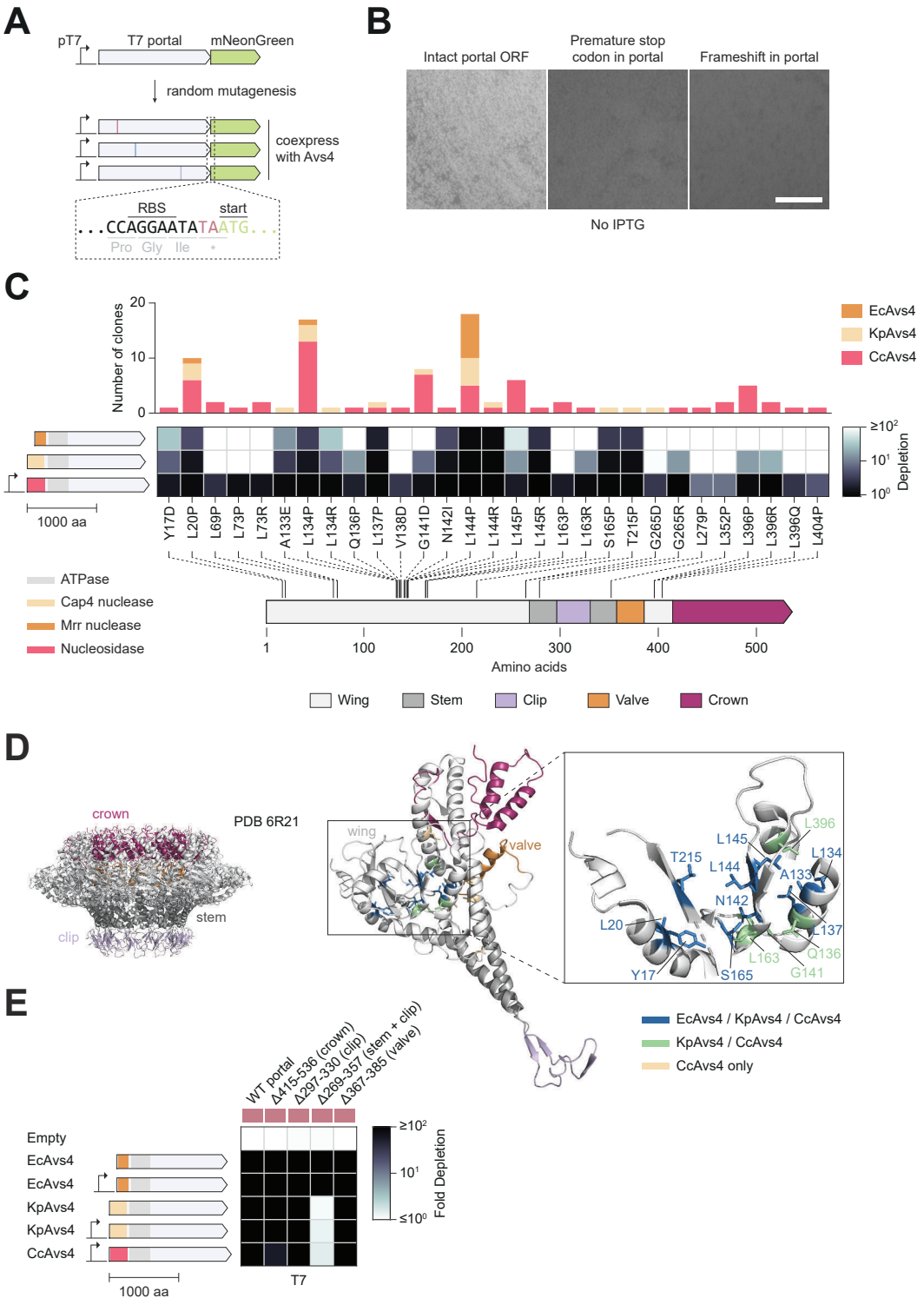


Figure S23: Identification of single amino acid substitutions in the T7 portal protein that abrogate Avs4 activation. (**A**) Design of a translation-reinitiation reporter system used to facilitate screening of Avs4 mutants. (**B**) Validation of reporter performance via mNeonGreen fluorescence from *E. coli* colonies. Scale bar: 1 cm. (**C**) Activity and location of the 29 identified portal mutants that abrogate Avs4 activation. (**D**) Locations of mutations in the T7 portal structure (PDB 6R21) (*82*). (**E**) Effect of T7 portal deletions on the activation of Avs4 as assessed by plasmid depletion. Arrows represent *lac* promoters.

