A bacterial NLR-related protein recognizes multiple unrelated phage triggers to sense infection

Emily M. Kibby[1], Laurel K. Robbins[1,2], Amar Deep[3], Nathan K. Min[3], Lindsay A. Whalen[1], Toni A. Nagy[1], Layla Freeborn[4], Kevin D. Corbett[3,5], Aaron T. Whiteley[1]*

[1] Department of Biochemistry, University of Colorado Boulder, Boulder, CO, USA
[2] Interdisciplinary Quantitative Biology Program (IQ Biology), BioFrontiers Institute, University of Colorado Boulder, Boulder, CO, USA
[3] Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, California, USA
[4] Research Computing, Office of Information Technology, University of Colorado Boulder, Boulder, CO, USA
[5] Department of Molecular Biology, University of California, San Diego, La Jolla, California, USA
*To whom correspondence should be addressed: aaron.whiteley@colorado.edu

Immune systems must rapidly sense viral infections to initiate antiviral signaling and protect the host. Bacteria encode >100 distinct viral (phage) defense systems and each has evolved to sense crucial components or activities associated with the viral lifecycle. Here we used a high-throughput AlphaFold-multimer screen to discover that a bacterial NLR-related protein directly senses multiple phage proteins, thereby limiting immune evasion. Phages encoded as many as 5 unrelated activators that were predicted to bind the same interface of a C-terminal sensor domain. Genetic and biochemical assays confirmed activators bound to the bacterial NLR-related protein at high affinity, induced oligomerization, and initiated signaling. This work highlights how *in silico* strategies can identify complex protein interaction networks that regulate immune signaling across the tree of life.

Keywords: phage defense, sensing, AlphaFold-multimer, STAND, NACHT, AVAST system

Introduction

Bacteria, like eukaryotes and archaea, have evolved diverse molecular strategies to protect themselves against viruses. In bacteria these are called phage defense systems and each typically encodes sensor and effector components to allow the system to recognize and respond to bacteriophage (phage) infection. Recent studies have identified many unrecognized defense systems, but a major gap in our understanding is the mechanism by which each senses infection[1–5]. This gap is potentiated by technical limitations of common genetic and biochemical approaches. Isolation of phage mutants that escape defense systems often identifies genes that indirectly impact sensing or phage-encoded inhibitors but can miss essential phage genes or redundant activators. Mass spectrometry-based experiments often suffer from high background, making it difficult to distinguish relevant signals from noise. This can be particularly challenging for activators that bind with lower affinity or are expressed at low levels.

Protein structure prediction software offers an opportunity to employ computational strategies to predict protein-protein interactions and understand how defense systems sense infection. Here, we used high-throughput AlphaFold-multimer screening to identify phage proteins that are recognized by a phage defense system[6,7]. This software package uses amino acid sequence, multiple sequence alignments, structural homology, and machine learning models to enable protein structure prediction based on input amino acid sequences[7]. The AlphaFold-multimer model incorporates species-of-origin information and other parameters to enable *in silico* prediction of structures for multichain complexes and protein-protein interactions[6]. The multimer model uses a weighted predicted template model score (pTM) that is calculated to measure interchain contacts and quantify the confidence of a predicted interaction[6,7], meaning a higher weighted pTM score represents a more confidently predicted interaction between the proteins in the model.

We sought to understand phage protein sensing by bacterial NACHT domain-containing (bNACHT) proteins. NACHT domains are P-loop NTPases that are part of the large and diverse family of STAND NTPases. In bacteria, many STAND NTPases are antiphage and are broadly named AVAST or Avs systems[3,8,9]. bNACHT defense systems are related to the well-characterized mammalian NLR immune proteins that form inflammasomes, other mammalian NACHT domain containing proteins involved in immune signaling such as CIITA, and fungal NLR-like immune proteins[9–14]. AVAST systems and bNACHT proteins are also distantly related to plant NLRs[15,16]. bNACHT systems are potently antiphage and widely distributed in the genomes of bacteria but have poorly characterized mechanisms for activation. We screened a series of bNACHT proteins for interactions with phage proteins using AlphaFold-multimer and identified that a bNACHT C-terminal domain is an exceptionally versatile sensor of many activators.

Results

An AlphaFold-multimer screen predicts phage proteins that bind bNACHT11

The bNACHT11 protein from *Klebsiella pneumoniae* provides robust defense against phages T2, T4, and T6, but the phage sensing mechanism is unknown (**Fig. 1B**)[9]. Prior characterization of bNACHT11 and closely related defense proteins revealed that bNACHT11 encodes a new type of pattern recognition domain

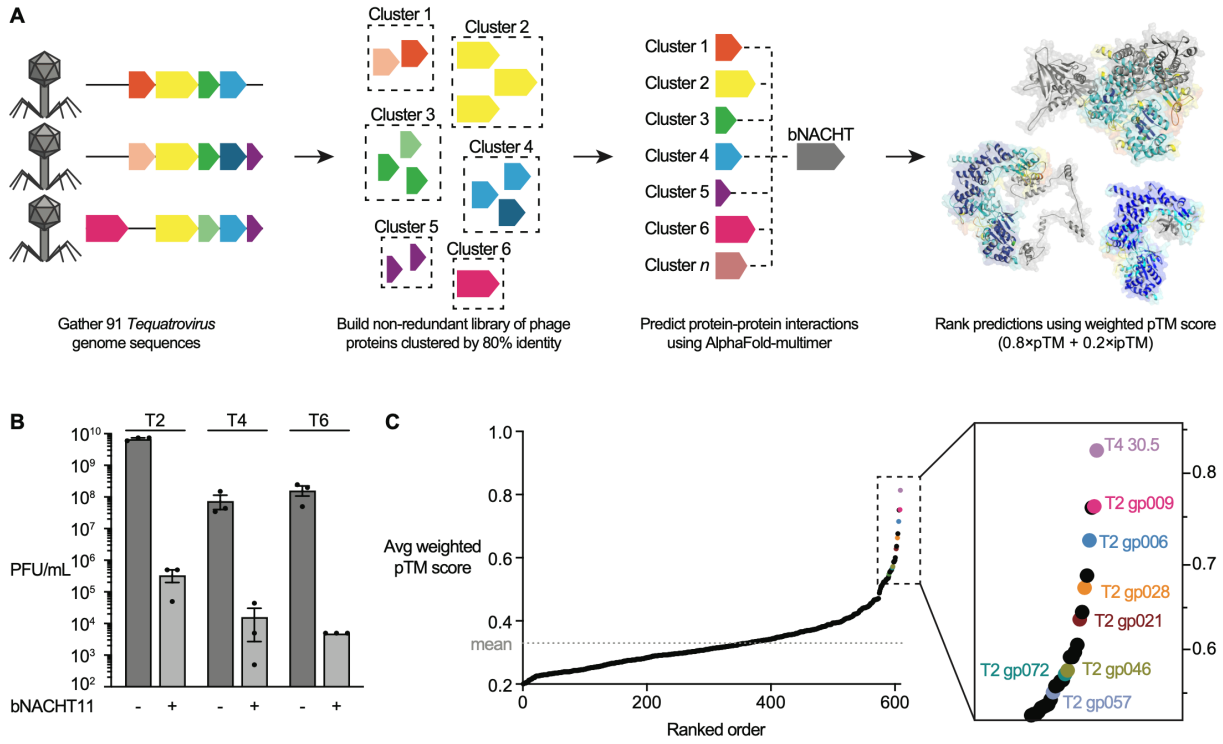

Figure 1. An *in silico* screen identifies phage proteins predicted to bind bNACHT11.
(A) Schematic of the workflow for an *in silico* screen to identify phage-encoded activators of clade 14 bNACHT proteins (see Methods). Proteins encoded by 91 phages in the *Tequatrovirus* family were clustered using MMSeq2 [20,21] and 629 representative proteins were then screened using AlphaFold-multimer to predict interactions with bNACHT proteins of interest. See **table S1** for a list of all proteins used to generate the clusters, and a final list of the cluster representatives. **(B)** Efficiency of plating of indicated phages infecting *E. coli* expressing bNACHT11 (+) or an empty vector (-). Data represent the mean ± standard error of the mean (SEM) of n = 3 biological replicates, shown as individual points. **(C)** Results of AlphaFold-multimer screening for bNACHT11 (WP_114260439.1). Average weighted pTM score (ipTM×0.8 + pTM×0.2) was calculated for each protein screened, then ranked such that the highest-scoring proteins are on the right of the graph (inset). High-scoring clusters are labeled with the name of the phage T2 or T4 protein in that cluster (T4 CDS 30.5: NP_049819.1; T2 gp009: YP_010073657.1; T2 gp006: YP_010073654.1; T2 gp028: YP_010073676.1; T2 gp021: YP_010073669.1; T2 gp046: YP_010073694.1; T2 gp072: YP_010073720.1; T2 gp057: YP_010073705.1). See **table S2** for the data used to generate this figure.

called a Short NACHT-associated C-terminal (SNaCT) domain. SNaCT domains are widespread in bacteria, rapidly evolving, and predicted to be crucial for phage sensing[9]. To understand bNACHT11 activators and the SNaCT domain, we used the AlphaPulldown and AlphaFold-multimer software packages to enable high-throughput computational screening for protein-protein interactions between a representative set of phage-encoded proteins and bNACHT11 (**Fig. 1A**, **table S1 and S2**, Methods)[6,7,17].

Phages T2, T4, and T6 are "T-even" phages and the best characterized members of the double-stranded DNA phage genus *Tequatrovirus*[18,19]. To capture the diversity of proteins encoded by this genus, we used the MMSeq2 sequence-based software to build a nonredundant library of clusters of phage proteins that represents the pangenome of 91 different *Tequatrovirus* members[20,21]. To decrease the complexity of this library, we excluded proteins that were shorter than 50 amino acids or found in only one genome, which generated a final list of 629 representative phage proteins (**table S1**, Methods).

We then used AlphaFold-multimer[6] to predict interactions between each representative *Tequatrovirus* protein and bNACHT11. 15 predicted structures were computed for each interaction and the weighted pTM scores were averaged. This *in silico* screen identified several phage-encoded proteins that were confidently predicted to interact with bNACHT11 (**Fig. 1C**).

