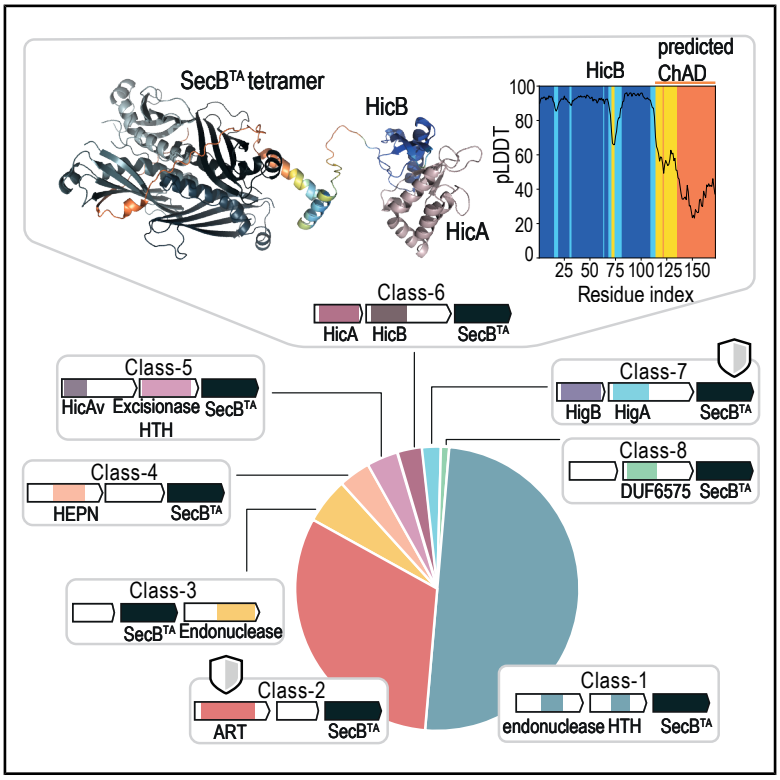
Cell Reports

Article

Structural and functional diversity of toxin-antitoxin-chaperone systems

Graphical abstract



Highlights

- Sensitive search identifies eight TAC classes, including five previously unknown types
- SecBTA chaperones show conserved folds despite deep sequence divergence
- SecBTA phylogeny reveals modular recombination with diverse TA families
- Class 6 TAC HicA toxin active site and toxicity are validated

Authors

Jose A. Nakamoto, Roni Odai, Toomas Mets, Tanel Tenson, Tatsuaki Kurata, Vasili Hauryliuk, Gemma C. Atkinson

Correspondence

jose.nakamoto@med.lu.se (J.A.N.), vasili.hauryliuk@med.lu.se (V.H.), gemma.atkinson@med.lu.se (G.C.A.)

In brief

Toxin-antitoxin-chaperone (TAC) systems contribute to bacterial phage defense. Surveying over 230 million proteins, Nakamoto et al. identify eight TAC classes, five previously undescribed. Conserved SecBTA structures persist despite extreme sequence divergence, whereas modular reshuffling, consistent with mobile, prophage-linked loci, drives diversification.

Nakamoto et al., 2026, Cell Reports *45*, 117024
March 24, 2026
https://doi.org/10.1016/j.celrep.2026.117024

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

Structural and functional diversity of toxin-antitoxin-chaperone systems

Jose A. Nakamoto,[1,*] Roni Odai,[1] Toomas Mets,[1,2] Tanel Tenson,[2] Tatsuaki Kurata,[1,3] Vasili Hauryliuk,[1,2,4,5,*] and Gemma C. Atkinson[1,4,6,7,*]

[1]Department of Experimental Medical Science, Lund University, 22100 Lund, Sweden
[2]Institute of Technology, University of Tartu, 50090 Tartu, Estonia
[3]RNA Systems Biochemistry Laboratory, Pioneering Research Institute, RIKEN, Wako, Saitama 351-0198, Japan
[4]Science for Life Laboratory, 22100 Lund, Sweden
[5]NanoLund, Lund University, 22100 Lund, Sweden
[6]Senior author
[7]Lead contact
*Correspondence: jose.nakamoto@med.lu.se (J.A.N.), vasili.hauryliuk@med.lu.se (V.H.), gemma.atkinson@med.lu.se (G.C.A.)
https://doi.org/10.1016/j.celrep.2026.117024

SUMMARY

Toxin-antitoxin-chaperone (TAC) systems are three-part gene clusters encoding a toxin, antitoxin, and specialized SecB-like chaperone (SecB^{TA}) with emerging roles in phage defense. To identify and classify SecB homologs and associated TACs across bacteria, we surveyed the full RefSeq database. Phylogenetic and gene neighborhood analyses reveal three major SecB subfamilies: two housekeeping groups and a diverse SecB^{TA} clade associated with eight TAC classes, five of which were previously unknown. Despite broad sequence divergence, structural predictions show conserved SecB tetrameric folds and toxin-antitoxin interfaces. The SecB chaperone phylogeny is incongruent with the identity of the TA component, suggesting modular shuffling during TAC evolution. We demonstrate toxicity of class 2 ART toxins from *Escherichia coli*, *Bacillus subtilis*, and *Streptococcus gordonii*, all of which we show inhibit protein synthesis. All TAC classes can be prophage encoded, indicative of phage-driven mobility and rapid diversification.

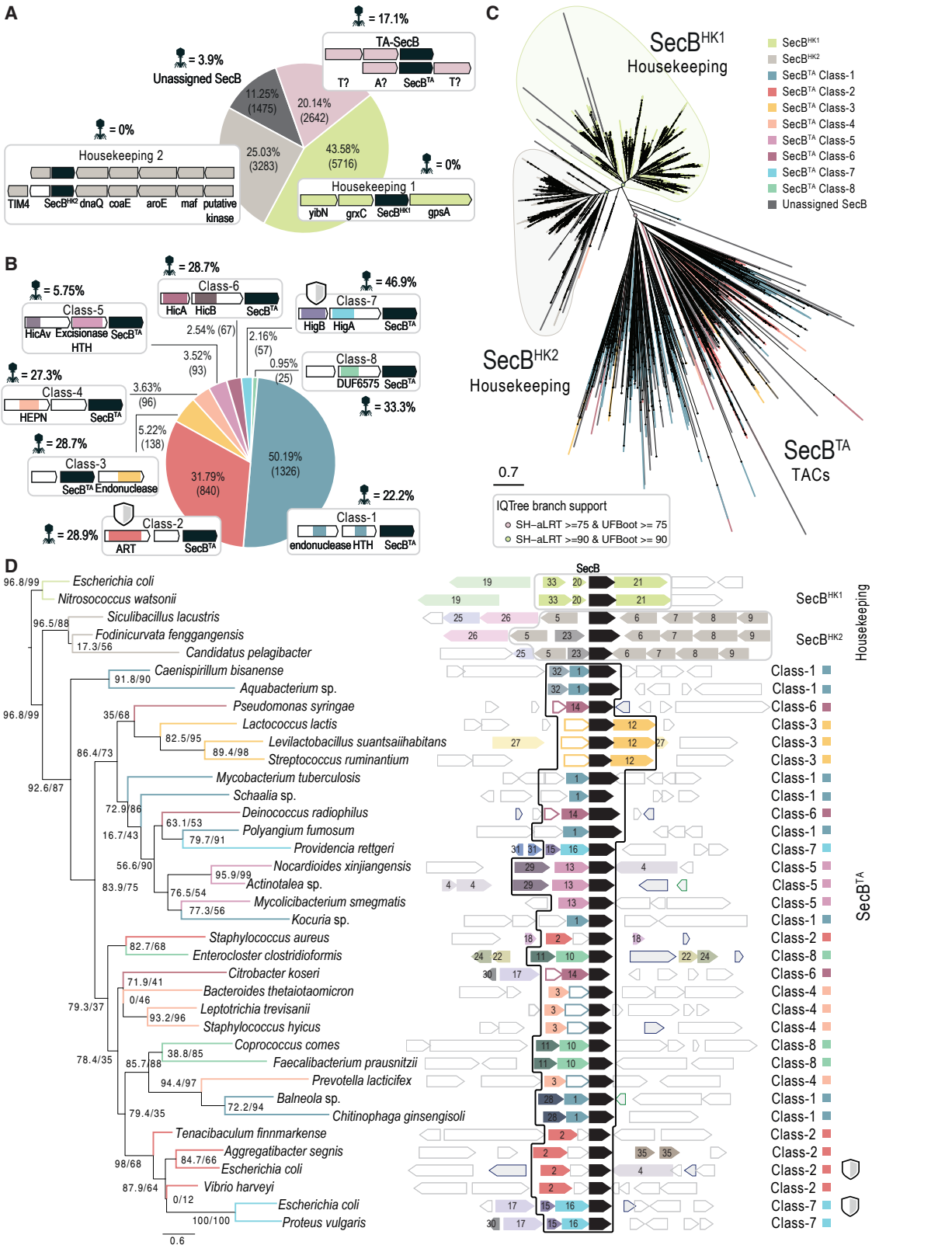
INTRODUCTION

Toxin-antitoxin (TA) systems are diverse two-component genetic modules present in archaea, bacteria, and temperate bacteriophages.[1] TA systems are classified into eight categories, types I–VIII, based on the nature of the antitoxin and its mechanism of toxin neutralization. Among these, type II TAs consist of proteinaceous antitoxins that directly neutralize their cognate toxin through the formation of a non-toxic TA complex.[1] TAs are modular in that they can swap protein partners and are evolutionarily connected to other phage defense systems through sharing and recombination of domains during evolution.[2] Type II TAs have diverse biological functions, including plasmid addiction[3] and antiphage defense.[4–6]

Some type II TA modules are encoded in tricistronic operons with a SecB-like chaperone, forming tripartite toxin-antitoxin-chaperone (TAC) systems.[7,8] Housekeeping SecB is a homotetrameric chaperone found in Pseudomonadota (Proteobacteria) that binds nascent precursor proteins co-translationally, supports their correct fold maturation by preventing premature folding, and delivers membrane-targeted or secretory substrates to SecA for export via the Sec translocon.[9,10] In contrast, the TAC-encoded SecB homologs (SecB^{TA}s) are highly specialized to safeguard the stability of TAC antitoxins.[7] An intrinsically disordered region (IDR) located at the C terminus of the antitoxin, called the chaperone addiction domain (ChAD), defines the TAC TA dependence on SecB^{TA} stabilization.[11,12] Removal of the ChAD region from the HigA antitoxin of *Mycobacterium tuberculosis* HigBAC converts this TAC system into a SecB^{TA}-independent TA.[12]

Recent studies have established the biological function of four *Escherichia coli* TAC systems in phage defense. These are all found in prophage regions of different *E. coli* strains: HigBAC (encoded in NT1F31 strain), CmdTAC PD-T4-9 (ECOR22), CmdTAC (O112ab:H26), and MqsRAC (C496_10).[13–15] CmdTAC PD-T4-9 and CmdTAC O112ab:H26 protect *E. coli* from Tevenvirinae phage T4, *E. coli* C496_10 MqsRAC protects from Tevenvirinae phage T2, and HigBAC from *E. coli* strain NT1F31 can protect against two siphoviruses (i.e., phages with long noncontractile tails), the Queuovirinae coliphage Bas2522 and Lambdavirus λ_{vir}. The ChAD element of HigBAC is critical for defense as it competes for chaperone binding with the phage protein trigger (tail protein gpV).[13] All of the validated TAC toxins act as translational inhibitors: HigB and MqsR are ribonucleases that degrade mRNAs,[16–18] while CmdT ADP-ribosylates mRNA.[13,14]

A foundational bioinformatic survey of TACs in 2013 resulted in the discovery of a diverse range of toxins with nuclease activity.[8] However, TAC systems employing ADP-ribosylating CmdT toxins have gone under the radar until recently,[13–15] suggesting that sensitive sequence searching of the rapidly expanding bacterial sequence space may reveal novel TAC architectures,



(legend on next page)

mechanisms of toxicity, and evolutionary relationships. Here, we surveyed 230 million RefSeq proteins to identify and classify TACs through large-scale SecB homolog detection and gene neighborhood analysis. We define three major SecB subfamilies: two housekeeping subfamilies, SecB^{HK1} and SecB^{HK2}, encoded within well-conserved operons, and a more diverse subfamily, SecB^{TA}, that includes known and novel putative TACs. We sub-classify TACs into eight classes based on the identity of the TA modules, including five that were not described previously. Structural analyses show fold conservation across all SecBs and within each toxin and antitoxin class. We confirmed toxicity for a previously uncharacterized class 6 HicA toxin, along additional class 2 ART toxins. Although novel TAC classes did not afford protection against coliphages in a heterologous *E. coli* system, an association of all TAC classes with prophages suggests that additional classes of phage-encoded TAC defense systems are still likely to be discovered. Strikingly, SecB^{TA} phylogeny breaks TAC class boundaries, revealing extensive modularity in which SecB chaperones recombine with different toxin-antitoxin modules during evolution.

RESULTS

Genomic context analysis classifies SecB homologs into two subfamilies of housekeeping SecBs as well as seven subfamilies of TAC-associated SecB^{TA}s

To discover the full diversity of TAC-associated SecBs encoded in publicly available genomes, we searched for homologs in the NCBI RefSeq protein database with the sensitive iterative sequence search method Jackhmmer[19] using the *E. coli* HigBAC SecB (RefSeq: WP_112844288.1) sequence as the query. We found 13,116 homologs encoded in phyla across all bacteria, with the exceptions of Cyanobacteriota and Planctomycetota (Data S1). Leveraging the observation that TAC-associated SecBs occur in compact three-gene neighborhoods, while housekeeping SecBs are found in longer conserved gene clusters,[8] we used our tool FlaGs2 (Flanking Genes version 2[20]) to analyze and classify their gene neighborhoods (Data S2) and thereby categorize the TAC systems. Four major groups were identified based on neighboring genes (Figure 1A). 43.58% of the *secB* sequences are flanked by two or more components of the classical SecB export operon, *yibN*, *grxC*, or *gpsA*. We refer to this subfamily as SecB housekeeping group 1 (SecB^{HK1}). The second largest group of SecBs, 25.03%, are flanked by a TIM4-like gene and a metabolic operon containing homologs of *dnaQ*, *coaAE*, *aroE*, *maf*, and a putative pyruvate phosphatase dikinase regulatory protein. We call this subfamily SecB housekeeping group 2 (SecB^{HK2}). 20.14% of the sequences were found in tripartite TAC-like systems and were grouped in the SecB^{TA} subfamily, and 11.25% did not have a conserved tripartite architecture nor had any of the conserved genes from the housekeeping operons. These were treated as unassigned SecB. The SecB^{HK1} and SecB^{HK2} subfamilies are mainly found in Pseudomonadota while the SecB^{TA} subfamily is more widespread across bacteria (Data S1) and shows diversity in the composition of its tripartite neighborhoods (Figure 1A).

To sub-classify the SecB^{TA} subfamily further, we used the sequence clustering of FlaGs2[20] to identify conserved homologs in the putative TA components of the TACs, and predicted protein domains with the HH-suite package.[22] Based on the clusters of neighborhood genes, eight classes were identified within the SecB^{TA} subfamily (Figure 1B). The most abundant architectures are TACs where the putative toxin contains an endonuclease domain and the antitoxin has an HTH domain (50.19%). Sequence similarity is lower for this group, but it becomes more cohesive at the structure level (see dedicated section below). This group is classified as class 1 TAC. The second most prevalent group contains putative toxin sequences containing ART domains (31.79%). Previously studied CmdTAC systems are found in this group, which we name class 2 TAC. The class 3 TAC group (5.22%) is defined by a putative endonuclease toxin located downstream of SecB^{TA}. Class 4 TACs (3.63%) are characterized by the presence of a conserved putative toxin with an HEPN domain. Class 5 TACs (3.52%) comprise a HicA domain-containing putative toxin (HicAv) and a putative antitoxin with a DNA-binding domain found in the excisionase RdfS.[23] Class 6 TACs (2.54%) are conserved HicAB TAs associated with SecB^{TA}, while class 7 TACs (2.16%) are HigBAC TACs with conserved HigB-HigA TA modules. Class 8 TACs (0.95%) encode a conserved putative TA module with unknown domains in the toxin and the DUF6575 domain of unknown function in the putative antitoxin, previously associated with phage defense.[14]

To investigate the diversity and relationships among SecB homologs, we aligned the 13,116 sequences to construct a phylogenetic tree using IQ-TREE[21] (Figures 1C and S1A). The tree roughly groups the SecB^{HK1} and SecB^{HK2} subfamilies in support of their gene neighborhood architectures. Long branches obscure relationships; however, there is some statistical support for monophyly of the two groups in the phylogeny (SH-aLRT and UFBoot values over 75; Figure 1C). Full branch support is shown in Figure S1A. The SecB^{TA} subfamily lacks resolution, with little statistical support for deep branches and no grouping of classes into monophyletic clades (Figures 1C and S1A). The lack of resolution at this level is not surprising given that SecB is a small protein, which, in the case of SecB^{TA}s, has high mobility and complex selection pressures arising from host-pathogen

Figure 1. SecB homologs have different genomic contexts, and these are incongruent with the chaperone's phylogenetic relationships

(A) The relative abundance of SecB homologs with different genomic contexts. Phage percentage indicates the incidence of systems found in predicted prophages.

(B) The relative abundance of TAC subfamily classes. Gray shields indicate previously validated defense.

(C) SecB phylogeny constructed by IQ-TREE.[21] Branches are colored by classes as per the inset legend. Branch support is shown with symbols as per the inset legend. Full branch support is found in Figures S2A and S2B.

(D) Gene neighborhood analysis using FlaGs2[20] with SecB homologs as queries (black), mapped on an IQ-TREE phylogeny. Sequences are ordered as per FlaGs2 output by SecB sequence similarity. Colors and numbers indicate protein family clusters, with the annotations of each cluster found in Data S2. Numbers on branches are support values (SH-aLRT/UFBoot). Uncolored genes are unconserved in this dataset. Gray shields indicate previously validated defense.

interactions. To try to better resolve inter-class relationships, we carried out phylogenetic analysis of a subset of SecB homologs, selecting representatives from each TAC class and HK subfamilies (Figure 1D). Here, we do recover some well-supported branches (UFBoot >90), and those relationships are, surprisingly, incongruent with the TAC architecture. For example, class 7 TAC (where the TAs are HigBA homologs) can be found in different parts of the SecB tree, separated by multiple well-supported branches. The same can be said for DUF6575-containing class 8 TACs, and class 6 TACs with HicBA-like components. This is a strong indication that, just as toxins and antitoxins can swap partners during evolution,[2] the chaperone component of TACs is also free to recombine into other TAC architecture classes during evolution.

Previously characterized HigBAC and CmdTAC systems are located in predicted prophage regions.[13] To determine the occurrence of TACs in prophages, we used geNomad[24] to predict if the SecBs are encoded in prophage-like regions. 504 (likely an underestimate due to contig-level genome incompleteness in some cases) were found in predicted prophages across the $SecB^{TA}$ subfamily, with prophage-encoded representatives in every class, while no $SecB^{HK1}$ or $SecB^{HK2}$ were predicted as prophage-encoded (Figure 1A; Table S1, Data S1). Thus, assuming the TACs are generally phage defense systems, the phage versus phage defense that we previously observed with HigBAC and CmdTAC[13] is likely a widespread phenomenon.

SecB homologs possess conserved structural fold, and TAs show structural conservation within TAC classes

The multiple sequence alignment of SecB representatives shows substantial variability at the sequence level, but with conserved residues key for the formation of the tetramer interface across all classes except for class 8 (Figure 2A; expanded alignment in Figure S1B). Importantly, residues responsible for the SecA-SecB interface[25] are conserved within the housekeeping groups, while they are variable within the TAC subfamily, supporting subfunctionalization of these paralogous families in the interaction regions (Figure 2A).

Previous studies have found that *E. coli* $SecB^{HK1}$ and *M. tuberculosis* $SecB^{TA}$ each form tetrameric complexes shaped as a cylinder or barrel (Figure 2B).[26,27] To evaluate the structural conservation of this structural architecture, we clustered the 13,116 homologs and predicted tetrameric structures of 5,094 cluster representatives with AlphaFold3 (Data S1), followed by structural clustering with Foldseek.[28] 80% of the sequences were grouped within 10 structural clusters (Figure 2C). A barrel-shaped tetramer complex prediction is favored for 95% of the clusters, while the remaining 5% include non-barrel-shaped complex predictions (Figure 2D).

To determine the structural diversity of toxins and antitoxins within and between TAC classes, we predicted the structures of class-representative TA components. Pairwise structural comparisons of all the predicted toxin and antitoxin structures show structural coherence within different classes, although to varying degrees (Figure S2; see class-specific results below). For all antitoxins, we see a substantial drop in pLDDT at the C termini (Figures S3A–S3I). Because pLDDT is a well-established indicator of intrinsically disordered regions (IDRs),[29] and ChADs are known IDRs, this pattern is unsurprising. Indeed, the regions after the pLDDT drop-off also correspond to the region predicted (although often with low ipTM) to interact with the SecB component (Figures S3B–S3I). We therefore suggest that pLDDT provides a useful proxy for estimating ChAD boundaries accurate to within a few amino acids. In support of this, the experimentally supported ChADs of HigA1[12] and HigBAC[13] both start at the point where the per-residue pLDDT score drops to between 70 and 50 (Figure S3A).

TA modules of class 1 TACs contain endonuclease toxin domains and antitoxins with HTH domains

TA modules of the most abundant TAC architecture, class 1, are characterized by the presence of an endonuclease domain in the putative toxin and a helix-turn-helix (HTH) domain in the antitoxin (Figure 3A). This class contains the classic TAC HigB1-A1-SecB from *Mycobacterium tuberculosis* strain H37Rv[12,27] and all the TACs identified in a previous systematic search.[8] HMM searches with HH-suite for Pfam protein domains get hits to a variety of related nuclease fold toxin domains belonging to RelE, ParE, YafQ, YoeB, and HigB families (Figure 3B), with the significance of the top hit not being substantially different from the hits to the other related domains, consistent with limited discriminatory power arising from high sequence divergence. Class 1 TAC members can be found across multiple phyla of bacteria (Figure S4A). Both toxins and antitoxins from class 1 TACs are diverse at the sequence level, as seen by the lack of coherent clustering in FlaGs2 analysis (Figure 3A). To explore the potential structural conservation of class 1 toxins and antitoxins, we predicted the structures of the clustered representatives with AlphaFold3 and aligned them with Foldseek (Figure S2). Both toxins and antitoxins show structural conservation, with an average pairwise TM alignment score of 0.64 ± 0.16 and 0.68 ± 0.14, respectively (Table S2), supporting the grouping of these systems into a single class, despite the sequence diversity. The antitoxin representative structure predicts the C-terminal region corresponding to the ChAD domain is unstructured, with a low pLDDT score, comprising the last 38 amino acids (Figure S3B). Structural clustering of the toxins resulted in 80% being grouped into four clusters (Figures 3C and 3D). Representatives of the three most abundant clusters, all with mRNase folds, are shown in Figure 3D. Thus, while class 1 TACs are very divergent at the sequence level, relationships are clearly apparent at the fold level, with indications that the nuclease activity of the toxins is conserved. As more genomes are sequenced in the future, this may enable further sequence-based resolution of this class into subclasses.

