

MIT Open Access Articles

Prokaryotic innate immunity through pattern recognition of conserved viral proteins

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Gao, Linyi Alex, Wilkinson, Max E, Strecker, Jonathan, Makarova, Kira S, Macrae, Rhiannon K et al. 2022. "Prokaryotic innate immunity through pattern recognition of conserved viral proteins." *Science*, 377 (6607).

As Published: 10.1126/science.abm4096

Publisher: American Association for the Advancement of Science (AAAS)

Persistent URL: <https://hdl.handle.net/1721.1/150378>

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Terms of use: Creative Commons Attribution-Noncommercial-Share Alike





HHS Public Access

Author manuscript

Science. Author manuscript; available in PMC 2023 March 21.

Published in final edited form as:

Science. 2022 August 12; 377(6607): eabm4096. doi:10.1126/science.abm4096.

Prokaryotic innate immunity via 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. Makarova⁷, Rhiannon K. Macrae^{1,2,3,4,5}, Eugene V. Koonin⁷, Feng Zhang^{1,2,3,4,5,*}

¹Howard Hughes Medical Institute, Cambridge, MA 02139, USA

²Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

³McGovern Institute for Brain Research

⁴Department of Brain and Cognitive Sciences

⁵Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

⁶Society of Fellows, Harvard University, Cambridge, MA 02138, USA

⁷National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA

Abstract

Organisms across life 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. While 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-EM 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.

Graphical Abstract

*Correspondence: zhang@broadinstitute.org (F.Z.) or linyigao@stanford.edu (L.A.G.).

†These authors contributed equally to this work.

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 phylogenetic analysis. L.A.G., M.E.W, J.S., K.S.M., E.V.K., and F.Z. analyzed data. F.Z. supervised 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.

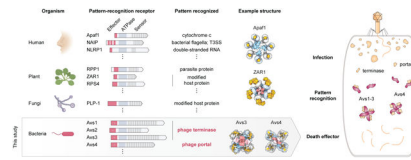
Supplementary Materials

Figs. S1–S27

Table S1

Data S1–S19

References (74–82)



One-Sentence Summary:

Bacteria and archaea possess innate immune receptors that recognize conserved viral proteins and prevent infection, extending the paradigm of protein pattern recognition across all three domains of life.

Bacteria and archaea have evolved numerous defense mechanisms against viral infections involving 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 NTPases, which include 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 containing 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 transcription regulators (22–24). We recently identified a group of 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 double-stranded DNA (dsDNA) 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. 1A, B), it is unclear whether they function via similar molecular mechanisms. We identified four distinct families of Avs proteins (Avs1-4) (fig. S3), each containing highly divergent tetratricopeptide repeat (TPR) 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 comprising 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 co-expression 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 co-expressed 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 gp8 and gp19 knockout phage strains, respectively (Fig. 1E and fig. S5), indicating that gp8 and gp19 are both necessary and sufficient for Avs activation.

To validate these findings, we transformed plasmids expressing gp8 or gp19 into *E. coli* harboring SeAvs3 or EcAvs4 and measured cell viability. Consistent with our previous results, we observed cell death following co-expression of SeAvs3 and gp19, as well as co-expression 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, as recoded gene sequences also led to cell death (Fig. 1F). Furthermore, the enzymatic activity of the phage terminase, which contains ATPase and nuclease domains unrelated to those of Avs proteins, was not required for SeAvs3-mediated toxicity (Fig. 1F).

Avs proteins are pattern recognition receptors recognizing 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 co-expressed these genes in *E. coli* with 15 Avs systems spanning all four Avs families (data S2 and S3). 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 exquisite target specificity: Avs1-3 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 varying the Avs promoter or the amount of terminase and portal induction, obtaining similar results (fig. S7). Surprisingly, Avs1 and

Avs2 also recognized terminases despite the lack of significant sequence similarity among the C-terminal TPRs of Avs1, Avs2, and 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 remarkable conservation of target recognition across Avs families and suggest that the portal and large terminase subunit are key PAMPs recognized by prokaryotes. Moreover, Avs systems recognize PAMPs from diverse phages; for example, SeAvs3 and EcAvs2 are strongly activated by 20 of 24 and 19 of 24 tested terminases, respectively, and EcAvs4 is 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 (fig. S9–11), this broad range of activity implies that Avs proteins are triggered by conserved structural features rather than by specific peptide sequences. Consistent with 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).

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 the recently reported 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 catalytic motifs (Fig. 3B), and in addition to the STAND NTPase, the SeAvs3 system contains a small ORF, the deletion of which reduced anti-phage 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 portal (gp8) and terminase (gp19) proteins (fig. S13A). We incubated SeAvs3 and the small ORF product with linear 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 ssDNA and RNA were not cleaved (fig. S13B). Moreover, SeAvs3 cleaves 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 ATP; however, ATP hydrolysis is not strictly required because nuclease activity was observed at a reduced level in the presence of the nonhydrolyzable ATP analogue 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 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, widely 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-EM structures of the SeAvs3-terminase and EcAvs4-portal complexes in the presence of ATP and Mg^{2+} (Fig. S14–17 and table S1). A reconstruction at 3.4 Å resolution revealed that SeAvs3 forms a tetramer, with each C-terminal TPR domain gripping the terminase ATPase and nuclease domains of the gp19 terminase (Fig. 4A). These TPR lobes are flexible and required symmetry expansion to improve their local resolution to 3.4 Å (Methods, fig. S17). For EcAvs4 bound to the PhiV-1 gp8 portal, 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 via 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 STAND 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 EcAvs4 STAND forms a C₂-symmetric dimer of dimers with a tighter interface (1232 Å² buried surface area cf. 436 Å² for SeAvs3), with adjacent WHDs and NBDs both interacting (fig. S18). The smaller interface in EcAvs3 is compensated for by additional contacts between its C-terminal TPR domains (Fig. 4C). 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 alpha 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 HindIII 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, Glu49, which is part of the conserved trio of active-site residues, is displaced (Fig. 4, M and N). Glu49 is found in the loop spanning residues 33–52, 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 HindIII (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 co-purification (fig. S19–20). 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 less than 5% sequence identity (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 Å² 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, Asp1710 in SeAvs3 forms a salt bridge with the highly conserved Arg61 within the Walker A motif of the terminase ATPase (Fig. 5C). An arginine in this position is found in most terminase ATPases that activate SeAvs3, but not in non-activating terminases (Fig. 5E and fig. S21). These observations suggest that Arg61 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 (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, Arg1196 and Lys1198 in SeAvs3 form salt bridges to the four conserved aspartates that comprise the active site of the terminase nuclease (Fig. 5, D and E), and mutation of Asp365 in the PhiV-1 terminase nuclease notably reduced SeAvs3 activation (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 His1770 and Tyr1714 of SeAvs3 (Fig. 5C). Targeting this ligand presumably helps avoid phage escape mutations, since 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 to Avs2.

The TPR domain of EcAvs4 also binds the PhiV-1 portal with a large interface, burying 5800 Å² of solvent-accessible surface area, that includes remarkably few residue-residue contacts (Fig. 5G). The portal protein is recognized via its stem, clip, and part of its wing domain. In an assembled dodecameric portal complex, these regions are found towards the interior and are therefore more constrained in their fold requirements (Fig. 5I). Consistent with this observation, we performed random mutagenesis by PCR to screen for portal mutations that abrogate Avs4 activation, finding that all 29 identified mutations were non-conservative and located in the core wing or stem regions of the portal (fig. S23), possibly disrupting the core fold. The clip domain, which contains a conserved anti-parallel beta sheet with an intervening alpha helix, is recognized by β-sheet augmentation with a hairpin of EcAvs4, a mode of fold recognition that does not depend on the amino acid sequence of the target (Fig. 5H). 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 Avs4 recognizes portal monomers before they assemble into the procapsid.