Multiple phage proteins bind and activate bNACHT11

We selected several proteins encoded by phage T2 and/or T4 as representatives of the highest scoring *Tequatrovirus* protein clusters and used a genetic assay to interrogate their ability to activate bNACHT11. We co-expressed genes from phage T2 (NCBI genome ID: NC_054931.1, *gp006*, *gp009*, *gp028*, *gp021*, *gp046*, *gp072*, *gp057*) and T4 (NCBI genome ID: AF158101.1, gene *30.5*) using an inducible vector with bNACHT11 under its endogenous promoter in *E. coli* MG1655. We and others have reported that AVAST and bNACHT defense systems typically initiate abortive infection in response to phage and activation of bNACHT11-like proteins inhibited bacterial growth[4,8,9,14,22]. Therefore we assayed for colony formation as a readout for bNACHT activation. Coexpression of bNACHT11 with T2 *gp006*, T2 *gp009*, *T2gp028*, or T2 *gp046* resulted in a large decrease in colony formation, suggesting that these proteins activate bNACHT11 (**Fig. 2A**, **fig. S1I**). Subsequent analyses revealed that the potent inhibition of colony formation for some

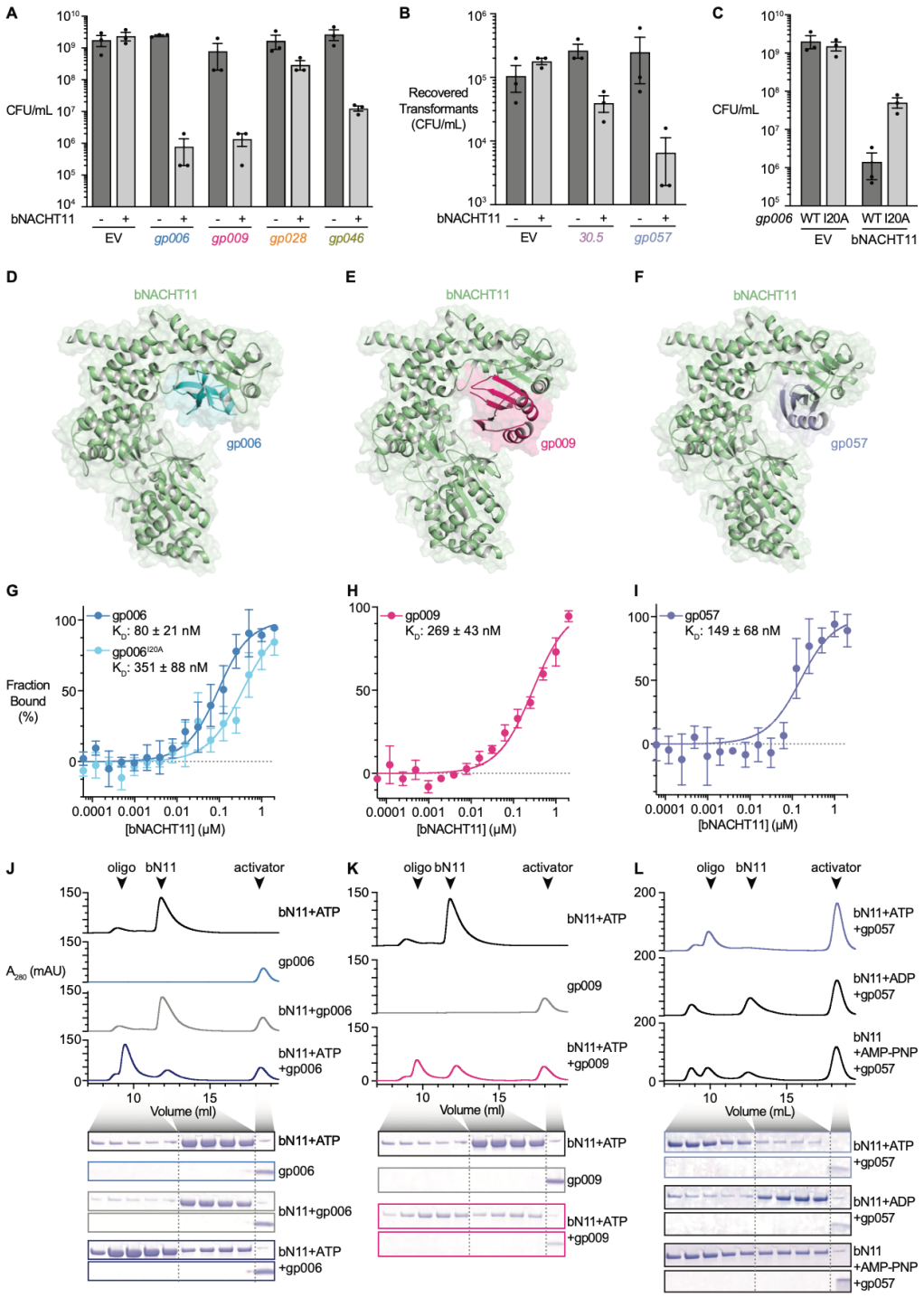

Figure 2. Multiple phage proteins bind and activate bNACHT11.
(A) Quantification of colony formation of *E. coli* expressing an Empty Vector (-) or bNACHT11 (+) on one plasmid and *gfp* (EV) or indicated phage protein on a second plasmid. The expression of *gfp* or phage proteins is IPTG-inducible. See **fig. S1** for predicted structures and PAE plots of proteins tested. **(B)** Quantification of recovered transformants of *E. coli* cotransformed with plasmids expressing an Empty Vector (-) or bNACHT11 (+) and a *gfp* empty vector (EV) or the indicated phage protein. The expression of *gfp* and phage proteins is IPTG-inducible. For A and B, experiments were performed using 500 μM IPTG induction in LB media. **(C)** Quantification of colony formation of *E. coli* expressing wild-type *gp006* (WT) or *gp006*^{I20A} (I20A) with a C-terminal VSV-G tag on one plasmid and empty vector (EV) or bNACHT11 on a second plasmid. For A–C, data represent the mean ± SEM of n = 3 biological replicates, shown as individual points. **(D–F)** AlphaFold-multimer predicted structure of bNACHT11 (green) binding the indicated phage protein. **(G–I)** Binding curve of purified His-tagged gp006 and gp006^{I20A} (G), gp009 (H), or gp057 (I) binding to purified bNACHT11, measured using microscale thermophoresis (MST). Data represent the average of n = 4 (G) or n = 3 (H and I) individual replicates ± SEM. See **fig. S2** for protein purification gels, untransformed individual replicates, and binding measurements in high salt conditions. **(J–L)** Above: Traces showing the absorbance at 280 nM (A_{280}) of the indicated samples being run over a size exclusion (SEC) column. Labels indicate the identity of proteins in each peak. Below: Coomassie staining of representative samples taken from in between the indicated volumes. For samples containing both bNACHT11 and an activator, portions of the gel corresponding to each are shown. Top: bNACHT11. Bottom: activator. Data in J are representative of n = 3 replicates. Data in K are representative of n = 1 replicates. Data in L are representative of n = 2 replicates.

plasmid pairs selected for mutations during strain construction (data not shown). We further tested plasmid pairs by cotransforming into *E. coli* and plating directly on inducing vs. non-inducing conditions. Expression of T4 *30.5* and T2 *gp057* with bNACHT11 resulted in fewer recovered transformants than phage proteins alone, suggesting that these proteins are also activators of bNACHT11 (**Fig. 2B**, **fig. S1J**).

To extend the results of our genetic assay, we purified recombinant bNACHT11 and three candidate activators (T2 gp006, T2 gp009, and T2 gp057), and measured the affinity of protein binding. We found that bNACHT11 binds gp006 with a K_D of 80 ± 21 nM, T2 gp009 with a K_D of 269 ± 43 nM, and T2 gp057 with a K_D of 149 ± 68 nM (**Fig. 2G–I**, **fig. S2I–L**). We hypothesized that this binding was dominated by hydrophobic interactions due to the number of hydrophobic residues at the predicted binding interface. To test whether ionic contacts were important in this interaction, we performed binding experiments using bNACHT11 and T2 gp006 under high salt conditions (1 M KCl) and found that the K_D of this interaction was not disrupted (**fig. S2G**, **H**).

Previous investigations have demonstrated that human NLRs are activated by directly binding a ligand, leading to a conformational change in the core NACHT domain that causes oligomerization[23,24]. The conformational change requires NTP binding. We hypothesized that the phage proteins identified in our screen may similarly lead to oligomerization upon binding bNACHT11. To test this hypothesis, we used size exclusion chromatography to detect the formation of higher molecular weight complexes in samples containing bNACHT11, T2 gp006, T2 gp009, T2 gp057, and/or NTP. We found that incubation of bNACHT11 with ATP and each of these activators resulted in the formation of a high molecular weight complex (**Fig. 2J–L**). Investigation of the complex showed this peak contained primarily the bNACHT11 protein. This data suggests that bNACHT11 oligomerizes upon activator binding, in a manner analogous to the formation of the inflammasome in humans or resistosome in plants[24,25]. The formation of this oligomer required ATP binding, but not hydrolysis, as incubating bNACHT11 with the nonhydrolyzable ATP analog AMP-PNP and T2 gp057 resulted in oligomerization (**Fig. 2L**). Using molecular weight standards, we estimate that this complex consists of 6–9 bNACHT11 monomers (**fig. S3N**).

To confirm the predicted binding interface of bNACHT11 and T2 gp006, we mutated isoleucine 20 in T2 gp006 to an alanine. This residue is predicted to interact with a hydrophobic interface on bNACHT11. We found that the gp006^{I20A} mutation resulted in decreased growth inhibition when coexpressed with bNACHT11 (**Fig. 2C**). We also found that purified T2 gp006^{I20A} had a reduced affinity for bNACHT11, with a K_D of 351 ± 88 nM (**Fig. 2G** and **S2J**). This data suggests that the predicted binding interface between bNACHT11 and its activators is likely accurate, but that additional residues play a role in T2 gp006 binding to bNACHT11.

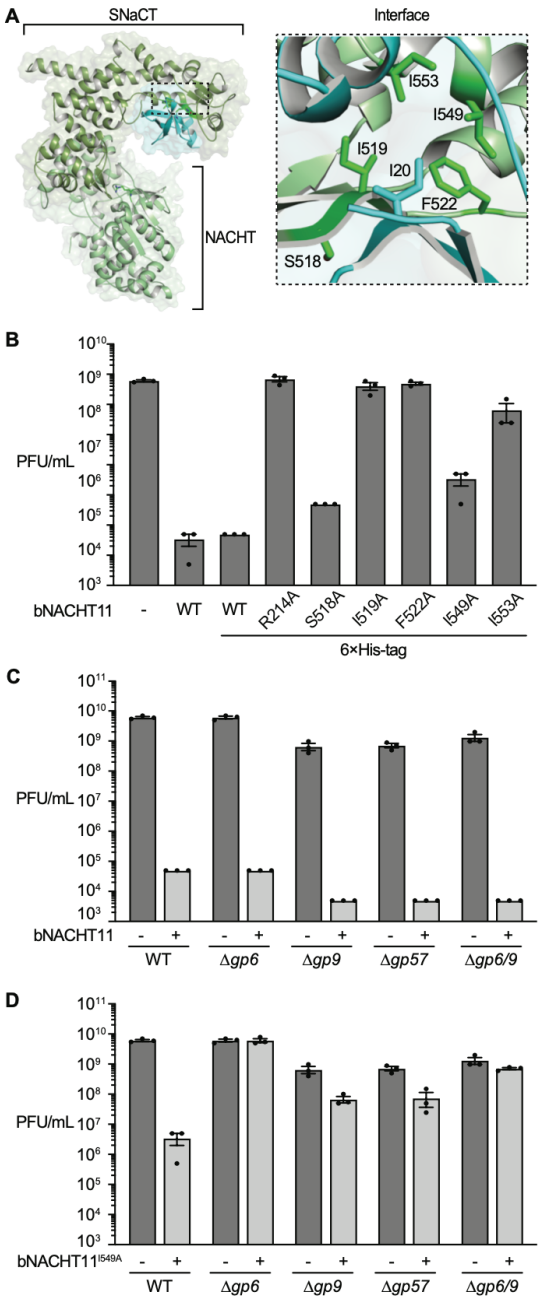

Figure 3. bNACHT11 SNaCT domain mutants allow phage escape. **(A)** Above: Zoomed in view of the AlphaFold-multimer predicted binding interface between bNACHT11 (green) and gp006 (cyan). Residues predicted to be involved in the interaction are labeled and shown as stick models. Below: Domain organization of bNACHT11, with lines indicating the position of mutated residues. **(B)** Efficiency of plating of phage T2 infecting *E. coli* expressing bNACHT11 of the indicated genotype. See **fig. S3** for phage T4 and T6 efficiency of plating. **(C)** Efficiency of plating of wild-type (WT) or mutant (Δ) T2 phages infecting *E. coli* expressing bNACHT11-6×His (+) or an empty vector (-). **(D)** Efficiency of plating for a panel of phage T2 mutants infecting *E. coli* expressing bNACHT11^{I549A}-6×His (+) or an empty vector (EV). For B, C, and D, data indicate the mean ± SEM of n = 3 biological replicates, shown as individual points.