Class 2 CmdTAC is a large and broadly distributed class encoding diverse antitoxins

Despite being overlooked as a TAC type until recently, when they were confirmed as phage-phage defense systems,[13–15] the second largest TAC class is defined by the presence of an ART toxin homolog within the two genes upstream of *secB* (Figures 1C and 4A). This class is distributed across multiple phyla, with a larger presence in Bacillota (Figure S4B). Class 2 TACs show high sequence diversity in potential antitoxins, meaning that, as with class 1 antitoxins, they do not get automatically clustered

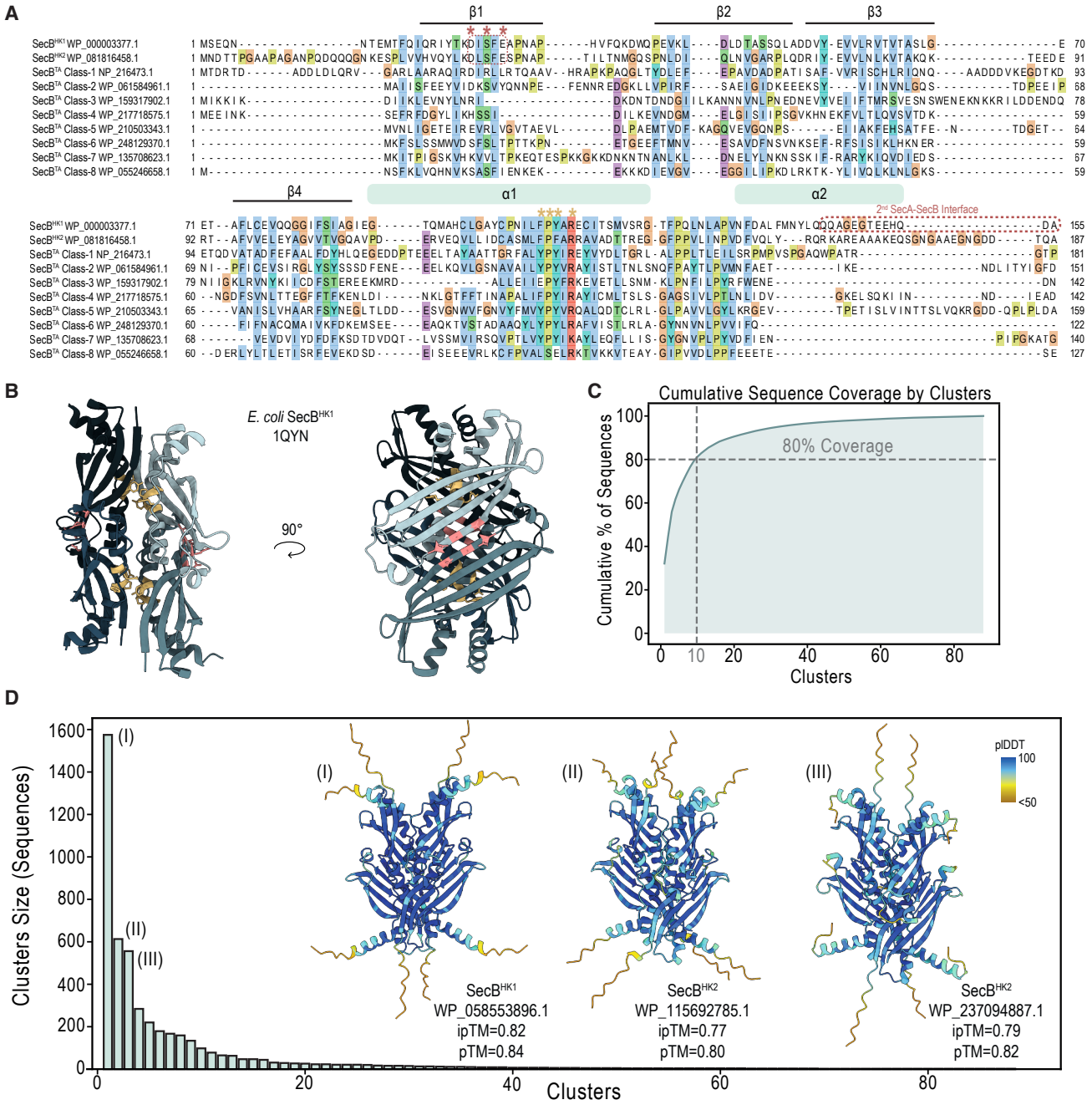


Figure 2. SecB homologs in TAC neighborhoods are divergent at the sequence level but conserved at the structural level

(A) Sequence alignment of SecB homologs from the three subfamilies. Secondary structure is shown above, and gold asterisks indicate key tetramer interface residues, while pink asterisks and the dotted CTD region indicate SecB^{HK1}:SecA interface residues.[25] See Figure S1B for a fuller alignment.

(B) Classical *E. coli* SecB^{HK1} structure (PDB: 1QYN).[26] Colors show individual chains, and conserved interfacing residues from the first alpha helix (marked in A) are shown in gold. SecA:SecB interface residues are shown in pink.

(C) Cumulative SecB homolog sequence coverage by cluster.

(D) Structural clusters ordered by size. Inset: structural predictions of representative SecB from the top three largest clusters. Structures are colored by pLDDT score according to the upper-right key, and the visible surfaces show the regions with a match to known endonuclease domains with HHpred or Foldseek, as indicated.

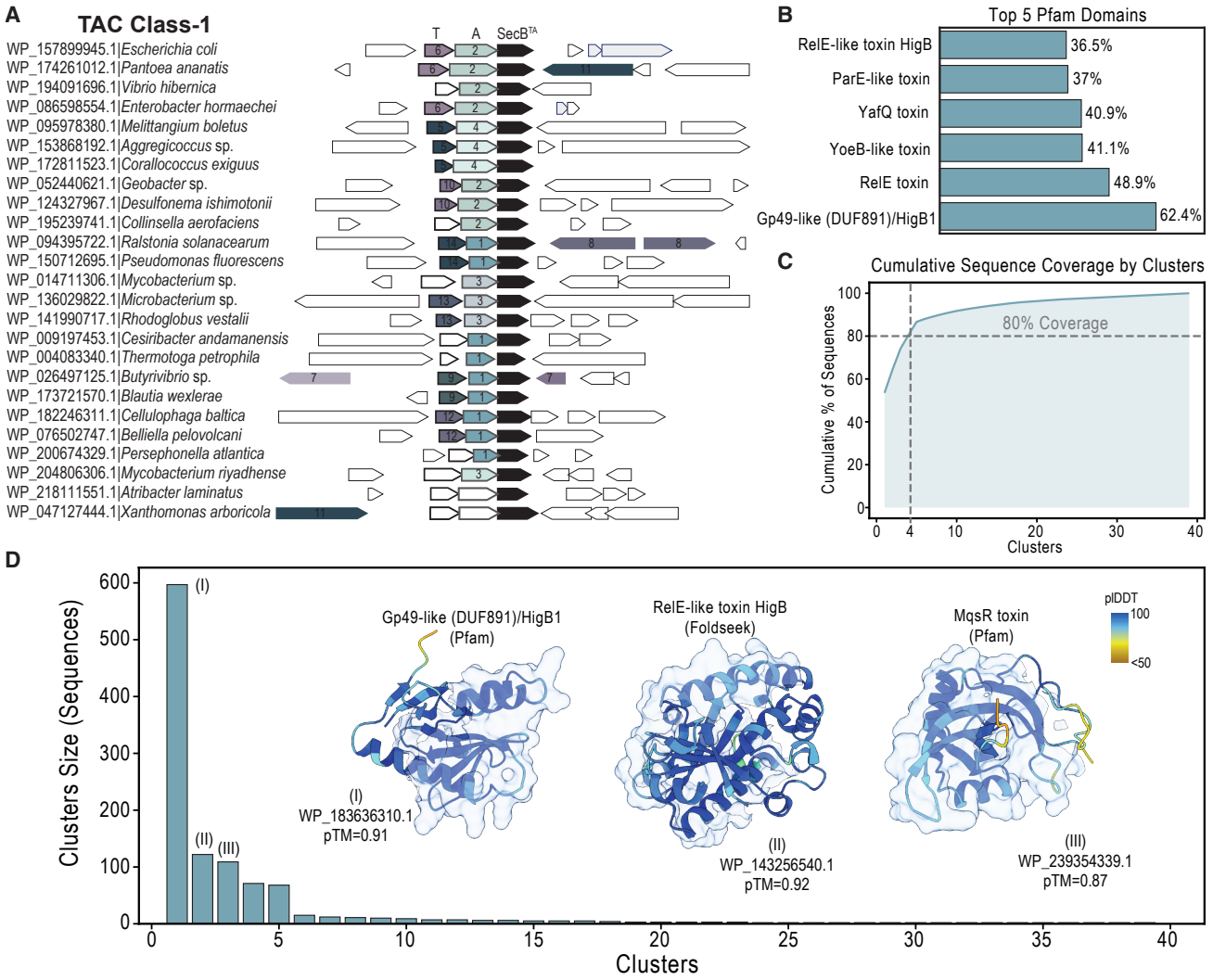


Figure 3. Class 1 TACs employ diverse but related RNase toxins

(A) Gene neighborhood analysis of class 1 TACs. Colors and numbers are as per Figure 1, with SecB homologs in black. Thicker border white genes indicate that a conserved structural fold is identifiable even though sequence similarity is too low to determine. Annotations of each cluster are found in Data S2.

(B) Relative abundance of putative toxin domains.

(C) Cumulative toxin sequence coverage by cluster.

(D) Toxin structural clusters ordered by size. Inset structural predictions of the representative sequence from the top three largest clusters. The structures are colored by pIDDT score. Surface indicates functional domain annotation.

together by Jackhmmer in the FlaGs2 pipeline (Figure 4A). Toxin and antitoxin pairwise structural alignments give a mean TM score of 0.69 ± 0.17 and 0.49 ± 0.13 for the toxin and antitoxins, respectively (Figure S2; Table S2). This indicates that the toxins in this class are likely homologs, while the antitoxins are more ambiguous, with greater structural diversity.[30] AlphaFold3 predicts a complex for the TAC components of the class representative, with the TA dimer interfacing with the $SecB^{TA}$ tetramer at the A:C through the C-terminal IDR of the putative antitoxin (Figure 4B). The low global interface quality of the complex (ipTM 0.36; Figure 4B) indicates unreliable interfaces; nevertheless, the T:A interface score (ipTM 0.66; Figure 4B) is good and indicates that the low ipTM score likely comes from the A:C interface, which is known to be challenging to model confidently.[13] The structural prediction of the *Streptococcus gordonii* TAC antitoxin suggests an unstructured ChAD on the last 38 C-terminal amino acids (Figure S3C).

While CmdTACs have previously been validated as genuine TAC systems from *E. coli*,[14,15] this has not previously been tested for CmdTACs from other bacterial lineages. We have tested the toxicity of three TAC ART toxins encoded by *E. coli* MOD1-EC6301, *Bacillus subtilis* BSn5, and *Streptococcus gordonii* SK184 (Figure 4C). When expressed from a pBAD33 derivative under the control of an arabinose-inducible P_{BAD} promoter and a strong Shine-Dalgarno motif, the *B. subtilis* BSn5 ART toxin fully suppresses the growth of the *E. coli*

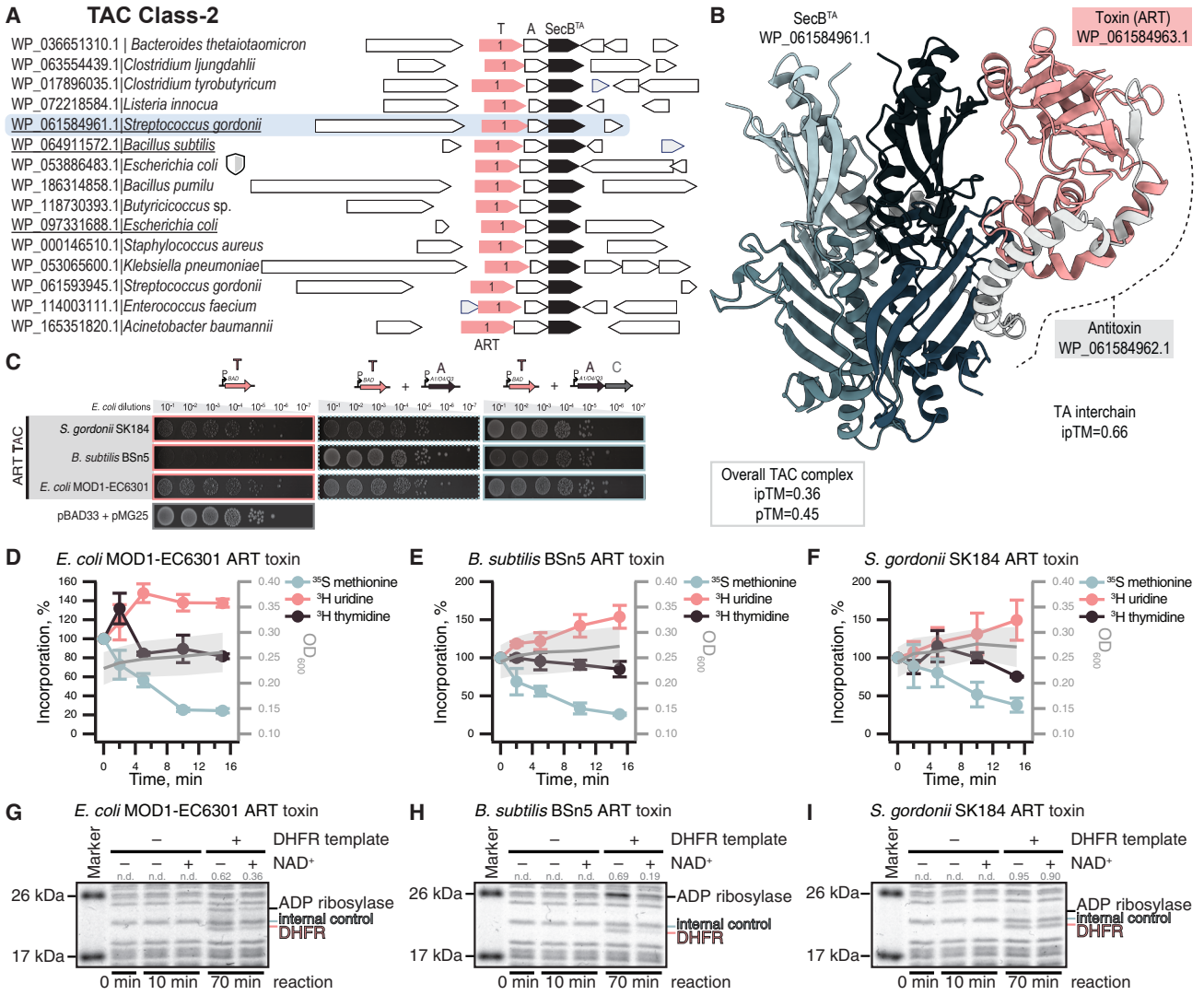


Figure 4. Class 2 TAC toxins are CmdTA-like ART domain-containing translational inhibitors

(A and B) (A) Gene neighborhoods of class 2 TAC *secB* genes. SecB homologs are shown in black, and ART toxin homologs in pink. Colors and numbers are as per Figure 1. Thicker borders indicate conserved structural folds. The highlighted system is shown as a predicted complex in (B). Gray shields indicate previously known phage defense. Annotations of each cluster are found in Data S2. (B) Structural prediction of the *S. gordonii* TAC class 2 complex.

(C) Validation of TAC systems through toxicity neutralization assays. Cultures of *E. coli* strains transformed with pBAD33 and pMG25 vectors or derivatives expressing putative TAC toxins and TAC antitoxins (or antitoxin-chaperon pairs), correspondingly, were adjusted to OD_{600} 1.0, serially diluted from 10^1- to 10^8-fold, and spotted on LB medium supplemented with appropriate antibiotics and inducers (0.2% arabinose for toxin induction as well as either 500 μM IPTG for induction of either antitoxin or the antitoxin-chaperon pair).

(D–F) Metabolic labeling assays with *E. coli* BW25113 expressing ART TAC toxins from *E. coli* MOD1-EC6301 (WP_097331689.1) (D), *B. subtilis* BSn5 (WP_015715417.1) (E), and *S. gordonii* SK184 (WP_061594493.1) (F).

(G–I) Cell-free expression assays with *E. coli* MOD1-EC6301 (G), *B. subtilis* BSn5 (H), and *S. gordonii* SK184 (I). Quantification of the DHFR band intensities relative to the internal control is shown in gray.

BW25113 *uspC-flhDC*::IS5 (VHB17) host strain.[31] *E. coli* MOD1-EC6301 and *S. gordonii* SK184 ART toxins also inhibit growth but not fully. We tested the neutralization of TAC toxins through ectopic co-expression of either antitoxin alone or antitoxin encoded in a bicistronic arrangement with the chaperone. In all cases the toxicity was, albeit to varied degrees, counteracted by co-expression of the antitoxin-chaperone pair. Furthermore, co-expression of the antitoxin alone neutralizes *B. subtilis* BSn5 and *E. coli* MOD1-EC6301 ARTs. Importantly, in these experiments the toxin was expressed from one plasmid (a pBAD33 derivative), while the antitoxin (either alone or together with the chaperone) was expressed from another (a pMG25 derivative). This architecture could interfere with efficient co-translational assembly of the three polypeptides encoded by the TAC operon, thus reducing the efficiency of toxin neutralization.

E. coli CmdT toxins have previously been found to inhibit translation through sequence-specific mRNA modification.[13,14] To assay the mechanism of toxicity of our newly validated CmdT toxins, we performed metabolic labeling assays. As with the *E. coli* O112ab:H26 ART toxin,[13] *E. coli* MOD1-EC6301, *B. subtilis* BSn5, and *S. gordonii* SK184 ART toxins specifically interfere with the incorporation of ^{35}S-methionine, suggesting an effect on translation (Figures 4D–4F). To establish that the effect is direct, we tested the effect of the toxin on the production of dihydrofolate reductase (DHFR) in a PURE (Protein synthesis Using Recombinant Elements) *E. coli* cell-free protein synthesis system.[32] The toxin was produced *in situ* in the PURE system both in the presence and absence of the NAD^+ cofactor, which is essential for ART enzymatic activity, and the DHFR production was tested. While *E. coli* MOD1-EC6301 (Figure 4G) and *B. subtilis* BSn5 (Figure 4H) ART toxins do inhibit DHFR production in the PURE system, we detect no inhibitory effect for *S. gordonii* SK184 (Figure 4I). A possible explanation for the lack of efficient inhibition by *S. gordonii* SK184 is that, as mRNA-targeting ART CmdTs are sequence-specific,[13,14] it is also possible that the DHFR-encoding mRNA is a poor substrate in this case.

Class 3 TAC systems contain conserved endonuclease domain-containing toxins and sequence-diverse antitoxins with structurally conserved folds

We found 138 SecB^{TA}s with a conserved downstream endonuclease-encoding gene and a diverse third component upstream of the chaperone (Figure 5A). This is an unusual TAC architecture and is relatively limited in taxonomy, being exclusively within Bacillota (Figure S4C). Structural predictions of the TAC complex assign the first component of this tripartite operon as the putative antitoxin, interfacing with the SecB^{TA} tetramer through the C-terminal ChAD (Figures 5B and S3D). The global interface quality of the complex (ipTM = 0.63) and the T:A interchain ipTM of 0.82 indicate high confidence in the overall model and the TA dimer.[30] Despite the lack of sequence conservation for the antitoxin, structural alignments show that there is fold similarity, with an average TM alignment score of 0.62 ± 0.15 (Figure S2; Table S2). Toxicity was tested but could not be confirmed in our system (Figure S5A). Although this system is particularly novel in terms of architecture, the lack of toxicity and an especially limited taxonomic distribution did not encourage us to move to phage defense assays for this system.

Class 4 TACs encode the HEPN-MNT TA module

We found 96 SecB^{TA}s associated with a putative TA pair defined by a conserved HEPN domain in the putative toxin (class 4; Figure 1C). While the HEPN gene is conserved in the class, the other component shows considerable sequence diversity (Figure 6A). Class 4 TACs can be found across bacteria, with the majority found within Bacteroidota (Figure S4D). AlphaFold3 predicts the putative antitoxin interfacing with the HEPN protein and through the predicted ChAD with the SecB^{TA} tetramer (Figures 6B and S3E). The global interface quality of the complex is low (ipTM = 0.27), while the interchain ipTM between the toxin and the antitoxin is much more confident at ipTM = 0.74 (Figure 6B). The antitoxins display structural conservation, with an average pairwise TM alignment score of 0.74 ± 0.14 (Figure S2; Table S2). Foldseek annotation indicates structural homology to the MNT antitoxin from the HEPN-MNT TA system that adenylates the cognate HEPN toxin to inactivate it.[33,34]

To test the toxicity of the HEPN toxin, we expressed the predicted toxins from *B. thetaiotaomicron* and *S. epidermidis* as above. However, neither predicted toxin showed activity in the *E. coli* surrogate host strain (Figure S5A). Because this is a particularly novel TAC class containing a known phage defense-associated domain (HEPN), we nevertheless tested the complete TAC system from *B. thetaiotaomicron* for an ability to defend against coliphages, in case toxicity might be triggered only during infection. However, we observed no protective activity (Figure S5B), consistent with a lack of HEPN-mediated toxicity that would be crucial for phage defense via abortive mechanisms.

Class 5 TACs encode an excisionase DNA-binding domain-containing protein and sometimes no predicted toxin component

93 SecB^{TA}s in our dataset are associated with a conserved upstream gene encoding a protein carrying a DNA-binding domain (HTH) found in excisionases, DNA-binding proteins that direct site-specific excision of integrated viral DNA[35,36] (class 5; Figure 1B). One-third of those are associated with a conserved HicAv toxin domain encoded upstream of the excisionase-like gene (Figure 5C). Structural prediction of the TAC complex places the excisionase-like protein as the likely antitoxin, binding to the HicAv domain gene and the chaperone through the C-terminal ChAD region (Figures 5D and S3F). However, the global interface quality of the complex is low (ipTM = 0.36), while the interchain ipTM between the toxin and the antitoxin is medium confidence at 0.52. Both the putative toxin and antitoxin show structural conservation within the class (Figure S2). The toxins have a mean TM alignment score of 0.73 ± 0.14, and the antitoxins have a mean value of 0.70 ± 0.20 (Table S2). Most of the members of this class are found within Actinomycetota (Figure S6A). Since most of the members in this group lack the HicAv-encoding toxin gene, we tested the toxicity of the single protein associated with SecB in this system from *Mycolicibacterium smegmatis.* Although the presence of a predicted C-terminal IDR suggests that it is an antitoxin, in the absence of a predicted co-directional toxin gene—and with predicted similarity to an excisionase—we can be less sure of its role in the system. Furthermore, we also tested a putative HicAv toxin adjacent to the excisionase in *N. xinjiangensis.* However, neither *M. smegmatis* nor *N. xinjiangensis* proteins showed toxicity upon expression in *E. coli* (Figure S5A).

Class 6 HicABC TACs have conserved toxin and antitoxin components

We found 67 HicABC TACs, and this architecture was categorized as class 6 (Figure 1C). All the TA components of this class are conserved enough at the sequence level to be clustered together by FlaGs (Figure 6A) and show structural conservation (Figure S2). Toxins have a mean TM alignment score of

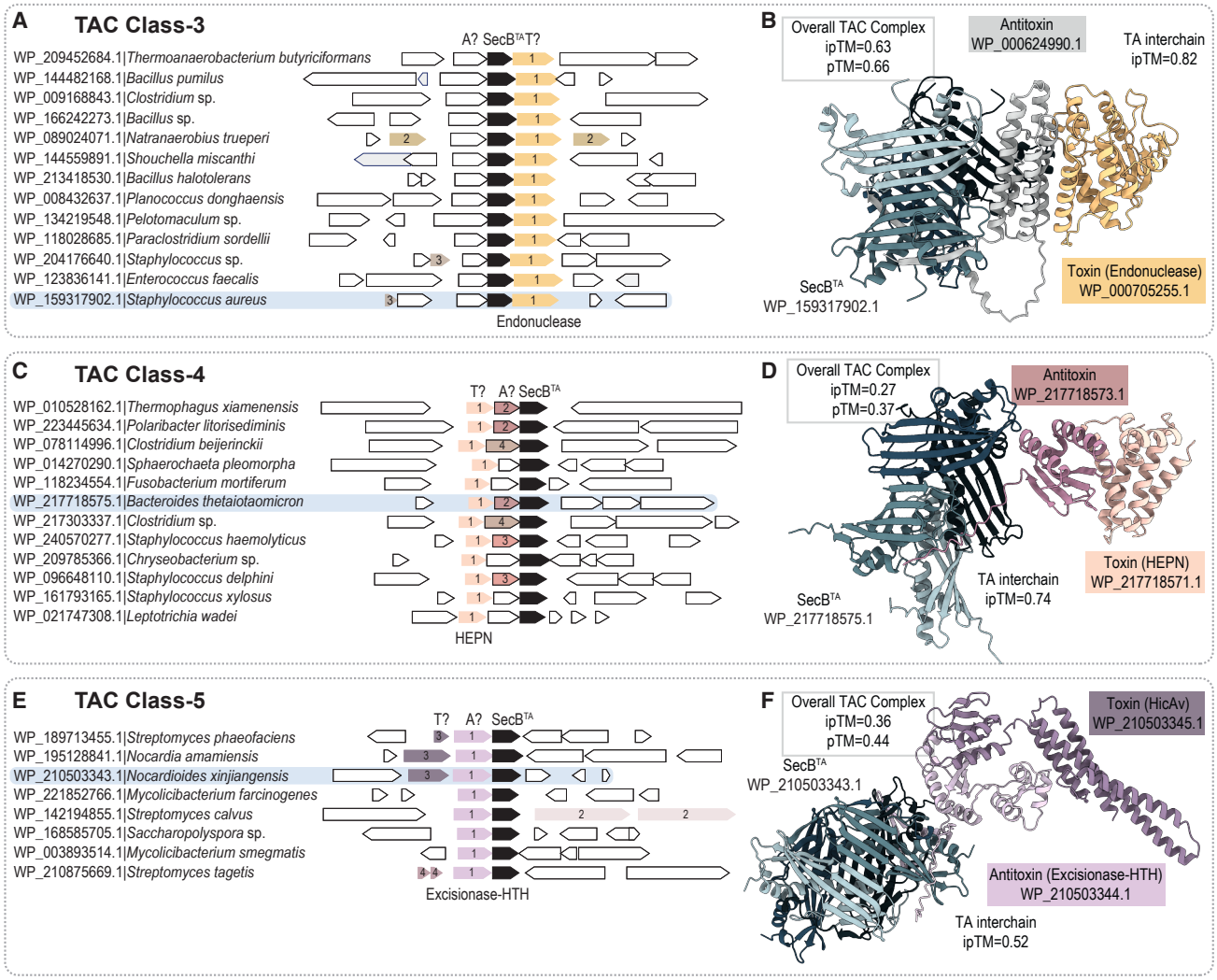


Figure 5. Class 3 TACs have a conserved endonuclease domain toxin with a structurally conserved antitoxin, class 4 TACs have a conserved HEPN toxin with a structurally conserved MNT antitoxin, and class 5 TACs comprise HicAv and excisionase homologs

(A) Gene neighborhood analysis of class 3 TACs. SecB^{TA} is in black, colors indicate protein family clusters, and black border genes indicate a conserved structural fold. Highlighted system is shown as a predicted structure in (B).

(B) Structural prediction of the *S. aureus* class 3 TAC complex.

(C) Gene neighborhood analysis of class 4 TACs. SecB^{TA} is shown in black, colors indicate protein family clusters, and black border genes indicate a conserved structural fold. Highlighted system shown as a predicted structure in (D).

(D) Structural prediction of the *B. thetaiotaomicron* class 4 TAC complex.