Avs proteins are widespread in prokaryotes and possess diverse, modular N-terminal effector domains

To assess the diversity of *avs* genes across prokaryotic lineages, we collected all intact homologs from each of the four families present in the NCBI non-redundant sequence database (data S4–8). The *avs* genes were identified in approximately 4–5% of sequenced prokaryotic genomes and are broadly distributed (Fig. 5A), with at least one *avs* gene detected in 27 of 29 and 3 of 10 well-represented bacterial and archaeal phyla, respectively (Fig. 5B and fig. S24A). Each Avs family has a characteristic protein size (Fig. 5C), consistent with their distinct mechanisms of target recognition. We next constructed phylogenetic trees of each of the four families (Fig. 5, 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, SIR2, TIR, CMP hydrolases, transmembrane helices, and domains with unknown functions (data S4–8). Some less common variants are predicted to participate in intracellular signaling networks via effector-associated domains (EADs) that 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 highly similar to those of EcAvs4, with 44% amino acid identity, but encompasses an N-terminal region with 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. 4F and fig. S26B), while retaining the ability to recognize the portal proteins from diverse phages (fig. S26C).

Phages circumvent Avs defense through diverse anti-defense 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 anti-defense genes (data S9). We performed a genetic screen for suppressors of Avs toxicity by co-expressing these genes with SeAvs3, EcAvs4, or KpAvs4 and their cognate phage trigger (Fig. 6A). We identified several early genes which rescued cell growth (Fig. 6B), most of which originate from a hypervariable region within a group of closely related phages isolated from wastewater (Fig. 6C) (40, 41).

To validate these observations, we produced three of the Avs inhibitors by cell-free translation and observed inhibition of SeAvs3 nuclease activity *in vitro* by Lidtsur-17, and to a lesser degree by Forsur-7 (Fig. 6D). 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 unique 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 (Avs1-3) 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. Although many Avs proteins contain predicted nucleases, we identified diverse N-terminal effector domains throughout Avs families, indicating unique 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 Avs3 and Avs4 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 proteins (46). Notably, oligomerization is also involved in the activation of the Cap4 nucleases in cyclic oligonucleotide-based anti-phage signaling systems (26).

Remarkably, 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.

Striking similarities, but also notable differences, exist between eukaryotic NLRs and prokaryotic Avs proteins. Both are intracellular receptors of the STAND superfamily that detect PAMPs via 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 two-fold 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). Some Avs homologs in multicellular prokaryotes, such as cyanobacteria, contain EADs mediating protein-protein interaction and might recruit caspase-like proteases encoded in the same genomic loci (fig. S25), consistent with the recent discovery that such signaling is a hallmark of multicellularity (35, 36).

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 S1) but observed no detectable toxicity when co-expressed with any of the 48 tested terminases or portals. In light of the high success rate of this assay for Avs homologs, these results suggest that they are triggered by other pathogen-related patterns that remain to be identified. Further investigation will shed light on 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 their eukaryotic NLR homologs, 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 can 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 Avs1-3 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 2,000 best hits from each run were taken and combined with 949 Avs4 homologs found in the NCBI non-redundant protein sequence database (nr) in 2021. A non-redundant 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 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 non-redundant 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 via 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 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).

Prior to computing pairwise sequence identity, 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) or Phusion Flash (Thermo Scientific) polymerase. Plasmids were assembled using the Gibson Assembly or NEBuilder HiFi DNA Assembly mix (New England Biolabs). Plasmid sequences were verified by Tn5 tagmentation and high-throughput sequencing, as previously described (4, 64).

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 open reading frames 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 lacks the ability to express from T7 promoters, was also included to assess the basal toxicity of the phage genes.

After 1 hr outgrowth in S.O.C. at 37°C, cells were plated on LB agar plates containing 25 µg/mL chloramphenicol and 100 µg/mL ampicillin in the absence of IPTG. Plates were incubated for an additional 12h 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 bp. Following addition of 0.5 volumes of 0.1% SDS and column purification, tagmented fragments were amplified over 8 cycles by Q5 DNA polymerase (NEB) with unique 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 homolog or a pACYC184 empty vector. After 1 hr outgrowth in S.O.C. 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 12h at 37°C, plasmids were isolated and sequenced, and depletion values were computed as described for the PhiV-1 depletion screen (Fig. 2, fig. S7, and fig. S26C).

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 via 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 re-transformed 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 unique 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 (fig. S22 and fig. S23E), as well as terminase domain mutations (Fig. 5F).

Anti-defense 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 via Gibson assembly into expression vectors containing either the portal or terminase from phage PhiV-1 driven by a pBAD promoter. Anti-defense 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 16h on LB agar plates containing 25 µg/mL chloramphenicol and 100 µg/mL ampicillin with no added arabinose. Following plasmid isolation, anti-defense candidates were amplified over two rounds of PCR to attach 8-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 Luria Broth (LB) or Terrific Broth. To 10 mL top agar (10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl, 7 g/L agar) was added chloramphenicol 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.

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.

Protein purification

Avs NTPases 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 BL21(DE3) cells (NEB #C2527H). Cells were grown in Terrific Broth to mid-log phase, and the temperature was lowered to 18°C. Expression was induced at OD₆₀₀ 0.6 with 0.25 mM IPTG for 16–20 h 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 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 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 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 MWCO Amicon spin concentrator.

SeAvs3-terminase complex purification

Avs-TwinStrep constructs were co-transformed with plasmids expressing either gp8 or gp19 into electrocompetent BL21 cells (Sigma Aldrich CMC0016) and grown and induced as before. Avs pulldowns using Strep-Tactin Superflow Plus resin were run on SDS-PAGE gels, and gp8 and gp19 bands were excised and confirmed by mass spectrometry (Taplin Biological Mass Spectrometry Facility, Harvard Medical School). Plasmids containing SeAvs3-6xHis and gp19-StrepTag were co-transformed into electrocompetent and grown and induced as before. An SeAvs3-gp19 complex was purified using a tandem affinity strategy, with Ni-NTA followed by Strep-Tactin resin. The final elution was run on a Superose 6 Increase column yielding 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 μ g of SeAvs3 was mixed with 8.3 μ g of PhiV-1 gp19 terminase in a total volume of 24 μ L with buffer composition 17 mM Tris-HCl pH 7.5, 280 mM NaCl, 0.8 mM DTT, 2% glycerol, 5 mM MgCl₂, and 1 mM ATP. The reaction was incubated at 37 °C for 30 min then placed on ice for approximately 1 hour prior to 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 μ L 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 and 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 co-transformed into BL21(DE3). A total of 6 L 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 MgCl₂, 0.1 mM ATP) supplemented with EDTA-free cComplete 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 μ g homemade bdSENP1 protease. Eluted protein was incubated with Strep-Tactin Superflow Plus (Qiagen) 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 MgCl₂, 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 was 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 2-fold binning, and an energy filter with slit width of 20 eV.

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

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 \times 5 patches and dose-weighting. 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 3D classification was used to train Topaz, and this trained model was used to pick 128,500 particles from the entire dataset. Extracted particles, downsampled 4-fold, 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 re-extracted 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 sub-particles 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 4-fold 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 re-extracted 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 sub-particles 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 4th 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 sub-particles (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 macro-cycle of global minimization and 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 final concentration). Typical reactions contained approximately 100 ng of DNA substrate, 100 ng Avs protein, and 100 ng gp8 or gp19 in a 10 μ L reaction volume. MgCl₂ was added at 5 mM where indicated, 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. BTH101 cells (*F⁻*, *cya-99*, *araD139*, *galE15*, *galK16*, *rpsL1 (Str^r)*, *hsdR2*, *mcrA1*, *mcrB1*) were co-transformed 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 μ g/mL ammonium iron (III) citrate, 300 μ g/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 (OD₆₀₀ 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, prior to cell lysis, infected cells were pelleted and resuspended in 200 μ L 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 μ L cultures by adding 200 μ L lysis buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 100 mM NaCl, 1% SDS, 2% Triton X-100), 100 μ L of glass beads, and 200 μ L 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 ice-cold 100% ethanol and centrifugation at 4°C. DNA pellets were resuspended in 200 μ L TE with 300 μ g RNase A (Sigma-Aldrich) and incubated at 37°C for 1 hour. DNA was once again precipitated with the addition of 1 mL ice-cold 100% ethanol and 20 μ L of 4M ammonium acetate, centrifuged, dried, and resuspended in TE.

DNA was digested with Eco47III and run on a 1% agarose gel in 1x 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 NaPO₄ 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments:

We thank David Li, Han Altae-Tran, and Guilhem Faure for valuable discussions; Ross Tomaino for assistance with mass spectrometry; Ed Brignole and Christopher 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 U.S. Department of Health and Human Services (to 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 by the Phillips family and J. and P. Poitras.

Data and materials availability:

Expression plasmids for all Avs systems and inhibitors are available from Addgene. 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), 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 PDB 8DGC (SeAvs3-gp19) and 8DGF (EcAvs4-gp8).