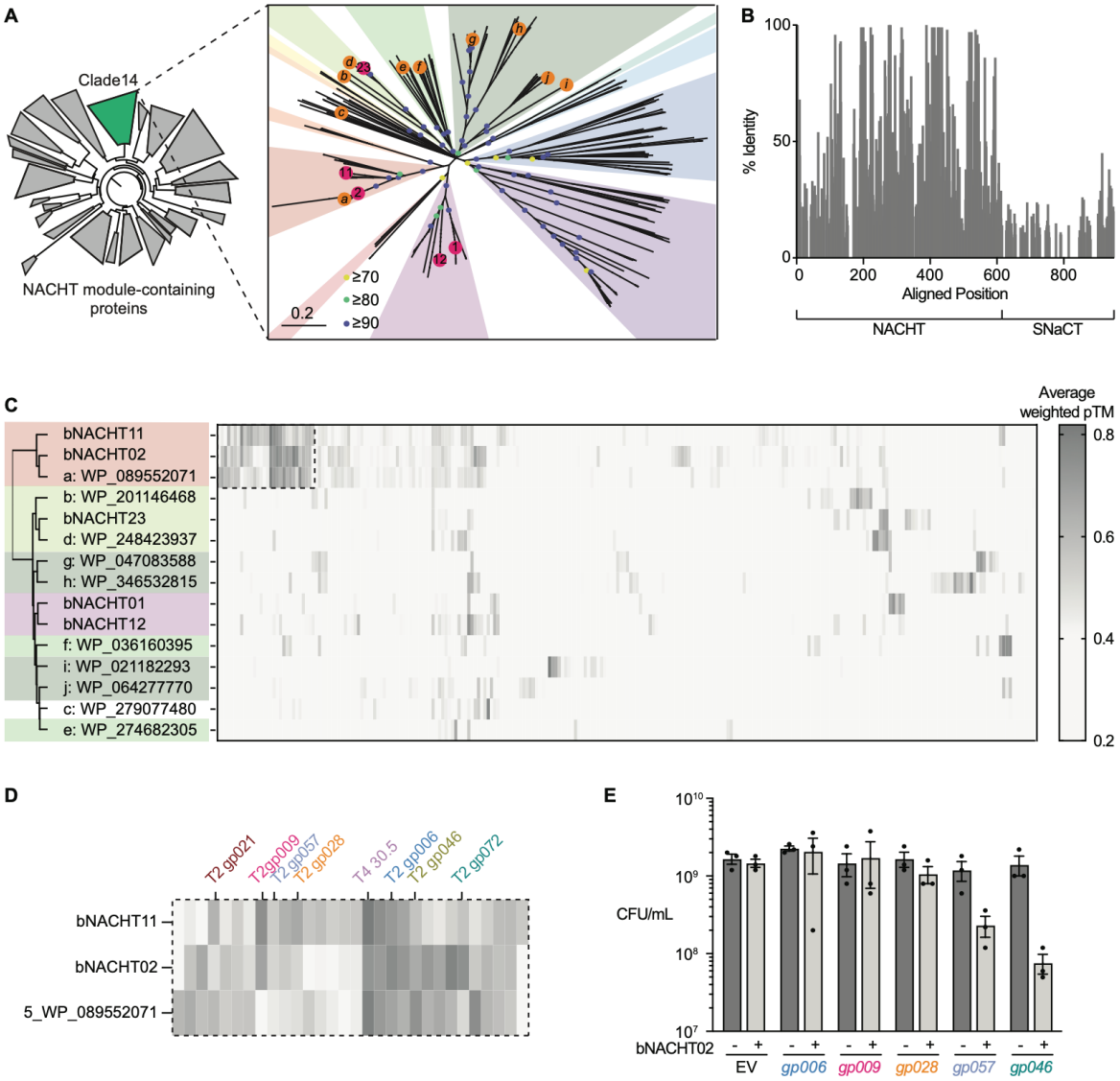

Figure 4. *In silico* screening of clade 14 bNACHT proteins reveals specificity features.
(A) Left: A sequence-based phylogenetic tree of NACHT modules generated using NACHT module-containing proteins from eukaryotes and prokaryotes. In green is highlighted the monophyletic clade 14. Adapted from Kibby et al., 2023. Right: A phylogenetic tree built on 544 bacterial NACHT-containing (bNACHT) proteins from clade 14 (see Methods). Yellow, green, and blue dots indicate the bootstrap values at those branch points. Pink dots indicate clade 14 bNACHT proteins with confirmed phage defense phenotypes. Orange dots indicate additional bNACHT proteins screened for interactions with *Tequatrovirus* proteins (See **table S2**). See **table S3** for a list of proteins included in the tree. bNACHT11: WP_114260439.1, bNACHT01: WP_015632533.1, bNACHT02: WP_021557529.1, bNACHT12: WP_021519735.1, bNACHT23: WP_000433597.1, a: WP_089552071, b: WP_201146468, c: WP_279077480, d: WP_248423937, e: WP_274682305, f: WP_036160395, g: WP_047083588, h: WP_346532815, i: WP_021182293, j: WP_064277770. **(B)** Percent identity of an alignment of all clade 14 bNACHT proteins used to assemble the tree on the right in (A). The position of the NACHT and SNaCT domains in the alignment is indicated below. **(C)** Heat map representing AlphaFold-multimer screening of interactions between SNaCT domain-containing bNACHT proteins with Tequatrovirus proteins. X- and y- axes were arranged using hierarchical clustering (see Methods, **table S4**). The bNACHT proteins are colored according to their position on the phylogenetic tree in A. Shown here is the data for all cluster representatives with an average weighted pTM score > 0.4 for at least one bNACHT protein screened. See **fig. S4F** for a heat map including every phage protein cluster screened. **(D)** Inset of the highest scoring proteins for bNACHT11, bNACHT02, and WP_089552071 from **Fig. 4C**. Phage proteins tested for bNACHT11 activity are labeled above. **(E)** Colony formation of *E. coli* expressing bNACHT02 on one plasmid and the indicated phage protein on a second plasmid. The expression of phage proteins (+) or *gfp* (-) is IPTG-inducible. Experiments were performed using 500 µM IPTG induction in LB media. Data represent the mean ± SEM of n = 3 biological replicates, shown as individual points.

The SNaCT domain interface mediates activator recognition

Surprisingly, the predicted structures of bNACHT11 interacting with T2 gp006, T2 gp009, or T2 gp057 show a single binding interface on bNACHT11 recognizing multiple unrelated phage proteins. This surface is modeled as a predicted β-sheet fold that helps to form a hydrophobic surface in the SNaCT domain (**Fig. 2D–F**, **Fig. 3A**, **fig. S1A–H**). We hypothesized that mutations to this surface would disrupt bNACHT11 activator binding and decrease defense system sensitivity. To test the role of this interface in activator recognition, we generated alanine mutants to bNACHT11 S518, I519, F522, I549, and I553 (**Fig. 3B**). We mutated R214 in the NACHT domain of bNACHT11 as a

negative control. This residue is predicted to be involved nucleotide binding[26], and previous work showed that mutation at this position abrogates activity in a related bNACHT protein without impacting expression[9]. We found that alanine mutations to I519, F522, and I553 strongly decreased phage protection against all phages tested. Mutations to S518 and I549 subtly decreased protection against phage T2, but strongly decreased protection against phages T4 and T6 (**Fig. 3B**, **fig. S3D**, **E**). Both the I519A and the F522A mutant phenotypes could be explained by decreased protein expression, however, other mutants were equivalently expressed (**fig. S3C**).

We further investigated the activity of these mutants in our genetic assay for activation and found that both the S518A mutant and I549A mutant maintained growth inhibition comparable to wild-type bNACHT11 when coexpressed with T2 *gp006* (**fig. S3F**). The I519A and I553A mutants disrupted the growth-inhibiting effects of coexpression. These results suggest that hydrophobic residues at the binding interface of the SNaCT domain in bNACHT11 are crucial for sensitive activator recognition.

bNACHT11 recognizes a large pool of phage-encoded activators

We investigated bNACHT11 activating proteins in the context of phage infection by generating a panel of knockout phages missing one or more putative activators. Individual and combined deletions of *gp006*, *gp009*, *gp028,* or *gp057* could be constructed using conventional methods, showing that these proteins are not essential for phage replication. None of these deletion mutants evaded bNACHT11-mediated defense **(Fig. 3C**, **fig. S3**). These results suggest that there are additional, high-affinity activators of bNACHT11 encoded by phage T2Δ*gp6/9/57/28*.

To demonstrate the role of *gp006*, *gp009*, *gp028*, and *gp057 in vivo*, we reasoned that we needed to weaken the ability of the SNaCT domain to recognize other unidentified phage activators, thereby making phage recognition more dependent on the known activators. Targeted mutations were constructed in the bNACHT11 SNaCT domain and these strains were infected with wild-type and T2 phages deleted for activating proteins. Strikingly, we found that while bNACHT11^{I549A} conferred over 1,000-fold protection against wild-type T2, this mutant completely lost protection against T2 phages lacking *gp006*, *gp009*, *gp028*, or *gp057* individually **(Fig. 3D**, **S3H)**. bNACHT11^{I553A} confers only weak protection against wild-type phage T2, but this protection is also abrogated by deletion of *gp006* (**fig. S3I**).

Taken together, these data suggest that bNACHT11 directly binds several phage proteins with high affinity to sensitively detect redundant signatures of infection. Point mutations in the SNaCT domain weaken the interaction between defense system and activator, amplifying the effect of mutations to escape detection.

High-throughput screening reveals SNaCT domain specificity features

We next sought to expand our screen to predict activators of related phage defense systems. Our previous analysis of NACHT domain-containing proteins across the tree of life divided these proteins in 25 monophyletic clades. bNACHT11 is a member of clade 14, a large group of proteins characterized by the presence of a C-terminal SNaCT domain that is widely distributed in bacteria (**Fig. 4A**)[9]. The sequence of the SNaCT domains found in this clade were highly variable, suggesting that they are under intense evolutionary pressure to diversify (**Fig. 4B**)[9].

To understand the breadth of *Tequatrovirus* phage proteins detected by clade 14 bNACHT proteins, we ran 14 related SNaCT-containing bNACHT proteins through our AlphaFold-multimer pipeline (**Fig. 4C**). We observe that closely related clade 14 bNACHT proteins, such as bNACHT02, bNACHT11, and WP_089552071, are predicted to interact with many of the same phage proteins as bNACHT11, but more distantly related bNACHT proteins have different predicted activators (**Fig. 4C**, **fig. S4**).

Based on these findings, we hypothesized that bNACHT02 recognizes many of the same proteins that activate bNACHT11. We tested this hypothesis using our genetic coexpression assay and found that bNACHT02 was indeed also activated by T2 *gp057* and T2 *gp046* **(Fig. 4E**). We next tested these phage proteins for their ability to activate other bNACHT proteins that they were not predicted to interact with. We found that T2 *gp006* and T2 *gp009* had no impact on colony formation in the presence of other clade 14 bNACHT systems. Unexpectedly, T2 *gp046* and T2 *gp057* did activate bNACHT01 (**fig. S4A–D**). Overall, these data suggest that diversification of the SNaCT domain has enabled specific detection of many different phage proteins across diverse bNACHT defense systems.