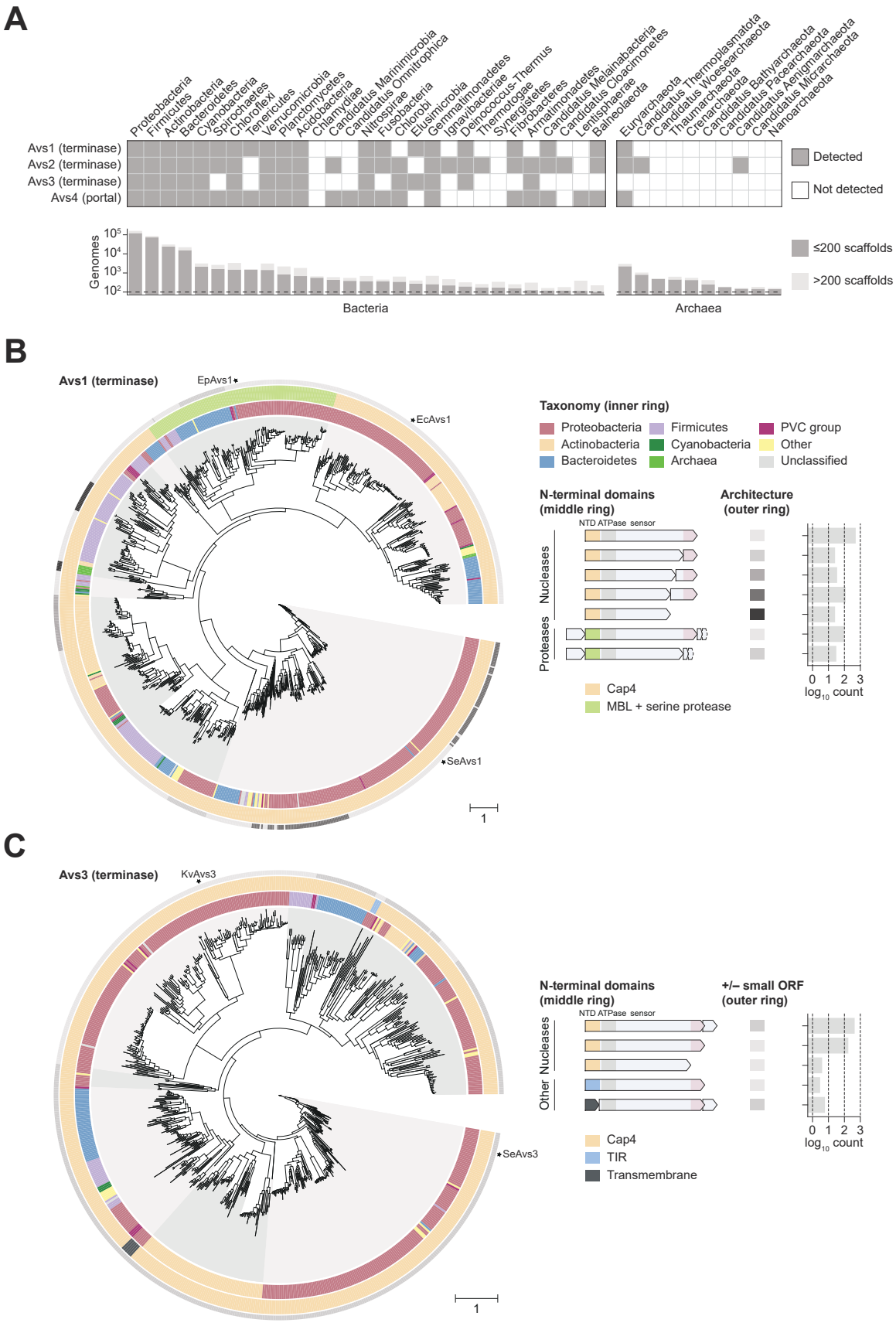


Figure S24: Related to Fig. 6. (**A**) Taxonomic distribution of Avs families, stratified by bacterial and archaeal phylum. The bar graphs show the number of genomes available for analysis. Maximum likelihood phylogenetic trees of representatives of the (**B**) Avs1 (*n* = 843) and (**C**) Avs3 (*n* = 630) families clustered at 95% sequence identity. Inner, middle, and outer rings indicate taxonomy, N-terminal effector domains, and locus architecture, respectively.

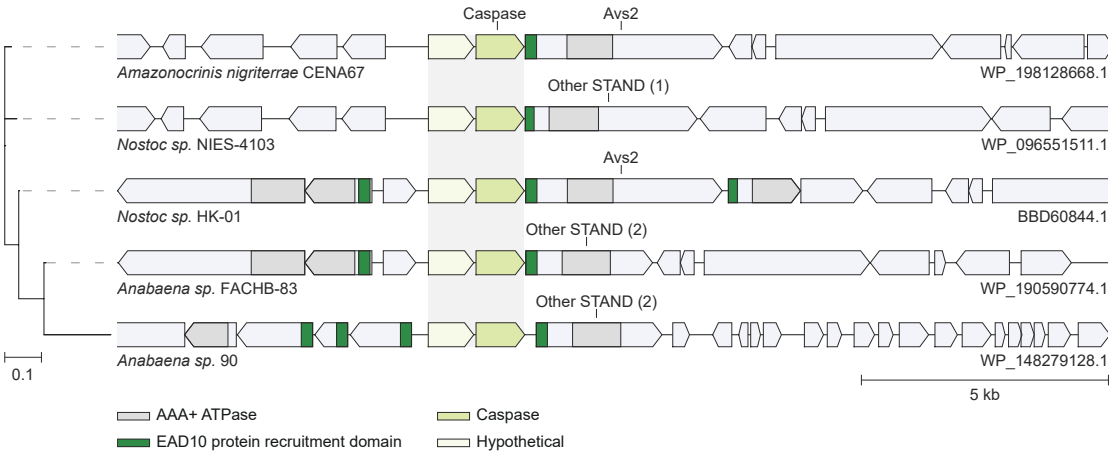


Figure S25: Examples of Avs proteins implicated in protein-protein signaling. Predicted caspase recruitment by cyanobacterial Avs2 homologs via an N-terminal EAD10 protein recruitment domain that is also shared by proteins encoded in the vicinity. The tree was constructed from a multiple sequence alignment of the caspase. Protein accession numbers refer to the STAND NTPase.