(E) Gene neighborhood analysis of class 5 TACs. SecB^{TA} is shown in black, and colors indicate protein family clusters. The highlighted system is shown as a predicted complex in (F).

(F) Structural prediction of the *Nocardioides* sp. Class 5 TAC complex. Annotations of each protein family cluster are found in Data S2.

0.82 ± 0.10, and the antitoxins have a mean value of 0.67 ± 0.11 (Table S2). The toxins contain the HicA RNase domain, while the antitoxins possess a conserved HicB domain followed by a variable and unstructured ChAD region (Figure S3G). TACs in this class can be found in multiple bacterial phyla, predominantly within Gammaproteobacteria (Figure S6B). Structural prediction of the TAC complex places the TA dimer interfacing with the SecB^{TA} tetramer through the ChAD (Figure 6B). This results in a low-confidence complex, with an overall global interface quality of 0.42. However, as with the other classes, the T:A interface score is much better supported with an ipTM of 0.86.

To test the toxicity of the putative HicA toxin, we expressed the predicted toxin from *C. koseri* and a predicted catalytically dead mutant H45A, based on previous HicAB TA studies[37,38] (Figure 6C), as above. Toxicity of the wild-type protein was confirmed, and the active site-substituted variant was non-toxic (Figure 6D). Given this positive result, we tested the complete TAC system for its ability to defend against coliphages. However, we observed no defense (Figure S5C).

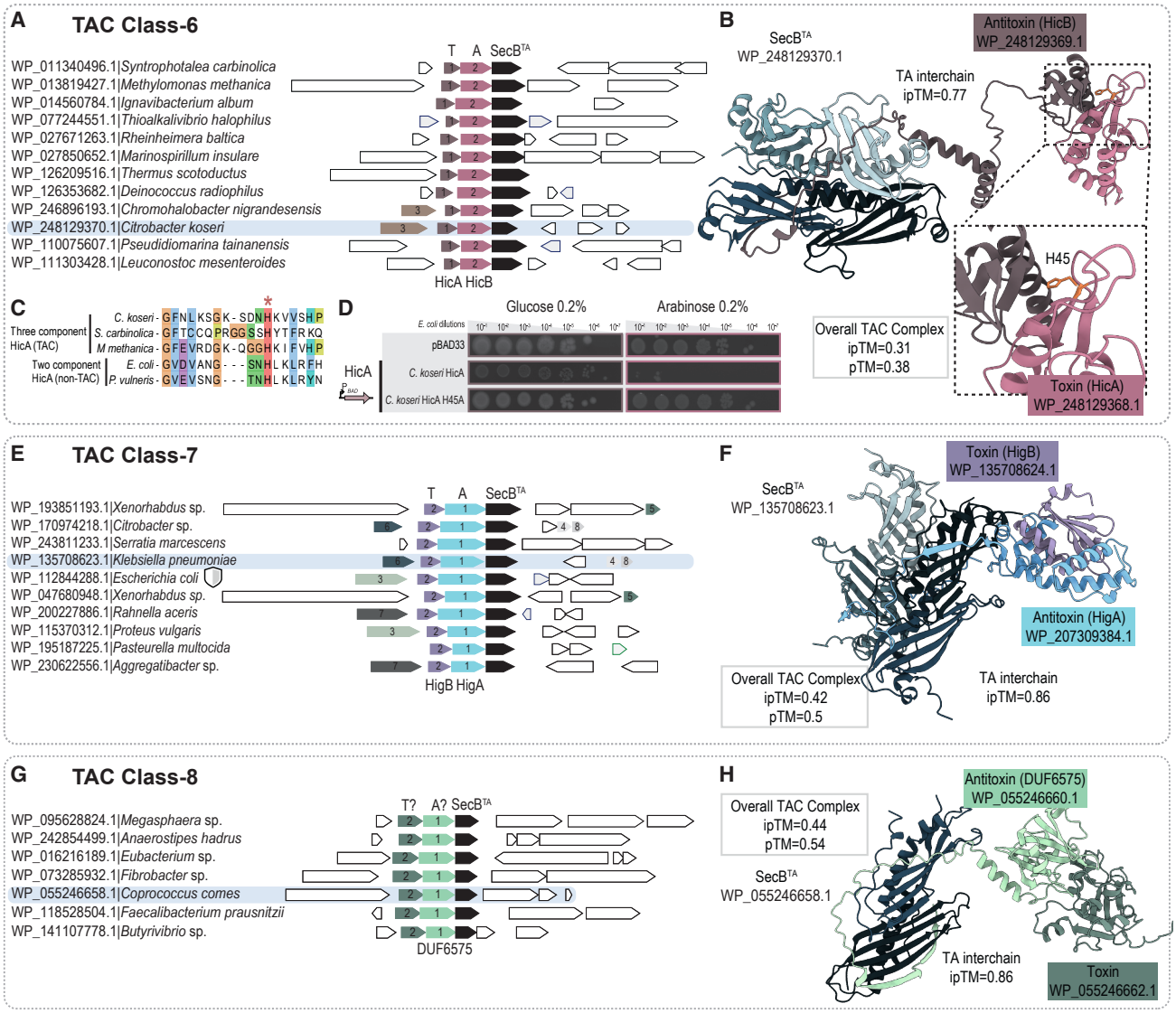


Figure 6. Class 6 TACs have a conserved HicAB TA pair, class 7 TACs are composed of a conserved HigBA TA system, and class 8 TACs show a conserved TA pair with a DUF6575-containing antitoxin

(A) Gene neighborhood analysis of class 6 TACs. SecB^{TA} is shown in black, colors indicate protein family clusters, and black border genes indicate a conserved structural fold. The highlighted system is shown as a predicted complex in (B).

(B) Structural prediction of the *C. koseri* class 6 TAC complex. TA interface is shown in the inset, with H45 colored in pink.

(C) Sequence alignment from the active site of HicA toxin homologs. The conserved histidine residue H45 of *C. koseri* HicA is marked with a red asterisk.

(D) Validation of *C. koseri* HicA toxicity. Overnight cultures of *E. coli* strains transformed with pBAD33 expressing either wild-type HicA or its H45A variant were adjusted to OD_{600} 1.0, serially diluted from 10^1- to 10^8-fold, and spotted on LB medium supplemented with appropriate antibiotics and inducers (0.2% arabinose).

(E) Gene neighborhood analysis of class 7 TACs. SecB^{TA} is shown in black, and colors indicate protein family clusters. The highlighted system shown is as a predicted complex in (F). Gray shields indicate previously validated defense.

(F) Structural prediction of the *K. pneumoniae* class 7 TAC complex.

(G) Gene neighborhood analysis of class 8 TACs. SecB^{TA} is shown in black, and colors indicate protein family clusters. The highlighted system is shown as a predicted complex in (H).

(H) Structural prediction of the class 8 TAC complex from *C. comes*. Annotations of each protein family cluster are found in Data S2.

Class 7 HigBAC TACs have conserved toxin and antitoxin components

We found 57 HigBAC TACs and categorized this architecture as class 7 (Figure 1C). All the TA components of this class are conserved enough at the sequence level for them to be clustered together with FlaGs (Figure 6D), despite this architecture being incongruent with the SecB phylogeny (Figure 1C). The FlaGs clustering also shows that these are distinct from the class 1 toxins, which are distant HigB relatives with similar folds (Figures 1C, 3D, and S2). The toxins contain the HigB RNase

domain, while the antitoxins possess a conserved HigA domain followed by a variable and unstructured ChAD region.[13] All the TACs in this class can be found in Gammaproteobacteria, predominantly within Enterobacteriales (Figure S6C). The experimentally validated HigBAC phage defense system[13] belongs to this class. Structural prediction of the class representative TAC complex places the TA dimer interfacing with the SecB^{TA} tetramer through the ChAD, with an overall global interface quality of 0.42 (Figures 6E and S3H). This results in a low-confidence complex; however, the T:A interface score of 0.86 suggests that the A:C interface between the IDR ChAD and SecB^{TA} reduces the overall score.

Class 8 TACs carry a conserved toxin with unknown function and a novel structural fold

Finally, we found 25 SecB^{TA}s associated with a DUF6575 domain containing gene. These are classified as class 8 TACs (Figure 1C). Both putative TA genes in this class were found to be conserved at the sequence and structural levels (Figures 6F and S2; Table S2). This group is found exclusively within Bacillota, with most of its members within Clostridia (Figure S6D), and a strong tendency toward gut bacteria. Structural prediction of the TAC complex from *C. comes* favors a SecB^{TA} dimer binding to the putative TA system through the DUF6575 protein (Figure 6G). However, the global interface quality score below 0.5 and a T:A high interface score of 0.88 indicative of low-confidence ChAD:SecB^{TA} interface (Figure S3I). Domain searches failed to find a known functional domain in the predicted toxin; however, Foldseek suggests fold similarity to the ADP ribosylase cholera toxin. To test the toxicity of the putative toxin, we expressed the predicted toxin from *C. comes* in *E. coli* BW25113 surrogate host as above; no toxicity was detected (Figure S5A).

DISCUSSION

Our broad survey of SecB homologs shows that classical housekeeping SecBs and TAC-associated SecB^{TA}s occupy fundamentally different genomic niches. Housekeeping SecBs (SecB^{HK1} and SecB^{HK2}) occur within conserved operons, retain the hallmarks expected of chaperones supporting SecA-dependent protein export, and are not predicted to be prophage encoded. In contrast, SecB^{TA}s are found in highly variable, often prophage-associated genomic regions and lack the conserved SecA-binding residues. Although the evolutionary order of events cannot be determined with certainty, the most parsimonious scenario, given the current taxonomic distribution of SecB^{HK}s and their conserved neighborhoods, is that housekeeping SecB originated at least before the diversification of Beta-, Gamma-, and Alphaproteobacteria within the Pseudomonadota (Figures S1A and S1B). Given this relatively ancient origin, and the clear mobility of TAC loci, a plausible scenario is that a SecB-like protein was later incorporated into mobile TA modules and subsequently specialized in stabilizing antitoxins as part of infection-triggered defense systems rather than functioning as a secretion chaperone. A shared feature that may have facilitated this transition is that both SecB^{HK}s and SecB^{TA}s recognize and bind intrinsically disordered protein regions.

As microbial conflicts are likely ancient,[39] an alternative interpretation is that a defense-related SecB function evolved first, with the secretion-related role in Pseudomonadota arising later through co-option. One piece of circumstantial support for this comes from the observation that the taxonomic distribution of TACs is broader (although patchier due to mobility) than that of housekeeping SecBs, with representatives in multiple major phyla of bacteria (Data S1).

Whatever the order of events, nearly all SecB homologs—including those within TACs, with the exception of class 8—retain the predicted conserved barrel-shaped tetrameric fold characteristic of SecB, and key tetramer-stabilizing residues remain preserved. In contrast, residues mediating the classical SecA:SecB interaction are not conserved in SecB^{TA}s, consistent with the absence of selection for protein-export functions.

AlphaFold predictions can provide valuable insights into the organization of TAC antitoxins, particularly the ChAD regions that mediate SecB^{TA} binding. Because ChADs are intrinsically disordered, they lack clear sequence markers that define the boundary between the structured antitoxin domain and the disordered ChAD, making this transition difficult to identify from sequence alone. In some antitoxins, the presence of a conserved domain found in non-TAC systems helps indicate where the structured region ends,[12] but this approach is less effective for more divergent or novel antitoxins. Across TAC classes, however, our predicted antitoxin structures consistently show a pronounced drop in pLDDT scores at the C terminus, corresponding to the expected position of the ChAD (Figure S3). Thus, the pLDDT signature provides a practical proxy for estimating the minimal ChAD region and its approximate extent.

One of the most striking findings of our study is the incongruence between the SecB phylogeny and the architecture of the associated TA units. This disconnection suggests extensive recombination, where SecB^{TA} chaperones are repeatedly paired with diverse toxin-antitoxin modules. Such modularity echoes previous observations in TA systems more broadly,[2] where recombination likely enables adaptation to different functional contexts or triggers. In the case of TACs, this modular flexibility may be particularly important for maintaining and adapting defense mechanisms under phage-driven evolutionary pressure. Additionally, while we were unable to validate phage defense here, the frequent prophage association of TAC loci suggests that TACs are often employed as phage-versus-phage warfare machinery. The rapid evolution of phages is likely what allows continual reshuffling and fine-tuning of TAC components to counter evolving threats.

Conclusion

The iterative sequence-search approach applied here identifies multiple previously unrecognized TAC classes and shows that TACs are far more diverse and widespread than earlier appreciated. We find that structural conservation often underpins sequence-level variability. These findings broaden our understanding of TA system diversity and evolution, provide functional validation for a previously unknown TACs, and show that there is much still to be discovered about TACs from across diverse bacterial lineages.

Limitations of the study

By surveying TAC systems broadly across the bacterial Tree of Life, our study uncovers a diversity of candidates that would likely remain undetected in screens limited to *E. coli* and its close relatives. Experimental testing, however, was necessarily carried out in a heterologous *E. coli* system in which many TACs, particularly those from distant lineages, are unlikely to function as they would in their native hosts. This likely explains why phage defense assays did not detect activity despite evidence that TACs can act in antiviral defenses.[13,14] Work in *B. subtilis* and its phages has been instrumental in validating numerous anti-phage defense systems,[40] underscoring the importance of host-phage systems that better match the native biological context. We therefore anticipate that alternative models, such as *B. subtilis*, will be particularly useful for future mechanistic studies of the diverse TACs identified here.

Technical limitations mean our survey likely captures only part of the true diversity and mobility of TAC systems. Prophage prediction is constrained by the quality and completeness of available genome assemblies, and phage-prediction tools may miss highly divergent or cryptic elements. Our analysis also reflects RefSeq data as of February 2023. As additional bacterial lineages are sequenced, especially from understudied clades, we anticipate that many more TAC variants and mobile systems will emerge.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Gemma C. Atkinson (gemma.atkinson@med.lu.se).

Materials availability

All unique/stable reagents, plasmids, bacterial strains, and phages used in this study are available from the lead contact without restriction.

Data and code availability

- This paper does not report original code.
- Alignments, FlaGs2 output, and phylogenetic tree are available for download from (https://github.com/GCA-VH-lab/TAC). Structures are available for download from (https://data-sharing.atkinson-lab.com/TAC_survey/tac_survey_af3_models.tar.gz).
- All data reported in this paper will be shared by the lead contact upon request.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

We thank Lena Shyrokova for the valuable help and guidance with figures. Computations were enabled by the Berzelius resource provided by the Knut and Alice Wallenberg Foundation at the National Supercomputer Centre and by resources provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS), partially funded by the Swedish Research Council through grant agreement no. 2022-06725. The Knut and Alice Wallenberg Foundation (project grant 2020-0037 to G.C.A. and V.H.); The Swedish Research Council (Vetenskapsrådet) grants 2022-01603, 2023-02353, and 2024-06071 to G.C.A.; 2021-01146 and 2024-06059 to V.H.; The Estonian Research Council (PRG2696 to V.H. and T.T.); Göran Gustafsson Foundation for Research in Natural Sciences and Medicine (the Göran Gustafsson Prize to V.H.); and Japan Society for the Promotion of Science (JSPS) KAKENHI (grant no. 25K10354 to T.K.)

AUTHOR CONTRIBUTIONS

Conceptualization, G.C.A. and J.A.N.; methodology, J.A.N., R.O., T.M., and T.K.; investigation, J.A.N., R.O., T.M., T.K., and G.C.A.; writing – original draft, G.C.A., J.A.N., and V.H.; writing – review & editing, all authors; funding acquisition, G.C.A., V.H., and T.K.; supervision, G.C.A., V.H., and T.T.

DECLARATION OF INTERESTS

The authors declare that they have no competing interests.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work, the authors used ChatGPT in order to write the draft of the manuscript. After using this tool or service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Bacterial strains, plasmids, and phages
- METHOD DETAILS
  - Sequence retrieval and gene neighborhood analysis
  - Sequence alignment and phylogenetic analysis
  - Prophage searching
  - Structural predictions and alignment
  - Bacterial growth and phage infections
  - Construction of plasmids
  - TA and TAC toxicity assays
  - Experimental phage infections
  - In vivo functional assays
  - In vitro translation assays
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2026.117024.

Received: August 15, 2025
Revised: December 17, 2025
Accepted: January 30, 2026
Published: March 3, 2026

REFERENCES

1. Jurenas, D., Fraikin, N., Goormaghtigh, F., and Van Melderen, L. (2022). Biology and evolution of bacterial toxin-antitoxin systems. Nat. Rev. Microbiol. *20*, 335–350. https://doi.org/10.1038/s41579-021-00661-1.
2. Ernits, K., Saha, C.K., Brodiazhenko, T., Chouhan, B., Shenoy, A., Buttress, J.A., Duque-Pedraza, J.J., Bojar, V., Nakamoto, J.A., Kurata, T., et al. (2023). The structural basis of hyperpromiscuity in a core combinatorial network of type II toxin-antitoxin and related phage defense systems. Proc. Natl. Acad. Sci. USA *120*, e2305393120. https://doi.org/10.1073/pnas.2305393120.
3. Fraikin, N., and Van Melderen, L. (2024). Single-cell evidence for plasmid addiction mediated by toxin-antitoxin systems. Nucleic Acids Res. *52*, 1847–1859. https://doi.org/10.1093/nar/gkae018.

4. LeRoux, M., and Laub, M.T. (2022). Toxin-Antitoxin Systems as Phage Defense Elements. Annu. Rev. Microbiol. *76*, 21–43. https://doi.org/10.1146/annurev-micro-020722-013730.

5. Song, S., and Wood, T.K. (2020). A Primary Physiological Role of Toxin/Antitoxin Systems Is Phage Inhibition. Front. Microbiol. *11*, 1895. https://doi.org/10.3389/fmicb.2020.01895.

6. Kelly, A., Arrowsmith, T.J., Went, S.C., and Blower, T.R. (2023). Toxin-antitoxin systems as mediators of phage defence and the implications for abortive infection. Curr. Opin. Microbiol. *73*, 102293. https://doi.org/10.1016/j.mib.2023.102293.

7. Bordes, P., Cirinesi, A.M., Ummels, R., Sala, A., Sakr, S., Bitter, W., and Genevaux, P. (2011). SecB-like chaperone controls a toxin-antitoxin stress-responsive system in Mycobacterium tuberculosis. Proc. Natl. Acad. Sci. USA *108*, 8438–8443. https://doi.org/10.1073/pnas.1101189108.

8. Sala, A., Calderon, V., Bordes, P., and Genevaux, P. (2013). TAC from Mycobacterium tuberculosis: a paradigm for stress-responsive toxin-antitoxin systems controlled by SecB-like chaperones. Cell Stress Chaperones *18*, 129–135. https://doi.org/10.1007/s12192-012-0396-5.

9. Eismann, L., Fijalkowski, I., Galmozzi, C.V., Koubek, J., Tippmann, F., Van Damme, P., and Kramer, G. (2022). Selective ribosome profiling reveals a role for SecB in the co-translational inner membrane protein biogenesis. Cell Rep. *41*, 111776. https://doi.org/10.1016/j.celrep.2022.111776.

10. Zhou, J., and Xu, Z. (2003). Structural determinants of SecB recognition by SecA in bacterial protein translocation. Nat. Struct. Biol. *10*, 942–947. https://doi.org/10.1038/nsb980.

11. Texier, P., Bordes, P., Nagpal, J., Sala, A.J., Mansour, M., Cirinesi, A.M., Xu, X., Dougan, D.A., and Genevaux, P. (2021). ClpXP-mediated Degradation of the TAC Antitoxin is Neutralized by the SecB-like Chaperone in Mycobacterium tuberculosis. J. Mol. Biol. *433*, 166815. https://doi.org/10.1016/j.jmb.2021.166815.

12. Bordes, P., Sala, A.J., Ayala, S., Texier, P., Slama, N., Cirinesi, A.M., Guillet, V., Mourey, L., and Genevaux, P. (2016). Chaperone addiction of toxin-antitoxin systems. Nat. Commun. *7*, 13339. https://doi.org/10.1038/ncomms13339.

13. Mets, T., Kurata, T., Ernits, K., Johansson, M.J.O., Craig, S.Z., Evora, G.M., Buttress, J.A., Odai, R., Wallant, K.C., Nakamoto, J.A., et al. (2024). Mechanism of phage sensing and restriction by toxin-antitoxin-chaperone systems. Cell Host Microbe *32*, 1059–1073.e8. https://doi.org/10.1016/j.chom.2024.05.003.

14. Vassallo, C.N., Doering, C.R., and Laub, M.T. (2024). Anti-viral defence by an mRNA ADP-ribosyltransferase that blocks translation. Nature *636*, 190–197. https://doi.org/10.1038/s41586-024-08102-8.

15. Vassallo, C.N., Doering, C.R., Littlehale, M.L., Teodoro, G.I.C., and Laub, M.T. (2022). A functional selection reveals previously undetected anti-phage defence systems in the E. coli pangenome. Nat. Microbiol. *7*, 1568–1579. https://doi.org/10.1038/s41564-022-01219-4.

16. Schureck, M.A., Dunkle, J.A., Maehigashi, T., Miles, S.J., and Dunham, C.M. (2015). Defining the mRNA recognition signature of a bacterial toxin protein. Proc. Natl. Acad. Sci. USA *112*, 13862–13867. https://doi.org/10.1073/pnas.1512959112.

17. Schureck, M.A., Maehigashi, T., Miles, S.J., Marquez, J., and Dunham, C.M. (2016). mRNA bound to the 30S subunit is a HigB toxin substrate. RNA *22*, 1261–1270. https://doi.org/10.1261/rna.056218.116.

18. Yu, V., Ronzone, E., Lord, D., Peti, W., and Page, R. (2022). MqsR is a non-canonical microbial RNase toxin that is inhibited by antitoxin MqsA via steric blockage of substrate binding. J. Biol. Chem. *298*, 102535. https://doi.org/10.1016/j.jbc.2022.102535.

19. Eddy, S.R. (2011). Accelerated Profile HMM Searches. PLoS Comput. Biol. *7*, e1002195. https://doi.org/10.1371/journal.pcbi.1002195.

20. Saha, C.K., Sanches Pires, R., Brolin, H., Delannoy, M., and Atkinson, G.C. (2021). FlaGs and webFlaGs: discovering novel biology through the analysis of gene neighbourhood conservation. Bioinformatics *37*, 1312–1314. https://doi.org/10.1093/bioinformatics/btaa788.

21. Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol. Biol. Evol. *37*, 1530–1534. https://doi.org/10.1093/molbev/msaa015.

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

23. Verdonk, C.J., Marshall, A.C., Ramsay, J.P., and Bond, C.S. (2022). Crystallographic and X-ray scattering study of RdfS, a recombination directionality factor from an integrative and conjugative element. Acta Crystallogr. D Struct. Biol. *78*, 1210–1220. https://doi.org/10.1107/S2059798322008579.

24. Camargo, A.P., Roux, S., Schulz, F., Babinski, M., Xu, Y., Hu, B., Chain, P.S.G., Nayfach, S., and Kyrpides, N.C. (2024). Identification of mobile genetic elements with geNomad. Nat. Biotechnol. *42*, 1303–1312. https://doi.org/10.1038/s41587-023-01953-y.

25. Suo, Y., Hardy, S.J.S., and Randall, L.L. (2015). The basis of asymmetry in the SecA:SecB complex. J. Mol. Biol. *427*, 887–900. https://doi.org/10.1016/j.jmb.2014.12.008.

26. Dekker, C., de Kruijff, B., and Gros, P. (2003). Crystal structure of SecB from Escherichia coli. J. Struct. Biol. *144*, 313–319. https://doi.org/10.1016/j.jsb.2003.09.012.

27. Guillet, V., Bordes, P., Bon, C., Marcoux, J., Gervais, V., Sala, A.J., Dos Reis, S., Slama, N., Mares-Mejía, I., Cirinesi, A.M., et al. (2019). Structural insights into chaperone addiction of toxin-antitoxin systems. Nat. Commun. *10*, 782. https://doi.org/10.1038/s41467-019-08747-4.

28. van Kempen, M., Kim, S.S., Tumescheit, C., Mirdita, M., Lee, J., Gilchrist, C.L.M., Söding, J., and Steinegger, M. (2024). Fast and accurate protein structure search with Foldseek. Nat. Biotechnol. *42*, 243–246. https://doi.org/10.1038/s41587-023-01773-0.

29. Vander Meersche, Y., Diharce, J., Gelly, J.C., and Galochkina, T. (2025). Flexibility or uncertainty? A critical assessment of AlphaFold 2 pLDDT. Structure *33*, 2157–2163.e2. https://doi.org/10.1016/j.str.2025.09.001.

30. Xu, J., and Zhang, Y. (2010). How significant is a protein structure similarity with TM-score = 0.5? Bioinformatics *26*, 889–895. https://doi.org/10.1093/bioinformatics/btq066.

31. Shyrokova, L., Egorov, A.A., Cole, A., Duque-Pedraza, J.J., Tyagi, A., Ernits, K., Mets, T., Kurata, T., Juozapaitis, J., Yang, A.L.J., et al. (2025). Characterization of five environmental phages infecting Escherichia coli K-12 isolated during a phage biology training course. Microbiol. Spectr. *13*, e0227425. https://doi.org/10.1128/spectrum.02274-25.

32. Shimizu, Y., Kuruma, Y., Kanamori, T., and Ueda, T. (2014). The PURE system for protein production. Methods Mol. Biol. *1118*, 275–284. https://doi.org/10.1007/978-1-62703-782-2_19.