Discussion

For many recently identified phage defense systems, the mechanism by which they sense phage infection remains unknown. Here we use an *in silico* pipeline to predict interactions between a bacterial defense system and phage-encoded proteins, revealing that bNACHT11 can recognize multiple distinct activators. This activation is the result of direct binding dominated by hydrophobic interactions and mutations in the C-terminal SNaCT domain impacted recognition of phage-encoded activators. Finally, we expand this *in silico* strategy to other defense systems to explore how diverse SNaCT sensor domains confer specificity to different phage proteins and predict likely activators of other bNACHT defense systems.

bNACHT11 joins an emerging group of defense systems that recognize multiple unrelated phage proteins to sense infection. The CapRel^{SJ46} and KpAvs2 defense systems also bind more than one phage protein during infection[22,27,28]. The CapRel^{SJ46} system

has overlapping but distinct binding sites for both of its identified activators, and KpAvs2 binds disparate proteins for different phages. However, bNACHT11 is distinct from other defense systems because it uses the same binding interface to recognize several proteins that differ in both their sequence and predicted structure.

Using computational, genetic, and biochemical approaches, we found that a single binding site in the C-terminal SNaCT domain of bNACHT11 recognized at least five distinct phage proteins from the T-even family. We showed that bNACHT11 binds multiple phage proteins with nanomolar affinity and that activator binding results in bNACHT11 oligomerization. A major outstanding question raised by our work is: what is the function of the many phage-encoded proteins that activate bNACHT11? We hypothesize that these phage proteins perform redundant or similar roles during infection and that bNACHT11 mimics a binding partner that these phage proteins share. This mode of innate immune signaling has been referred to as the "integrated decoy model", whereby immune sensors detect pathogen activities by serving as a "decoy" of a core host pathway. One example comes from the plant immune sensor RRS1 in *Arabidopsis thaliana*, which recognizes the AvrRps4 avirulence protein from pathogenic *Pseudomonas syringae* by the C-terminus functioning as a decoy of the WRKY-like transcription factors that are targeted by ArRps4[29]. Understanding the function bNACHT11 activators may lead to new insights into fundamental steps within the phage life cycle.

Understanding the mechanism by which defense systems sense phage infection gives us critical mechanistic insight into the NLR molecular machine. Identifying protein activators of bNACHT11 now enables studying the molecular dynamics of the SNaCT domain, bNACHT oligomer assembly, and the effector function initiated by the bNACHT N-terminus. This work further demonstrates the power of using computational approaches to discover novel protein-protein interactions. Like all techniques, AlphaFold-multimer produces false positive and false negative results, however, the trade-off is exceptionally high throughput compared to conventional techniques such as immunoprecipitation followed by mass spectrometry or two-hybrid approaches. We demonstrated the power of *in silico* screening by expanding our analysis to predict activators for diverse bNACHT systems. We anticipate similar *in silico* strategies will continue to provide a detailed understanding of the complex signaling that occurs between virus and host and provide a framework for illuminating immune signaling events throughout the tree of life.

References

1. Doron, S. *et al.* Systematic discovery of antiphage defense systems in the microbial pangenome. *Science* **359**, eaar4120 (2018).
2. Millman, A. *et al.* An expanded arsenal of immune systems that protect bacteria from phages. *Cell Host Microbe* **30**, 1556-1569.e5 (2022).
3. Gao, L. *et al.* Diverse enzymatic activities mediate antiviral immunity in prokaryotes. *Science* **369**, 1077–1084 (2020).
4. Rousset, F., Dowding, J., Bernheim, A., Rocha, E. P. C. & Bikard, D. *Prophage-Encoded Hotspots of Bacterial Immune Systems*. http://biorxiv.org/lookup/doi/10.1101/2021.01.21.427644 (2021) doi:10.1101/2021.01.21.427644.
5. Vassallo, C. N., Doering, C. R., Littlehale, M. L., Teodoro, G. I. C. & Laub, M. T. A functional selection reveals previously undetected anti-phage defence systems in the E. coli pangenome. *Nat. Microbiol.* **7**, 1568–1579 (2022).
6. Evans, R. *et al.* Protein complex prediction with AlphaFold-Multimer. 2021.10.04.463034 Preprint at https://doi.org/10.1101/2021.10.04.463034 (2022).
7. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
8. Gao, L. A. *et al.* Prokaryotic innate immunity through pattern recognition of conserved viral proteins. *Science* **377**, eabm4096 (2022).
9. Kibby, E. M. *et al.* Bacterial NLR-related proteins protect against phage. *Cell* **186**, 2410-2424.e18 (2023).
10. Koonin, E. V. & Aravind, L. The NACHT family – a new group of predicted NTPases implicated in apoptosis and MHC transcription activation. *Trends Biochem. Sci.* **25**, 223–224 (2000).
11. Duncan, J. A. & Canna, S. W. The NLRC4 Inflammasome. *Immunol. Rev.* **281**, 115–123 (2018).
12. Kufer, T. A. & Sansonetti, P. J. NLR functions beyond pathogen recognition. *Nat. Immunol.* **12**, 121–128 (2011).
13. Dyrka, W. *et al.* Diversity and Variability of NOD-Like Receptors in Fungi. *Genome Biol. Evol.* **6**, 3137–3158 (2014).
14. Conte, A. N. *et al.* Phage detection by a bacterial NLR-related protein is mediated by DnaJ. *bioRxiv* 2024.06.04.597415 (2024) doi:10.1101/2024.06.04.597415.
15. Arya, P. & Acharya, V. Plant STAND P-loop NTPases: a current perspective of genome distribution, evolution, and function: Plant STAND P-loop NTPases: genomic organization, evolution, and molecular mechanism models contribute broadly to plant pathogen defense. *Mol. Genet. Genomics* **293**, 17–31 (2018).
16. van Wersch, S., Tian, L., Hoy, R. & Li, X. Plant NLRs: The Whistleblowers of Plant Immunity. *Plant Commun.* **1**, 100016 (2020).
17. Yu, D., Chojnowski, G., Rosenthal, M. & Kosinski, J. AlphaPulldown—a python package for protein–protein interaction screens using AlphaFold-Multimer. *Bioinformatics* **39**, btac749 (2023).
18. Miller, E. S. *et al.* Bacteriophage T4 Genome. *Microbiol. Mol. Biol. Rev.* **67**, 86–156 (2003).
19. Schoch, C. L. *et al.* NCBI Taxonomy: a comprehensive update on curation, resources and tools. *Database J. Biol. Databases Curation* **2020**, baaa062 (2020).
20. Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, 1026–1028 (2017).
21. Mirdita, M., Steinegger, M. & S"oding, J. MMseqs2 desktop and local web server app for fast, interactive sequence searches. *Bioinformatics* **35**, 2856–2858 (2019).
22. Béchon, N. *et al.* Diversification of molecular pattern recognition in bacterial NLR-like proteins. *Nat. Commun.* **15**, 9860 (2024).
23. Tenthorey, J. L., Kofoed, E. M., Daugherty, M. D., Malik, H. S. & Vance, R. E. Molecular Basis for Specific Recognition of Bacterial Ligands by NAIP/NLRC4 Inflammasomes. *Mol. Cell* **54**, 17–29 (2014).
24. Broz, P. & Dixit, V. M. Inflammasomes: mechanism of assembly, regulation and signalling. *Nat. Rev. Immunol.* **16**, 407–420 (2016).
25. Martin, R. *et al.* Structure of the activated ROQ1 resistosome directly recognizing the pathogen effector XopQ. *Science* **370**, eabd9993 (2020).
26. Zurek, B., Proell, M., Wagner, R. N., Schwarzenbacher, R. & Kufer, T. A. Mutational analysis of human NOD1 and NOD2 NACHT domains reveals different modes of activation. *Innate Immun.* **18**, 100–111 (2012).
27. Zhang, T. *et al.* Direct activation of a bacterial innate immune system by a viral capsid protein. *Nature* **612**, 132–140 (2022).
28. Zhang, T. *et al.* A bacterial immunity protein directly senses two disparate phage proteins. *Nature* **635**, 728–735 (2024).
29. Cesari, S., Bernoux, M., Moncuquet, P., Kroj, T. & Dodds, P. N. A novel conserved mechanism for plant NLR protein pairs: the "integrated decoy" hypothesis. *Front. Plant Sci.* **5**, 606 (2014).
30. Kropinski, A. M., Mazzocco, A., Waddell, T. E. & Johnson, R. P. Enumeration of Bacteriophages by Double Agar Overlay Plaque Assay. *Bacteriophages* **501**, 69–76 (2009).

31. Maffei, E. *et al.* Systematic exploration of Escherichia coli phage–host interactions with the BASEL phage collection. *PLOS Biol.* **19**, e3001424 (2021).
32. Darling, A. C. E., Mau, B., Blattner, F. R. & Perna, N. T. Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements. *Genome Res.* **14**, 1394–1403 (2004).
33. Darling, A. E., Mau, B. & Perna, N. T. progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. *PLoS ONE* **5**, e11147 (2010).
34. Darling, A. E., Tritt, A., Eisen, J. A. & Facciotti, M. T. Mauve Assembly Metrics. *Bioinformatics* **27**, 2756–2757 (2011).
35. Whiteley, A. T. *et al.* Bacterial cGAS-like enzymes synthesize diverse nucleotide signals. *Nature* **567**, 194–199 (2019).
36. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
37. Adler, B. A. *et al.* Broad-spectrum CRISPR-Cas13a enables efficient phage genome editing. *Nat. Microbiol.* **7**, 1967–1979 (2022).
38. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
39. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).
40. Thévenot, E. A., Roux, A., Xu, Y., Ezan, E. & Junot, C. Analysis of the Human Adult Urinary Metabolome Variations with Age, Body Mass Index, and Gender by Implementing a Comprehensive Workflow for Univariate and OPLS Statistical Analyses. *J. Proteome Res.* **14**, 3322–3335 (2015).
41. Jerabek-Willemsen, M., Wienken, C. J., Braun, D., Baaske, P. & Duhr, S. Molecular Interaction Studies Using Microscale Thermophoresis. *ASSAY Drug Dev. Technol.* **9**, 342–353 (2011).
42. Jarmoskaite, I., AlSadhan, I., Vaidyanathan, P. P. & Herschlag, D. How to measure and evaluate binding affinities. *eLife* **9**, e57264.

Acknowledgements

The authors thank the CU Boulder Department of Biochemistry Shared Instruments Pool core facility (RRID:SCR_018986), Annette Erbse, and its staff; Jennifer Doudna and Ben Adler for generously sharing CRISPR-Cas13 editing plasmids; Connor Keane for assistance with Python; and members of the Whiteley lab for their advice and helpful discussion. This work used the Alpine high performance computing resource at the University of Colorado Boulder. Alpine is jointly funded by the University of Colorado Boulder, the University of Colorado Anschutz, and Colorado State University and with support from NSF grants OAC-2201538 and OAC-2322260. This work was funded by the National Institutes of Health R35 GM144121 (KDC) and the NIH Director's New Innovator award DP2AT012346 (ATW), the Howard Hughes Medical Institutes Emerging Pathogens Initiative (KDC), the National Institutes of Health Boettcher Foundation's Webb-Waring Biomedical Research Program (ATW), the PEW Charitable Trust Biomedical Scholars Award (ATW), and the Burroughs Wellcome Fund PATH Award (ATW). EMK was funded in part by the NIH T32 Signaling and Cellular Regulation training grant (T32 GM008759 and T32 GM142607). LKR is supported in part by the NIH Biophysics training grant (T32 GM145437), the Interdisciplinary Quantitative Biology PhD Program at the BioFrontiers Institute, University of Colorado Boulder, and the NRT Integrated Data Science Program NSF grant 2022138 at the BioFrontiers Institute, University of Colorado Boulder. LAW was supported in part by the Biological Sciences Initiative funded by the University of Colorado Boulder and an Undergraduate Research Opportunities Program Student Assistantship Grant funded by the University of Colorado Boulder.