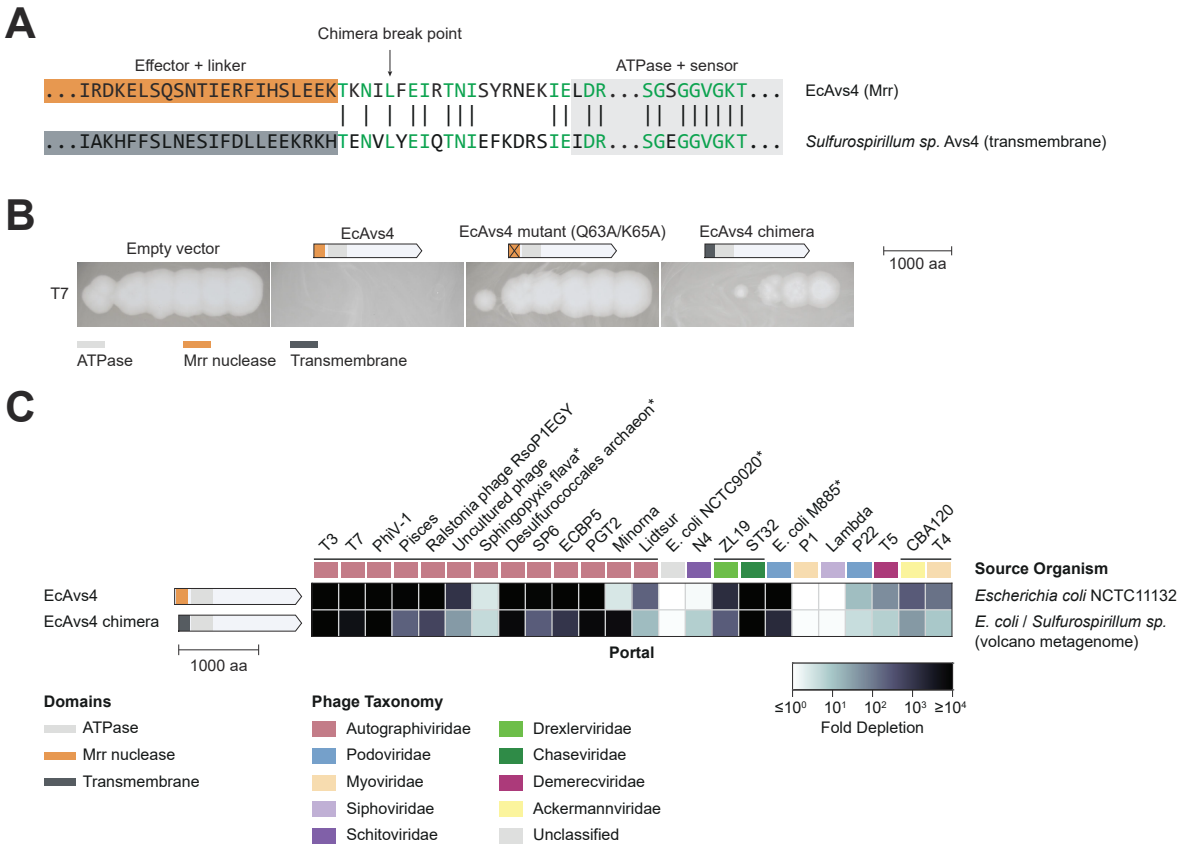


Figure S26: Related to Fig. 6. (**A**) Amino acid sequence surrounding the EcAvs4 chimera break point. (**B**) Chimera activity against phage T7. (**C**) Plasmid depletion assay for the target recognition specificity of the EcAvs4 chimera in comparison with EcAvs4.

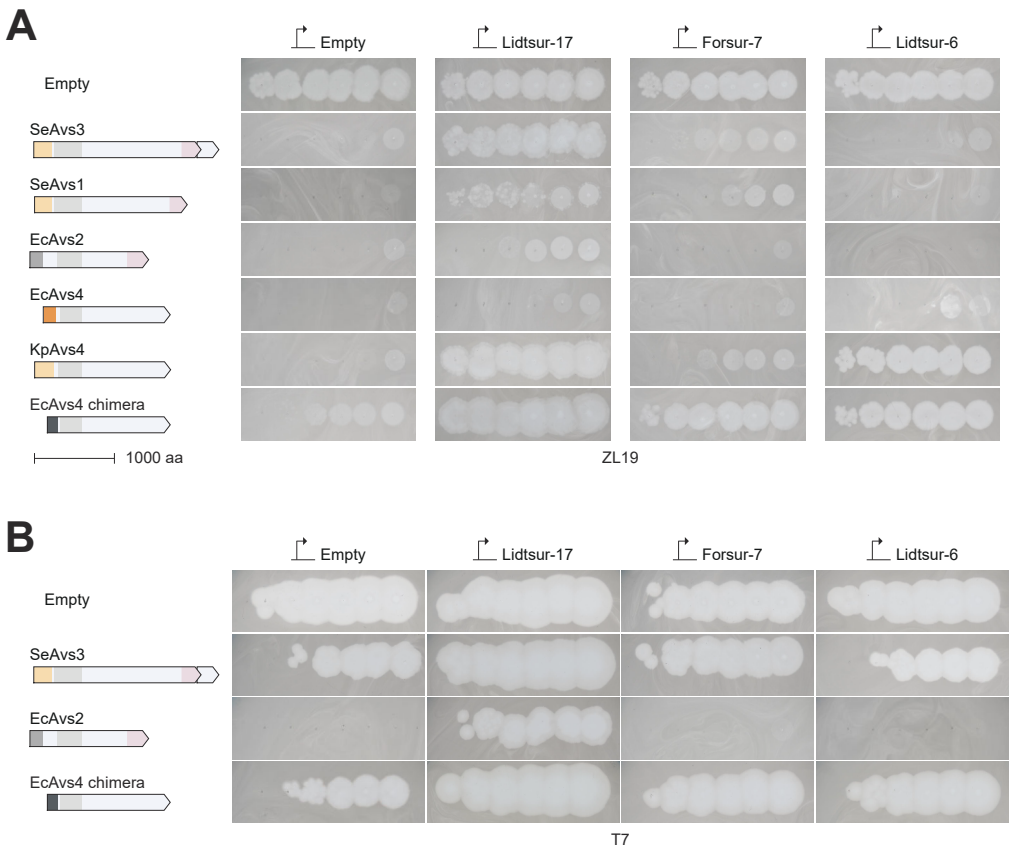


Figure S27: Related to Fig. 7. Antidefense genes inhibit Avs activity in bacterial cells. Plaque assays against (**A**) phage ZL19 and (**B**) phage T7 with *E. coli* strain C containing both an Avs plasmid and an antidefense plasmid. Antidefense genes were expressed under the control of a J23105 promoter. Spots correspond to 10-fold dilutions from right to left.