References and Notes

1. Makarova KS, Wolf YI, Snir S, Koonin EV, Defense islands in bacterial and archaeal genomes and prediction of novel defense systems. *J. Bacteriol* 193, 6039–6056 (2011). [PubMed: 21908672]
2. Koonin EV, Makarova KS, Wolf YI, Evolutionary Genomics of Defense Systems in Archaea and Bacteria. *Annu. Rev. Microbiol* 71, 233–261 (2017). [PubMed: 28657885]
3. Doron S, Melamed S, Ofir G, Leavitt A, Lopatina A, Keren M, Amitai G, Sorek R, Systematic discovery of antiphage defense systems in the microbial pangenome. *Science*. 359 (2018), doi:10.1126/science.aar4120.
4. Gao L, Altae-Tran H, Böhning F, Makarova KS, Segel M, Schmid-Burgk JL, Koob J, Wolf YI, Koonin EV, Zhang F, Diverse enzymatic activities mediate antiviral immunity in prokaryotes. *Science*. 369, 1077–1084 (2020). [PubMed: 32855333]
5. Hampton HG, Watson BNJ, Fineran PC, The arms race between bacteria and their phage foes. *Nature*. 577, 327–336 (2020). [PubMed: 31942051]
6. Klaiman D, Steinfels-Kohn E, Kaufmann G, A DNA break inducer activates the anticodon nuclease RloC and the adaptive immunity in *Acinetobacter baylyi* ADPI. *Nucleic Acids Res.* 42, 328–339 (2014). [PubMed: 24062157]
7. Guegler CK, Laub MT, Shutoff of host transcription triggers a toxin-antitoxin system to cleave phage RNA and abort infection. *Mol. Cell* 81, 2361–2373.e9 (2021). [PubMed: 33838104]
8. Cheng R, Huang F, Wu H, Lu X, Yan Y, Yu B, Wang X, Zhu B, A nucleotide-sensing endonuclease from the Gabija bacterial defense system. *Nucleic Acids Res.* 49, 5216–5229 (2021). [PubMed: 33885789]

9. Bingham R, Ekunwe SI, Falk S, Snyder L, Kleanthous C, The major head protein of bacteriophage T4 binds specifically to elongation factor Tu. *J. Biol. Chem* 275, 23219–23226 (2000). [PubMed: 10801848]
10. Millman A, Bernheim A, Stokar-Avihail A, Fedorenko T, Voichek M, Leavitt A, Oppenheimer-Shaanan Y, Sorek R, Bacterial Retrons Function In Anti-Phage Defense. *Cell*. 183, 1551–1561.e12 (2020). [PubMed: 33157039]
11. Kronheim S, Daniel-Ivad M, Duan Z, Hwang S, Wong AI, Mantel I, Nodwell JR, Maxwell KL, A chemical defence against phage infection. *Nature*. 564 (2018), pp. 283–286. [PubMed: 30518855]
12. Bernheim A, Millman A, Ofir G, Meitav G, Avraham C, Shomar H, Rosenberg MM, Tal N, Melamed S, Amitai G, Sorek R, Prokaryotic vaperins produce diverse antiviral molecules. *Nature*. 589, 120–124 (2021). [PubMed: 32937646]
13. Makarova KS, Wolf YI, Koonin EV, Comparative genomics of defense systems in archaea and bacteria. *Nucleic Acids Res.* 41, 4360–4377 (2013). [PubMed: 23470997]
14. Koonin EV, Aravind L, Origin and evolution of eukaryotic apoptosis: the bacterial connection. *Cell Death Differ.* 9, 394–404 (2002). [PubMed: 11965492]
15. Leipe DD, Koonin EV, Aravind L, 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). [PubMed: 15381417]
16. Zhao Y, Yang J, Shi J, Gong Y-N, Lu Q, Xu H, Liu L, Shao F, The NLRC4 inflammasome receptors for bacterial flagellin and type III secretion apparatus. *Nature*. 477 (2011), pp. 596–600. [PubMed: 21918512]
17. Kofoed EM, Vance RE, Innate immune recognition of bacterial ligands by NAIPs determines inflammasome specificity. *Nature*. 477, 592–595 (2011). [PubMed: 21874021]
18. Caruso R, Warner N, Inohara N, Núñez G, NOD1 and NOD2: signaling, host defense, and inflammatory disease. *Immunity*. 41, 898–908 (2014). [PubMed: 25526305]
19. Jones JDG, Vance RE, Dangl JL, Intracellular innate immune surveillance devices in plants and animals. *Science*. 354 (2016), doi:10.1126/science.aaf6395.
20. Heller J, Clavé C, Gladieux P, Saupe SJ, Glass NL, 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). [PubMed: 29463729]
21. Bauernfried S, Scherr MJ, Pichlmair A, Duderstadt KE, Hornung V, Human NLRP1 is a sensor for double-stranded RNA. *Science*. 371 (2021), doi:10.1126/science.abd0811.
22. Danot O, 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). [PubMed: 11209048]
23. Horinouchi S, Kito M, Nishiyama M, Furuya K, Hong SK, Miyake K, Beppu T, Primary structure of AfsR, a global regulatory protein for secondary metabolite formation in *Streptomyces coelicolor* A3(2). *Gene*. 95, 49–56 (1990). [PubMed: 2253887]
24. Ye R, Rehemtulla SN, Wong SL, Glucitol induction in *Bacillus subtilis* is mediated by a regulatory factor, GutR. *J. Bacteriol* 176, 3321–3327 (1994). [PubMed: 8195087]
25. Koonin EV, Dolja VV, Krupovic M, Varsani A, Wolf YI, Yutin N, Zerbini FM, Kuhn JH, Global Organization and Proposed Megataxonomy of the Virus World. *Microbiol. Mol. Biol. Rev* 84 (2020), doi:10.1128/MMBR.00061-19.
26. Lowey B, Whiteley AT, Keszei AFA, Morehouse BR, Mathews IT, Antine SP, Cabrera VJ, Kashin D, Niemann P, Jain M, Schwede F, Mekalanos JJ, Shao S, Lee ASY, Kranzusch PJ, 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). [PubMed: 32544385]
27. Makarova KS, Timinskas A, Wolf YI, Gussow AB, Siksnys V, Venclovas , Koonin EV, Evolutionary and functional classification of the CARF domain superfamily, key sensors in prokaryotic antiviral defense. *Nucleic Acids Res.* 48, 8828–8847 (2020). [PubMed: 32735657]
28. Heitman J, Model P, Site-specific methylases induce the SOS DNA repair response in *Escherichia coli*. *J. Bacteriol* 169, 3243–3250 (1987). [PubMed: 3036779]