Author contributions

Conceptualization: EMK, LKR, and ATW. Methodology: EMK, LKR, AD, and ATW. Software: EMK, LKR, and LF. Investigation: all authors. Writing – original draft: EMK and ATW. Writing – review & editing: all authors. Visualization: AD and EMK. Supervision: KDC and ATW. Funding acquisition: KDC and ATW.

Competing interests

The authors declare no competing interests.

Data and materials availability

All data are available in the main text or the supplementary materials. Strains, phages, and plasmids used in this study are available upon request.

Supplementary Materials

Materials and Methods
Figs. S1–S4
Tables S1–S6

Experimental Model and Subject Details

Bacterial Strains and culture conditions

All *E. coli* strains used in this study are listed in **table S5**. *E. coli* were grown as described previously [9]. Briefly, bacteria were cultured in LB medium (1% tryptone, 0.5% yeast extract, and 0.5% NaCl) shaking at 37 °C and 220 rpm in 1-3 mL of media in 14 mL culture tubes, unless otherwise indicated. When applicable, chloramphenicol (20 µg/mL) or carbenicillin (100 µg/mL) were added. Constructed strains were frozen for storage in LB with 30% glycerol at -70 °C. OmniPir *E. coli* was used for construction and propagation of all plasmids [9]. *E. coli* MG1655 (CGSC6300) was used to collect all experimental data. *E. coli* BL21 (DE3) (NEB cat#: C2527H) or *E. coli* Rosetta2 with the pRARE2 plasmid (Millipore Sigma cat#: 71400-3) were used to express proteins for purification.

When indicated, experimental data was collected using MMCG medium (47.8 mM Na_2HPO_4, 22 mM KH_2PO_4, 18.7 mM NH_4Cl, 8.6 mM NaCl, 22.2 mM Glucose, 2 mM $MgSO_4$, 100 µM $CaCl_2$, 3 µM Thiamine, Trace Metals at 0.1× (Trace Metals Mixture T1001, Teknova, final concentration: 5 mM Ferric chloride, 2 mM Calcium chloride, 1 mM Manganese chloride, 1 mM Zinc Sulfate, 0.2 mM Cobalt chloride, 0.2 mM Cupric chloride, 0.2 mM Nickel chloride, 0.2 mM Sodium molybdate, 0.2 mM Sodium selenite, 0.2 mM Boric acid)). When experiments using MMCG required bacteria expressing two plasmids, strains were grown using reduced antibiotic concentrations (MMCG with 20 µg/mL carbenicillin and 4 µg/mL chloramphenicol).

Phage Amplification and Storage

All phages used this study are listed in **table S6**. Phages were amplified via either liquid or plate amplification following the protocol for a modified double agar overlay [30]. For liquid amplification, 5 mL mid-log cultures of *E. coli* MG1655 in LB plus 10 mM $MgCl_2$, 10 mM $CaCl_2$, and 100 µM $MnCl_2$ were infected with phage at an MOI of 0.1 and grown, shaking, for 2–16 hours. Cultures were then pelleted to remove cellular debris. The supernatant was harvested by decanting and ~100 µL chloroform was added to the cultures to remove any bacterial contamination.

For plate amplification, 400 µL of mid-log MG1655 were mixed with 3.5 mL LB soft agar mix (LB with 0.35% agar and 10 mM $MgCl_2$, 10 mM $CaCl_2$ and 100 µM $MnCl_2$) and 100-1,000 PFU. Plates were then incubated for 16 hours at 37 °C. 5 mL of SM buffer (100 mM NaCl, 8 mM $MgSO_4$, 50 mM Tris-HCl pH 7.5, 0.01% gelatin) was added to the plate and allowed to soak out the phages for 1 hour before SM buffer was collected and passed through a 0.2 µm filter or treated with 1–3 drops of chloroform to remove viable bacteria. All phages were stored at 4 °C in SM buffer (100 mM NaCl, 8 mM $MgSO_4$, 50 mM Tris-HCl pH 7.05, 0.01% gelatin) or LB.

Method Details

Generation of Representative Phage gene clusters

Representative phage gene clusters were generated using all Tequatrovirus family reference genomes from NCBI, 16 *Tequatrovirus* genome sequences found in the BASEL collection [31], and the genomes representing the common lab phages T2, T4, and T6, for a total of 91 phage genome sequences. A complete list of the input genomes used can be found in **table S1**. The CDS regions for each of these genomes were extracted using Geneious. Excluding proteins that were less than 50 amino acids, we used the MMSeq2 clustering algorithm[20,21] (default parameters: 80% identity and 80% coverage) to cluster the remaining 22,719 proteins into 862 representative clusters. Any clusters representing only one input protein were removed, leaving us with a final set of 629 phage proteins representing those commonly expressed by phages in the genus *Tequatrovirus.* For a full list of the proteins represented in each cluster, see **table S1**.

To order clusters on the X-axis based on their position in the phage genome, the Mauve alignment plugin was used in Geneious Prime to generate a genome alignment of all 91 genomes represented in the clusters[32–34]. To make this alignment possible, the origin of each genome used was set to 7 nucleotides upstream of the gene *rIIA,* commonly used as the beginning of *Tequatrovirus* genomes. We then assumed that there were no major genome rearrangements and

used otherwise default settings in Mauve to generate the alignment and extract the position of each cluster representative in the genome.

MSA generation and protein structure prediction
Multiple sequence alignments for each gene were generated using AlphaPulldown[17]. Generated MSAs were then used as input for AlphaFold-multimer[6]. In AlphaFold, each model generated three predictions for a total of 15 predicted structures per interaction. After the predictions were completed, the ipTM + pTM scores ("weighted pTM") for each model were averaged, and the average weighted pTM score was used to identify confidently predicted interactions for further follow-up.

Plasmid construction
The plasmids used in this study are listed in **table S5**. DNA manipulations and cloning were performed as previously described[9,35]. Briefly, target genes were amplified from plasmid, phage or bacterial genomic DNA using Q5 Hot Start High Fidelity Master Mix (NEB, M0494L) flanked by 18 base pairs of homology to the vector backbone. Vectors were digested using restriction digest and genes were ligated into vectors using modified Gibson Assembly[36]. Gibson reactions were then transformed via heat shock or electroporation into competent OmniPir[9] and plated onto appropriate antibiotic selection. When possible, phage gene coding sequences were amplified from the genomic DNA of the indicated *E. coli* phages. bNACHT and phage protein point mutations were generated by amplifying out the gene of interest in two parts from a plasmid template, with the desired mutation occurring in the overlapping region between the two amplicons. Unless otherwise indicated, all enzymes were purchased from New England Biolabs.

For all vectors using the pLOCO2 backbone, pEK0252 was amplified and purified from OmniPir. Purified plasmid was then linearized using SbfI-HF and NotI-HF.

For all vectors using the pTACxc backbone, pAW1608 was amplified and purified from OmniPir. Purified plasmid was then linearized using BmtI-HF and NotI-HF.

For all vectors using the pETSUMO2 backbone, pAW1642 was amplified and purified from OmniPir. Purified plasmid was then linearized using BamHI-HF and NotI-HF.

For all vectors using the pET backbone, pAW1642 was amplified and purified from OmniPir. Purified plasmid was then linearized using NdeI-HF and NotI-HF.

Plasmid pEK0278 was purchased from TWIST Biosciences and cloned into the BamHI and XhoI restriction sites of the pET21(+) vector. When needed, DNA sequences were randomly generated to ensure that inserts had a minimum length of 300bp, per manufacturer suggestions.

For constructing vectors encoding the eLbuCas13a system [37] with appropriate spacers, the backbone was amplified from pBA681 using PCR (forward primer oEK0228: ATGCTTGGGCCCGAA. Reverse primer oEK0229: GGGCGGAGCCTATGGAAAAACGGCTTTGCCGCG). Gibson ligation was used to circularize the vector with the new insert.

Sanger sequencing (Azenta, Quintara) was used to validate the correct sequence within the multiple cloning site. When needed, nanopore sequencing (Quintara, Plasmidsaurus) was used to obtain the sequence of the entire plasmid.

Efficiency of plating/phage replication analysis
Efficiency of plating (EOP) was used to determine phage titer and replication. To do this, we employed a modified double agar overlay assay [30]. Briefly, overnight cultures of *E. coli* MG1655 expressing the indicated plasmids in MMCG plus appropriate antibiotics were diluted 1:10 into the same media and cultivated for an additional two to three hours to reach logarithmic growth phase (OD600 0.1–0.8). 400 µL of the culture was then mixed with 3.5 mL 0.35% agar MMCG, plus an additional 5 mM $MgCl_2$ and 100 mM $MnCl_2$. The mixture was then poured onto a 1.6% agar MMCG plate and cooled for 15 minutes. 2 µL of a phage dilution series in SM buffer was spotted onto the overlay and allowed to adsorb for 10 minutes before the plate was incubated overnight at 37 °C.

Plaque formation was analyzed the following day. Instances with a hazy zone of clearance rather than individual plaque formation counted the lowest phage concentration at which clearance was observed as ten plaques. 0.9 plaques at the least dilute spot were used as the limit of detection in instances where no zone of clearance or plaque formation was visible.

The inverse of EOP was used to calculate fold protection. The PFU of a phage lysate on sensitive host bacteria expressing an empty vector was divided by the PFU for the same phage lysate measure on test bacterial strains. In this way, a 10-fold decrease in EOP is a 10-fold increase in phage protection.

Colony formation/growth inhibition analysis
The impact of coexpression of different alleles of bNACHT and phage proteins on bacterial growth was quantified using a colony formation assay. Briefly, *E. coli* was cultivated overnight in LB with appropriate antibiotics. Cultures were diluted in a 10-fold series into LB and 5 µL of each dilution was spotted onto an LB or MMCG agar plate containing the appropriate antibiotics, as well as IPTG as indicated. Spotted bacteria were allowed to dry for 10 minutes before the plates were incubated overnight at 37 °C.

Growth inhibition was measured the following day by enumerating the colony forming units of each strain, reported as CFU/mL for the starting culture. For instances where bacteria were growing but no individual colonies could be counted, the lowest bacterial concentration at which growth was observed was counted as ten CFU. In instances where no growth was visible, 0.9 CFU at the least dilute spot was used as the limit of detection.

Transformation efficiency analysis
The impact of coexpression of bNACHT alleles and phage proteins on bacterial growth was quantified using a cotransformation assay. Briefly, electrocompetent *E. coli* MG1655 was transformed with 50 ng of each purified plasmid in the cotransformation assay, recovered for 1 hour in SOC (2% Tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose), and spot-plated to measure colony formation as described above.

Construction of phage gene deletions
Phage T2 and T4 knockout mutations were generated following the protocol described in Adler et al., 2022 [37]. Briefly, wild-type phages were amplified via either plate or liquid culture (see above) on *E. coli* MG1655 strains expressing a pET vector encoding the template for homologous repair. Phage lysates amplified in this way were then mixed with 400 µL mid-log bacteria expressing eLbuCas13a constructs with spacers targeting the gene of interest and poured onto an MMCG agar plate as in the method described in solid plate amplification detailed above. Individual plaques were isolated and spot-plated onto *E. coli* MG1655 expressing the same spacer to confirm that phages were able to evade the spacers and to plaque-purify each clone. Deletion of the target gene was validated by PCR. Validated phage T2 and T4 mutants were subsequently amplified via either liquid or plate amplification on *E. coli* MG1655 in MMCG.