Table S1: Cryo-EM data collection, refinement, and validation statistics

	SeAvs3-gp19 PDB ID: 8DGC C2 refinement of Cap4 nuclease domain EMDB ID: EMD-27421	 Symmetry-expanded C1 refinement of TPR-terminase domain EMDB ID: EMD-27424
Data collection and Processing		
Microscope	Thermo Scientific Titan Krios G3i cryo TEM	
Voltage (keV)	300	
Camera	Gatan K3	
Magnification	105,000	
Pixel size at detector (Å/pixel)	0.8697	
Total electron exposure (e^{-}/Å^2)	30	
Exposure rate (e-/pixel/sec)	25	
Number of frames collected during exposure	30	
Defocus range (μm)	-1 to -3.5	
Automation software	EPU	
Energy filter slit width (if used)	20 eV	
Micrographs collected (no.)	15,422	
Total extracted particles (no.)	128,500	
Refined particles (no.)	44,479	177,916 subparticles
Symmetry imposed	C2	C1
Estimated angular accuracy	0.74	1.305
Estimated translation accuracy (Å)	0.56	0.60
Resolution (global, Å)		
FSC 0.5 (unmasked/masked)	4.82 / 3.82	4.74 / 3.87
FSC 0.143 (unmasked/masked)	4.01 / 3.40	4.01 / 3.40
Resolution range (local, Å)	3.40 – 4.13	3.34 – 4.34
Map sharpening *B* factor (Å^2)	-93	-103
Model composition		
Protein residues	10,216	
Ligands	16	
Model Refinement		
Refinement package	phenix.real_space_refine	
- resolution cutoff	3.40	3.40
Model-Map scores		
-CC	0.82	0.84
-FSC 0.5 (Å)	3.50	3.50
B factors (Å^2)		
Protein residues	87	
Ligands	77	
R.m.s. deviations from ideal values		
Bond lengths (Å)	0.006	
Bond angles (°)	0.956	
Validation		
MolProbity score	0.89	
CaBLAM outliers (%)	1.13	
Clashscore	1.47	
Poor rotamers (%)	0.14	
C-beta deviations (%)	0.02	
EMRinger score	2.48	
Ramachandran plot		
Favored (%)	98.12	
Outliers (%)	0	

	EcAvs4-gp8 PDB ID: 8DGF Overall C2 reconstruction EMDB ID: EMD-27425	C2 refinement of Mrr nuclease domain EMDB ID: EMD-27422	Symmetry-expanded C1 refinement of TPR-portal domain EMDB ID: EMD-27426
Data collection and Processing			
Microscope	Thermo Scientific Titan Krios G3i cryo		
Voltage (keV)	TEM		
Camera	300		
Magnification	Gatan K3		
Pixel size at detector (Å/pixel)	130,000		
Total electron exposure (e^{-}/Å^2)	0.6788		
Exposure rate (e-/pixel/sec)	30.8		
Number of frames collected during exposure	25		
Defocus range (μm)	24		
Automation software	-1 to -2.5		
Energy filter slit width (if used)	EPU		
Micrographs collected (no.)	20 eV		
Total extracted particles (no.)	22,902 444,626		
Refined particles (no.)	169,977	169,977	679,908 subparticles
Symmetry imposed	C2	C2	C1
Estimated angular accuracy	0.488	0.770	0.448
Estimated translation accuracy (Å)	0.363	0.409	0.500
Resolution (global, Å)			
FSC 0.5 (unmasked/masked)	4.54 / 3.92	4.14 / 3.27	4.18 / 3.42
FSC 0.143 (unmasked/masked)	3.96 / 3.48	3.51 / 2.90	3.62 / 3.05
Resolution range (local, Å)	3.30 – 6.20	2.85 – 3.73	3.00 – 3.90
Map sharpening *B* factor (Å^2)	-112	-80	-86
Model composition			
Protein residues	7,902		
Ligands	8		
Model Refinement			
Refinement package	phenix.real_space_refine		
- resolution cutoff	3.5	3.0	3.0
Model-Map scores			
-CC	0.80	0.80	0.85
-FSC 0.5 (Å)	3.5	3.0	3.1
B factors (Å^2)			
Protein residues	89		
Ligands	53		
R.m.s. deviations from ideal values			
Bond lengths (Å)	0.010		
Bond angles (°)	1.249		
Validation			
MolProbity score	0.91		
CaBLAM outliers (%)	0.71		
Clashscore	1.36		
Poor rotamers (%)	0.67		
C-beta deviations (%)	0.01		
EMRinger score	3.29		
Ramachandran plot			
Favored (%)	97.79		
Outliers (%)	0		

References and Notes

1. K. S. Makarova, Y. I. Wolf, S. Snir, E. V. Koonin, Defense islands in bacterial and archaeal genomes and prediction of novel defense systems. *J. Bacteriol.* **193**, 6039–6056 (2011). doi:10.1128/JB.05535-11 Medline

2. E. V. Koonin, K. S. Makarova, Y. I. Wolf, Evolutionary genomics of defense systems in archaea and bacteria. *Annu. Rev. Microbiol.* **71**, 233–261 (2017). doi:10.1146/annurev-micro-090816-093830 Medline

3. S. Doron, S. Melamed, G. Ofir, A. Leavitt, A. Lopatina, M. Keren, G. Amitai, R. Sorek, Systematic discovery of antiphage defense systems in the microbial pangenome. *Science* **359**, eaar4120 (2018). doi:10.1126/science.aar4120 Medline