33. Yao, J., Zhen, X., Tang, K., Liu, T., Xu, X., Chen, Z., Guo, Y., Liu, X., Wood, T.K., Ouyang, S., and Wang, X. (2020). Novel polyadenylylation-dependent neutralization mechanism of the HEPN/MNT toxin/antitoxin system. Nucleic Acids Res. *48*, 11054–11067. https://doi.org/10.1093/nar/gkaa855.

34. Songailiene, I., Juozapaitis, J., Tamulaitiene, G., Ruksenaite, A., Šulčius, S., Sasnauskas, G., Venclovas, Č., and Siksnys, V. (2020). HEPN-MNT Toxin-Antitoxin System: The HEPN Ribonuclease Is Neutralized by OligoAMPylation. Mol. Cell *80*, 955–970.e7. https://doi.org/10.1016/j.molcel.2020.11.034.

35. Swalla, B.M., Cho, E.H., Gumport, R.I., and Gardner, J.F. (2003). The molecular basis of co-operative DNA binding between lambda integrase and excisionase. Mol. Microbiol. *50*, 89–99. https://doi.org/10.1046/j.1365-2958.2003.03687.x.

36. Van Duyne, G.D., and Landy, A. (2024). Bacteriophage lambda site-specific recombination. Mol. Microbiol. *121*, 895–911. https://doi.org/10.1111/mmi.15241.

37. Manav, M.C., Turnbull, K.J., Jurenas, D., Garcia-Pino, A., Gerdes, K., and Brodersen, D.E. (2019). The E. coli HicB Antitoxin Contains a Structurally

Stable Helix-Turn-Helix DNA Binding Domain. Structure *27*, 1675–1685 e1673. https://doi.org/10.1016/j.str.2019.08.008.

38. Winter, A.J., Williams, C., Isupov, M.N., Crocker, H., Gromova, M., Marsh, P., Wilkinson, O.J., Dillingham, M.S., Harmer, N.J., Titball, R.W., and Crump, M.P. (2018). The molecular basis of protein toxin HicA-dependent binding of the protein antitoxin HicB to DNA. J. Biol. Chem. *293*, 19429–19440. https://doi.org/10.1074/jbc.RA118.005173.
39. Koonin, E.V., Senkevich, T.G., and Dolja, V.V. (2006). The ancient Virus World and evolution of cells. Biol. Direct *1*, 29. https://doi.org/10.1186/1745-6150-1-29.
40. Doron, S., Melamed, S., Ofir, G., Leavitt, A., Lopatina, A., Keren, M., Amitai, G., and Sorek, R. (2018). Systematic discovery of antiphage defense systems in the microbial pangenome. Science *359*, eaar4120. https://doi.org/10.1126/science.aar4120.
41. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera–a visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612. https://doi.org/10.1002/jcc.20084.
42. Remmert, M., Biegert, A., Hauser, A., and Söding, J. (2011). HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. Nat. Methods *9*, 173–175. https://doi.org/10.1038/nmeth.1818.
43. Söding, J. (2005). Protein homology detection by HMM-HMM comparison. Bioinformatics *21*, 951–960. https://doi.org/10.1093/bioinformatics/bti125.
44. Breitwieser, F.P., and Salzberg, S.L. (2020). Pavian: interactive analysis of metagenomics data for microbiome studies and pathogen identification. Bioinformatics *36*, 1303–1304. https://doi.org/10.1093/bioinformatics/btz715.
45. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780. https://doi.org/10.1093/molbev/mst010.
46. Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., and Barton, G.J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics *25*, 1189–1191. https://doi.org/10.1093/bioinformatics/btp033.
47. Capella-Gutierrez, S., Silla-Martinez, J.M., and Gabaldon, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics *25*, 1972–1973. https://doi.org/10.1093/bioinformatics/btp348.
48. R Core Team (2021). R: A language and environment for statistical computing (Vienna, Austria: R Foundation for Statistical Computing). https://www.R-project.org/.
49. Yu, G., Smith, D.K., Zhu, H., Guan, Y., and Lam, T.T.Y. (2017). GGTREE: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol. Evol. *8*, 28–36. https://doi.org/10.1111/2041-210x.12628.
50. Wang, L.G., Lam, T.T.Y., Xu, S., Dai, Z., Zhou, L., Feng, T., Guo, P., Dunn, C.W., Jones, B.R., Bradley, T., et al. (2020). Treeio: An R Package for Phylogenetic Tree Input and Output with Richly Annotated and Associated Data. Mol. Biol. Evol. *37*, 599–603. https://doi.org/10.1093/molbev/msz240.
51. Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A.J., Bambrick, J., et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature *630*, 493–500. https://doi.org/10.1038/s41586-024-07487-w.
52. Steinegger, M., and Söding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nat. Biotechnol. *35*, 1026–1028. https://doi.org/10.1038/nbt.3988.
53. O'Leary, N.A., Cox, E., Holmes, J.B., Anderson, W.R., Falk, R., Hem, V., Tsuchiya, M.T.N., Schuler, G.D., Zhang, X., Torcivia, J., et al. (2024). Exploring and retrieving sequence and metadata for species across the tree of life with NCBI Datasets. Sci. Data *11*, 732. https://doi.org/10.1038/s41597-024-03571-y.
54. Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., and Vinh, L.S. (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. Mol. Biol. Evol. *35*, 518–522. https://doi.org/10.1093/molbev/msx281.
55. Zhang, Y., and Skolnick, J. (2005). TM-align: a protein structure alignment algorithm based on the TM-score. Nucleic Acids Res. *33*, 2302–2309. https://doi.org/10.1093/nar/gki524.
56. Meng, E.C., Goddard, T.D., Pettersen, E.F., Couch, G.S., Pearson, Z.J., Morris, J.H., and Ferrin, T.E. (2023). UCSF ChimeraX: Tools for structure building and analysis. Protein Sci. *32*, e4792. https://doi.org/10.1002/pro.4792.
57. Wickham, H. (2016). ggplot2 : Elegant Graphics for Data Analysis. Use R, 2nd ed. (Springer International Publishing : Imprint: Springer).
58. Quan, J., and Tian, J. (2014). Circular polymerase extension cloning. Methods Mol. Biol. *1116*, 103–117. https://doi.org/10.1007/978-1-62703-764-8_8.
59. Quan, J., and Tian, J. (2011). Circular polymerase extension cloning for high-throughput cloning of complex and combinatorial DNA libraries. Nat. Protoc. *6*, 242–251. https://doi.org/10.1038/nprot.2010.181.
60. Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., 3rd, and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods *6*, 343–345. https://doi.org/10.1038/nmeth.1318.
61. Neidhardt, F.C., Bloch, P.L., and Smith, D.F. (1974). Culture medium for enterobacteria. J. Bacteriol. *119*, 736–747. https://doi.org/10.1128/jb.119.3.736-747.1974.
62. Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods *9*, 671–675. https://doi.org/10.1038/nmeth.2089.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
For strains used see Data S3	N/A	N/A
For bacteriophages used see Data S3	N/A	N/A
Chemicals, peptides, and recombinant proteins		
Phusion High-Fidelity PCR Master Mix	Thermo Fisher Scientific	Cat# 10402678
FastDigest DpnI	Thermo Fisher Scientific	Cat# 10819410
NEBuilder HiFi DNA Assembly Master Mix	New England Biolabs	Cat# E2621L
T4 Polynucleotide Kinase	New England Biolabs	Cat# M0201L
T4 DNA Ligase	Thermo Fisher Scientific	Cat# 10723941
DreamTaq PCR Master Mix	Thermo Fisher Scientific	Cat#11816843
BD Bacto Dehydrated Agar	Thermo Fisher Scientific	Cat#214010
Bacto Tryptone	Thermo Fisher Scientific	Cat#211705
Bacto Yeast Extract	Thermo Fisher Scientific	Cat#212750
Sodium chloride	VWR Chemicals	Cat#27808.366
IPTG	ITW Reagents Panreac	Cat#A1008
Ampicillin sodium salt	Sigma-Aldrich	Cat#A8351
L-(+)-Arabinose	Sigma-Aldrich	Cat#A3256
^{35}S methionine	Hartmann	Cat# ARS0104A
^{3}H uridine	Hartmann	Cat# MT6036
^{3}H thymidine	Hartmann	Cat# MT799
Ammonium chloride	Sigma-Aldrich	Cat#21236.267
Calcium chloride	Sigma-Aldrich	Cat#C1016
Potassium sulfate	Sigma-Aldrich	Cat#P9458
Magnesium chloride hexahydrate	VWR Chemicals	Cat#M2670
Sodium chloride	VWR Chemicals	Cat#27808.366
Iron (II) sulfate heptahydrate	Sigma-Aldrich	Cat#F8633
Dipotassium hydrogen phosphate	VWR Chemicals	Cat#33612.268
Ammonium molybdate tetrahydrate	Sigma-Aldrich	Cat#M1019
Boric acid	Sigma-Aldrich	Cat#B0394
Cobalt (II) chloride hexahydrate	Sigma-Aldrich	Cat#255599
Copper (II) sulfate pentahydrate	Sigma-Aldrich	Cat#209198
Manganese (II) chloride tetrahydrate	VWR Chemicals	Cat#25222.233
Zinc sulfate heptahydrate	Sigma-Aldrich	Cat#Z0251
MOPS	Sigma-Aldrich	Cat#A1076
Tricine	Sigma-Aldrich	Cat#T9784
Trichloroacetic acid	Thermo Fisher Scientific	Cat#A11156.0B
Glycerol	VWR Chemicals	Cat#24388.364
Magnesium sulfate heptahydrate	VWR Chemicals	Cat#25167.298
Imidazole	Thermo Fisher Scientific	Cat#A10221.36
β-mercaptoethanol	Sigma-Aldrich	Cat#M6250
Anhydrotetracycline hydrochloride (aTc)	Thermo Fisher Scientific	Cat#J66688-MB
Brilliant Blue G250	Sigma-Aldrich	Cat#27815
Critical commercial assays		
Monarch PCR & DNA Cleanup Kit	New England Biolabs	Cat# T1030
QIAprep Spin Miniprep Kit	QIAGEN	Cat# 27104

(*Continued on next page*)

Continued		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
PURExpress *in vitro* protein synthesis kit	New England Biolabs	Cat# E6800
RNase Inhibitor Murine	New England Biolabs	Cat# M0314S
Oligonucleotides		
For primers used for cloning of *E. coli* plasmids see Data S3	N/A	N/A
Recombinant DNA		
For *E. coli* plasmids see Data S3	This paper	N/A
Software and algorithms		
SnapGene v8.2.1	GSL Biotech LLC	RRID:SCR_015052
UCSF ChimeraX v1.9	Pettersen et al.[41]	RRID:SCR_015872
Igor Pro v7.08	WaveMetrics, Inc.	RRID:SCR_000325
GraphPad Prism v10.4.1	GraphPad Software, LLC	RRID:SCR_002798
Adobe Illustrator v29.8.2	Adobe Inc.	RRID:SCR_010279
HMMER v3.3.2	Eddy[19]	RRID:SCR_005305
FlaGs2	Saha et al.[20]	https://github.com/GCA-VH-lab/FlaGs
HHblits HHsuite v3.3.0	Remmert et al.[42]	https://github.com/soedinglab/hh-suite
HHsearch HHsuite v3.3.0	Söding[43]	https://github.com/soedinglab/hh-suite
Pavian	Breitwieser and Salzberg.[44]	RRID:SCR_016679
MAFFT L-INS-i v7.505	Katoh and Standley.[45]	N/A
Jalview v.2.11.2.0	Waterhouse et al.[46]	RRID:SCR_006459
trimAl v1.2	Capella-Gutierrez et al.[47]	RRID:SCR_017334
IQ-TREE 2 v2.1.4	Minh et al.[21]	RRID:SCR_017254
R v4.3.0	R Core Team[48]	N/A
GGTREE v3.10.0	Yu et al.[49]	RRID:SCR_018560
Treeio v1.26.0	Wang et al.[50]	N/A
geNomad v1.11.2	Camargo et al.[24]	N/A
AlphaFold v3.0.1	Abramson et al.[51]	N/A
MMseqs2 Release 15-6f452	Steinegger and Söding[52]	N/A
NCBI-Datasets v16.41.0	O'leary et al.[53]	N/A
Other		
Grant Instruments OLS Aqua Pro Orbital and Linear Shaking Bath	Thermo Fisher	Cat#15345802
MaxQ 6000 Incubated/Refrigerated Stackable Shaker	Thermo Scientific	Cat#SHKE6000-8CE
Cytiva Whatman Binder-Free Glass Microfiber Filters, Grade GF/C Sheets	Thermo Fisher	Cat#11340594
Tri-Carb 4910TR-scintillation counter	Perkin Elmer	Cat#425-209
EcoLite(+) Liquid Scintillation Cocktail	MP Biomedicals	Cat#IC882475
Scintillation vials	Sarstedt	Cat#73.662.500
Alignments, FlaGs2 output, and phylogenetic tree	This paper	https://github.com/GCA-VH-lab/TAC
Predicted alphafold structures	This paper	https://data-sharing.atkinson-lab.com/TAC_survey/tac_survey_af3_models.tar.gz

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Bacterial strains, plasmids, and phages

Bacterial strains, plasmids and phages used in this study are listed in Data S3. Unless stated otherwise, microbiology experiments were performed using motile *E. coli* BW25113 strain, BW25113 *uspC-flhDC*::IS5 (VHB17).[31]

METHOD DETAILS

Sequence retrieval and gene neighborhood analysis

SecB homologues sequences were identified in bacterial proteomes from NCBI RefSeq database (downloaded in February 2023) using Jackhmmer from HMMER version 3.3.2[19] using an E-value threshold of $1e^{-10}$ and HigBAC SecB (WP_112844288.1) as the query. To avoid the inclusion of aberrant predicted fusion proteins, we excluded sequences longer than 250 residues (SecB homologues are ~150 residues) as well as sequences annotated as partial because they occur at contig boundaries. The remaining 13,116 homologues were grouped in clusters with a minimum identity of 0.4, coverage of 0.8, and coverage mode 0, using MMseqs cluster from MMseqs2 Release 15-6f452.[52] The 849 representative sequences were used as queries for FlaGs2[20] (https://github.com/GCA-VH-lab/FlaGs2) (Data S2). HHblits and HHsearch from HH-suite[22] were used to identify protein domains from FlaGs2 clustered families from the representative subset. FlaGs2 analysis was performed with the full set of homologues with two flanking genes, and TAC classes were defined by the identified protein families from the preliminary analysis. Taxonomic distribution was visualised with Pavian.[44]

Sequence alignment and phylogenetic analysis

Sequence alignments with fewer than 200 sequences were generated using MAFFT v7.505[45] with the L-INS-i strategy and otherwise default parameters. Larger datasets were aligned using MAFFT v7.505 with default settings. Alignments were visualised using Jalview v.2.11.2.0.[46] Alignment positions with >50% gaps were removed with trimAl v1.2.[47] Phylogenetic analysis was carried out using IQ-TREE 2 v2.1.4 with parameters -B 1000 -alrt 1000.[21,54] The 'LG + I + G' substitution model was determined as the best-fit by the program. Trees were visualised with attached data in R v4.3.0 using GGTREE v3.10.0,[49] Treeio v1.26.0.[50]

Prophage searching

The genomic file for each assembly from the SecB homologue search (Data S1) was downloaded from NCBI using NCBI-Datasets v16.41.0 (datasets download genome).[53] Prophage prediction was performed with geNomad version 1.11.2 using end-to-end annotation.[24]

Structural predictions and alignment

The 13,116 SecB homologues were grouped into 5,094 clusters using MMseqs cluster (Identity 0.9, Coverage 0.9, and Coverage mode 0). Protein structure predictions of representative SecBs were made with AlphaFold v3.0.1 with standard runtime parameters and databases.[51] We share the structures at https://github.com/GCA-VH-lab/TAC but note that the use of AlphaFold3 output is subject to the terms found at https://github.com/google-deepmind/alphafold3/blob/main/OUTPUT_TERMS_OF_USE.md. The quality of predictions was assessed using the predicted template-modelling (pTM) scores: pTM for scoring all residues of the model, and ipTM for scoring only the interfacing residues. Models with scores $\geq$0.5 are likely to have the correct overall fold and we considered models with scores $\geq$0.7 to be of good quality. Antitoxin structural alignments were performed with TM-align (last update: 2022/4/12)[55] and the Foldseek server[28] was used for the structural homology search. Protein structures were visualized using ChimeraX-1.9.[56] Statistical analysis of results was performed in R with ggplot2 library used for visualization.[57]

Bacterial growth and phage infections

All microbiology experiments were performed using. coli BW25113 strain (VHB987) or the motile strain BW25113 uspC-flhDC::IS5 (VHB17).[31] *E. coli* strains were grown at 37°C on solid (1.5% w/v agar) or in liquid LB medium. To select for plasmids, the medium was when needed supplemented with 25 μg/mL chloramphenicol, 100 μg/mL carbenicillin/ampicillin, and/or 50 μg/mL kanamycin. The expression from inducible promoters were controlled by supplementing the medium with appropriate concentrations of glucose/arabinose and/or IPTG. For phage infection of liquid cultures, the medium was supplemented with 10 mM $MgSO_4$ and 2.5 mM $CaCl_2$.

Phages were propagated using the double-agar overlay ' using *E. coli* BW25113 *uspC-flhDC*::IS5 (VHB17) strain[31] as the host. Top agar was prepared as LB agar (0.5% w/v agar) supplemented with 20 mM $MgSO_4$, and 5 mM $CaCl_2$. Phage stocks were prepared from the top agar layer and stored in SM buffer (0.1 M NaCl, 10 mM $MgSO_4$, and 0.05 M Tris-HCl pH 7.5) at 4°C.

Construction of plasmids

All plasmids were constructed with either circular polymerase extension cloning (CPEC)[58,59] with Phusion polymerase (Thermo Scientific) or Gibson assembly.[60] To clone TAC genes, we ordered the synthesized DNA fragments shown in Data S3. pBAD33 plasmid (arabinose inducible P_{BAD} promoter) was used to express toxins, TAs and TACs, pMG25 plasmid (IPTG inducible $P_{A1/O4/O3}$ promoter) was used for expression of chaperones and antitoxins and pET24d plasmid (T7 promoter) was used to produce toxin proteins *in vitro* translation reaction. In the plasmid used for *in vivo* gene expression, TAC genes were cloned with a strong Shine-Dalgarno and a spacer sequence (AGGAGGAATTAA).

TA and TAC toxicity assays

The toxin regions of Class-2 TAC systems from *E. coli* MOD1-EC6301 (gene: WP_097331689.1; plasmid: VHp1144), *S. gordonii* SK184 (RefSeq: WP_061594493.1, VHp1133), and *B. subtilis* BSn5 (RefSeq: WP_015715417.1, VHp1137); Class-3 TAC from S. aureus T126_A17_C01 (RefSeq: WP_000705255.1, VHp2223); Class-4 TAC from *B. thetaiotaomicron* MSK.16.57

(WP_217718571.1, VHp1712) and *S. epidermidis* 4320 (RefSeq: WP_002491514.1, VHp1713); Class-5 TAC from *N. xinjiangensis* SYSUD00778 (RefSeq: WP_210503345.1, VHp2224); Class-6 TAC from *C. koseri* CK1008 (RefSeq: WP_013819429.1, VHp2222) and the single site mutant H45A (VHp2260); Class-8 system from *C. comes* 2789STDY5608832 (RefSeq: WP_055246662.1, VHp1710); and antitoxin region from Class-5 TAC system from *M. smegmatis* atR9 (WP_003893515.1, VHp1711) were expressed from arabinose inducible pBAD33 with a strong Shine-Dalgarno sequence and spacer (AGGAGGAATTAA). Antitoxins and chaperones from *E. coli* MOD1-EC6301 (WP_064363165.1, WP_097331688.1 - VHp1146, VHp1147), *S. gordonii* SK184 (RefSeq: WP_061594494.1, WP_061594495.1 - VHp1135, VHp1136), and *B. subtilis* BSn5 (RefSeq: WP_015715418.1, WP_015715419.1 - VHp1138, VHp1139) were expressed with a strong SD from IPTG inducible pMG25 plasmid. Plasmids were transformed into *E. coli* strain VHB17 or, for TAC Class-2 assays, VHB987[31] Bacterial cultures from single colonies were grown for five to six hours, adjusted to OD_{600} 1.0, serially diluted (10^1- to 10^8-fold) and spotted on LB medium supplemented with 100 μg/mL carbenicillin (Fisher Bioreagents), 25 μg/mL chloramphenicol (AppliChem) and inducers (0.2% arabinose for TA induction, 0.2% glucose for TA repression, 50 μM or 500 μM IPTG for chaperone induction). Plates were incubated 16–17 h at 37°C and scored.

Experimental phage infections

To assess the effect of the *B. thetaiotaomicron* Class-5 TAC (RefSeq: WP_217718571.1, WP_217718573.1, WP_217718575.1; plasmid VHp2118) and the *C. koseri* class 6 TAC on phage infectivity (RefSeq: WP_248129368.1 - WP_248129369.1 - WP_248129370.1; plasmid VHp2261), we performed efficiency of plating assays using *E. coli* VHB17 strain[31] as a surrogate host. The experiments were performed essentially as described previously,[13] using a subset of common laboratory phages. Briefly, overnight cultures of *E. coli* VHB17 cells carrying either an empty vector (pJD1423 (VHp1423), a pBR322 derivative lacking the tetracycline resistance cassette) or the TAC-expressing pJD1423 derivatives (VHp2118 and Vhp2261) were mixed with top agar (LB with 0.5% agar, 20 mM $MgSO_4$, and 5mM $CaCl_2$), to a final concentration of 0.075 OD_{600}, and overlayed on LB-agar plates (1.5% agar). Individual phage stocks were 10-fold serially diluted in SM buffer (0.1 M NaCl, 10 mM $MgSO_4$, and 0.05 M Tris-HCl pH 7.5) and 2.5 μL of each of eight dilutions spotted on solidified top agar plates. Plaque formation was monitored after 6 and 24h of incubation at 37°C.

In vivo functional assays

Metabolic labeling with ^{35}S-methionine, ^{3}H-uridine and ^{3}H-thymidine

The experiments were performed essentially as described previously.[13] For metabolic labeling experiments the *E. coli* VHB987 strain was co-transformed with the D33 plasmid carrying the toxin gene of interest for L-arabinose-inducible expression and the empty pMG25 vector. Transformed cells were initially plated on LB plates supplemented with 100 μg/mL carbenicillin, 25 μg/mL chloramphenicol, and 0.2% glucose (to suppress leaky toxin expression). Using individual *E. coli* colonies for inoculation, 2 mL liquid cultures were prepared in defined Neidhardt MOPS minimal media,[61] supplemented with 100 μg/mL carbenicillin, 25 μg/mL chloramphenicol, 0.1% casamino acids, and 0.2% glucose, and grown overnight at 37°C with shaking. Subsequently, experimental 15 mL media were prepared in 125 mL conical flasks in MOPS medium, supplemented with 0.5% glycerol, 100 μg/mL carbenicillin, 25 μg/mL chloramphenicol, as well as a set of 19 amino acids (lacking methionine), each at a final concentration of 25 μg/mL. The overnight cultures were inoculated to the experimental media at a final OD_{600} of 0.05 and grown at 37°C with shaking until the OD_{600} reached 0.2. At this point, 1 mL aliquots (designated as the pre-induction zero time-point) were transferred to 1.5 mL Eppendorf tubes containing 10 μL of the respective radioisotope (4.35 μCi ^{35}S methionine [Hartmann], 0.65 μCi ^{3}H uridine [Hartmann] or 2 μCi ^{3}H thymidine [Hartmann]) and placed in a heat block at 37°C. Toxin expression in the remaining 14 mL culture was induced by adding L-arabinose to a final concentration of 0.2%. Throughout the toxin induction time course, 1 mL aliquots were taken from the experimental culture and transferred to 1.5 mL Eppendorf tubes containing 10 μL of the appropriate radioisotope (^{35}S methionine, ^{3}H uridine, or ^{3}H thymidine). Radioisotope incorporation was halted after 8 min of incubation at 37°C by adding 200 μL of ice-cold 50% trichloroacetic acid (TCA) to the 1 mL cultures. Additionally, 1 mL aliquots were periodically sampled for OD_{600} measurements. The resultant 1.2 mL culture/TCA samples were loaded onto GF/C filters (Whatman) prewashed with 5% TCA and unincorporated label was removed by washing the filter twice with 5 mL of ice-cold TCA followed by a 5 mL wash with 95% EtOH (twice). The filters were placed in scintillation vials, dried for at least two hours at room temperature, followed by the addition of EcoLite-scintillation cocktail (5 mL per vial; MP Biomedicals). After shaking for 15 min, radioactivity was quantified using Tri-Carb 4910TR-scintillation counter (PerkinElmer). Isotope incorporation was quantified by normalizing radioactivity counts (CPM) to OD_{600}, with the pre-induction zero time-point serving as the reference (set to 100%). All experiments were conducted in triplicates (n = 3), using three independent cultures initiated from distinct colonies.