The homologous repair templates were designed to encode 250 bp on either side of the target gene, and the first and last 6 amino acids of the target gene were maintained in the knockout to minimize polar effects. Two 31-nt spacers were selected to target the beginning of each gene and induced as needed using anhydrotetracycline at 5 nM in the top agar of the soft agar overlay.

Validation of protein expression
To analyze the expression of bNACHT and phage protein alleles, 3 mL of *E. coli* MG1655 expressing the indicated plasmid were grown to mid-logarithmic phage in MMCG and 5×10^8 CFU were pelleted. Bacterial pellets were resuspended in 100 µL of 1 × LDS buffer (106 mM Tris-HCl pH7.4, 141 mM Tris Base, 2% w/v Lithium dodecyl sulfate, 10% v/v Glycerol, 0.51 mM EDTA, 0.05% Orange G). Samples were incubated at 95 °C for 10 minutes followed by a 5-minute centrifugation at 20,000 × g to remove debris. Samples in LDS were loaded at equal volumes and resolved using SDS-PAGE, then transferred to PVDF membranes charged in methanol. Membranes were blocked in Licor Intercept Buffer for one hour at 24 °C, followed by incubation with primary antibodies diluted in Intercept buffer overnight at 4 °C with rocking.

α6×His antibody (Thermo) was used at 1:5,000 to detect bNACHT11-6×His and α*E. coli* RNA polymerase B antibody (Biolegend) was used at 1:5,000 as a loading control. Blots were then incubated with Licor infrared (800CW/680RD) αRabbit/Mouse secondary antibodies at 1:40,000 dilution in TBS-T (0.1% Triton-X) for one hour at 24 °C and visualized using the Licor Odyssey CLx. Representative images were assembled using Adobe Illustrator CC 2024.

Clade 14 phylogenetic analysis
To construct the tree in **Fig. 4**, we used the NCBI protein-protein BLAST plugin on Geneious to collect the top 500 BLAST results for bNACHT01, bNACHT02, bNACHT11, bNACHT12, and bNACHT23, which generated a list of 600

nonredundant protein sequences [38]. We excluded protein sequences with fewer than 300 amino acids, which often had truncated SNaCT domains, resulting in 544 proteins. We then aligned these sequences in Geneious Prime using the MAFFT alignment strategy [39] with default settings. The percent identity plotted in **Fig. 4B** is from this alignment of 544 proteins. We excluded the last value of this protein as it represented one individual amino acid. This alignment was then used to build the clade 14 consensus tree in **Fig. 4A** using the Geneious Tree Builder (Genetic Distance Model: Jukes-Cantor, Tree Build Method: Neighbor-joining, no outgroup, resampled 100 times to generate bootstrap values). From this tree, we sampled 14 proteins distributed throughout the tree to robustly predict the diverse sensing capabilities of SNaCT domains in enterobacteria.

Hierarchical clustering of AlphaFold Screen results

To generate the dendrograms and heat maps displayed in **Fig. 4C** and **fig. S4F**, data from all the AlphaFold screens reported in this study was organized using hierarchical clustering. To do this, we used the "Heatmap of the dataMatrix" [40] workflow available on Galaxy (dissimilarity for clustering = euclidean, number of sample and variable clusters to identify = 1, correlation method = pearson, agglomeration method = ward).

Protein expression

Vectors for expressing gp006-6×His and gp006^{I20A}-6×His were transformed into BL21 (DE3) and plated onto 1.6% LB agar plates with 100 µg/mL carbenicillin. An individual colony was picked the following day and inoculated into 5 mL of liquid LB media plus 100 µg/mL carbenicillin. The culture was then grown overnight shaking at 37 °C and 220 rpm. The following day, the culture was used to inoculate 0.5–1 L of the same media, then grown to an OD600 of ~0.6 before IPTG was added to 500 µM to induce protein expression. The culture was then moved to a 16 °C shaking incubator and allowed to grow overnight.

The vector for expressing 6×His-SUMO-bNACHT11 was transformed into Rosetta2 expressing the pRARE2 plasmid and plated onto 1.6% LB agar plates + 100 µg/mL carbenicillin and 20 µg/mL chloramphenicol. An individual colony was picked the following day and inoculated into 5 mL of liquid LB media plus 100 µg/mL carbenicillin and 20 µg/mL chloramphenicol. The culture was then grown overnight shaking at 37 °C and 220 rpm. The following day, the culture was used to inoculate 0.5–1 L of the same media, then grown to an OD_{600} of ~0.6 before IPTG was added to 500 µM to induce protein expression. The culture was then moved to a 16 °C shaking incubator and allowed to grow overnight.

Vectors for expressing gp009-6×His, gp009I^{34A}-6×His, gp046-6×His, gp057-6×His, and 6×His-SUMO-bNACHT11^{I553A} were transformed into Rosetta2 expressing the pRARE2 plasmid and plated onto 1.6% MMCG agar plates + 100 µg/mL carbenicillin and 20 µg/mL chloramphenicol. An individual colony was picked the following day and inoculated into 20 mL of M9ZB media (47.8 mM Na_2HPO_4, 22 mM KH_2PO_4, 18.7 mM NH_4Cl, 85.6 mM NaCl, 1% Casamino acids (VWR), 0.5% v/v Glycerol, 2 mM $MgSO_4$, Trace Metals at 0.5 × (Trace Metals Mixture T1001, see above) plus 100 µg/mL carbenicillin and 20 µg/mL chloramphenicol. The culture was then grown overnight shaking at 37 °C and 220 rpm. The following day, the culture was used to inoculate 0.5–1 L of the same media to an OD_{600} of 0.05, then grown to an OD600 of ~1.5. Cultures were crash-cooled on ice for 20 minutes before IPTG was added to 500 µM to induce protein expression. The culture was then moved to a 16 °C shaking incubator and allowed to grow overnight. Proteins expressed in this way were purified and used in microscale thermophoresis experiments.

Protein purification

After overnight induction with IPTG, cultures were harvested by centrifugation for 30 minutes at 5,000 rpm and 4 °C in an Avanti JXN-26 Floor Centrifuge using the JXN 12.500 rotor (Beckman). The resulting pellets were resuspended in 40 mL Lysis buffer (20 mM HEPES pH 7.5, 400 mM NaCl, 10% v/v Glycerol, 30 mM Imidazole, 0.1 mM Dithiothreitol (DTT)). After resuspension, cells were lysed by sonication at 80% amplitude, with 30 second pulses for a total processing time of 10 minutes using a Sonicator 4000 (Misonix). Debris was removed from sonicated lysates by centrifugation for 60 minutes at 4 °C and 14,000 × g in a 5910 R centrifuge (Eppendorf). The soluble lysate was then decanted and protein was purified using immobilized metal affinity chromatography. Briefly, the soluble lysate was run over 1 mL of Ni-NTa resin (Fisher Sci) equilibrated in Lysis Buffer. The resin was then washed with 2 × 25 mL of Wash Buffer (20 mM HEPES pH 7.5, 1 M NaCl, 10% v/v glycerol, 30 mM Imidazole, 0.1 mM DTT) and protein was eluted in 5-10 mL of Elution Buffer (20 mM HEPES pH 7.5, 400 mM NaCl, 10% v/v glycerol, 300 mM Imidazole, 0.1 mM DTT). Proteins were then dialyzed against 2 × 1 L of Dialysis Buffer (20 mM HEPES pH 7.5, 250 mM KCl, 0.1 mM DTT), overnight at 4 °C using either 10 kDa MWCO tubing (VWR), 3 kDa MWCO Snakeskin Dialysis Tubing (VWR), or 3 kDa MWCO Slide-A-Lyzer Dialysis cassettes (VWR) as appropriate for the molecular weight of the protein.

The 6×His-SUMO-tag was cleaved from alleles of bNACHT11 using 6×His-ULP1 (produced in-house) during the overnight dialysis step. After dialysis, proteins were run over 1 mL Ni-Nta beads equilibrated in dialysis buffer to remove any uncleaved 6×His-SUMO tagged proteins.

After dialysis, proteins were concentrated as needed using 3 kDa or 30 kDa MWCO Nanosep spin concentration columns (Pall Labs) and stored 200–500 µL aliquots in Dialysis Buffer at –70 °C. Protein concentrations were measured using A_{280} on a Nanodrop OneC (Thermo) and protein purity was visualized using SDS-PAGE followed by Coomassie staining. Proteins purified in this way were used in microscale thermophoresis experiments.

Microscale thermophoresis

Protein binding affinities were measured using microscale thermophoresis [41]. For these experiments, alleles of phage proteins containing C-terminal 6×His tags were labeled using a Monolith His-Tag Labeling Kit RED-tris-NTA 2nd Generation (Nanotemper: Cat# MO-L018) following manufacturer instructions. Samples were allowed to equilibrate for 30 minutes at room temperature before MST measurement. Experiments were performed using independently labeled proteins and independently pipetted ligand titrations. Measurements were performed using 60% laser excitation, medium MST power, and a chamber temperature of 25 °C on a Nano-BLUE/RED Monolith NT.115 (NanoTemper). All data was analyzed using a hot time of 9–10 seconds. Fraction bound values were calculated by Mo.AffinityAnalysis software (NanoTemper). Binding data from three independent experiments were fit using the quadratic binding equation[42].

$$fraction\ bound = \frac{[L] + [T] + K_d - \sqrt{([L] + [T] + K_d)^2 - 4 \cdot [L] \cdot [T]}}{2 \cdot [T]}$$

In this equation [L] is the concentration of ligand, [T] is the concentration of labeled target (50 nM for all experiments reported here). The dissociation constants reported here represent the average of the dissociation constants calculated for each biological replicate ± standard error of the mean.

All reactions used MST Buffer (20 mM HEPES pH 7.5, 250 mM KCl, 5 mM $MgCl_2$, 0.05% v/v Tween-20, 0.1 mM DTT, 1 mM ATP). For high-salt measurements, salt concentrations were increased to 1 M KCl.

Oligomerization Analysis

To express 6×His-SUMO-bNACHT11, gp006-6×His, gp009-6×His, and 6×His-SUMO-gp057, the respective constructs were transformed into E. coli Rosetta2 pLysS (EMD Millipore). The primary cultures (10 mL) were grown in the presence of appropriate antibiotics (see **table S5** for construct details and antibiotic usage). The following day, primary cultures were used to inoculate 1 L of 2XYT medium (1.6% tryptone, 1% yeast extract, 0.5% NaCl) in 2 L flasks and grown at 37 °C until the OD_{600} reached approximately 0.75. Protein expression was induced by adding 0.33 mM IPTG and the cultures were incubated overnight at 20 °C to promote protein production. After 14–16 hours, the cells were harvested by centrifugation, and the bacterial pellets were resuspended in ice-cold resuspension buffer (50 mM Tris-HCl (pH 7.5), 300 mM NaCl, 10 mM imidazole, 10% glycerol, and 2 mM β-mercaptoethanol). Resuspended cells were lysed using a sonicator and the lysate was clarified by centrifugation. Proteins were purified through Ni^{2+} affinity chromatography (Ni-NTA Superflow, Qiagen). For bNACHT11 and gp057, the N-terminal 6×His-SUMO tags were removed by buffer exchange to eliminate imidazole using a buffer containing 50 mM Tris (pH 7.5), 300 mM NaCl, 10% glycerol, and 2 mM β-mercaptoethanol, followed by digestion with Sentrin-specific protease 2 (SENP2) overnight at 4 °C. The cleaved tags were removed the next day by passing the proteins through Ni^{2+} beads again. All purified proteins were concentrated and further purified using a Superdex 200 Increase 10/300 GL size exclusion column (Cytiva) in a buffer containing 20 mM HEPES (pH 7.4), 200 mM KCl, and 1 mM dithiothreitol (DTT). The purity of the proteins was confirmed by analyzing the samples using SDS-PAGE, followed by staining with Coomassie blue.