4. L. Gao, H. Altae-Tran, F. Böhning, K. S. Makarova, M. Segel, J. L. Schmid-Burgk, J. Koob, Y. I. Wolf, E. V. Koonin, F. Zhang, Diverse enzymatic activities mediate antiviral immunity in prokaryotes. *Science* **369**, 1077–1084 (2020). doi:10.1126/science.aba0372 Medline

5. H. G. Hampton, B. N. J. Watson, P. C. Fineran, The arms race between bacteria and their phage foes. *Nature* **577**, 327–336 (2020). doi:10.1038/s41586-019-1894-8 Medline

6. D. Klaiman, E. Steinfels-Kohn, G. Kaufmann, A DNA break inducer activates the anticodon nuclease RloC and the adaptive immunity in *Acinetobacter baylyi* ADP1. *Nucleic Acids Res.* **42**, 328–339 (2014). doi:10.1093/nar/gkt851 Medline

7. C. K. Guegler, M. T. Laub, Shutoff of host transcription triggers a toxin-antitoxin system to cleave phage RNA and abort infection. *Mol. Cell* **81**, 2361–2373.e9 (2021). doi:10.1016/j.molcel.2021.03.027 Medline

8. R. Cheng, F. Huang, H. Wu, X. Lu, Y. Yan, B. Yu, X. Wang, B. Zhu, A nucleotide-sensing endonuclease from the Gabija bacterial defense system. *Nucleic Acids Res.* **49**, 5216–5229 (2021). doi:10.1093/nar/gkab277 Medline

9. R. Bingham, S. I. Ekunwe, S. Falk, L. Snyder, C. Kleanthous, The major head protein of bacteriophage T4 binds specifically to elongation factor Tu. *J. Biol. Chem.* **275**, 23219–23226 (2000). doi:10.1074/jbc.M002546200 Medline

10. A. Millman, A. Bernheim, A. Stokar-Avihail, T. Fedorenko, M. Voichek, A. Leavitt, Y. Oppenheimer-Shaanan, R. Sorek, Bacterial retrons function in anti-phage defense. *Cell* **183**, 1551–1561.e12 (2020). doi:10.1016/j.cell.2020.09.065 Medline

11. S. Kronheim, M. Daniel-Ivad, Z. Duan, S. Hwang, A. I. Wong, I. Mantel, J. R. Nodwell, K. L. Maxwell, A chemical defence against phage infection. *Nature* **564**, 283–286 (2018). doi:10.1038/s41586-018-0767-x Medline

12. A. Bernheim, A. Millman, G. Ofir, G. Meitav, C. Avraham, H. Shomar, M. M. Rosenberg, N. Tal, S. Melamed, G. Amitai, R. Sorek, Prokaryotic viperins produce diverse antiviral molecules. *Nature* **589**, 120–124 (2021). doi:10.1038/s41586-020-2762-2 Medline

13. K. S. Makarova, Y. I. Wolf, E. V. Koonin, Comparative genomics of defense systems in archaea and bacteria. *Nucleic Acids Res.* **41**, 4360–4377 (2013). doi:10.1093/nar/gkt157 Medline

14. E. V. Koonin, L. Aravind, Origin and evolution of eukaryotic apoptosis: The bacterial connection. *Cell Death Differ.* **9**, 394–404 (2002). doi:10.1038/sj.cdd.4400991 Medline

15. D. D. Leipe, E. V. Koonin, L. Aravind, STAND, a class of P-loop NTPases including animal and plant regulators of programmed cell death: Multiple, complex domain architectures, unusual phyletic patterns, and evolution by horizontal gene transfer. *J. Mol. Biol.* **343**, 1–28 (2004). doi:10.1016/j.jmb.2004.08.023 Medline

16. Y. Zhao, J. Yang, J. Shi, Y.-N. Gong, Q. Lu, H. Xu, L. Liu, F. Shao, The NLRC4 inflammasome receptors for bacterial flagellin and type III secretion apparatus. *Nature* **477**, 596–600 (2011). doi:10.1038/nature10510 Medline

17. E. M. Kofoed, R. E. Vance, Innate immune recognition of bacterial ligands by NAIPs determines inflammasome specificity. *Nature* **477**, 592–595 (2011). doi:10.1038/nature10394 Medline

18. R. Caruso, N. Warner, N. Inohara, G. Núñez, NOD1 and NOD2: Signaling, host defense, and inflammatory disease. *Immunity* **41**, 898–908 (2014). doi:10.1016/j.immuni.2014.12.010 Medline

19. J. D. G. Jones, R. E. Vance, J. L. Dangl, Intracellular innate immune surveillance devices in plants and animals. *Science* **354**, aaf6395 (2016). doi:10.1126/science.aaf6395 Medline

20. J. Heller, C. Clavé, P. Gladieux, S. J. Saupe, N. L. Glass, NLR surveillance of essential SEC-9 SNARE proteins induces programmed cell death upon allorecognition in filamentous fungi. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E2292–E2301 (2018). doi:10.1073/pnas.1719705115 Medline

21. S. Bauernfried, M. J. Scherr, A. Pichlmair, K. E. Duderstadt, V. Hornung, Human NLRP1 is a sensor for double-stranded RNA. *Science* **371**, eabd0811 (2021). doi:10.1126/science.abd0811 Medline

22. O. Danot, A complex signaling module governs the activity of MalT, the prototype of an emerging transactivator family. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 435–440 (2001). doi:10.1073/pnas.98.2.435 Medline

23. S. Horinouchi, M. Kito, M. Nishiyama, K. Furuya, S. K. Hong, K. Miyake, T. Beppu, Primary structure of AfsR, a global regulatory protein for secondary metabolite formation in *Streptomyces coelicolor* A3(2). *Gene* **95**, 49–56 (1990). doi:10.1016/0378-1119(90)90412-K Medline