In vitro translation assays

The experiments were performed essentially as described previously.[13] PURExpress *in vitro* protein synthesis kit (NEB, E6800) supplemented with 0.8 U/μL RNase Inhibitor Murine (NEB, M0314S) was used for reactions as per the manufacturer instructions. All reactions contained each ART template plasmids (10 ng/μL, VHp1219, VHp1220 or VHp1222) with or without 0.1 mM NAD^+. After a 10-min incubation at 37°C, a 1.3 μL aliquot of the reaction mixture was taken and quenched by addition of 13.7 μL of 2× SDS-PAGE sample buffer (100 mM Tris:HCl pH = 6.8, 4% SDS, 0.02% bromophenol blue, 20% glycerol, 20 mM DTT and 4% β-mercaptoethanol), and DHFR Control Template plasmid was added to the remaining reaction mixture at a final concentration of

20 ng/μL. After further incubation at 37°C for 1 h, the reaction mixture was mixed with 9-fold volume of 2× sample buffer, denatured at 95°C for 5 min and resolved on SDS-PAGE gel (18% acrylamide/bis-acrylamide = 37.5:1). The SDS-PAGE gel was fixed by incubating for 5 min at room temperature in 50% ethanol solution supplemented with 2% phosphoric acid, washed three times with water for 20 min at room temperature, and stained with "blue silver" solution (0.12% Brilliant Blue G250 (Sigma-Aldrich, 27815), 10% ammonium sulfate, 10% phosphoric acid, and 20% methanol) overnight at room temperature. After washing with water for 3 h at room temperature, the gel was imaged on an Amersham ImageQuant 800 (Cytiva) imaging system. The intensities of reporter DHFR band that of the ribosomal protein used as internal control were quantified using ImageJ v. 1.54g.[62]

QUANTIFICATION AND STATISTICAL ANALYSIS

Precision measures (mean, SEM, and SD) were computed using R v4.3.0 and are shown in Table S2.

MICROBIOLOGY

The Viral AlphaFold Database of monomers and homodimers reveals conserved protein folds in viruses of bacteria, archaea, and eukaryotes

Roni Odai[1], Michèle Leemann[2], Tamim Al-Murad[1], Minhal Abdullah[1,3], Lena Shyrokova[1], Tanel Tenson[3], Vasili Hauryliuk[1,3,4,5,6]*, Janani Durairaj[2,7]*, Joana Pereira[2,7]*, Gemma C. Atkinson[1,4,6]*

Viruses are the most abundant and genetically diverse entities on Earth, yet the functions and evolution of most viral proteins remain poorly understood. Their rapid evolution often obscures evolutionary relationships, limiting the ability to assign functions using sequence-based methods. Although the conservation of protein fold can reveal deep homologies, viral proteins remain underrepresented in structural databases. We address this by clustering viral sequences from RefSeq and predicting the structures of ~27,000 representative proteins using AlphaFold2 to create the Viral AlphaFold Database (VAD). We uncover conserved folds in diverse viruses infecting bacteria, archaea, and eukaryotes. We predict homodimers and make comparisons to the Protein Data Bank, providing data on oligomerization potential. We reveal considerable functional darkness in the viral protein universe and report the discovery and validation of an uncharacterized toxin-antitoxin system. The VAD provides a foundation for exploring viral structure-function relationships, including ancient folds shaping viral interactions across all life.

INTRODUCTION

Modern deep learning structural prediction methods, such as AlphaFold and RosettaFold, have revolutionized the ability to use structural information in research (*1*–*3*). As well as the use of inferred folds to make predictions about the function that can be tested experimentally, structure versus structure comparison methods allow the discovery of homology that is not detectable at the sequence level using tools such as Foldseek (*4*) and distance-matrix alignment (Dali) (*5*). This can be especially informative for proteins of viruses, which evolve fast because of short generation times, large population sizes, high mutation rates, and host-virus arms races (*6*, *7*). The taxonomic domain of viruses includes both eukaryote-infecting viruses and bacteriophages that infect bacteria and archaea. Despite the fundamental differences between their hosts—including cell structure, gene regulation, and evolutionary history—viruses across these domains can share conserved protein folds. The single jellyroll is one of the most common capsid folds and can be found in various phages and eukaryote-infecting viruses (*8*), while the capsid protein of bacteriophage HK97 has homologs in herpes viruses, adenoviruses, and viruses of archaea and protists (*8*, *9*). Most recently, phosphodiesterases of both phages and eukaryote-infecting viruses have been found to contain the same fold and been experimentally confirmed to have a shared role in immune evasion (*10*, *11*).

The ability to search for homologous proteins and infer the function at the fold level depends upon the availability of large and diverse protein structure databases. The Protein Data Bank (PDB) has been an essential resource for structural comparisons for decades, storing and organizing experimentally determined structures of proteins and complexes (*12*). A massive increase in the numbers and diversity of publicly available structures came with the release of the EBI AlphaFold database (AFDB) (*13*), which includes predicted structures for most proteins in the UniProt database, vastly expanding the number of available protein structures beyond what is present in the PDB. The structural comparison of these millions of protein folds has enabled the construction of networks of structural communities, annotated with their functional "brightness" or "darkness," depending on how well annotated the biochemical and biological functions of those proteins are (*14*).

A substantial portion of the known protein fold universe remains functionally dark, and this is likely an underestimate of the true extent of structural and functional dark matter. One group that remains underexplored in structural biology is viruses. Despite being regarded as the most abundant biological entity on Earth and being an immense reservoir of genetic diversity (*15*, *16*), viruses have so far been excluded from the EBI AFDB. This is primarily due to the problem of polyproteins. Polyproteins are composed of multiple viral peptides that are translated as a single concatenated peptide and subsequently proteolytically cleaved into individual mature proteins using viral or host proteases (*17*, *18*). While these polyprotein structures complicate structural prediction, the potential benefits of uncovering viral protein structural diversity outweigh these challenges. Recent viral structure databases, such as the BFVD (*19*) and Nomburg24 (*10*) projects, have addressed this gap to a large extent. However, these resources rely on less accurate methods than the reference AlphaFold2 implementation and are limited to monomeric structural predictions.

Here, we have clustered 647,000 unique viral sequences from the National Center for Biotechnology Information (NCBI) RefSeq database (*20*), predicted the structure of ~27,000 representatives using AlphaFold2 (*1*), and predicted higher-order oligomeric states where possible. This represents a high-quality resource for exploring viral protein diversity, including in the context of host taxonomy. We have identified 1142 clusters of viral proteins that share the same fold across viruses infecting two or more domains of cellular life (bacteria, eukaryotes, or archaea). We have characterized the functional

[1]Department of Experimental Medical Science, Lund University, Lund, Sweden. [2]Biozentrum, University of Basel, Basel, Switzerland. [3]Institute of Technology, University of Tartu, Tartu, Estonia. [4]Science for Life Laboratory, Lund, Sweden. [5]NanoLund, Lund University, Lund, Sweden. [6]Lund University Virus Centre, Lund, Sweden. [7]SIB Swiss Institute of Bioinformatics, Basel, Switzerland.
*Corresponding author. Email: vasili.hauryliuk@med.lu.se (V.H.); janani.durairaj@unibas.ch (J.D.); joana.pereira@kuleuven.be (J.P.); gemma.atkinson@med.lu.se (G.C.A.)

darkness of viral folds, predicted homodimerization tendencies and structures for almost all proteins in our database, and predicted higher-order multimeric states where possible. Through searching for functionally dark protein structures in conserved neighborhoods, we have uncovered and experimentally validated a previously unidentified type II toxin-antitoxin (TA) system, KreTA, found in prophages of bacteria and mapped conserved and variable folds across virus-host interactions, including defense and antidefense systems. All structures are available at data-sharing.atkinson-lab.com/vad/, and the structural clusters can be searched and explored at vad.atkinson-lab.com.

RESULTS

Clustering the diversity of viral proteins at the sequence level reveals homology among proteins from viruses that infect hosts in different domains of life

Because of the computational challenges of predicting structures for all available viral protein sequences, we aimed to assemble and fold a representative subset that captures broad diversity across the sequence space. To do this, we first downloaded protein sequences from the NCBI RefSeq database (*20*), limiting by taxonomy to virus (NCBI taxonomy ID 10239). The taxonomic distribution of these 647,000 sequences is depicted in fig. S1, spanning the 10 most abundant viral taxa across the kingdom, phylum, and family levels. The host of each virus was assigned by querying the Virus-Host database (*21*). Protein sequences were then clustered using MMseqs2 (*22*), with relatively lenient identity and coverage thresholds (30% and 0.01, respectively) to minimize the number of clusters while maintaining sufficient separation between them (Fig. 1A). This clustering process yielded 117,479 clusters and 61,868 singletons, with an average cluster size of 5.5 members (table S1 and dataset S1).

The assignments from the Virus-Host database (*21*) allowed us to group proteins by host domain (bacteria, archaea or eukaryotes), where a cluster with multiple hosts associated with the proteins within is referred to as a heterohost cluster. It is important to note that this does not imply that the same protein can be found in viruses infecting vastly different hosts; rather, the cluster is composed of homologous proteins that can be found in different host-infecting viruses. Therefore, heterohost cluster composition reflects molecular patterns that are common across viral life. Reflecting the original taxonomic distributions, most clusters comprise proteins from viruses infecting bacteria (Fig. 1B). In total, 388 clusters were assigned a sequence-level heterohost label, most of which comprise proteins from viruses with either bacterial or eukaryotic hosts (Fig. 1C). The distribution of host domain (eukaryota, bacteria, or archaea) before and after clustering is shown in Fig. 1D. To reduce ambiguity, we excluded from these counts any viruses annotated as infecting other viruses, such as virophages that parasitize giant viruses and phage satellites (*23*). Most unique viral protein hosts are bacteria (79.9%), followed by eukaryotes (15.4%). Archaeal hosts are assigned for only 1.1% of the proteins (Fig. 1D).

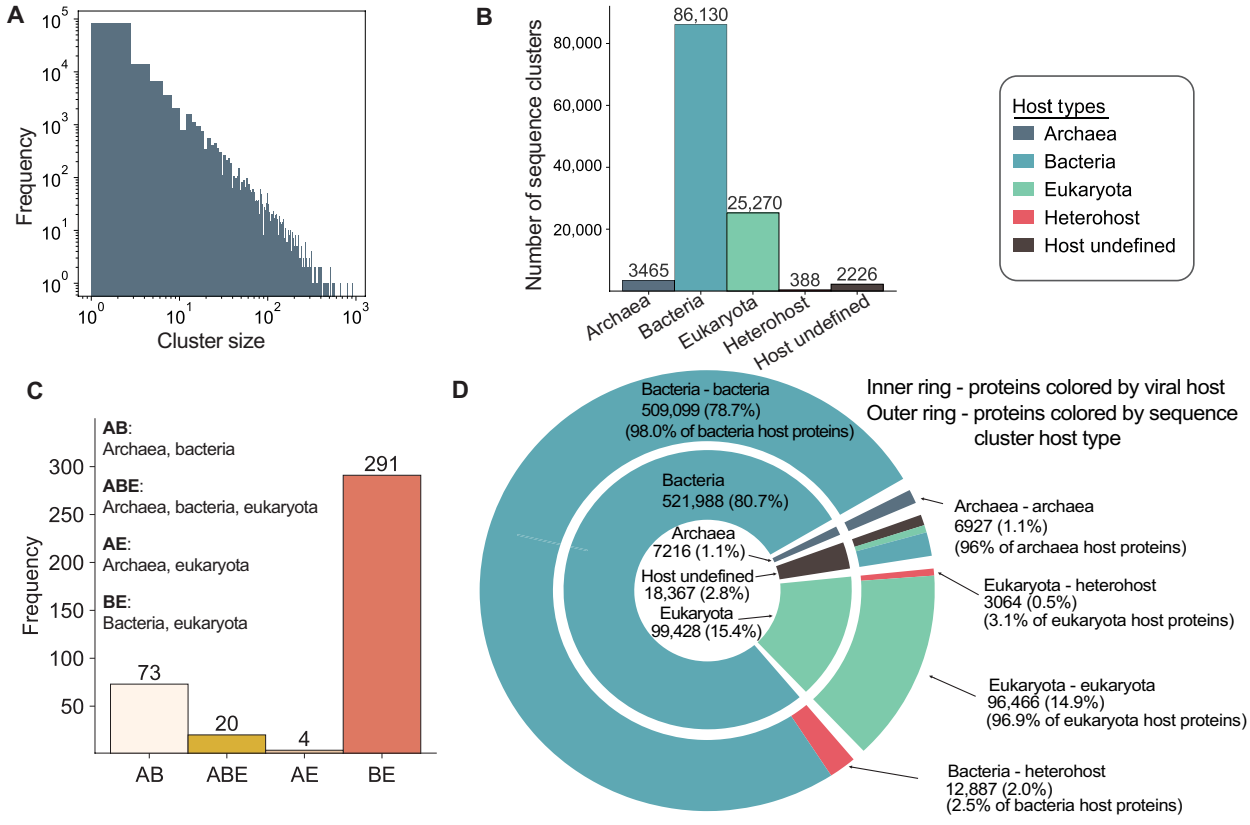


Fig. 1. Sequence clustering and host-type assignment of viral proteins and clusters. (**A**) Distribution of sequence clusters by size. (**B**) Distribution of sequence clusters by host type. (**C**) Distribution of sequence heterohost clusters based on host domain composition. (**D**) Distribution of proteins by viral host (inner ring) and sequence cluster host type (outer ring), where viral host segments correspond to cluster type segments, displaying the cluster host-type membership of proteins by viral host.

While some of the sequence-level heterohost clusters contain proteins annotated as structural proteins, i.e., tail fibers, capsids, or their chaperones, most heterohost clusters represent proteins that are annotated to interact with or modify nucleotides or nucleic acids (dataset S1). The most prevalent are proteins involved in replication, nucleases, nucleotide hydrolases, nucleotide modification enzymes, helicases, nucleic acid ligases, and methyltransferases. The largest heterohost sequence cluster (567 proteins, found in viruses infecting either bacteria or eukaryotes) contains homologs of cytidine and deoxycytidylate deaminase—an enzyme that viruses can use to modulate nucleotide pools to favor their own replication (*24*) and that sometimes serves as a component of antiphage defense (*25*). The second biggest sequence cluster contains proteins homologous to the clamp loader of DNA polymerase, which enables efficient and processive replication and is found in bacteriophages and viruses of protists. Last, there are 20 sequence clusters spanning bacteria, eukaryote, and archaeal hosts. The largest of these contains DNA methyltransferases (dataset S1). Thus, at the sequence level, conserved nucleotide-processing functions are often the most readily identifiable across host domains.

Prediction of representative protein structures to create the VAD

Representatives from the viral protein sequence clusters were folded with AlphaFold2 (*1*), limiting to clusters with a size of 5 or greater. To sample sequence diversity in smaller clusters, we also randomly selected a set of 7243 sequence clusters, with an average cluster size of 1.6. We did not set a strict amino acid length limit but rather folded all proteins that we could, given our computational resources. Compared to the two other predicted structure databases for viral proteins, BFVD (*19*) and Nomburg24 (*10*), our length distribution is most similar to BFVD, with an average length of 176 amino acids (fig. S2). The length distribution in our database has a longer tail because of not using a length cutoff; our largest predicted fold is from a 3595 amino acid–long protein. Nomburg24 predictions tend to be longer on average, probably due to eukaryote-infecting viruses having longer proteins on average than phages. The final Viral AlphaFold Database (VAD) structure dataset consists of 26,962 predictions for monomers and 26,754 homodimer cofolding attempts (of which only a subset is confident homodimers; see below) (dataset S1). The taxonomic distribution of VAD is depicted in fig. S3 (full lineage information is found in dataset S1) and covers the full diversity of viral kingdoms, phyla, and families seen in the starting sequence set (figs. S1 and S3).

The quality of the predictions in the dataset is assessed with the predicted local difference test (pLDDT) metric (*1*). The VAD dataset has a global average pLDDT of 78, with most predictions reaching an average pLDDT of ~90, reflecting high overall model confidence (Fig. 2A). Compared to Nomburg24 (*10*) and BFVD (*19*), the VAD demonstrates improvements in prediction quality. Among near-identical proteins (>90% sequence identity and coverage) shared between the VAD and BFVD, 74% of VAD models have higher average pLDDT scores, and 54% show an increase of more than 5 units. This highlights the accuracy gains of the standard AlphaFold2 pipeline, with differences in prediction quality likely attributable to BFVD's use of faster, but potentially less sensitive, multiple sequence alignment construction approaches. Examples of such cases are shown in Fig. 2B.

Structure-based all-versus-all searches using the ~27,000 VAD cluster proteins identify more than twice as many unique cluster-to-cluster

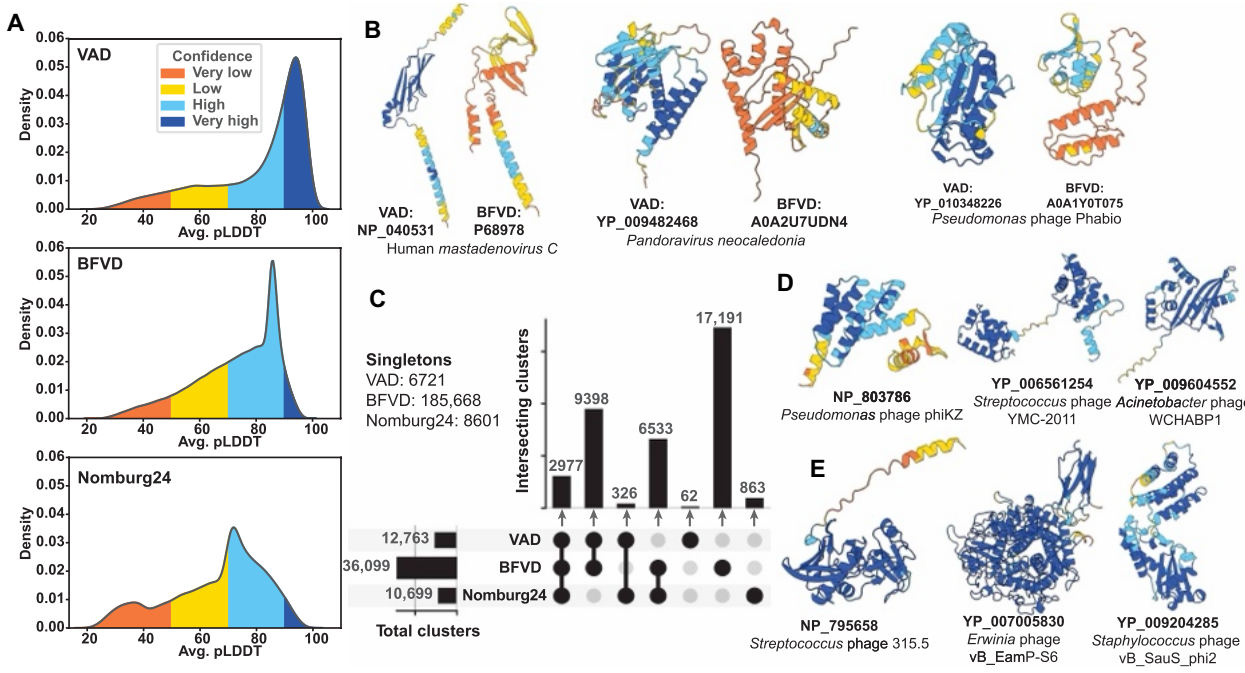


Fig. 2. VAD comparison with other databases of viral predicted structures. (**A**) Density plot of average pLDDT distributions for the VAD, BFVD, and Nomburg24. pLDDT is averaged per prediction in each dataset. (**B**) Three examples where the VAD structure pLDDT is higher than the corresponding BFVD structure pLDDT. (**C**) UpSet plot detailing overlaps of nonsingleton structural clusters (with Foldseek easy-cluster coverage of 0.7 and default parameters) across the VAD, BFVD, and Nomburg24 with singleton cluster counts listed. (**D**) Three examples of VAD cluster singletons with >80 pLDDT. (**E**) Three examples of clusters unique to the VAD with >80 pLDDT.

relationships as sequence-based searches can find across all 647,000 viral proteins. This indicates that this relatively small, structure-representative set effectively captures relationships within the viral NCBI RefSeq database. The VAD is structurally diverse, as seen by the 12,763 structural clusters and 6721 singletons encompassing the 26,962 VAD cluster proteins (Fig. 2C). In addition, despite its smaller size, the VAD dataset spans a substantial portion of the nonsingleton structural clusters found in both BFVD and Nomburg24, which contain 2.5× and 13× more structures, respectively (Fig. 2C). Examples of singletons and unique VAD clusters are shown in Fig. 2 (D and E, respectively). Together, these results show that the VAD dataset both is structurally diverse and has high confidence, making it a valuable resource for structural searches and for advancing our understanding of viral protein function and evolution.

Proteins with the same structural fold in the VAD are found in viruses infecting bacteria, archaea, and eukaryotes

Sequence clustering shows that the vast majority of cluster representatives comes from viruses infecting bacterial hosts (Fig. 1B). To test whether structure-based clustering reveals broader relationships across host types, we examined whether proteins from viruses infecting different domains of life share similar folds even when their sequences diverge. We therefore clustered the 26,962 structures in the VAD using Foldseek (*4*) in TMalign mode, which groups proteins by structural similarity. This resulted in 12,894 clusters with 9753 singletons, with an average cluster size of 2.1 (table S2 and dataset S1). Host domains (bacteria, archaea, or eukaryotes) were then assigned to structural clusters using the same host annotations and criteria as for sequence clusters.

Structural clusters show a shift in host distribution: Although bacterial host-type clusters still dominate, the proportion of eukaryotic host-type clusters increases (Fig. 3A). Heterohost clusters also expand markedly; 913 clusters include proteins from both bacterial and eukaryotic viruses, with additional clusters spanning archaea-bacteria, archaea-eukaryota, and all three domains (Fig. 3B). While fewer than 1% of VAD proteins belong to heterohost sequence clusters, this jumps to 35.8% in structural clusters (Fig. 3C). This notable increase, visualized in the network of fig. S4, suggests that folds are conserved across greater evolutionary distances and are more frequently shared across viral life than sequence similarity alone would indicate.

Next, we examined how proteins from viruses infecting specific host domains are distributed across structural clusters (Fig. 3D), as we did for sequence clusters in Fig. 1D. Most proteins from bacteriophages (62.9%) fall into clusters that also contain proteins from other bacterial viruses, but a substantial fraction (25.5%) is in heterohost clusters. Similarly, 31.1% of proteins from viruses infecting eukaryotes are in heterohost clusters, suggesting widespread structural reuse across viral lineages despite divergent host range. Archaeal viruses also show notable structural overlaps, with 62.3% of their proteins falling into archaeal-only clusters and 37.6% into heterohost clusters.

To identify common structural themes across diverse viruses, we examined the largest structural clusters in more detail (Fig. 3D). Most large structural heterohost cluster representatives are small mixed α/β folds, comprising ~4% of all proteins belonging to structural heterohost clusters (dataset S1). The largest, third largest, and fifth largest cluster representatives are α-β sandwiches: two meander folds and one split fold. The fourth largest cluster representative is an α-β OB (oligonucleotide/oligosaccharide–binding)–fold (Fig. 3E). These small folds are typical "urfolds"—three-dimensional architectures that can be formed from vastly different sequences, allowing for extensive functional innovation (*26*). The second largest structural heterohost cluster representative features an HK97 fold found in capsid proteins of various viruses. Overall, apart from the capsid fold, the largest clusters seem to represent common ancient core folds, not limited to viruses but also found in proteins across all life on Earth.

Most of the large structural eukaryotic monohost cluster representatives are α-domain structures. The two largest of these clusters are likely pentatricopeptide and ankyrin-repeat proteins, where both repeat four-helix bundles. These tandem repeat clusters comprise ~1% of all proteins belonging to structural eukaryotic clusters and ~3% of proteins belonging to nonsingleton clusters (dataset S1). The third largest cluster is a singular α helix with unstructured termini; the members of this cluster are similarly simple, with a minority exhibiting a turn between two α helices. The fourth largest of these clusters is a globular α-domain protein, the M11L-like apoptosis regulator (Fig. 3E). Thus, most of these eukaryote host–specific protein folds are likely all involved in hallmark eukaryote-like protein-protein and protein-RNA interactions (*27*, *28*). This may reflect viral adaptation to the complex and compartmentalized regulation characteristic of eukaryotic hosts (*29*).

Bacterial monohost cluster representatives consist mostly of small β-barrel folds. The largest, fourth largest, and fifth largest of these clusters have SH3/Tudor-like folds (Fig. 3E). These putative Tudor fold clusters cover ~1.3% of all proteins in bacterial structural clusters and ~2.6% of proteins belonging to nonsingleton clusters (dataset S1). The second largest cluster contains tail tube proteins from λ-like phage tail tube proteins, while the third largest cluster contains wHTH (winged helix-turn-helix) folds that are often involved in DNA binding (Fig. 3E) (*30*). Archaeal structural clusters consist almost entirely of singletons, where member proteins are uncharacterized, reflecting the paucity of data for viruses infecting this domain of life.