To study the effect of phage trigger addition on the oligomeric state of bNACHT11, size exclusion chromatography was performed. Briefly, 25 µM bNACHT11 was incubated with or without 1 mM ATP, 1 mM ATPγS, 1 mM AMP-PNP, or 1 mM ADP in a buffer containing 20 mM HEPES (pH 7.4), 200 mM KCl, 10 mM $MgCl_2$, and 1 mM DTT. The reaction mixtures were incubated at room temperature for 2–3 minutes. A 2–3 molar excess of either gp006, gp009, or gp057 purified protein was then added to the reaction mixture, and the samples were further incubated at room temperature for 5 minutes. The samples were immediately transferred to ice and kept on ice until injected into a Superdex 200 Increase 10/300 GL size exclusion column (using buffer 20 mM HEPES pH 7.4, 200 mM KCl, 10 mM $MgCl_2$, and 1 mM DTT). The corresponding protein fractions were collected in a 96-well block, changes in the elution volume were analyzed, and indicated samples from the peaks were subjected to SDS-PAGE analysis.

Quantification and Statistical Analyses

Experiments were performed in biological triplicate using cultures grown on three separate days unless otherwise indicated. Data was plotted using Graphpad Prism 10.3.1 at an n of 3 with error bars indicating standard error of the mean. Figures were created using Adobe Illustrator CC 2024 v28.6.0.

Accession numbers

T2 phage genome: NC_054931.1
T4 phage genome: AF158101.1
T4 CDS 30.5: NP_049819.1
T2 gp006: YP_010073654.1
T2 gp009: YP_010073657.1
T2 gp021: YP_010073669.1
T2 gp028: YP_010073676.1
T2 gp046: YP_010073694.1
T2 gp057: YP_010073705.1
T2 gp072: YP_010073720.1
bNACHT01: WP_015632533.1
bNACHT02: WP_021557529.1
bNACHT11: WP_114260439.1
bNACHT12: WP_021519735.1
bNACHT23: WP_000433597.1

Supplementary Data Tables

Supplementary Table 1. Cluster generation
- A. All phage genomes used to build clusters for AlphaFold screening
- B. List of proteins input into MMSeq for clustering (all greater than 50 aa)
- C. Complete MMSeq2 output
- D. Clusters used for AlphaFold-screening

Supplementary Table 2. AlphaFold-multimer screening results
Supplementary Table 3. Clade 14 bNACHT proteins used to generate phylogenetic tree
Supplementary Table 4. Hierarchical clustering of AlphaFold screening results.
Supplementary Table 5. E. coli strains and plasmids used in this study
Supplementary Table 6. Phages used in this study

Supplementary Figures

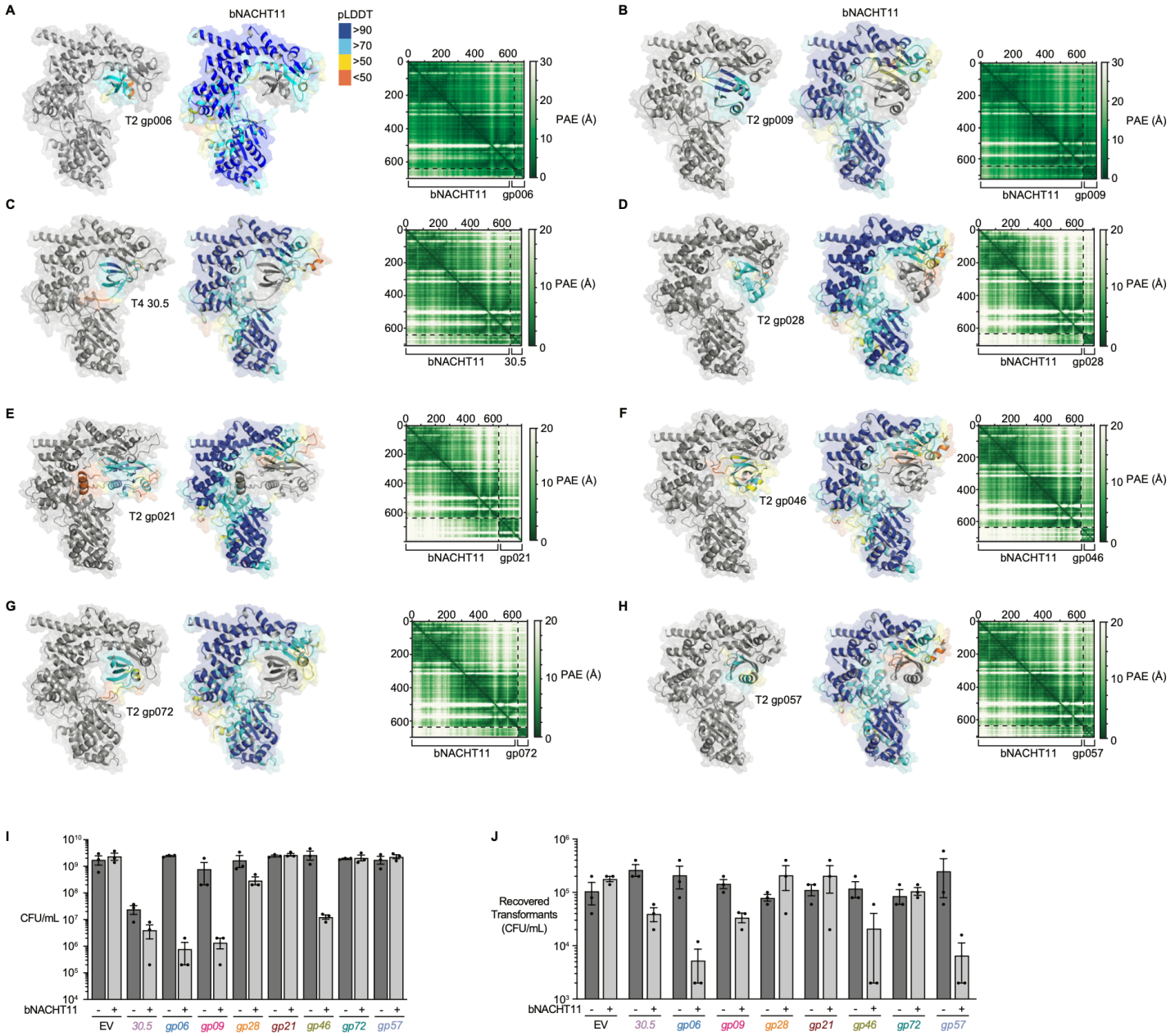

Figure S1. Predicted activators of bNACHT11.
(A–H) Predicted structure of bNACHT11 binding the indicated phage proteins. Left: The phage proteins screened are colored according to pLDDT. Middle: bNACHT11 is colored according to pLDDT. Right: Predicted aligned error (PAE) of the predicted interaction between bNACHT11 and the indicated phage protein. (A) T2 gp006 (B) T2 gp009 (C) T4 30.5 (D) T2 gp028 (E) T2 gp021 (F) T2 gp046 (G) T2 gp072 (H) T2 gp057. **(I)** Quantification of colony formation of *E. coli* expressing an Empty Vector (-) or bNACHT11 (+) on one plasmid and *gfp* (EV) or indicated phage protein on a second plasmid. The expression of *gfp* or phage proteins is IPTG-inducible. **(J)** Quantification of recovered transformants of *E. coli* cotransformed with plasmids expressing an Empty Vector (-) or bNACHT11 (+) and a *gfp* empty vector (EV) or the indicated phage protein. The expression of *gfp* and phage proteins is IPTG-inducible. For I and J, experiments were performed using 500 μM IPTG induction in LB media, and data represent the mean ± SEM of n = 3 biological replicates, shown as individual points.

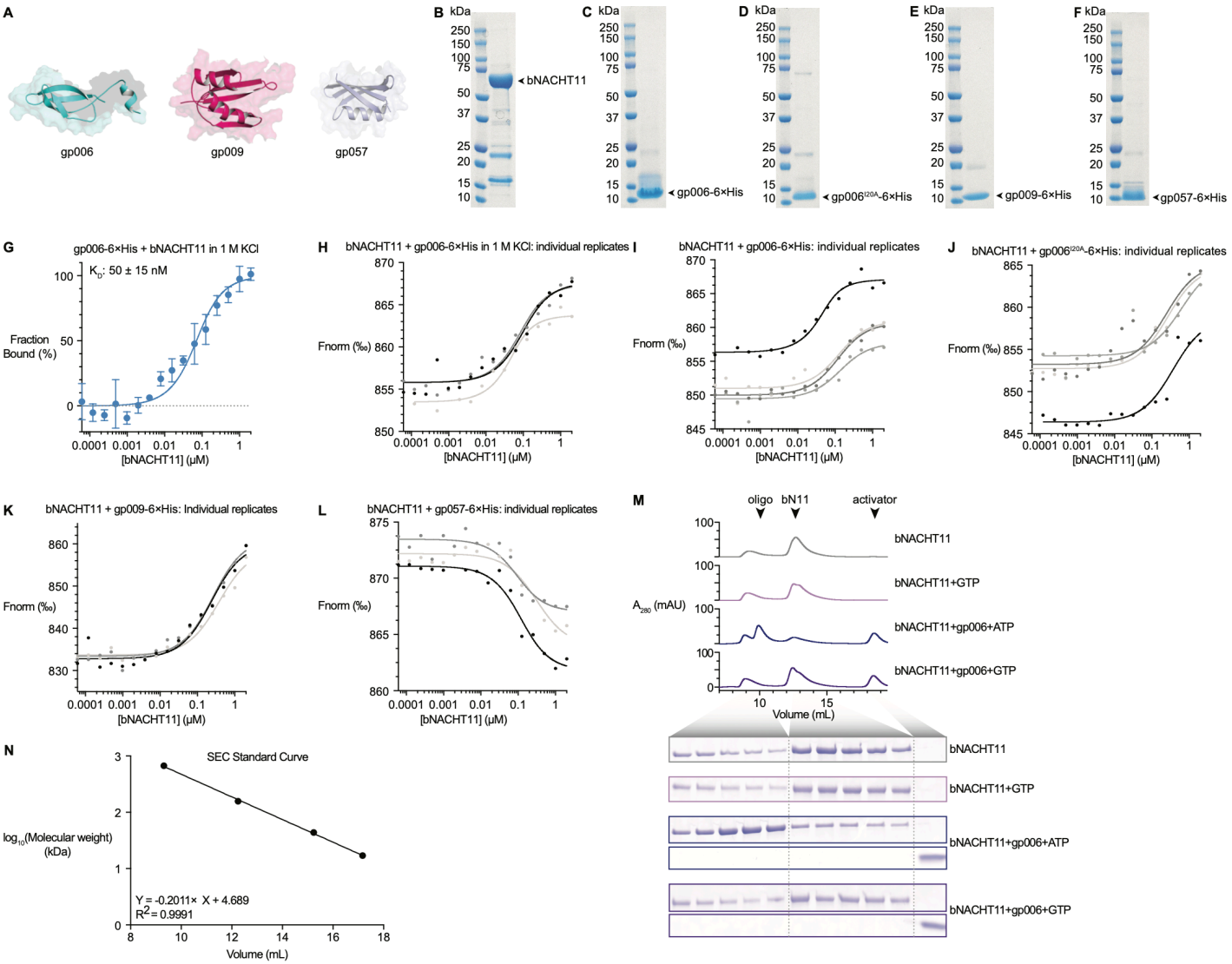

Figure S2. Binding and oligomerization of bNACHT11 with activators.
(A) AlphaFold2 predicted structures of the indicated phage proteins. **(B–F)** Coomassie stained SDS-PAGE gels of the purified protein indicated with an arrow. **(G)** Binding curve of purified His-tagged gp006 binding to purified bNACHT11, measured using microscale thermophoresis (MST) in 1 M KCl. Data represent the average of n = 3 individual replicates ± SEM. **(H–L)** Normalized fluorescence data for individual MST experiments used to generate the binding curves shown in **fig. S2G** (H), **Fig. 2G** (I, J) **Fig. 2H** (K) and **Fig. 2I** (L). **(M)** Above: Traces showing the absorbance at 280 nM (A_{280}) of the indicated samples being run over a size exclusion (SEC) column. Labels indicate the identity of proteins in each peak. Below: Coomassie staining of representative samples taken from in between the indicated volumes. For samples containing both bNACHT11 and gp006, portions of the gel corresponding to each are shown. Top: bNACHT11. Bottom: gp006. Data are representative of n = 3 replicates. **(N)** Standard curve of proteins with known molecular weights run over the SEC column used in (M).