24. R. Ye, S. N. Rehemtulla, S. L. Wong, Glucitol induction in *Bacillus subtilis* is mediated by a regulatory factor, GutR. *J. Bacteriol.* **176**, 3321–3327 (1994). doi:10.1128/jb.176.11.3321-3327.1994 Medline

25. E. V. Koonin, V. V. Dolja, M. Krupovic, A. Varsani, Y. I. Wolf, N. Yutin, F. M. Zerbini, J. H. Kuhn, Global organization and proposed megataxonomy of the virus world. *Microbiol. Mol. Biol. Rev.* **84**, e00061-19 (2020). doi:10.1128/MMBR.00061-19 Medline

26. B. Lowey, A. T. Whiteley, A. F. A. Keszei, B. R. Morehouse, I. T. Mathews, S. P. Antine, V. J. Cabrera, D. Kashin, P. Niemann, M. Jain, F. Schwede, J. J. Mekalanos, S. Shao, A. S. Y. Lee, P. J. Kranzusch, 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). doi:10.1016/j.cell.2020.05.019 Medline

27. K. S. Makarova, A. Timinskas, Y. I. Wolf, A. B. Gussow, V. Siksnys, Č. Venclovas, E. V. Koonin, Evolutionary and functional classification of the CARF domain superfamily, key sensors in prokaryotic antivirus defense. *Nucleic Acids Res.* **48**, 8828–8847 (2020). doi:10.1093/nar/gkaa635 Medline

28. J. Heitman, P. Model, Site-specific methylases induce the SOS DNA repair response in *Escherichia coli*. *J. Bacteriol.* **169**, 3243–3250 (1987). doi:10.1128/jb.169.7.3243-3250.1987 Medline

29. R. Martin, T. Qi, H. Zhang, F. Liu, M. King, C. Toth, E. Nogales, B. J. Staskawicz, Structure of the activated ROQ1 resistosome directly recognizing the pathogen effector XopQ. *Science* **370**, eabd9993 (2020). doi:10.1126/science.abd9993 Medline

30. M. Zhou, Y. Li, Q. Hu, X.-C. Bai, W. Huang, C. Yan, S. H. W. Scheres, Y. Shi, Atomic structure of the apoptosome: Mechanism of cytochrome *c*- and dATP-mediated activation of Apaf-1. *Genes Dev.* **29**, 2349–2361 (2015). doi:10.1101/gad.272278.115 Medline

31. K. Steczkiewicz, A. Muszewska, L. Knizewski, L. Rychlewski, K. Ginalski, Sequence, structure and functional diversity of PD-(D/E)XK phosphodiesterase superfamily. *Nucleic Acids Res.* **40**, 7016–7045 (2012). doi:10.1093/nar/gks382 Medline

32. A. Pingoud, M. Fuxreiter, V. Pingoud, W. Wende, Type II restriction endonucleases: Structure and mechanism. *Cell. Mol. Life Sci.* **62**, 685–707 (2005). doi:10.1007/s00018-004-4513-1 Medline

33. N. Watanabe, Y. Takasaki, C. Sato, S. Ando, I. Tanaka, Structures of restriction endonuclease HindIII in complex with its cognate DNA and divalent cations. *Acta Crystallogr. D Biol. Crystallogr.* **65**, 1326–1333 (2009). doi:10.1107/S0907444909041134 Medline

34. A. M. Burroughs, L. M. Iyer, L. Aravind, Comparative genomics and evolutionary trajectories of viral ATP dependent DNA-packaging systems. *Genome Dyn.* **3**, 48–65 (2007). doi:10.1159/000107603 Medline

35. G. Kaur, L. M. Iyer, A. M. Burroughs, L. Aravind, Bacterial death and TRADD-N domains help define novel apoptosis and immunity mechanisms shared by prokaryotes and metazoans. *eLife* **10**, e70394 (2021). doi:10.7554/eLife.70394 Medline

36. G. Kaur, A. M. Burroughs, L. M. Iyer, L. Aravind, Highly regulated, diversifying NTP-dependent biological conflict systems with implications for the emergence of multicellularity. *eLife* **9**, e52696 (2020). doi:10.7554/eLife.52696 Medline

37. J. E. Samson, A. H. Magadán, M. Sabri, S. Moineau, Revenge of the phages: Defeating bacterial defences. *Nat. Rev. Microbiol.* **11**, 675–687 (2013). doi:10.1038/nrmicro3096 Medline

38. A. Pawluk, A. R. Davidson, K. L. Maxwell, Anti-CRISPR: Discovery, mechanism and function. *Nat. Rev. Microbiol.* **16**, 12–17 (2018). doi:10.1038/nrmicro.2017.120 Medline

39. J. Bondy-Denomy, A. Pawluk, K. L. Maxwell, A. R. Davidson, Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature* **493**, 429–432 (2013). doi:10.1038/nature11723 Medline

40. N. S. Olsen, L. Forero-Junco, W. Kot, L. H. Hansen, Exploring the remarkable diversity of culturable *Escherichia coli* phages in the Danish wastewater environment. *Viruses* **12**, 986 (2020). doi:10.3390/v12090986 Medline

41. M. Schmerer, I. J. Molineux, J. J. Bull, Synergy as a rationale for phage therapy using phage cocktails. *PeerJ* **2**, e590 (2014). doi:10.7717/peerj.590 Medline

42. G. F. Hatfull, R. W. Hendrix, Bacteriophages and their genomes. *Curr. Opin. Virol.* **1**, 298–303 (2011). doi:10.1016/j.coviro.2011.06.009 Medline

43. H.-W. Ackermann, 5500 Phages examined in the electron microscope. *Arch. Virol.* **152**, 227–243 (2007). doi:10.1007/s00705-006-0849-1 Medline