29. Martin R, Qi T, Zhang H, Liu F, King M, Toth C, Nogales E, Staskawicz BJ, Structure of the activated ROQ1 resistosome directly recognizing the pathogen effector XopQ. *Science*. 370 (2020), doi:10.1126/science.abd9993.
30. Zhou M, Li Y, Hu Q, Bai X-C, Huang W, Yan C, Scheres SHW, Shi Y, Atomic structure of the apoptosome: mechanism of cytochrome c- and dATP-mediated activation of Apaf-1. *Genes Dev*. 29, 2349–2361 (2015). [PubMed: 26543158]
31. Steczkiewicz K, Muszewska A, Knizewski L, Rychlewski L, Ginalski K, Sequence, structure and functional diversity of PD-(D/E)XK phosphodiesterase superfamily. *Nucleic Acids Res*. 40, 7016–7045 (2012). [PubMed: 22638584]
32. Pingoud A, Fuxreiter M, Pingoud V, Wende W, Type II restriction endonucleases: structure and mechanism. *Cell. Mol. Life Sci* 62, 685–707 (2005). [PubMed: 15770420]
33. Watanabe N, Takasaki Y, Sato C, Ando S, Tanaka I, Structures of restriction endonuclease HindIII in complex with its cognate DNA and divalent cations. *Acta Crystallogr. D Biol. Crystallogr* 65, 1326–1333 (2009). [PubMed: 19966419]
34. Burroughs AM, Iyer LM, Aravind L, Comparative genomics and evolutionary trajectories of viral ATP dependent DNA-packaging systems. *Genome Dyn*. 3, 48–65 (2007). [PubMed: 18753784]
35. Kaur G, Iyer LM, Burroughs AM, Aravind L, Bacterial death and TRADD-N domains help define novel apoptosis and immunity mechanisms shared by prokaryotes and metazoans. *Elife*. 10 (2021), doi:10.7554/eLife.70394.
36. Kaur G, Burroughs AM, Iyer LM, Aravind L, Highly regulated, diversifying NTP-dependent biological conflict systems with implications for the emergence of multicellularity. *Elife*. 9 (2020), doi:10.7554/eLife.52696.
37. Samson JE, Magadán AH, Sabri M, Moineau S, Revenge of the phages: defeating bacterial defences. *Nat. Rev. Microbiol* 11, 675–687 (2013). [PubMed: 23979432]
38. Pawluk A, Davidson AR, Maxwell KL, Anti-CRISPR: discovery, mechanism and function. *Nat. Rev. Microbiol* 16, 12–17 (2018). [PubMed: 29062071]
39. Bondy-Denomy J, Pawluk A, Maxwell KL, Davidson AR, Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature*. 493, 429–432 (2013). [PubMed: 23242138]
40. Olsen NS, Forero-Junco L, Kot W, Hansen LH, Exploring the Remarkable Diversity of Culturable Phages in the Danish Wastewater Environment. *Viruses*. 12 (2020), doi:10.3390/v12090986.
41. Schmerer M, Molineux IJ, Bull JJ, Synergy as a rationale for phage therapy using phage cocktails. *PeerJ*. 2, e590 (2014). [PubMed: 25279269]
42. Hatfull GF, Hendrix RW, Bacteriophages and their genomes. *Current Opinion in Virology*. 1 (2011), pp. 298–303. [PubMed: 22034588]
43. Ackermann H-W, 5500 Phages examined in the electron microscope. *Arch. Virol* 152, 227–243 (2007). [PubMed: 17051420]
44. Dedeo CL, Cingolani G, Teschke CM, Portal Protein: The Orchestrator of Capsid Assembly for the dsDNA Tailed Bacteriophages and Herpesviruses. *Annu Rev Virol*. 6, 141–160 (2019). [PubMed: 31337287]
45. Casjens SR, The DNA-packaging nanomotor of tailed bacteriophages. *Nat. Rev. Microbiol* 9, 647–657 (2011). [PubMed: 21836625]
46. Danot O, Marquet E, Vidal-Ingigliardi D, Richet E, Wheel of Life, Wheel of Death: A Mechanistic Insight into Signaling by STAND Proteins. *Structure*. 17 (2009), pp. 172–182. [PubMed: 19217388]
47. Ma S, Lapin D, Liu L, Sun Y, Song W, Zhang X, Logemann E, Yu D, Wang J, Jirschtzka J, Han Z, Schulze-Lefert P, Parker JE, Chai J, Direct pathogen-induced assembly of an NLR immune receptor complex to form a holoenzyme. *Science*. 370 (2020), doi:10.1126/science.abe3069.
48. Morehouse BR, Govande AA, Millman A, Keszei AFA, Lowey B, Ofir G, Shao S, Sorek R, Kranzusch PJ, STING cyclic dinucleotide sensing originated in bacteria. *Nature*. 586, 429–433 (2020). [PubMed: 32877915]
49. Shmakov SA, Makarova KS, Wolf YI, Severinov KV, Koonin EV, 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). [PubMed: 29784811]

50. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402 (1997). [PubMed: 9254694]
51. Steinegger M, Söding J, MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol* 35, 1026–1028 (2017). [PubMed: 29035372]
52. Edgar RC, MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics.* 5, 113 (2004). [PubMed: 15318951]
53. Steinegger M, Meier M, Mirdita M, Vöhringer H, Haunsberger SJ, Söding J, HH-suite3 for fast remote homology detection and deep protein annotation,, doi:10.1101/560029.
54. Makarova KS, Wolf YI, Iranzo J, Shmakov SA, Alkhnbashi OS, Brouns SJJ, Charpentier E, Cheng D, Haft DH, Horvath P, Moineau S, Mojica FJM, Scott D, Shah SA, Siksnys V, Terns MP, Venclovas , White MF, Yakunin AF, Yan W, Zhang F, Garrett RA, Backofen R, van der Oost J, Barrangou R, Koonin EV, Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol* 18, 67–83 (2020). [PubMed: 31857715]
55. Price MN, Dehal PS, Arkin AP, FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One.* 5, e9490 (2010). [PubMed: 20224823]
56. Katoh K, Misawa K, Kuma K-I, Miyata T, MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30, 3059–3066 (2002). [PubMed: 12136088]
57. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ, IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol* 32, 268–274 (2015). [PubMed: 25371430]
58. Zimmermann L, Stephens A, Nam S-Z, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas AN, Alva V, A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J. Mol. Biol* 430, 2237–2243 (2018). [PubMed: 29258817]
59. Letunic I, Bork P, Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49, W293–W296 (2021). [PubMed: 33885785]
60. Van de Weyer A-L, Monteiro F, Furzer OJ, Nishimura MT, Cevik V, Witek K, Jones JDG, Dangl JL, Weigel D, Bemm F, A Species-Wide Inventory of NLR Genes and Alleles in *Arabidopsis thaliana*. *Cell.* 178, 1260–1272.e14 (2019). [PubMed: 31442410]
61. Dyrka W, Lamacchia M, Durrens P, Kobe B, Daskalov A, Paoletti M, Sherman DJ, Saupe SJ, Diversity and variability of NOD-like receptors in fungi. *Genome Biol. Evol* 6, 3137–3158 (2014). [PubMed: 25398782]
62. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D, Highly accurate protein structure prediction with AlphaFold. *Nature.* 596, 583–589 (2021). [PubMed: 34265844]
63. Pei J, Kim BH, Grishin NV, PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res.* 36 (2008), doi:10.1093/nar/gkn072.
64. Picelli S, Björklund AK, Reinius B, Sagasser S, Winberg G, Sandberg R, Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.* 24, 2033–2040 (2014). [PubMed: 25079858]
65. Huber M, Faure G, Laass S, Kolbe E, Seitz K, Wehrheim C, Wolf YI, Koonin EV, Soppa J, Translational coupling via termination-reinitiation in archaea and bacteria. *Nat. Commun* 10, 4006 (2019). [PubMed: 31488843]
66. Qimron U, Marintcheva B, Tabor S, Richardson CC, Genomewide screens for *Escherichia coli* genes affecting growth of T7 bacteriophage. *Proc. Natl. Acad. Sci. U. S. A* 103, 19039–19044 (2006). [PubMed: 17135349]
67. Grigonyte AM, Harrison C, MacDonald PR, Montero-Blay A, Tridgett M, Duncan J, Sagona AP, Constantinidou C, Jaramillo A, Millard A, Comparison of CRISPR and Marker-Based Methods for the Engineering of Phage T7. *Viruses.* 12 (2020), doi:10.3390/v12020193.

68. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H, Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol* 2, 2006.0008 (2006).
69. Kimanius D, Dong L, Sharov G, Nakane T, Scheres SHW, New tools for automated cryo-EM single-particle analysis in RELION-4.0. *Biochem. J* 478, 4169–4185 (2021). [PubMed: 34783343]
70. Bepler T, Morin A, Rapp M, Brasch J, Shapiro L, Noble AJ, Berger B, Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nat. Methods* 16, 1153–1160 (2019). [PubMed: 31591578]
71. Croll TI, ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr D Struct Biol.* 74, 519–530 (2018). [PubMed: 29872003]
72. Casañal A, Lohkamp B, Emsley P, Current developments in Coot for macromolecular model building of Electron Cryo-microscopy and Crystallographic Data. *Protein Sci.* 29, 1069–1078 (2020). [PubMed: 31730249]
73. Liebschner D, Afonine PV, Baker ML, Bunkóczi G, Chen VB, Croll TI, Hintze B, Hung LW, Jain S, McCoy AJ, Moriarty NW, Oeffner RD, Poon BK, Prisant MG, Read RJ, Richardson JS, Richardson DC, Sammito MD, Sobolev OV, Stockwell DH, Terwilliger TC, Urzhumtsev AG, Videau LL, Williams CJ, Adams PD, Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol.* 75, 861–877 (2019). [PubMed: 31588918]
74. Wong N, John S, Nussenzweig A, Canela A, 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). [PubMed: 32840769]
75. Holm L, Using Dali for Protein Structure Comparison. *Methods Mol. Biol* 2112, 29–42 (2020). [PubMed: 32006276]
76. Wang J, Hu M, Wang J, Qi J, Han Z, Wang G, Qi Y, Wang H-W, Zhou J-M, Chai J, Reconstitution and structure of a plant NLR resistosome conferring immunity. *Science.* 364 (2019), doi:10.1126/science.aav5870.
77. Pang Y, Bai X-C, Yan C, Hao Q, Chen Z, Wang J-W, Scheres SHW, Shi Y, Structure of the apoptosome: mechanistic insights into activation of an initiator caspase from *Drosophila*. *Genes Dev.* 29, 277–287 (2015). [PubMed: 25644603]
78. Zhang L, Chen S, Ruan J, Wu J, Tong AB, Yin Q, Li Y, David L, Lu A, Wang WL, Marks C, Ouyang Q, Zhang X, Mao Y, Wu H, Cryo-EM structure of the activated NAIP2-NLRC4 inflammasome reveals nucleated polymerization. *Science.* 350, 404–409 (2015). [PubMed: 26449474]
79. Tenthorey JL, Haloupek N, López-Blanco JR, Grob P, Adamson E, Hartenian E, Lind NA, Bourgeois NM, Chacón P, Nogales E, Vance RE, The structural basis of flagellin detection by NAIP5: A strategy to limit pathogen immune evasion. *Science.* 358, 888–893 (2017). [PubMed: 29146805]
80. Esterman ES, Wolf YI, Kogay R, Koonin EV, Zhaxybayeva O, Evolution of DNA packaging in gene transfer agents. *Virus Evol.* 7, veab015 (2021). [PubMed: 33732503]
81. Sun S, Kondabagil K, Draper B, Alam TI, Bowman VD, Zhang Z, Hegde S, Fokine A, Rossmann MG, Rao VB, The structure of the phage T4 DNA packaging motor suggests a mechanism dependent on electrostatic forces. *Cell.* 135, 1251–1262 (2008). [PubMed: 19109896]
82. Cuervo A, Fàbrega-Ferrer M, Machón C, Conesa JJ, Fernández FJ, Pérez-Luque R, Pérez-Ruiz M, Pous J, Vega MC, Carrascosa JL, Coll M, Structures of T7 bacteriophage portal and tail suggest a viral DNA retention and ejection mechanism. *Nat. Commun* 10, 3746 (2019). [PubMed: 31431626]

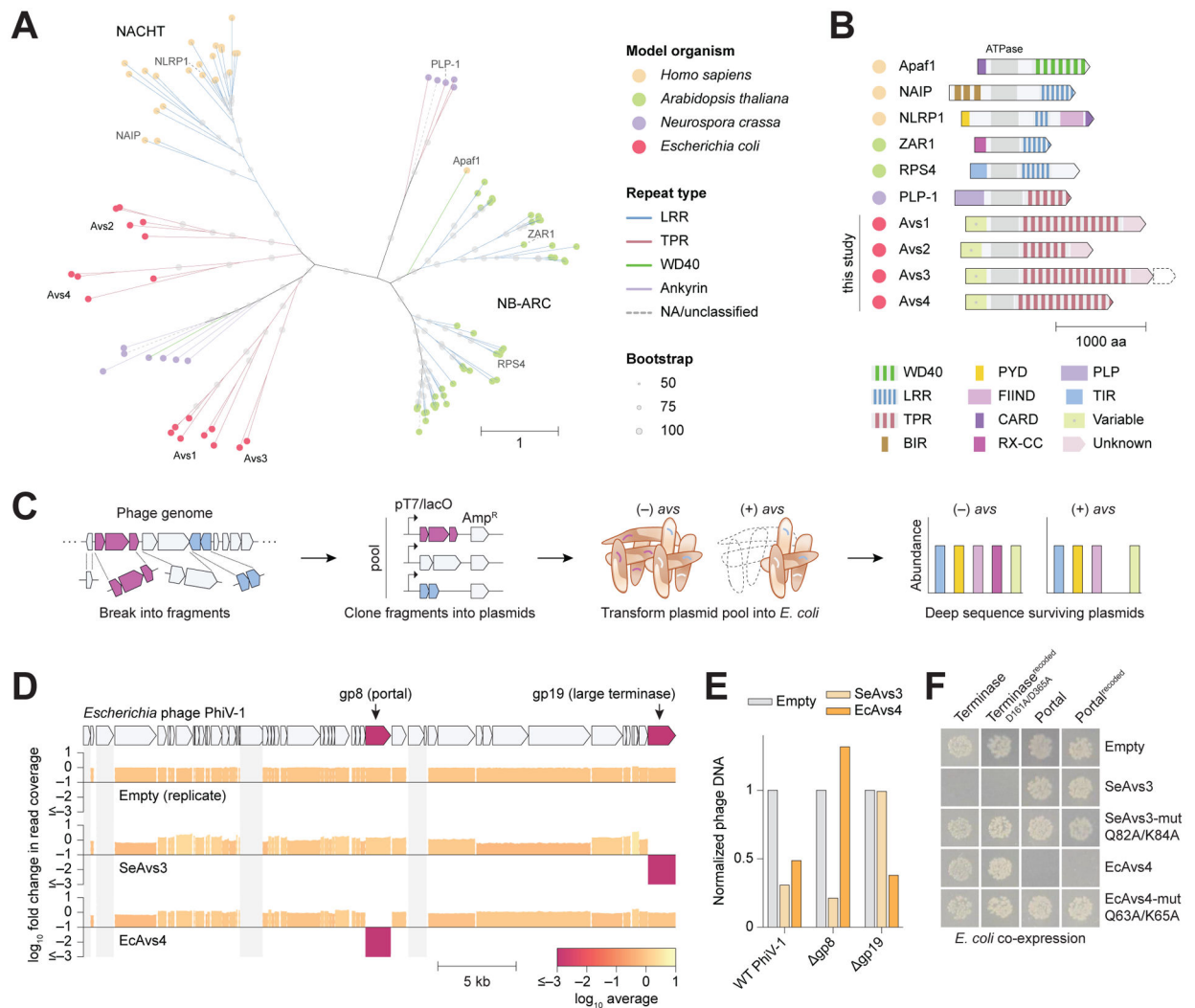
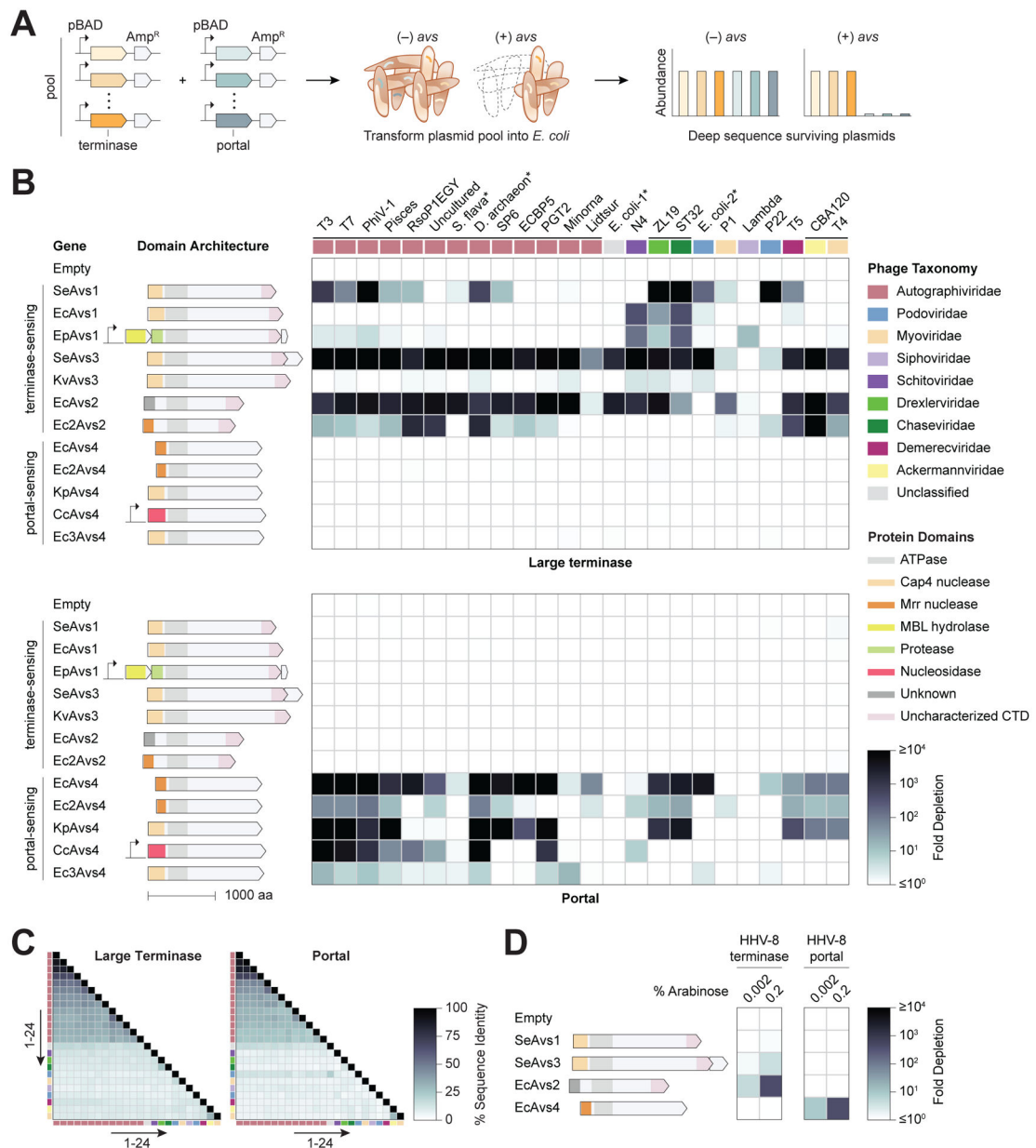