Oligomerization states across viral proteomes

Many viral proteins exert their biological functions as oligomers rather than monomers. To explore oligomerization tendencies, we predicted homodimeric structures for all but the ~200 largest proteins of VAD monomers with AlphaFold-Multimer (*31*). These homodimers include 26,754 predictions, of which 11,313 exhibit a confident fold and 2957 exhibit a good fold [with interface predicted template modeling (ipTM) thresholds set to 0.5 for "confident" and to 0.8 for "good"] (Fig. 4A). Multiple VAD monomers are not predicted to form homodimers; the average ipTM score for VAD homodimers is 0.44 (dataset S1), below the cutoff for "good" models. To determine whether low-quality monomeric predictions skewed the quality of homodimer counterparts, we filtered monomers with pLDDT scores lower than 50 (fig. S5). The effect of this filtering was marginal without much change in the distribution density of predictions by the ipTM score. In 1188 cases, the pLDDT of the dimer (max across the two chains) is >5 units higher than the monomer pLDDT, with Fig. 4B showing three extreme examples, where the monomer and dimer structures have significant differences. Where we do see good scores for homodimerization, this can potentially include homodimers in the strictest sense (two monomers interacting alone), as well as larger homo- and

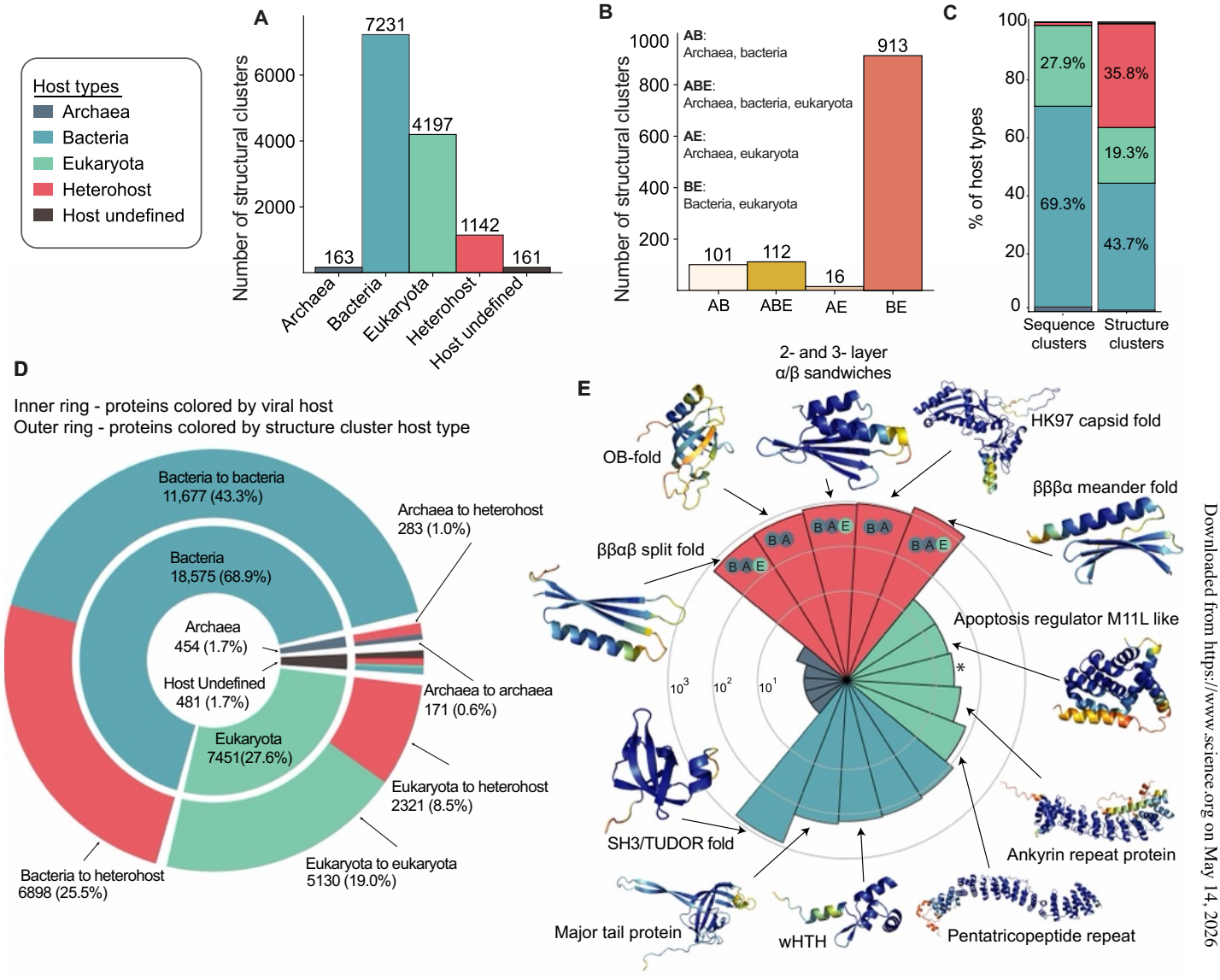


Fig. 3. Structural clustering of VAD proteins reveals core folds in viruses infecting different domains of life. (**A**) Distribution of structural clusters by host type. (**B**) Distribution of structural heterohost clusters based on host domain composition. (**C**) Sequence and structure cluster host-type distribution of VAD proteins. (**D**) Distribution of proteins by viral host (inner ring) and structural cluster host type (outer ring), where viral host segments correspond to cluster type segments, displaying the cluster host-type membership of proteins by viral host. (**E**) Radial histogram of the five largest structural clusters, their sizes, and the structures of select cluster representatives colored by pLDDT. The contributing host taxa are shown within the heterohost segments (B, bacteria; A, archaea; E, eukaryota). The third largest eukaryote specific cluster (15 proteins; marked with an asterisk) consists of a single helix topology and is not shown.

hetero-oligomeric complexes that include an interface between two identical protein subunits. To address potentially larger complexes of VAD proteins, we predicted oligomeric arrangements of monomers in the VAD. We compared monomeric structures with those in the PDB: Structural hits of monomers found more than once in a complex in a PDB structure were classified as homo-oligomeric (3864 VAD clusters), while those found with other monomers were classified as hetero-oligomeric (2623 VAD clusters) (dataset S1). Figure 4A, indicating turquoise density, shows the dimer ipTM distribution of all VAD clusters compared to those having homomer PDB hits, showing a shift toward higher ipTM scores for the latter. AlphaFold-Multimer seems able to predict dimeric structures in their higher-order oligomeric state, as shown by the two examples in Fig. 4C where the closest Foldseek-Multimer PDB hit is a homotrimer and the predicted dimer adopts the trimeric configuration. As AlphaFold-Multimer was trained on complexes from the PDB, predicted oligomeric conformations may simply reproduce those present in the training data for close homologs. In addition, PDB structures used as templates can project oligomeric states onto homodimeric predictions. However, oligomeric awareness can also be seen at low sequence identities, as seen in the deoxyuridine 5′-triphosphate nucleotidohydrolase (dUTPase) examples in Fig. 4C, which share only 28% sequence identity.

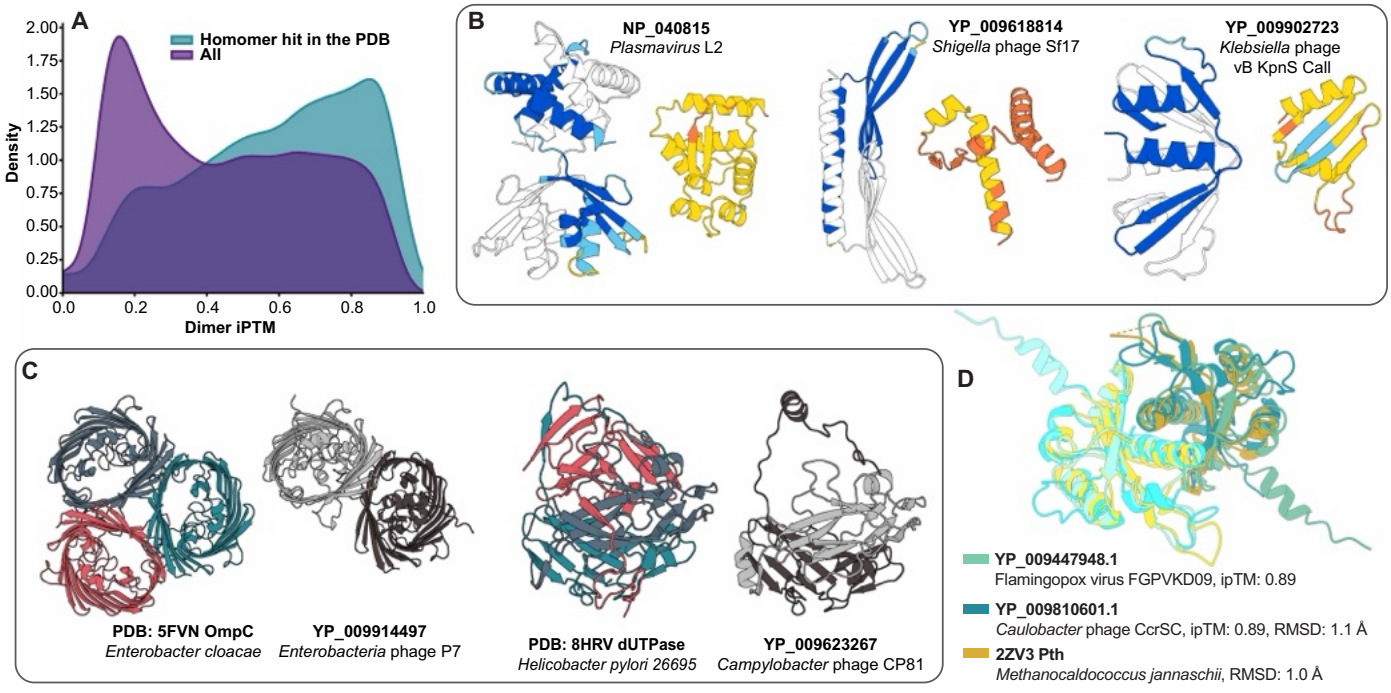


Fig. 4. A large proportion of proteins in the VAD is predicted to form homodimeric interfaces. (**A**) Density plot of ipTM distribution for all VAD proteins and those with a homomer (homo-oligomer or homodimer) hit in the PDB. (**B**) Cases where the prediction quality markedly increases in the homodimer compared to the monomer. (**C**) Examples of cases where VAD homodimers are found in oligomer-aware conformations. Pairwise sequence identities are 65% in the case of the OmpC homologs and 28% in the case of the deoxyuridine 5′-triphosphate nucleotidohydrolases (dUTPases). (**D**) Confident dimer predictions of Pth homologs from viruses that infect both bacteria and eukaryotes. RMSD, root mean square deviation.

One unexpected observation in the VAD is the prevalence of peptidyl tRNA hydrolases (Pths) across a variety of different viruses and phages. By cleaving peptides bound to tRNA, Pth plays a crucial role in protein synthesis by both rescuing stalled ribosomes and freeing up the pool of free tRNAs (*32*, *33*). We have found that Pth is encoded in various bacteriophages and viruses that infect birds and insects, and its dimeric structure is conserved (Fig. 4D). A functional Pth encoded in the bacterial host has previously been found to be important for λ phage translation of two-codon minigenes that are otherwise toxic because of accumulation of "dropped-off" peptidyl-tRNA (*34*). Thus, by carrying their own Pths, viruses may ensure efficient translation of their proteins and avoid ribosome stalling and toxicity. Our predictions suggest that the Pth sequence-level hetero-host clusters are primarily homodimers. Superposition of AlphaFold Pth predictions from the *Caulobacter* phage and Flamingopox virus with the PDB structure of Pth from archaeon *Methanocaldococcus jannaschii* shows excellent structural alignment (Fig. 4D).

Structural matches to the Protein Atlas; functional darkness and brightness of viral protein structures

We asked how similar or dissimilar viral proteins are to other proteins with known structures. A Foldseek search of VAD structures against UniProt3D community representatives (*14*) and the AFDB50 Foldseek database revealed 12,473 representatives (covering 292,376 viral sequences) falling into connected components in UniProt3D, 209 representatives (3635 sequences) into unconnected "dust" UniRef50 clusters, 2627 representatives (55,922 sequences) having matches to AFDB proteins not present in UniProt3D v1 because of the pLDDT threshold of 90 used, and 9279 representatives (115,750 sequences) having no good match to either database (Fig. 5A). Among those with matches to known structural communities, most are functionally bright—a term used to describe proteins with known or inferred biological function (*14*). However, a substantial fraction falls into dark communities, meaning that they lack functional annotation and remain poorly understood despite having a predicted structure (Fig. 5B). When examined by host taxonomy, proteins from viruses infecting bacteria and archaea were more likely to have high-confidence structural matches to the AFDB, whereas proteins from eukaryotic viruses had the highest proportion of no matches (Fig. 5, C and D). Among proteins that did yield a structural hit, the confidence (as measured by LDDT) was broadly similar across host types, suggesting that once a structural analog is found, the quality of the alignment does not depend strongly on host taxonomy.

VAD proteins are related to functionally dark bacterial protein communities that include prophage-encoded TA systems

A previous exploration of the universe of functionally dark protein folds led to the discovery of previously unknown type II TA systems (*14*). Given that we have uncovered additional dark proteins in the VAD, we asked whether there could be further undiscovered TAs in our data. We focused on VAD proteins in Protein Atlas communities that also include bacteria, because TAs can be frequently found on prophages integrated into bacterial chromosomes (*35*, *36*). Specifically, we focused on small operon–encoded proteins encoded in communities

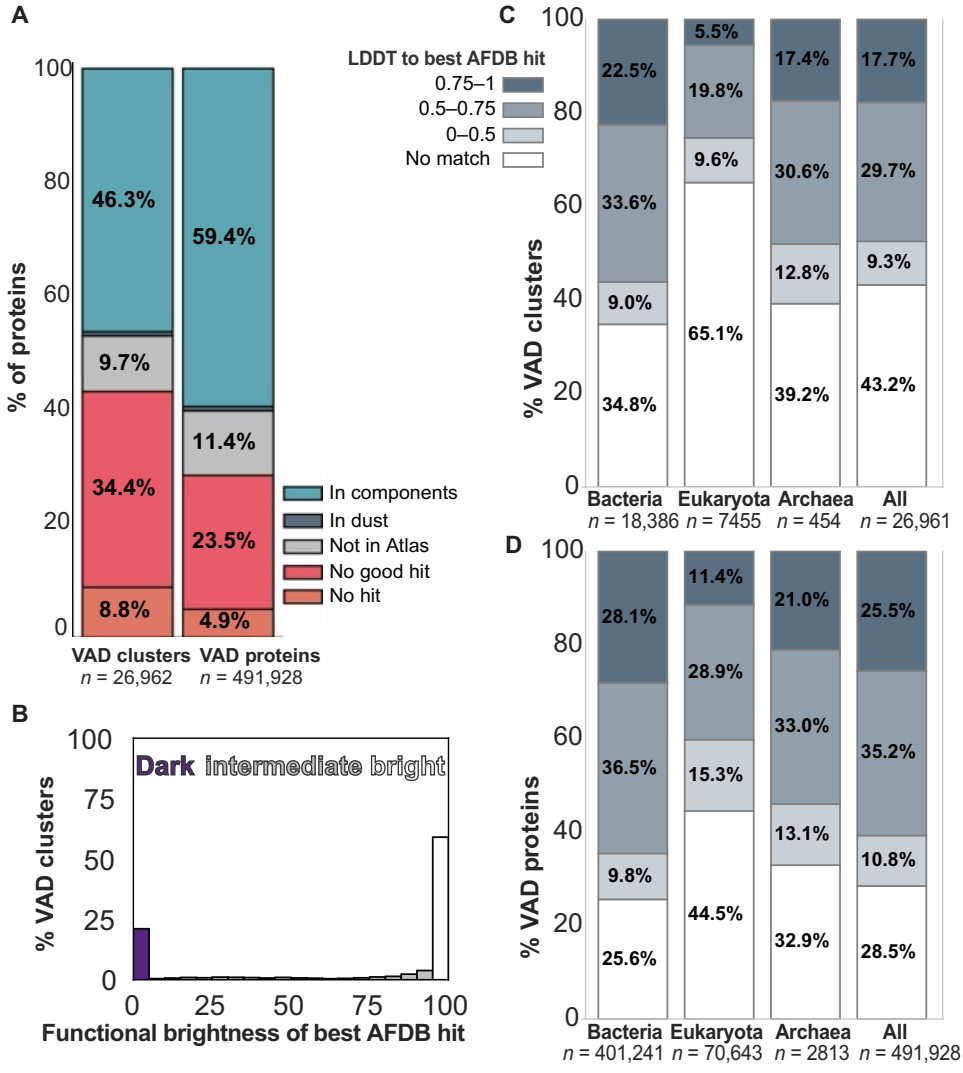


Fig. 5. Structural similarity of VAD to the AlphaFold database. (**A**) Division of VAD representatives and proteins into those without any hits to AFDB50 or UniProt3D, without hits with TM-score >0.5, without hits to UniProt3D (i.e., only hit AFDB structures with pLDDT <90), with hits only to dust, and with hits to connected components in UniProt3D. "Representatives" mean cluster representatives that we have predicted the structure for and therefore are present in the VAD. "Proteins" extends the count to all proteins in the clusters. (**B**) Distribution of functional brightness values for the brightest AFDB hit for each VAD representative. (**C**) LDDT of best hits across different host superkingdoms as per lower inset box. (**D**) LDDT of best hits across different host superkingdoms expanded to proteins in each representative cluster.

containing *Escherichia coli*, the model system for our validations. Using GCsnap and FlaGs, we analyzed gene neighborhoods, searching for conserved bicistronic gene operons characteristic of TA systems. We found a pair with homologs across Gammaproteobacteria, consistently localized in prophage-like regions of Enterobacteriales (Fig. 6A). This includes annotated pathogenic strains isolated from clinical and food sources. The conserved two-gene architecture is strongly suggestive of a TA system. We name this putative TA system KreTA after the mythical heroic twins in the Albanian folk epic *Kângë Kreshnikësh*.

KreT and KreA proteins are predicted confidently to dimerize (Fig. 6B). The proteins are classed as functionally dark in that they belong to communities that have no functional domain hits (dataset S1) (*14*). Foldseek (*4*) also does not identify any homologous proteins with known function. However, Dali (*5*) indicates distant but significant fold similarity of KreT to ribonuclease RegB, an endoribonuclease that controls the expression of multiple phage early genes and which has no identifiable sequence similarity to anything else of known function (fig. S6A) (*37*, *38*). The closest structural relative of KreT in the VAD is annotated as RegB, but the two proteins are not confidently alignable at the sequence level. In the KreTA dimer, KreA binding sequesters the predicted ribonuclease (RNase) active site of KreT (Fig. 6B), indicative of KreA acting as an antitoxin. Furthermore, Dali suggests that KreA may have a broadly similar fold to the immunity protein Tri^{Tu} that neutralizes the adenosine 5′-diphosphate (ADP) ribosylase toxin Tre^{Tu} (fig. S6B) (*39*).

To test our prediction that KreTA is a TA, we carried out a toxicity neutralization assay. While the expression of KreT is toxic to *E. coli*, this toxicity is efficiently neutralized by the KreA antitoxin, validating that this is TA system (Fig. 6C). To investigate the mechanism of KreT toxicity, we performed metabolic labeling using radioactive precursors for translation, transcription, and replication. As manifested by the selective inhibition of ^{3}H-methionine incorporation,

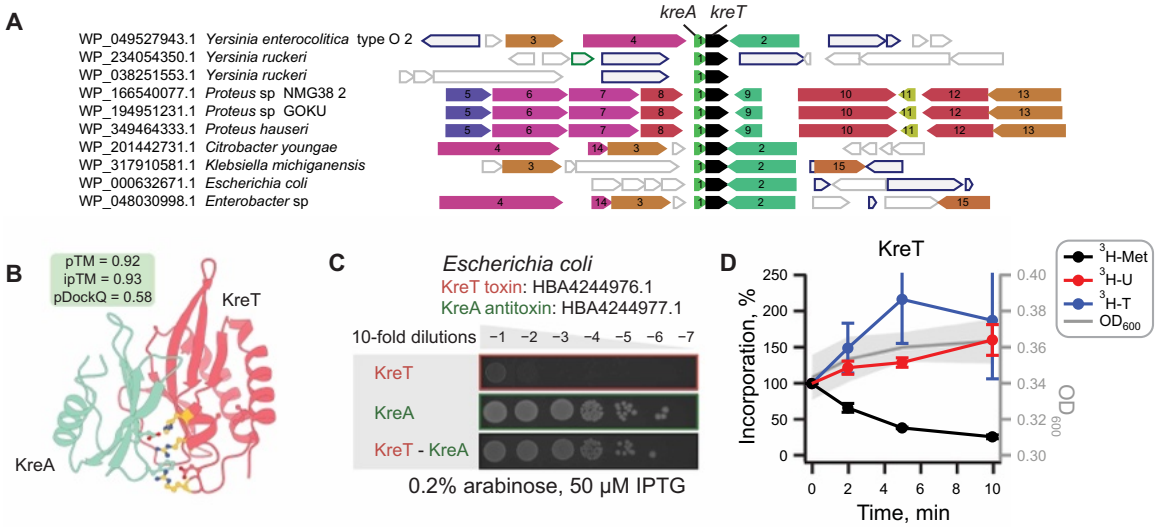


Fig. 6. KreTA is a prophage-encoded TA system. (**A**) Gene neighborhood analysis identifies putative TA system *kreTA* encoded in prophage regions of bacterial genomes (visualized here with FlaGs). The *kreT* gene used as a FlaGs query is shown in black. Proteins encoded by the numbered genes are annotated as follows: 1: KreA; 2: recombinase/integrase; 3: major capsid protein; 4: tape measure protein; 5: deoxyribose-phosphate aldolase; 6: thymidine phosphorylase; 7: phosphopentomutase; 8: purine-nucleoside phosphorylase; 9, 13, and 14: hypothetical proteins; 10: pyridoxal-dependent decarboxylase; 11: DksA/TraR family zinc finger protein; 12: MFS transporter; 15: LysR family transcriptional regulator. (**B**) KreA and KreT are predicted to form a dimer. Putative RNase active site residues determined from (*37*) and interfacing residues are shown as sticks. Interfacing active site residues are colored yellow. (**C**) Expression of KreT is toxic to *E. coli* BW25113 cells, but the toxicity is efficiently rescued by the antitoxin KreA. (**D**) Metabolic labeling assays follow the incorporation of ^{3}H-methionine (black traces), ^{3}H-uridine (red), and ^{3}H-thymidine (blue) upon the expression of KreT.

KreT inhibits translation, which is characteristic of RNase TA toxins (Fig. 6D) (*40*). Together, the fold similarity of KreT to endonucleases and its inhibition of translation suggest that it may function by cleaving mRNA, as observed for many known TA toxins (*40*). While KreT and KreA are viral in origin, being encoded on prophages, there are no relatives of either in the VAD that are identifiable at the sequence level. This highlights another challenge for understanding viral proteome diversity: that viruses can go under the radar by hiding within the genome of their hosts. While "cryptic" infections are most well known for phages, this is even the case for some eukaryotic viruses (*41*, *42*).

Conserved enzymatic folds are co-opted in antiviral defense and counterdefense

Viruses are territorial; when they have infected a host cell, it is in their interest to keep out competitor viruses. This is called superinfection exclusion and is observed in viruses that infect bacteria as well as eukaryotes (*43*, *44*). Temperate bacteriophages have been found to carry a rich diversity of phage defense systems (*45*). To analyze defense-like folds in the VAD, we used Foldseek to search our structures against the DefenseFinder database (*46*).

The largest number of hits to one defense system protein is for the DarG protein from the DarTG system with nine hits to the same heterohost structural cluster with a eukaryotic host-type representative (Fig. 7A; pink bar). Proteins in this cluster are all annotated as macrodomain or ADP-ribosyl glycohydrolase proteins (dataset S1). The alignments of these proteins to DarG align solely to their macrodomain and not to the C-terminal region implicated in binding DarT (fig. S7) (*47*). These hits may actually be counterdefense-related via the reversal of host ADP-ribosylation catalyzed in eukaryotic antiviral responses (*48*). However, given that macrodomains have multiple functions, we cannot rule out other roles beyond defense or counterdefense. Notably, six of the structural heterohost cluster proteins with hits to DarG belong to sequence monohost clusters, indicating a conservation of macrodomain folds across viruses infecting different domains of life, even when sequence similarity is not apparent (Fig. 7A).

Another notable hit to defense proteins is that of gp28 in the Panchino system. This protein contains restriction, modification, and specificity domains of type I restriction modification proteins (*49*). The VAD protein, which matches Panchino_gp28, is a hypothetical protein belonging to a heterohost structural cluster with a eukaryotic host-type representative. Within this cluster, the best-aligning protein is a bacterial host-type methyltransferase (Fig. 7A and dataset S1). Proteins in this cluster all have PDB hits to methyltransferase proteins (dataset S1). The alignments to Panchino_gp28 cover mostly the methyltransferase domain but also extend to the restriction domain (Fig. 7A). As before with DarG, the VAD protein with a hit to Panchino_gp28 belongs to a sequence monohost cluster and a structural heterohost cluster, providing another example of cross-domain conservation of folds not apparent on the sequence level (Fig. 7A).