44. C. L. Dedeo, G. Cingolani, C. M. Teschke, Portal protein: The orchestrator of capsid assembly for the dsDNA tailed bacteriophages and herpesviruses. *Annu. Rev. Virol.* **6**, 141–160 (2019). doi:10.1146/annurev-virology-092818-015819 Medline

45. S. R. Casjens, The DNA-packaging nanomotor of tailed bacteriophages. *Nat. Rev. Microbiol.* **9**, 647–657 (2011). doi:10.1038/nrmicro2632 Medline

46. O. Danot, E. Marquenet, D. Vidal-Ingigliardi, E. Richet, Wheel of life, wheel of death: A mechanistic insight into signaling by STAND proteins. *Structure* **17**, 172–182 (2009). doi:10.1016/j.str.2009.01.001 Medline

47. S. Ma, D. Lapin, L. Liu, Y. Sun, W. Song, X. Zhang, E. Logemann, D. Yu, J. Wang, J. Jirschitzka, Z. Han, P. Schulze-Lefert, J. E. Parker, J. Chai, Direct pathogen-induced assembly of an NLR immune receptor complex to form a holoenzyme. *Science* **370**, eabe3069 (2020). doi:10.1126/science.abe3069 Medline

48. B. R. Morehouse, A. A. Govande, A. Millman, A. F. A. Keszei, B. Lowey, G. Ofir, S. Shao, R. Sorek, P. J. Kranzusch, STING cyclic dinucleotide sensing originated in bacteria. *Nature* **586**, 429–433 (2020). doi:10.1038/s41586-020-2719-5 Medline

49. S. A. Shmakov, K. S. Makarova, Y. I. Wolf, K. V. Severinov, E. V. Koonin, Systematic prediction of genes functionally linked to CRISPR-Cas systems by gene neighborhood analysis. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E5307–E5316 (2018). doi:10.1073/pnas.1803440115 Medline

50. S. F. Altschul, T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D. J. Lipman, Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997). doi:10.1093/nar/25.17.3389 Medline

51. M. Steinegger, J. Söding, MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, 1026–1028 (2017). doi:10.1038/nbt.3988 Medline

52. R. C. Edgar, MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 113 (2004). doi:10.1186/1471-2105-5-113 Medline

53. M. Steinegger, M. Meier, M. Mirdita, H. Vöhringer, S. J. Haunsberger, J. Söding, HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinformatics* **20**, 473 (2019). doi:10.1186/s12859-019-3019-7 Medline

54. K. S. Makarova, Y. I. Wolf, J. Iranzo, S. A. Shmakov, O. S. Alkhnbashi, S. J. J. Brouns, E.

Charpentier, D. Cheng, D. H. Haft, P. Horvath, S. Moineau, F. J. M. Mojica, D. Scott, S. A. Shah, V. Siksnys, M. P. Terns, Č. Venclovas, M. F. White, A. F. Yakunin, W. Yan, F. Zhang, R. A. Garrett, R. Backofen, J. van der Oost, R. Barrangou, E. V. Koonin, Evolutionary classification of CRISPR-Cas systems: A burst of class 2 and derived variants. *Nat. Rev. Microbiol.* **18**, 67–83 (2020). doi:10.1038/s41579-019-0299-x Medline

55. M. N. Price, P. S. Dehal, A. P. Arkin, FastTree 2—Approximately maximum-likelihood trees for large alignments. *PLOS ONE* **5**, e9490 (2010). doi:10.1371/journal.pone.0009490 Medline

56. K. Katoh, K. Misawa, K. Kuma, T. Miyata, MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002). doi:10.1093/nar/gkf436 Medline

57. L.-T. Nguyen, H. A. Schmidt, A. von Haeseler, B. Q. Minh, IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015). doi:10.1093/molbev/msu300 Medline

58. L. Zimmermann, A. Stephens, S.-Z. Nam, D. Rau, J. Kübler, M. Lozajic, F. Gabler, J. Söding, A. N. Lupas, V. Alva, a completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J. Mol. Biol.* **430**, 2237–2243 (2018). doi:10.1016/j.jmb.2017.12.007 Medline

59. I. Letunic, P. Bork, Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021). doi:10.1093/nar/gkab301 Medline

60. A.-L. Van de Weyer, F. Monteiro, O. J. Furzer, M. T. Nishimura, V. Cevik, K. Witek, J. D. G. Jones, J. L. Dangl, D. Weigel, F. Bemm, A species-wide inventory of NLR genes and alleles in *Arabidopsis thaliana*. *Cell* **178**, 1260–1272.e14 (2019). doi:10.1016/j.cell.2019.07.038 Medline

61. W. Dyrka, M. Lamacchia, P. Durrens, B. Kobe, A. Daskalov, M. Paoletti, D. J. Sherman, S. J. Saupe, Diversity and variability of NOD-like receptors in fungi. *Genome Biol. Evol.* **6**, 3137–3158 (2014). doi:10.1093/gbe/evu251 Medline

62. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021). doi:10.1038/s41586-021-03819-2 Medline

63. J. Pei, B. H. Kim, N. V. Grishin, PROMALS3D: A tool for multiple protein sequence and structure alignments. *Nucleic Acids Res.* **36**, 2295–2300 (2008). doi:10.1093/nar/gkn072 Medline

64. S. Picelli, A. K. Björklund, B. Reinius, S. Sagasser, G. Winberg, R. Sandberg, Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.* **24**, 2033–2040 (2014). doi:10.1101/gr.177881.114 Medline

65. M. Huber, G. Faure, S. Laass, E. Kolbe, K. Seitz, C. Wehrheim, Y. I. Wolf, E. V. Koonin, J. Soppa, Translational coupling via termination-reinitiation in archaea and bacteria. *Nat. Commun.* **10**, 4006 (2019). doi:10.1038/s41467-019-11999-9 Medline