Fig. 1. Prokaryotic STAND NTPases recognize phage terminase and portal proteins. **(A)** Maximum likelihood tree of the ATPase domain of selected NLR-like STAND NTPases in four model organisms across kingdoms of life. **(B)** Domain architectures of representative NLR-like genes in (A). LRR, leucine-rich repeat; TPR, tetratricopeptide repeat; WD40, WD40 repeat; ankyrin, ankyrin repeat; BIR, baculoviral inhibitor of apoptosis repeat; PYD, pyrin domain; FIIND, function to find domain; CARD, caspase activation and recruitment domain; RX-CC, potato virus X resistance protein coiled-coil domain; PLP, patatin-like phospholipase; TIR, toll/interleukin-1 receptor homology domain. **(C)** Schematic of genetic screening approach to identify phage-encoded activators of Avs proteins that induce cell death. **(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*. **(F)** Photographs of *E. coli* co-transformation assays with Avs genes and phage activators identified in (D).

**Fig. 2.**

Avs proteins are pattern-recognition receptors for the terminase and portal of diverse tailed phages. (A) Schematic of 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 bars indicate groups of terminase proteins with at least 20% pairwise sequence identity. Asterisks indicate prophages. *S. flava*, *Shingopyxis flava* R11H; *D. archaeon*, *Desulfurococcales archaeon* ex4484_217_2; *E. coli-1*, *Escherichia coli* NCTC9020; *E. coli-2*, *Escherichia coli* M885. (C) Pairwise amino acid sequence identity between the core folds of the terminases and portals in (B), excluding

non-conserved regions. **(D)** Activity of four Avs proteins against the human herpesvirus 8 (HHV-8) terminase and portal.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

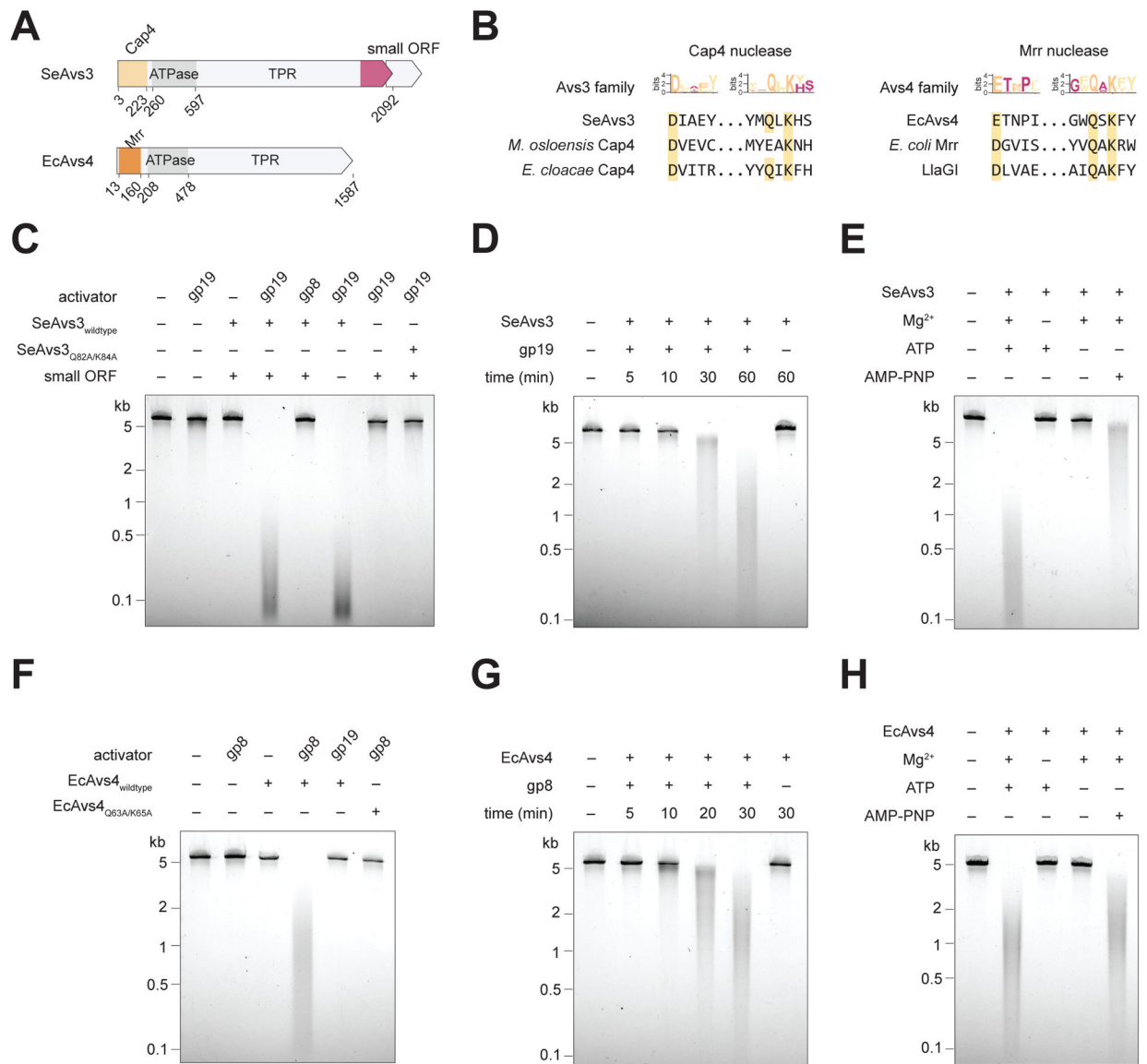


Fig. 3. SeAvs3 and EcAvs4 are phage-activated DNA endonucleases. **(A)** Domain architecture of SeAvs3 and EcAvs4. **(B)** Alignment of Avs D-QxK nuclease motifs with characterized Cap4 and Mrr representatives. **(C, D)** Agarose gel analysis of SeAvs3 nuclease function in vitro with a linear dsDNA substrate and **(E)** cofactor requirements. **(F, G)** Agarose gel analysis of EcAvs4 nuclease function in vitro with a linear dsDNA substrate and **(H)** cofactor requirements.