As well as defending against superinfection, viruses carry counterdefense systems. To investigate whether counterdefense folds are found in the VAD, we searched against antidefense proteins in the DefenseFinder database (*46*) and dbAPIS (*50*). The antidefense protein with the most hits to VAD proteins is AcrIIA8 (Fig. 7B), a phage head-tail adapter protein that—unexpectedly—is implicated in antidefense by inhibiting Cas9 activity (*51*). AcrIIA8 has seven hits within the same structural cluster containing bacterial and archaeal host-type proteins, along with one hit to a bacterial host-type singleton cluster (Fig. 7B). When clustered by sequence, AcrIIA8 hits belong to either archaeal or bacterial host-type clusters, and these clusters merge at the structure level, reflecting a conserved fold

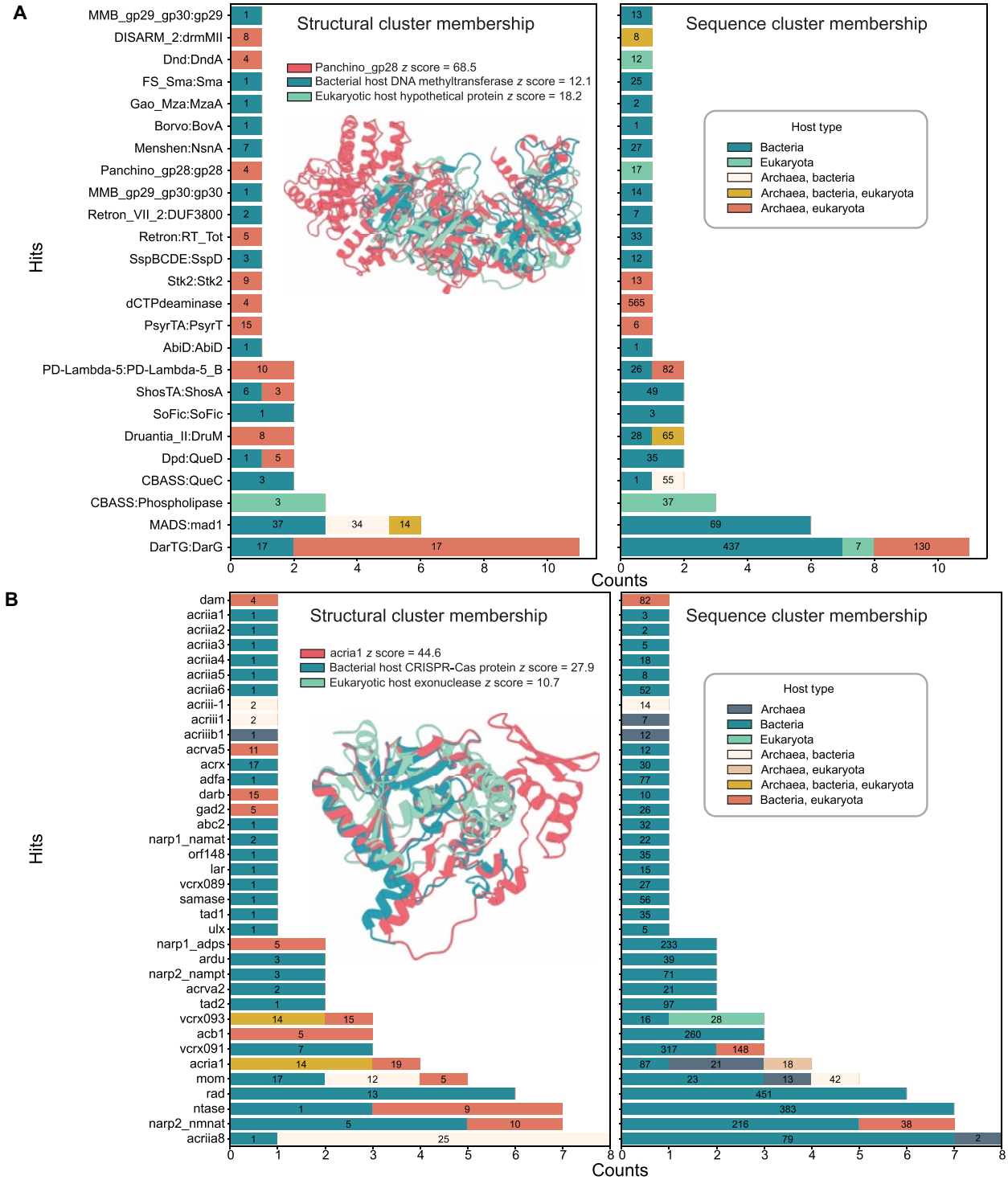


Fig. 7. Antiphage defense–related proteins in the VAD. Hits are colored by structural cluster host type and numbered by structural cluster size (left) and by sequence cluster host type and sequence cluster size (right). Each panel includes an example defense system protein, a VAD hit aligned to it, and a protein from a different host type within the same structural cluster, aligned using Dali. (**A**) DefenseFinder structural hits, with the example alignment showing Panchino_gp28 (*z* score = 68.5), a VAD hit (eukaryotic host hypothetical protein; *z* score = 18.2), and a different host type protein (bacterial host DNA methyltransferase; *z* score = 12.1). Numbers on bars refer to the size of the cluster that the protein with a hit to a defense protein belongs to. For example, DarG has 11 hits to two different cluster host types, each of which has a size of 17. (**B**) Combined structural hits from AntiDefenseFinder and dbAPIS, with the example alignment showing AcrIA1 (*z* score = 44.6), a VAD hit (phage CRISPR-Cas protein; *z* score = 27.9), and a different host type protein [viral exonuclease (eukaryotic host); *z* score = 10.7].

across host types (Fig. 7B). Proteins in these clusters are similarly annotated as phage head, neck, and tail structural proteins, and it is unclear whether they could also have additional moonlighting antidefense functions (dataset S1).

AcrIA1, an anti-CRISPR exonuclease affecting spacer acquisition in CRISPR-Cas–mediated defense (*52*), has hits to VAD proteins belonging to sequence and structure heterohost clusters (Fig. 7B). Structural representatives of these clusters are annotated as CRISPR-Cas associated and/or exonucleases, including a PD-(D/E)XK superfamily nuclease (dataset S1) (*53*). AcrIA1 aligns well with its bacterial host-type CRISPR-Cas–associated hit, as well as a eukaryote-host exonuclease found in the same structural cluster (Fig. 7B). This eukaryote-host exonuclease is remarkably similar in both sequence and structure to phage exonucleases (dataset S1). Furthermore, it has a hit to another CRISPR-Cas–associated antidefense protein, VCRX093 (Fig. 7B and dataset S1) (*54*). Given that eukaryotes do not carry CRISPR-Cas immune systems, this nuclease likely has an as-yet-undefined role in nucleic acid cleavage during the viral life cycle. In addition, its Foldseek hits to R354—a Cas4-like nuclease from the MIMIVIRE virophage resistance system in mimiviruses (*55*)—raise the possibility that it may represent a functionally analogous, virophage defense or genome processing element.

Last, phosphodiesterases have been found to be conserved across viruses that infect bacteria and eukaryotes (*10*). This class of proteins includes the antidefense protein anti-CBASS protein (Acb1), which degrades cyclic nucleotides (*11*, *56*). Acb1 also has hits to VAD proteins belonging to a structural heterohost cluster (Fig. 7B), further supporting its functional role in cross-domain counterdefense.

DISCUSSION

Comprising high-quality structural predictions of monomers and dimers, the VAD provides a resource for exploring viral diversity, evolution, and biology. There are significant functional "darkness" within viral proteomes and much novel biology awaiting discovery. Through searching for similarities of viral protein structures to communities of proteins in the dark corners of the protein universe, we have uncovered a previously unidentified Enterobacteriales prophage–encoded TA system, KreTA.

Our large-scale structural analysis reveals that despite vast sequence diversity, viruses often rely on conserved protein folds that cross host boundaries and recur across the tree of life. The structural similarities between viral proteins and defense-related systems highlight the repeated co-option of enzymatic folds in microbial conflicts. The observed conservation in viral defense and counterdefense systems is primarily confined to relatively large enzymes such as methyltransferases, ADP-ribose–processing macrodomains, and exonucleases. In contrast, small inhibitors, which are particularly common in counterdefense, appear to be less conserved. However, it is important to note that small proteins are inherently more difficult to detect confidently in structural homology searches, especially if their function depends on—and the structure changes upon—oligomerization. Overall, while viruses use diverse strategies to evade or counteract host defenses, the structural folds of enzymatic proteins involved in these processes are more likely to be preserved—or at least detectably conserved—across different domains of life. Notably, these conserved enzymatic folds are not unique to viruses but are repurposed from universal protein architectures that span across bacteria, archaea, and eukaryota. This structural conservation highlights the reuse of molecular mechanisms in host-virus interactions, underscoring that core and likely ancient protein folds remain essential components of defense and counterdefense systems. Together, our findings show that conserved structural frameworks underpin key functions in infection, defense, and adaptation across all domains of life.

MATERIALS AND METHODS

Sequence acquisition and host assignment of proteins

A total of 647,000 virus protein sequences was downloaded from the NCBI RefSeq (*20*) database (ncbi.nlm.nih.gov; April 2023). NCBI records for each protein sequence in the RefSeq virus dataset are used to attain taxonomy IDs for the virus the protein is found in. These IDs are then matched to entries in the Virus-Host database (*21*) to assign a host lineage to each protein. Approximately 10,000 RefSeq virus records lack taxonomy IDs, and about 3000 virus taxonomy IDs are not found in the Virus-Host database. Proteins associated with these missing or unrecognized taxonomy IDs, as well as those from viruses that infect other viruses and proteins without assigned hosts in the Virus-Host database, are labeled as "Host undefined." While these proteins are included in the dataset, they do not contribute to host-type classifications. Through manual curation, certain Virus-Host database hosts were determined to be wrongly or insufficiently classified. Viral lineages *Flyfo siphovirus Tbat1_6*, *uncultured phage*, and *Leviviridae* sp. were set to bacteria-infecting. Viral lineages *Dunaliella viridis virus* and *Tetraselmis viridis virus* were set to "Host undefined."

Sequence clustering and host assignment of clusters

MMSeqs2 (version 15.6f452) (*22*) was used to cluster RefSeq viral sequences and select representatives for structural analysis. The sequence identity threshold was set to 30%, and the query to target coverage threshold was set to 10%. This resulted in 117,479 clusters with 61,868 singletons. The clusters were assigned a host type on the basis of the host types of the proteins within them. To limit the effect of possible isolated cases of virus-host misannotation on heterohost assignment, we set a minimum threshold on the number of mixed hosts in order for a cluster to be assigned a heterohost label. We required that at least three proteins or 1% of the total cluster size (whichever is bigger) be of different host origin. The remaining clusters are classified as archaea, bacteria, or eukaryota monohost clusters. Only clusters composed solely of proteins without a defined host are classified as "Host undefined."

Structural prediction of monomers and homodimers

Protein structure predictions were made with AlphaFold2 version 2.3.2 (*1*) with default parameters. Homodimers were predicted with the AlphaFold-Multimer protocol (*31*). The structural template cutoff date was set to 14 May 2020 (--max_template_date = 2020-05-14). Sixty-three sequence cluster representatives were not predicted because of out-of-memory and database asynchronicity errors in the AlphaFold2 pipeline.

The quality of structural predictions was assessed with pLDDT (*1*) scores for monomers and predicted template modeling (pTM + ipTM) (*31*) and predicted DockQ (pDockQ) (*57*) scores for homodimers. The best models, determined by pLDDT scores for monomers and pTM + ipTM scores for homodimers, were selected for further analysis. Foldseek (*4*) exhaustive search

without thresholds was used to perform all-versus-all alignments of VAD monomers.

Structural clustering and host assignment of clusters

The VAD was structurally clustered with Foldseek (version 8.ef4e960) (*4*) in TMalign mode. Threshold parameters for the clustering were set at 90% for alignment coverage, 0.4 for the template modeling score (TM-score), and 0.001 for the *E*-value. Structural clusters were assigned a host type as was done with sequence clusters.

Searches against other databases

VAD structures and Foldseek (*4*) were used to search against the following structural databases: AFDB (*13*), PDB (*12*), and UniProt3D community representatives (*14*). The Dali webserver (*58*) was also used to search the PDB (*12*).

Genomic context analysis

For the 871 viral protein clusters belonging to prokaryotic hosts (i.e., bacteria, archaea, and bacteria and archaea) with a match to a dark community in the Protein Universe Atlas network (i.e., median functional brightness below 5%) with at least five protein members, genomic context analysis was conducted using GCsnap (*59*) with default parameters. For the 2313 tested Protein Universe Atlas community (*14*), all UniProt IDs linked to that community were used as input targets. To aid the visual inspection of gene neighborhoods, FlaGs2 (*60*) was also used.

Conserved genomic windows identified by GCsnap were analyzed for the distribution of protein family names (as automatically assigned by GCsnap) around viral protein targets. To uncover communities containing conserved bicistronic gene arrangements with characteristics of TA systems, genomic contexts containing target proteins from at least six different species were systematically screened. This approach was based on criteria adapted from NetFlax (*35*) and further generalized through the analysis of GCsnap outputs across diverse datasets. The gene pattern prediction relied on the following criteria: (i) Genes within a pattern must appear in the same order; (ii) intergenic distances between genes within the pattern should not exceed 100 nucleotides, while distances to conserved genes outside the pattern should exceed 100 nucleotides; (iii) neighborhood conservation should be restricted to the genes within the pattern; and (iv) the relative frequency difference between a pattern and the most frequent longer variant (i.e., one additional gene) should be maximal within the genomic context. This screening approach identified the most promising dark communities potentially harboring TA systems, providing a focused set of candidates for further inspection and functional analysis.

Predicting higher-order oligostates

Oligomeric states for PDB structures were obtained from the Swiss-Model Template Library (*61*). To predict the oligomeric state of a VAD cluster protein, we ran a Foldseek search against all chains in the PDB with a TM-score threshold of 0.5 and returned the most common oligomeric state of all the hits found weighted by the TM-score.

Searching against defense and antidefense databases

DefenseFinder (*46*) models were downloaded from https://defense-finder.mdmlab.fr/wiki/structure on 16 January 2025. For each defense system protein, one monomeric model was used in analyses. Antidefense protein hidden Markov models were manually downloaded from https://github.com/mdmparis/defense-finder-models (commit ca8f119). The PyHMMER (version 0.10.15) (*62*) function most_probable_sequence was used to extract sequences from hidden Markov models. AlphaFold2 (*1*) version 2.3.2 was used to model sequences extracted with the same runtime parameters as VAD structure prediction. This dataset is referred to as AntiDefenseFinder. dbAPIS (*50*) models were downloaded from https://bcb.unl.edu/dbAPIS/downloads/ on 9 February 2024. Foldseek search without prefiltering and with a TM-score threshold of 0.5 was used to compare VAD structures to DefenseFinder, AntiDefenseFinder, and dbAPIS models. For each VAD protein, the hit with the best identity fraction time query coverage was kept. After exhaustive searching, further filters were applied when plotting hits. DefenseFinder: alignment TM-score >0.65, *E*-value <0.001, average alignment coverage >0.6, cluster average TM-score >0.5. AntiDefenseFinder + dbAPIS: alignment TM-score >0.65, *E*-value <0.001 (for AntiDefenseFinder only), average alignment coverage >0.5, cluster average TM-score >0.5, intracluster alignment length >100. In addition, when plotting, AntiDefenseFinder and dbAPIS hits were combined where the hit with the best TM-score per VAD protein was kept.

Construction of plasmids

All primer and plasmid map designs were generated using SnapGene (GSL Biotech LLC) and Geneious Prime (Biomatters). All plasmids were constructed using circular polymerase extension cloning (*63*, *64*) using Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific). Detailed cloning strategies, including primer sequences and assembly schemes, are provided in table S3.

TA neutralization assays

The experiments were performed as described earlier, with minor modifications (*14*). The genes encoding the candidate toxin *kreT* (WP_000632671.1) and candidate antitoxin *kreA* (WP_001008346.1) were cloned into pBAD33 and pMG25 backbones, respectively, yielding VHP1906 (*kreT*) and VHP1970 (*kreA*) plasmids. *E. coli* BW25113 cells were cotransformed either with the plasmid pair expressing the TA system in trans or with one of the two protein-expressing plasmids being swapped for the appropriate empty vector. The cells were grown for 5 hours at 37°C with shaking at 200 rpm in an LB medium supplemented with carbenicillin (100 μg/ml; to maintain pMG25 and its derivatives), chloramphenicol (25 μg/ml; to maintain pBAD33 and its derivatives), and 0.2% glucose for the repression of the toxin expression. Next, the cells were diluted in an LB medium to a final optical density at 600 nm (OD_{600}) of 1.0, and serial 10-fold dilutions (from 10^{-1} to 10^{-7}) were made in LB. The dilutions were spotted on LB agar plates supplemented with carbenicillin (100 μg/ml), chloramphenicol (25 μg/ml), 50 μM IPTG (isopropyl-β-D-thiogalactopyranoside; for antitoxin expression), and 0.2% arabinose (for toxin expression). The plates were scored after an overnight incubation at 37°C.

Metabolic labeling assays

The labeling assays were also performed as described earlier, with minor modifications (*14*). A single colony of *E. coli* BW25113 cells expressing the KreT toxin from the pBAD33 derivative plasmid under the control of arabinose-inducible P_{BAD} promoter (VHP1976) was grown overnight in 2 ml of Neidhardt Mops minimal medium (*65*) supplemented with 1% glucose, 0.1% casein hydrolysate, and chloramphenicol (25 μg/ml) at 37°C with shaking at 160 rpm in MaxQ 6000 Shaker (Thermo Fisher Scientific) . The overnight culture was

used to inoculate a 20-ml culture in MOPS minimal medium supplemented with 0.5% glycerol as well as a set of 19 amino acids lacking methionine (each at 25 μg/ml) to the final OD_{600} of 0.05.

The cells were grown in a 100-ml flask at 37°C with shaking (160 rpm) in OLS Aqua Pro Shaking Water Bath (Grant Instruments) until recovery after the diauxic shift was observed (OD_{600} ~ 0.2 to 0.3). At this moment, 1-ml zero-time-point aliquots were taken and combined in 1.5-ml sterile Eppendorf tubes with either ^{3}H-thymidine (PerkinElmer) (10 μl per point, 4.5 μCi "hot" and 4.8 μM "cold" nucleotides in autoclaved distilled water), ^{3}H-uridine (PerkinElmer) (10 μl per point, 0.56 μCi "hot" and 4.97 μM "cold" nucleotides in autoclaved distilled water), or ^{3}H-methionine (Revvity) (10 μl per point, 0.67 μCi "hot" and 14.99 μM "cold" amino acids in autoclaved distilled water). After an 8-min incubation at 37°C, the zero-time-point labeling reactions were quenched by the addition of 200 μl of ice-cold 50% trichloroacetic acid (TCA) and transferred on ice.

Immediately after the collection of the zero time, the KreT expression was induced by adding L-arabinose to the culture to a final concentration of 0.2%. At 2, 5, 10, and 15 min postinduction, 1-ml culture aliquots were taken, combined with the "hot" label and processed analogously to the zero-time-point samples. OD_{600} values were recorded for each time point. The TCA-quenched samples were filtered through GF/C filters (Whatman) prewashed with 5 ml of 5% TCA. The filters were washed with 5 ml (twice for ^{3}H-methionine and ^{3}H-uridine and three times for ^{3}H-thymidine) of ice-cold 5% TCA and then with 95% ice-cold ethanol (5-ml washes, twice for ^{3}H-methionine and ^{3}H-uridine and three times for ^{3}H-thymidine). The filters were placed in 20-ml scintillation vials (Sarstedt) and air dried until they were dry (at least for 2 hours) at room temperature. Five milliliters of EcoLite Liquid (MP Biomedicals) scintillation cocktail was added to each vial, followed by shaking for 15 min before counting. The radioactivity was quantified in CPM (counts per minute) using a Hidex 600 SLe automatic liquid scintillation counter (Hidex). CPM values were normalized to OD_{600} at each time point, and incorporation percentages were calculated by dividing the normalized CPM/OD_{600} values by the corresponding value at zero-time-point values. All experiments were performed using three independent biological replicates and presented as the mean values ± SD.

Data visualization

The VAD network was built using NetworkX (version 3.4.2) (*66*). The network was drawn and visualized with Gephi (version 0.10.0) (*67*). Structures were visualized and superimposed with Chimera (*68*), and metabolic labeling data were visualized using Igor Pro 7 (WaveMetrics). The Pavian package (*69*) was used to plot Sankey diagrams. The VAD network was built using NetworkX (version 3.4.2), drawn and visualized with Gephi, and deployed with sigma.js. Web application protein structures are displayed using Mol* (*70*).

Supplementary Materials

The PDF file includes:

Figs. S1 to S7
Tables S1 and S2
Legend for table S3
Legend for dataset S1

Other Supplementary Material for this manuscript includes the following:

Table S3
Dataset S1

REFERENCES AND NOTES

1. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Zidek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
2. M. Baek, F. DiMaio, I. Anishchenko, J. Dauparas, S. Ovchinnikov, G. R. Lee, J. Wang, Q. Cong, L. N. Kinch, R. D. Schaeffer, C. Millan, H. Park, C. Adams, C. R. Glassman, A. DeGiovanni, J. H. Pereira, A. V. Rodrigues, A. A. van Dijk, A. C. Ebrecht, D. J. Opperman, T. Sagmeister, C. Buhlheller, T. Pavkov-Keller, M. K. Rathinaswamy, U. Dalwadi, C. K. Yip, J. E. Burke, K. C. Garcia, N. V. Grishin, P. D. Adams, R. J. Read, D. Baker, Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876 (2021).
3. Z. Yang, X. Zeng, Y. Zhao, R. Chen, AlphaFold2 and its applications in the fields of biology and medicine. *Signal Transduct. Target. Ther.* **8**, 115 (2023).
4. M. van Kempen, S. S. Kim, C. Tumescheit, M. Mirdita, J. Lee, C. L. M. Gilchrist, J. Soding, M. Steinegger, Fast and accurate protein structure search with Foldseek. *Nat. Biotechnol.* **42**, 243–246 (2024).
5. L. Holm, A. Laiho, P. Toronen, M. Salgado, DALI shines a light on remote homologs: One hundred discoveries. *Protein Sci.* **32**, e4519 (2023).
6. K. M. Peck, A. S. Lauring, Complexities of viral mutation rates. *J. Virol.* **92**, e01031-17 (2018).
7. J. L. Tenthorey, M. Emerman, H. S. Malik, Evolutionary landscapes of host-virus arms races. *Annu. Rev. Immunol.* **40**, 271–294 (2022).
8. M. Krupovic, E. V. Koonin, Multiple origins of viral capsid proteins from cellular ancestors. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E2401–E2410 (2017).
9. R. Khayat, L. Tang, E. T. Larson, C. M. Lawrence, M. Young, J. E. Johnson, Structure of an archaeal virus capsid protein reveals a common ancestry to eukaryotic and bacterial viruses. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 18944–18949 (2005).
10. J. Nomburg, E. E. Doherty, N. Price, D. Bellieny-Rabelo, Y. K. Zhu, J. A. Doudna, Birth of protein folds and functions in the virome. *Nature* **633**, 710–717 (2024).
11. S. J. Hobbs, J. Nomburg, J. A. Doudna, P. J. Kranzusch, Animal and bacterial viruses share conserved mechanisms of immune evasion. *Cell* **187**, 5530–5539.e8 (2024).
12. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
13. M. Varadi, S. Anyango, M. Deshpande, S. Nair, C. Natassia, G. Yordanova, D. Yuan, O. Stroe, G. Wood, A. Laydon, A. Zidek, T. Green, K. Tunyasuvunakool, S. Petersen, J. Jumper, E. Clancy, R. Green, A. Vora, M. Lutfi, M. Figurnov, A. Cowie, N. Hobbs, P. Kohli, G. Kleywegt, E. Birney, D. Hassabis, S. Velankar, AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, D439–D444 (2022).
14. J. Durairaj, A. M. Waterhouse, T. Mets, T. Brodiazhenko, M. Abdullah, G. Studer, G. Tauriello, M. Akdel, A. Andreeva, A. Bateman, T. Tenson, V. Hauryliuk, T. Schwede, J. Pereira, Uncovering new families and folds in the natural protein universe. *Nature* **622**, 646–653 (2023).
15. C. A. Suttle, Marine viruses — Major players in the global ecosystem. *Nat. Rev. Microbiol.* **5**, 801–812 (2007).
16. D. Paez-Espino, E. A. Eloe-Fadrosh, G. A. Pavlopoulos, A. D. Thomas, M. Huntemann, N. Mikhailova, E. Rubin, N. N. Ivanova, N. C. Kyrpides, Uncovering Earth's virome. *Nature* **536**, 425–430 (2016).
17. T. Crepin, C. Swale, A. Monod, F. Garzoni, M. Chaillet, I. Berger, Polyproteins in structural biology. *Curr. Opin. Struct. Biol.* **32**, 139–146 (2015).
18. S. A. Yost, J. Marcotrigiano, Viral precursor polyproteins: Keys of regulation from replication to maturation. *Curr. Opin. Virol.* **3**, 137–142 (2013).
19. R. S. Kim, E. Levy Karin, M. Mirdita, R. Chikhi, M. Steinegger, BFVD—A large repository of predicted viral protein structures. *Nucleic Acids Res.* **53**, D340–D347 (2025).
20. N. A. O'Leary, M. W. Wright, J. R. Brister, S. Ciufo, D. Haddad, R. McVeigh, B. Rajput, B. Robbertse, B. Smith-White, D. Ako-Adjei, A. Astashyn, A. Badretdin, Y. Bao, O. Blinkova, V. Brover, V. Chetvernin, J. Choi, E. Cox, O. Ermolaeva, C. M. Farrell, T. Goldfarb, T. Gupta, D. Haft, E. Hatcher, W. Hlavina, V. S. Joardar, V. K. Kodali, W. Li, D. Maglott, P. Masterson, K. M. McGarvey, M. R. Murphy, K. O'Neill, S. Pujar, S. H. Rangwala, D. Rausch, L. D. Riddick, C. Schoch, A. Shkeda, S. S. Storz, H. Sun, F. Thibaud-Nissen, I. Tolstoy, R. E. Tully, A. R. Vatsan, C. Wallin, D. Webb, W. Wu, M. J. Landrum, A. Kimchi, T. Tatusova, M. DiCuccio, P. Kitts, T. D. Murphy, K. D. Pruitt, Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–D745 (2016).
21. T. Mihara, Y. Nishimura, Y. Shimizu, H. Nishiyama, G. Yoshikawa, H. Uehara, P. Hingamp, S. Goto, H. Ogata, Linking virus genomes with host taxonomy. *Viruses* **8**, 66 (2016).
22. M. Steinegger, J. Soding, MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, 1026–1028 (2017).