66. U. Qimron, B. Marintcheva, S. Tabor, C. C. Richardson, Genomewide screens for Escherichia coli genes affecting growth of T7 bacteriophage. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 19039–19044 (2006). doi:10.1073/pnas.0609428103 Medline

67. A. M. Grigonyte, C. Harrison, P. R. MacDonald, A. Montero-Blay, M. Tridgett, J. Duncan, A. P. Sagona, C. Constantinidou, A. Jaramillo, A. Millard, Comparison of CRISPR and marker-based methods for the engineering of phage T7. *Viruses* **12**, 193 (2020). doi:10.3390/v12020193 Medline

68. T. Baba, T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K. A. Datsenko, M. Tomita, B. L. Wanner, H. Mori, Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: The Keio collection. *Mol. Syst. Biol.* **2**, 0008 (2006). Medline

69. D. Kimanius, L. Dong, G. Sharov, T. Nakane, S. H. W. Scheres, New tools for automated cryo-EM single-particle analysis in RELION-4.0. *Biochem. J.* **478**, 4169–4185 (2021). doi:10.1042/BCJ20210708 Medline

70. T. Bepler, A. Morin, M. Rapp, J. Brasch, L. Shapiro, A. J. Noble, B. Berger, Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nat. Methods* **16**, 1153–1160 (2019). doi:10.1038/s41592-019-0575-8 Medline

71. T. I. Croll, ISOLDE: A physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr. D Struct. Biol.* **74**, 519–530 (2018). doi:10.1107/S2059798318002425 Medline

72. A. Casañal, B. Lohkamp, P. Emsley, Current developments in Coot for macromolecular model building of electron cryo-microscopy and crystallographic data. *Protein Sci.* **29**, 1069–1078 (2020). doi:10.1002/pro.3791 Medline

73. D. Liebschner, P. V. Afonine, M. L. Baker, G. Bunkóczi, V. B. Chen, T. I. Croll, B. Hintze, L. W. Hung, S. Jain, A. J. McCoy, N. W. Moriarty, R. D. Oeffner, B. K. Poon, M. G. Prisant, R. J. Read, J. S. Richardson, D. C. Richardson, M. D. Sammito, O. V. Sobolev, D. H. Stockwell, T. C. Terwilliger, A. G. Urzhumtsev, L. L. Videau, C. J. Williams, P. D. Adams, Macromolecular structure determination using x-rays, neutrons and electrons: Recent developments in Phenix. *Acta Crystallogr. D Struct. Biol.* **75**, 861–877 (2019). doi:10.1107/S2059798319011471 Medline

74. N. Wong, S. John, A. Nussenzweig, A. Canela, END-seq: An unbiased, high-resolution, and genome-wide approach to map DNA double-strand breaks and resection in human cells. *Methods Mol. Biol.* **2153**, 9–31 (2021). doi:10.1007/978-1-0716-0644-5_2 Medline

75. L. Holm, Using Dali for protein structure comparison. *Methods Mol. Biol.* **2112**, 29–42 (2020). doi:10.1007/978-1-0716-0270-6_3 Medline

76. J. Wang, M. Hu, J. Wang, J. Qi, Z. Han, G. Wang, Y. Qi, H.-W. Wang, J.-M. Zhou, J. Chai, Reconstitution and structure of a plant NLR resistosome conferring immunity. *Science* **364**, eaav5870 (2019). doi:10.1126/science.aav5870 Medline

77. Y. Pang, X.-C. Bai, C. Yan, Q. Hao, Z. Chen, J.-W. Wang, S. H. W. Scheres, Y. Shi, Structure of the apoptosome: Mechanistic insights into activation of an initiator caspase from *Drosophila*. *Genes Dev.* **29**, 277–287 (2015). doi:10.1101/gad.255877.114 Medline

78. L. Zhang, S. Chen, J. Ruan, J. Wu, A. B. Tong, Q. Yin, Y. Li, L. David, A. Lu, W. L. Wang, C. Marks, Q. Ouyang, X. Zhang, Y. Mao, H. Wu, Cryo-EM structure of the activated NAIP2-NLRC4 inflammasome reveals nucleated polymerization. *Science* **350**, 404–409 (2015). doi:10.1126/science.aac5789 Medline

79. J. L. Tenthorey, N. Haloupek, J. R. López-Blanco, P. Grob, E. Adamson, E. Hartenian, N. A. Lind, N. M. Bourgeois, P. Chacón, E. Nogales, R. E. Vance, The structural basis of flagellin detection by NAIP5: A strategy to limit pathogen immune evasion. *Science* **358**, 888–893 (2017). doi:10.1126/science.aao1140 Medline

80. E. S. Esterman, Y. I. Wolf, R. Kogay, E. V. Koonin, O. Zhaxybayeva, Evolution of DNA packaging in gene transfer agents. *Virus Evol.* **7**, veab015 (2021). doi:10.1093/ve/veab015 Medline

81. S. Sun, K. Kondabagil, B. Draper, T. I. Alam, V. D. Bowman, Z. Zhang, S. Hegde, A. Fokine, M. G. Rossmann, V. B. Rao, The structure of the phage T4 DNA packaging motor suggests a mechanism dependent on electrostatic forces. *Cell* **135**, 1251–1262 (2008). doi:10.1016/j.cell.2008.11.015 Medline

82. A. Cuervo, M. Fàbrega-Ferrer, C. Machón, J. J. Conesa, F. J. Fernández, R. Pérez-Luque, M. Pérez-Ruiz, J. Pous, M. C. Vega, J. L. Carrascosa, M. Coll, Structures of T7 bacteriophage portal and tail suggest a viral DNA retention and ejection mechanism. *Nat. Commun.* **10**, 3746 (2019). doi:10.1038/s41467-019-11705-9 Medline