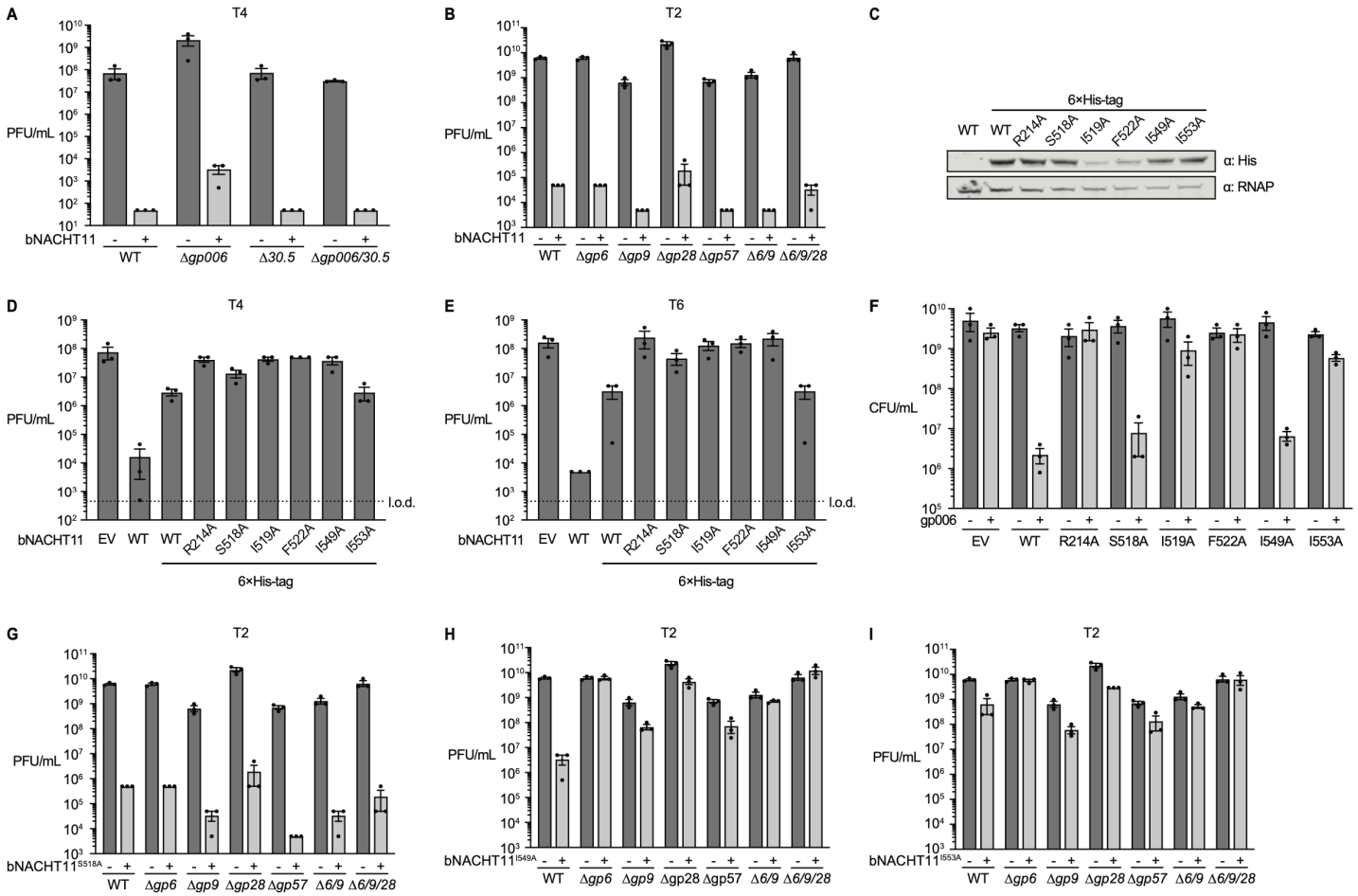

Figure S3. Phage gene deletions and bNACHT11 SNaCT domain mutations

(A–B) Efficiency of plating of the indicated strains of phage T4 (A) or T2 (B) infecting *E. coli* expressing bNACHT11 (+) or an empty vector (-). **(C)** Western blot analysis of *E. coli* expressing empty vector, untagged bNACHT11, or 6×His-tagged bNACHT11 of the indicated genotype. Representative image of n = 2 biological replicates. **(D–E)** Efficiency of plating of phage T4 (D) or T6 (E) infecting *E. coli* expressing bNACHT11 of the indicated genotype. **(F)** Quantification of colony formation of *E. coli* expressing an Empty Vector (-) or 6×His-tagged bNACHT11 of the indicated genotype on one plasmid and *gp006* on a second plasmid. The expression of *gp006* (+) or *gfp* (-) is IPTG-inducible. Experiments were performed using 500 μM IPTG induction in LB media. **(G–I)** Efficiency of plating for a panel of phage T2 mutants infecting *E. coli* expressing the indicated allele of bNACHT11-6×His (+) or an empty vector (EV). For A–B and D–I, data indicate the mean ± SEM of n = 3 biological replicates, shown as individual points.

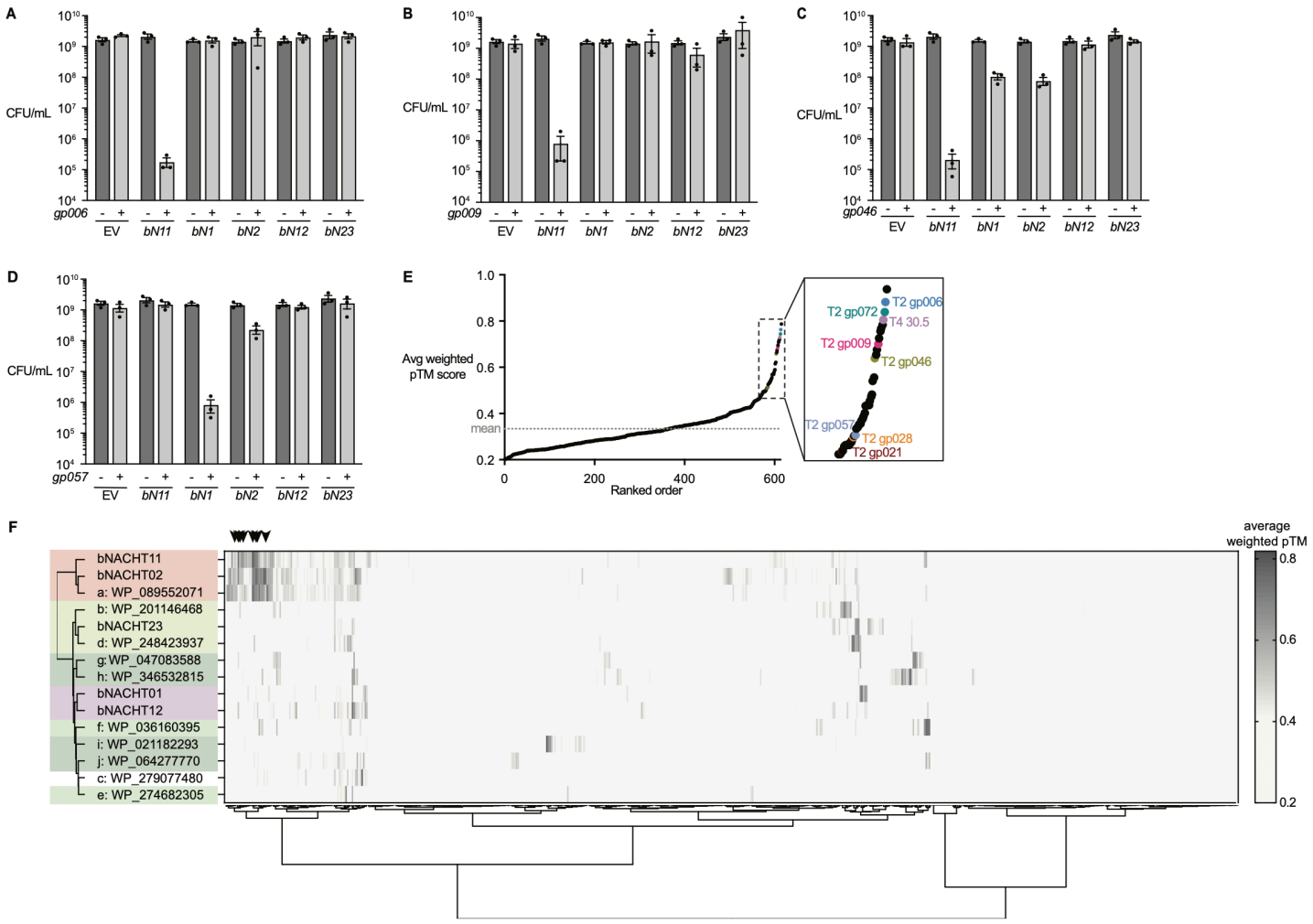

Figure S4. SNaCT domain specificity.

(A–D) Colony formation of *E. coli* expressing the indicated bNACHT protein on one plasmid and T2 *gp006* (A), T2 *gp009* (B), T2 *gp046* (C), or T2 *gp057* (D), on a second plasmid. The expression of phage proteins (+) and *gfp* (-) is IPTG-inducible. Experiments were performed using 500 µM IPTG induction in LB media. Data represent the mean ± SEM of n = 3 biological replicates, shown as individual points. **(E)** Results of AlphaFold-multimer screening for bNACHT02 (WP_021557529.1). Average weighted pTM score (ipTM×0.8 + pTM×0.2) was calculated for each protein screened, then ranked such that the highest-scoring proteins are on the right of the graph (inset). Clusters corresponding to genes tested for activity with bNACHT11 are labeled as in **Fig. 1**. See **table S2** for the data used to generate this figure. **(F)** Heat map representing the results of AlphaFold-multimer screens of many SNaCT domain-containing bNACHT proteins. Both x- and y- axis were arranged using hierarchical clustering (see Methods, **table S4**). Proteins are colored as in **Fig. 4A**.