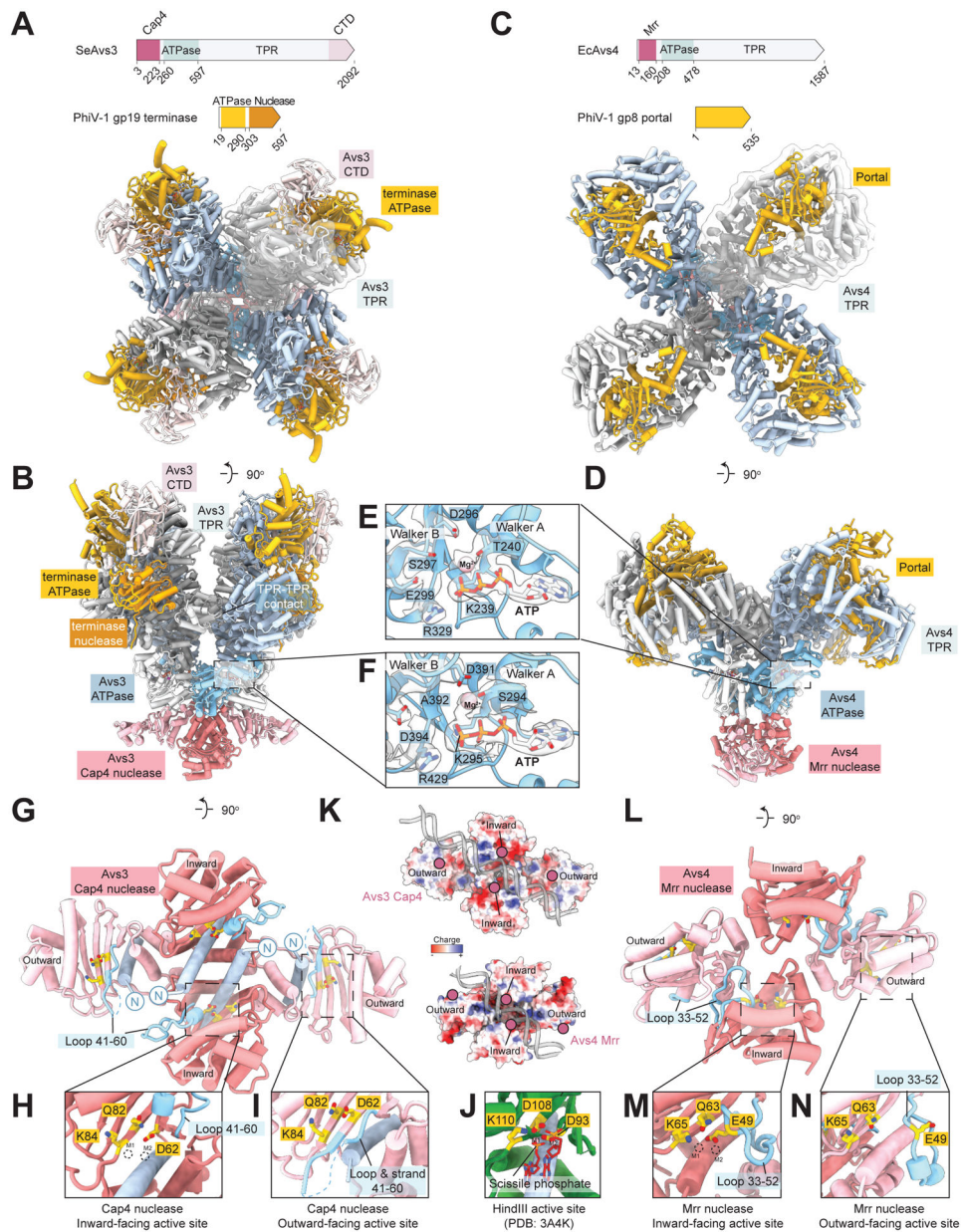


Fig. 4. Cryo-EM structures of SeAvs3 and EcAvs4 in complex with their cognate triggers. **(A, B)** Structure of the SeAvs3-terminase complex. **(C, D)** Structure of the EcAvs4-portal complex. **(E, F)** ATP molecule in the STAND ATPase active site of EcAvs4 and SeAvs3. Cryo-EM density is shown as a transparent surface. **(G)** SeAvs3 Cap4-like nuclease effector domain. **(H, I)** Active sites for the inward and outward facing protomers of the SeAvs3 Cap4-like nuclease. **(J)** Equivalent view of the active site of HindIII bound to target DNA with two divalent metal ions (PDB 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 HindIII bound to its target (PDB 3A4K). **(L)** EcAvs4 Mrr-like nuclease effector domain.

(M, N) Active sites for the inward and outward facing protomers of the EcAys4 Mrr-like nuclease. TPR, tetratricopeptide repeat.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

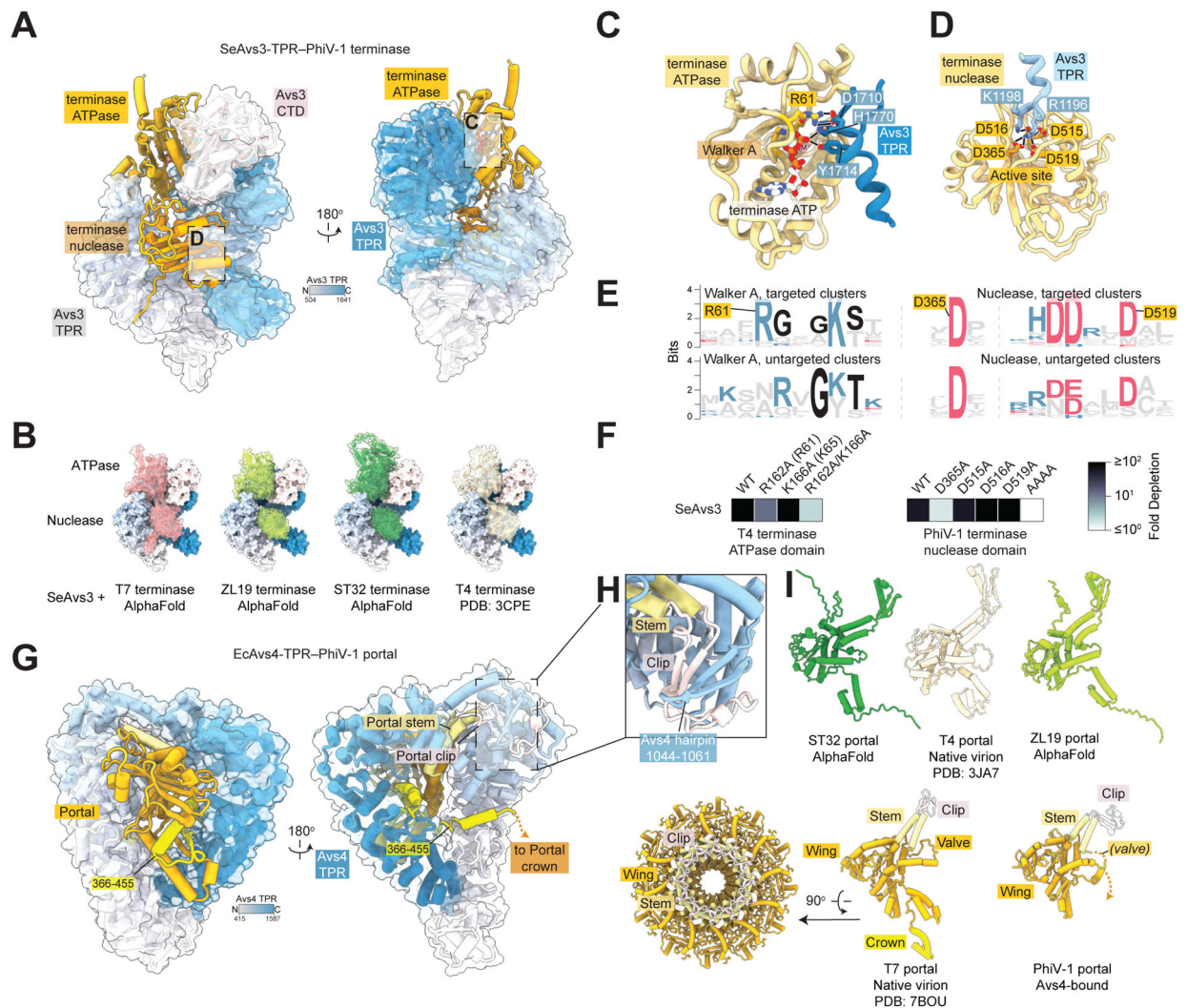


Fig. 5. Structural basis for viral fold recognition by SeAvs3 and EcAvs4. **(A)** The interface between SeAvs3 and the PhiV-1 terminase. An SeAvs3 surface view is shown in transparency. SeAvs3 is colored from N to C terminus according to the key. **(B)** AlphaFold or crystal structures of different terminases modeled into SeAvs3. The ATPase and nuclease domains were individually aligned to the PhiV-1 terminase domains. **(C, 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 co-expressed 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 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. TPR, tetratricopeptide repeat; CTD, C-terminal domain.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

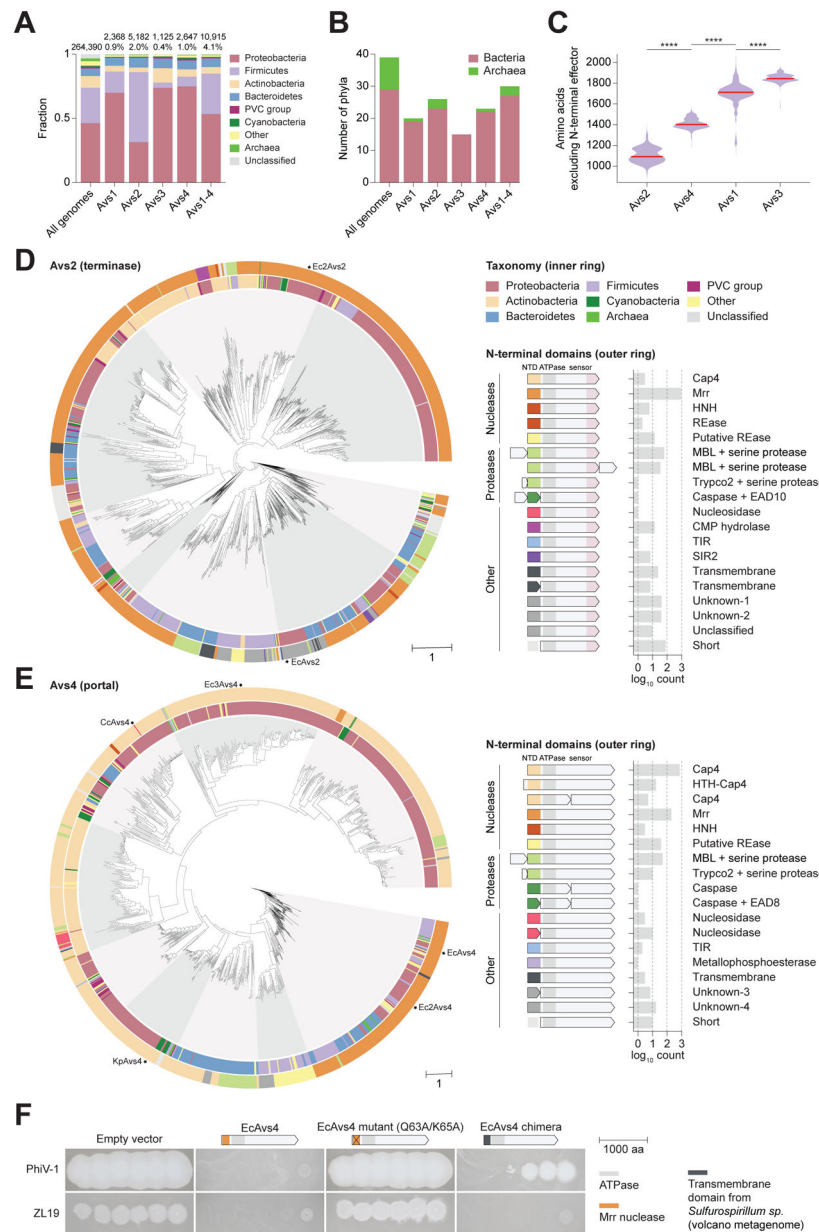


Fig. 6. Taxonomic distribution and domain architectures of Avs families. **(A)** Distribution of *avs* genes across phyla. The values above the bars indicate the number and percentage of genomes containing each gene. **(B)** Number of bacterial and archaeal phyla (minimum 100 sequenced isolates) with 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). Maximum likelihood tree of representatives of the ATPase + C-terminal domain of **(D)** Avs2 terminase sensors ($n = 1,255$) and **(E)** Avs4 portal sensors ($n = 1,089$) 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. MBL, metallo- β -lactamase; REase, restriction

endonuclease; TIR, Toll/interleukin-1 receptor homology domain; SIR2, sirtuin; CMP, cytidine monophosphate; HTH, helix-turn-helix. (F) Anti-phage defense activity of a chimeric Avs4 with transmembrane N-terminal helices from *Sulfurospirillum sp.* replacing the nuclease domain of EcAvs4.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

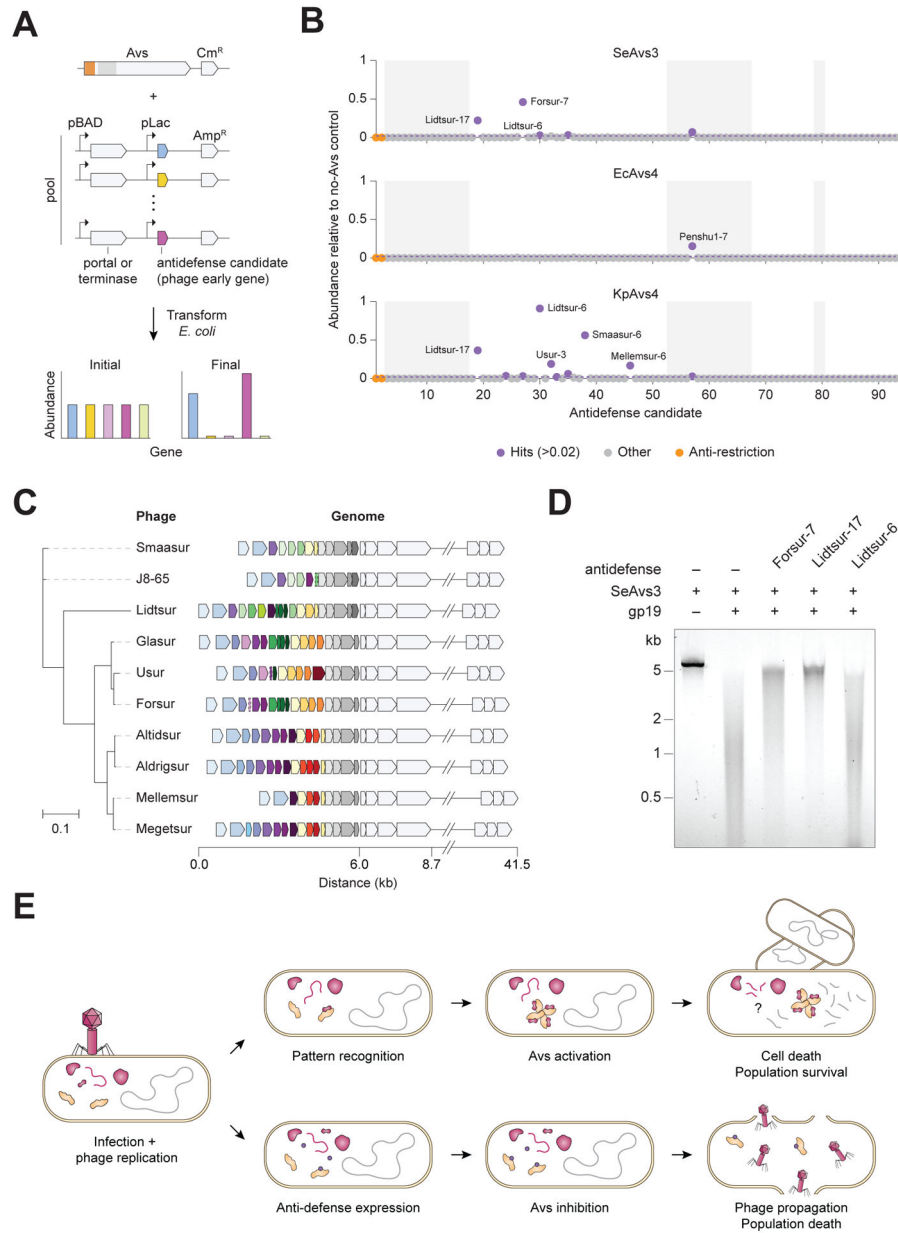


Fig. 7: Phage-encoded genes inhibit Avs activity. **(A)** Schematic of a pooled screen in *E. coli* for phage early genes that rescue Avs-mediated toxicity. **(B)** Deep sequencing readout of anti-defense candidate genes co-expressed with SeAvs3, EcAvs4, or KpAvs4. **(C)** A hypervariable early gene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