23. M. Bekliz, P. Colson, B. La Scola, The expanding family of virophages. *Viruses* **8**, 317 (2016).
24. Y. Zhang, F. Maley, G. F. Maley, G. Duncan, D. D. Dunigan, J. L. Van Etten, Chloroviruses encode a bifunctional dCMP-dCTP deaminase that produces two key intermediates in dTTP formation. *J. Virol.* **81**, 7662–7671 (2007).
25. B. Y. Hsueh, G. B. Severin, C. A. Elg, E. J. Waldron, A. Kant, A. J. Wessel, J. A. Dover, C. R. Rhoades, B. J. Ridenhour, K. N. Parent, M. B. Neiditch, J. Ravi, E. M. Top, C. M. Waters, Phage defence by deaminase-mediated depletion of deoxynucleotides in bacteria. *Nat. Microbiol.* **7**, 1210–1220 (2022).
26. P. Youkharibache, S. Veretnik, Q. Li, K. A. Stanek, C. Mura, P. E. Bourne, The small β-barrel domain: A survey-based structural analysis. *Structure* **27**, 6–26 (2019).
27. A. M. Monzon, P. N. Arrias, A. Elofsson, P. Mier, M. A. Andrade-Navarro, M. Bevilacqua, D. Clementel, A. Bateman, L. Hirsh, M. S. Fornasari, G. Parisi, D. Piovesan, A. V. Kajava, S. C. E. Tosatto, A STRP-ed definition of structured tandem repeats in proteins. *J. Struct. Biol.* **215**, 108023 (2023).
28. A. E. Douglas, K. D. Corbett, J. M. Berger, G. McFadden, T. M. Handel, Structure of M11L: A myxoma virus structural homolog of the apoptosis inhibitor, Bcl-2. *Protein Sci.* **16**, 695–703 (2007).
29. E. M. Marcotte, M. Pellegrini, T. O. Yeates, D. Eisenberg, A census of protein repeats. *J. Mol. Biol.* **293**, 151–160 (1999).
30. K. S. Gajiwala, S. K. Burley, Winged helix proteins. *Curr. Opin. Struct. Biol.* **10**, 110–116 (2000).
31. R. Evans, M. O'Neill, A. Pritzel, N. Antropova, A. Senior, T. Green, A. Žídek, R. Bates, S. Blackwell, J. Yim, O. Ronneberger, S. Bodenstein, M. Zielinski, A. Bridgland, A. Potapenko, A. Cowie, K. Tunyasuvunakool, R. Jain, E. Clancy, P. Kohli, J. Jumper, D. Hassabis, Protein complex prediction with AlphaFold-Multimer. bioRxiv 463034 (2022).
32. M. S. Svetlov, C. F. Dunand, J. A. Nakamoto, G. C. Atkinson, H. A. Safdari, D. N. Wilson, N. Vazquez-Laslop, A. S. Mankin, Peptidyl-tRNA hydrolase is the nascent chain release factor in bacterial ribosome-associated quality control. *Mol. Cell* **84**, 715–726.e5 (2024).
33. F. Cuzin, N. Kretchmer, R. E. Greenberg, R. Hurwitz, F. Chapeville, Enzymatic hydrolysis of N-substituted aminoacyl-tRNA. *Proc. Natl. Acad. Sci. U.S.A.* **58**, 2079–2086 (1967).
34. N. A. Oviedo, H. Salgado, J. Collado-Vides, G. Guarneros, Distribution of minigenes in the bacteriophage lambda chromosome. *Gene* **329**, 115–124 (2004).
35. T. Mets, T. Kurata, K. Ernits, M. J. O. Johansson, S. Z. Craig, G. M. Evora, J. A. Buttress, R. Odai, K. C. Wallant, J. A. Nakamoto, L. Shyrokova, A. A. Egorov, C. R. Doering, T. Brodiazhenko, M. T. Laub, T. Tenson, H. Strahl, C. Martens, A. Harms, A. Garcia-Pino, G. C. Atkinson, V. Hauryliuk, Mechanism of phage sensing and restriction by toxin-antitoxin-chaperone systems. *Cell Host Microbe* **32**, 1059–1073.e8 (2024).
36. T. Zhang, H. Tamman, K. Coppieters 't Wallant, T. Kurata, M. LeRoux, S. Srikant, T. Brodiazhenko, A. Cepauskas, A. Talavera, C. Martens, G. C. Atkinson, V. Hauryliuk, A. Garcia-Pino, M. T. Laub, Direct activation of a bacterial innate immune system by a viral capsid protein. *Nature* **612**, 132–140 (2022).
37. B. Odaert, F. Saida, P. Aliprandi, S. Durand, J. B. Crechet, R. Guerois, S. Laalami, M. Uzan, F. Bontems, Structural and functional studies of RegB, a new member of a family of sequence-specific ribonucleases involved in mRNA inactivation on the ribosome. *J. Biol. Chem.* **282**, 2019–2028 (2007).
38. F. Saida, M. Uzan, F. Bontems, The phage T4 restriction endoribonuclease RegB: A cyclizing enzyme that requires two histidines to be fully active. *Nucleic Acids Res.* **31**, 2751–2758 (2003).
39. D. Jurenas, M. Rey, D. Byrne, J. Chamot-Rooke, L. Terradot, E. Cascales, Salmonella antibacterial Rhs polymorphic toxin inhibits translation through ADP-ribosylation of EF-Tu P-loop. *Nucleic Acids Res.* **50**, 13114–13127 (2022).
40. K. Ernits, C. K. Saha, T. Brodiazhenko, B. Chouhan, A. Shenoy, J. A. Buttress, J. J. Duque-Pedraza, V. Bojar, J. A. Nakamoto, T. Kurata, A. A. Egorov, L. Shyrokova, M. J. O. Johansson, T. Mets, A. Rustamova, J. Dzigurski, T. Tenson, A. Garcia-Pino, H. Strahl, A. Elofsson, V. Hauryliuk, G. C. Atkinson, The structural basis of hyperpromiscuity in a core combinatorial network of type II toxin-antitoxin and related phage defense systems. *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2305393120 (2023).
41. H. Zhao, L. Meng, H. Hikida, H. Ogata, Eukaryotic genomic data uncover an extensive host range of mirusviruses. *Curr. Biol.* **34**, 2633–2643.e3 (2024).
42. M. P. Erazo-Garcia, U. Sheyn, Z. K. Barth, R. J. Craig, P. Wessman, A. M. Jivaji, W. K. Ray, M. Svensson-Coelho, C. K. Cornwallis, K. Rengefors, C. P. D. Brussaard, M. Moniruzzaman, F. O. Aylward, Cryptic infection of a giant virus in a unicellular green alga. *Science* **388**, eads6303 (2025).
43. S. Y. Folimonova, Superinfection exclusion is an active virus-controlled function that requires a specific viral protein. *J. Virol.* **86**, 5554–5561 (2012).
44. M. Hunter, D. Fusco, Superinfection exclusion: A viral strategy with short-term benefits and long-term drawbacks. *PLoS Comput. Biol.* **18**, e1010125 (2022).
45. P. H. Patel, K. L. Maxwell, Prophages provide a rich source of antiphage defense systems. *Curr. Opin. Microbiol.* **73**, 102321 (2023).
46. F. Tesson, R. Planel, A. A. Egorov, H. Georjon, H. Vaysset, B. Brancotte, B. Néron, E. Mordret, G. C. Atkinson, A. Bernheim, J. Cury, A comprehensive resource for exploring antiphage defense: DefenseFinder Webservice, Wiki and Databases. *Peer Community J.* **4**, e91 (2024).
47. M. LeRoux, S. Srikant, G. I. C. Teodoro, T. Zhang, M. L. Littlehale, S. Doron, M. Badiee, A. K. L. Leung, R. Sorek, M. T. Laub, The DarTG toxin-antitoxin system provides phage defence by ADP-ribosylating viral DNA. *Nat. Microbiol.* **7**, 1028–1040 (2022).
48. A. R. Fehr, G. Jankevicius, I. Ahel, S. Perlman, Viral macrodomains: Unique mediators of viral replication and pathogenesis. *Trends Microbiol.* **26**, 598–610 (2018).
49. R. M. Dedrick, D. Jacobs-Sera, C. A. Bustamante, R. A. Garlena, T. N. Mavrich, W. H. Pope, J. C. Reyes, D. A. Russell, T. Adair, R. Alvey, J. A. Bonilla, J. S. Bricker, B. R. Brown, D. Byrnes, S. G. Cresawn, W. B. Davis, L. A. Dickson, N. P. Edgington, A. M. Findley, U. Golebiewska, J. H. Grose, C. F. Hayes, L. E. Hughes, K. W. Hutchison, S. Isern, A. A. Johnson, M. A. Kenna, K. K. Klyczek, C. M. Mageeney, S. F. Michael, S. D. Molloy, M. T. Montgomery, J. Neitzel, S. T. Page, M. C. Pizzorno, M. K. Poxleitner, C. A. Rinehart, C. J. Robinson, M. R. Rubin, J. N. Teyim, E. Vazquez, V. C. Ware, J. Washington, G. F. Hatfull, Prophage-mediated defence against viral attack and viral counter-defence. *Nat. Microbiol.* **2**, 16251 (2017).
50. Y. Yan, J. Zheng, X. Zhang, Y. Yin, dbAPIS: A database of anti-prokaryotic immune system genes. *Nucleic Acids Res.* **52**, D419–D425 (2024).
51. R. V. Uribe, E. van der Helm, M. A. Misiakou, S. W. Lee, S. Kol, M. O. A. Sommer, Discovery and characterization of Cas9 inhibitors disseminated across seven bacterial phyla. *Cell Host Microbe* **25**, 233–241 e235 (2019).
52. Z. Zhang, S. Pan, T. Liu, Y. Li, N. Peng, Cas4 nucleases can effect specific integration of CRISPR spacers. *J. Bacteriol.* **201**, e00747-18 (2019).
53. K. Steczkiewicz, A. Muszewska, L. Knizewski, L. Rychlewski, K. Ginalski, Sequence, structure and functional diversity of PD-(D/E)XK phosphodiesterase superfamily. *Nucleic Acids Res.* **40**, 7016–7045 (2012).
54. D. Roy, K. T. Huguet, F. Grenier, V. Burrus, IncC conjugative plasmids and SXT/R391 elements repair double-strand breaks caused by CRISPR-Cas during conjugation. *Nucleic Acids Res.* **48**, 8815–8827 (2020).
55. C. Dou, M. Yu, Y. Gu, J. Wang, K. Yin, C. Nie, X. Zhu, S. Qi, Y. Wei, W. Cheng, Structural and mechanistic analyses reveal a unique Cas4-like protein in the mimivirus virophage resistance element system. *iScience* **3**, 1–10 (2018).
56. S. J. Hobbs, T. Wein, A. Lu, B. R. Morehouse, J. Schnabel, A. Leavitt, E. Yirmiya, R. Sorek, P. J. Kranzusch, Phage anti-CBASS and anti-Pycsar nucleases subvert bacterial immunity. *Nature* **605**, 522–526 (2022).
57. P. Bryant, G. Pozzati, A. Elofsson, Improved prediction of protein-protein interactions using AlphaFold2. *Nat. Commun.* **13**, (2022).
58. L. Holm, Dali server: Structural unification of protein families. *Nucleic Acids Res.* **50**, W210–W215 (2022).
59. J. Pereira, GCsnap: Interactive snapshots for the comparison of protein-coding genomic contexts. *J. Mol. Biol.* **433**, 166943 (2021).
60. C. K. Saha, R. Sanches Pires, H. Brolin, M. Delannoy, G. C. Atkinson, FlaGs and webFlaGs: Discovering novel biology through the analysis of gene neighbourhood conservation. *Bioinformatics* **37**, 1312–1314 (2021).
61. S. Bienert, A. Waterhouse, T. A. de Beer, G. Tauriello, G. Studer, L. Bordoli, T. Schwede, The SWISS-MODEL Repository—New features and functionality. *Nucleic Acids Res.* **45**, D313–D319 (2017).
62. M. Larralde, G. Zeller, PyHMMER: A Python library binding to HMMER for efficient sequence analysis. *Bioinformatics* **39**, btad214 (2023).
63. J. Quan, J. Tian, Circular polymerase extension cloning for high-throughput cloning of complex and combinatorial DNA libraries. *Nat. Protoc.* **6**, 242–251 (2011).
64. J. Quan, J. Tian, Circular polymerase extension cloning of complex gene libraries and pathways. *PLOS ONE* **4**, e6441 (2009).
65. F. C. Neidhardt, P. L. Bloch, D. F. Smith, Culture medium for enterobacteria. *J. Bacteriol.* **119**, 736–747 (1974).
66. A. Hagberg, P. J. Swart, D. A. Schult, "Exploring network structure, dynamics, and function using NetworkX," in *Proceedings of the 7th Python in Science Conference (SciPy 2008)* (SciPy Proceedings, 2008), pp. 11–16.
67. M. Bastian, S. Heymann, M. Jacomy, Gephi: An open source software for exploring and manipulating networks. *Proc. Int. AAAI Conf. Weblogs Soc. Media* **3**, 361–362 (2009).
68. E. C. Meng, T. D. Goddard, E. F. Pettersen, G. S. Couch, Z. J. Pearson, J. H. Morris, T. E. Ferrin, U. C. S. F. ChimeraX, Tools for structure building and analysis. *Protein Sci.* **32**, e4792 (2023).
69. F. P. Breitwieser, S. L. Salzberg, Pavian: interactive analysis of metagenomics data for microbiome studies and pathogen identification. *Bioinformatics* **36**, 1303–1304 (2020).
70. D. Sehnal, S. Bittrich, M. Deshpande, R. Svobodova, K. Berka, V. Bazgier, S. Velankar, S. K. Burley, J. Koca, A. S. Rose, Mol* Viewer: Modern web app for 3D visualization and analysis of large biomolecular structures. *Nucleic Acids Res.* **49**, W431–W437 (2021).

Acknowledgments: Computations were enabled by the Berzelius resource provided by the Knut and Alice Wallenberg Foundation at the National Supercomputer Centre and by resources provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS), partially funded by the Swedish Research Council through grant agreement no. 2022-06725. Also, we acknowledge sciCORE at the University of Basel (https://scicore.unibas.ch/) for providing computational resources and system administration support. We thank

A. Egorov for help in deploying the web server tool. We thank the members of the Schwede group at the University of Basel for valuable support. **Funding:** This work was supported by the following: The Knut and Alice Wallenberg Foundation (project grant 2020-0037 to G.C.A. and V.H.), The Swedish Research Council (Vetenskapsrådet) grants (2022-01603, 2023-02353, and 2024-06071 to G.C.A.; 2021-01146 and 2024-06059 to V.H.), The Estonian Research Council (PRG2696 to V.H.), Göran Gustafsson Foundation for Research in Natural Sciences and Medicine (the Göran Gustafsson Prize to V.H.), The University of Basel Research Fund for Excellent Junior Researchers (U.570.0006 to J.P.), The Swiss National Science Foundation (SNSF; Ambizione grant 223634 to J.D.), The Royal Physiographic Society of Lund (Endowments for the Natural Sciences, Medicine and Technology, number 45414 to R.O.), and Erasmus+ traineeship grant to M.A. **Author contributions:** Conceptualization: G.C.A., J.D., J.P., R.O., and V.H. Data curation: G.C.A., J.D., M.L., R.O., T.A.-M., and V.H. Project administration: G.C.A., J.P., R.O., T.T., and V.H. Methodology: G.C.A., J.D., J.P., L.S., M.A., M.L., R.O., T.A.-M., and V.H. Investigation: G.C.A., J.D., J.P., L.S., M.A., M.L., R.O., and T.A.-M. Visualization: G.C.A., J.D., M.A., R.O., T.A.-M., and V.H. Supervision: G.C.A., J.P., R.O., T.T., and V.H. Writing—original draft: G.C.A., J.D., J.P., M.A., M.L., R.O., and T.A.-M. Writing—review and editing: G.C.A., J.D., J.P., L.S., M.A., R.O., T.A.-M., T.T., and V.H. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. VAD- and AntiDefenseFinder-predicted structures are available at https://data-sharing.atkinson-lab.com/vad/. The structural clusters can be searched and explored at https://vad.atkinson-lab.com/.

Submitted 17 June 2025
Accepted 29 August 2025
Published 1 October 2025
10.1126/sciadv.adz8560

ScienceAdvances

The Viral AlphaFold Database of monomers and homodimers reveals conserved protein folds in viruses of bacteria, archaea, and eukaryotes

Roni Odai, Michèle Leemann, Tamim Al-Murad, Minhal Abdullah, Lena Shyrokova, Tanel Tenson, Vasili Hauryliuk, Janani Durairaj, Joana Pereira, and Gemma C. Atkinson

Sci. Adv. **11** (40), eadz8560. DOI: 10.1126/sciadv.adz8560

NAR Genomics and Bioinformatics, 2025, **7**, lqaf009
https://doi.org/10.1093/nargab/lqaf009
Application Notes

OXFORD

LoVis4u: a locus visualization tool for comparative genomics and coverage profiles

Artyom A. Egorov * **and Gemma C. Atkinson** *

Department of Experimental Medical Science, Lund University, 221 84, Lund, Sweden

*To whom correspondence should be addressed. Email: artem.egorov@med.lu.se
Correspondence may also be addressed to Gemma C. Atkinson. Email: gemma.atkinson@med.lu.se

Abstract

Comparative genomic analysis often involves visualization of alignments of genomic loci. While several software tools are available for this task, ranging from Python and R libraries to stand-alone graphical user interfaces, a tool is lacking that offers fast, automated usage and the production of publication-ready vector images. Here we present LoVis4u, a command-line tool and Python API designed for highly customizable and fast visualization of multiple genomic loci. LoVis4u generates vector images in PDF format based on annotation data from GenBank or GFF files. It is capable of visualizing entire genomes of bacteriophages as well as plasmids and user-defined regions of longer prokaryotic genomes. Additionally, LoVis4u offers optional data processing steps to identify and highlight accessory and core genes in input sequences. Finally, LoVis4u supports the visualization of genomic signal track profiles from sequencing experiments. LoVis4u is implemented in Python3 and runs on Linux and MacOS. The command-line interface covers most practical use cases, while the provided Python API allows usage within a Python program, integration into external tools, and additional customization. The source code is available at the GitHub page: github.com/art-egorov/lovis4u. Detailed documentation that includes an example-driven guide is available from the software home page: art-egorov.github.io/lovis4u.

Introduction

The exponential growth of microbial genome databases has unlocked numerous opportunities for comparative genomic analyses [1]. Various tasks such as analysis of gene neighbourhood conservation [2, 3], annotation of functional short open reading frames (ORFs) [4, 5], and investigation of genomic variability hotspots [6–8] often require visualization of multiple genomic loci. Several software tools have been developed for this purpose. A subset of these have graphical user interfaces (GUIs), such as the Artemis Comparison Tool [9], Easyfig [10], GeneSpy [11], and Geneious Prime (geneious.com). Another category comprises web-based applications like Gene Graphics [12]. Additionally, there are libraries such as the R packages genoPlotR [13] and gggenes [14], as well as the Python package GenomeDiagram [15]. Some tools integrate multiple approaches, creating hybrid solutions. For example, GEnView is a Python pipeline combined with an interactive web application [16], and Clinker & clustermap.js [17] is a popular tool with a command-line interface and an interactive web application that can generate vector graphics. While many of these tools feature interactivity through GUIs or web applications, there is a lack of a user-friendly command-line tool suitable for handling multiple input genomes, containing data analysis steps, flexible customization options, and with fast production of aesthetically pleasing publication-ready figures.

Here we present LoVis4u (Locus Visualization), a scalable software tool designed for customizable and fast visualization of multiple genomic loci. LoVis4u offers a command-line interface without requiring user-side scripting and provides a Python API for additional customization and integration within Python programs. In addition to visualization features, our tool has optional data analysis steps that include protein clustering to find groups of protein homologues with subsequent identification of accessory and core genes that can be highlighted with visualization. In addition, hidden Markov model searches enable fast functional annotation of proteins using a broad set of model databases. At the nucleotide level, LoVis4u can visualize sequencing track data such as GC content and GC skew, along with next-generation sequencing coverage profiles. While LoVis4u was designed for annotating and visualizing multiple bacteriophage genomes or their loci, it can also be used for visualizing defined regions of any prokaryotic genome.

Materials and methods

Design and implementation

The LoVis4u pipeline includes several default data processing steps, which are optional depending on user needs (Fig. 1A). LoVis4u supports input data in either GenBank or extended GFF (concatenated with the corresponding nucleotide sequence in fasta format) file format. GFF files in this format are produced by the widely used prokka [18] and pharokka [19] genome annotation tools.

By default, LoVis4u applies the MMseqs2 [20] protein clustering algorithm to all encoded protein sequences to identify groups of homologous proteins. Alternatively, a table with predefined protein groups can be used as input. Based on the defined protein groups, LoVis4u constructs a matrix of pairwise proteome composition similarity scores that reflect the fraction of shared homologous proteins between sequences, and a corresponding proteome composition

Received: November 24, 2024. **Revised:** January 16, 2025. **Editorial Decision:** February 3, 2025. **Accepted:** February 5, 2025

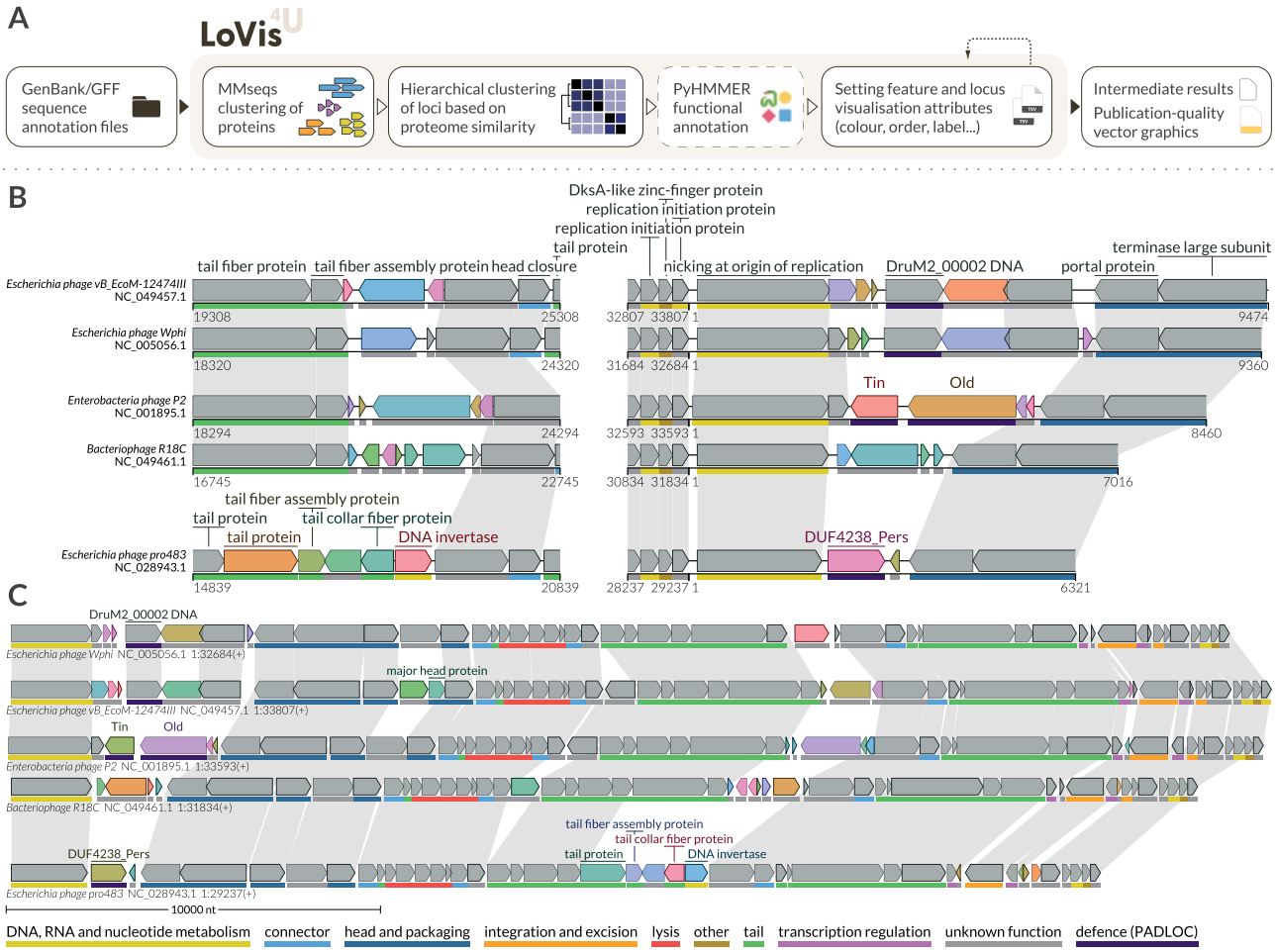


Figure 1. LoVis4u workflow and visualization example. (**A**) Schematic description of the LoVis4u pipeline and its input and output. (**B**) Visualization of multiple regions for a set of P2-like phage genomes. Conserved genes are shown in grey, while variable protein groups are highlighted with distinct colours. Homologous protein groups defined by MMSeqs2 clustering and located on different genomes are connected by grey homology lines. As can be set in the configuration file, protein labels are hidden for hypothetical proteins or proteins of unknown function. Labels for conserved proteins are shown only for the first occurrence since homology lines indicate additional occurrences. Label positions are arranged automatically by the algorithm to avoid overlaps of text. PHROG and PyHMMER functional annotations [38] are indicated by coloured lines beneath each ORF, according to the colour code at the bottom of the panel. (**C**) Visualization of a set of P2-like phage genomes that showcases a compact visualization of full-length sequences, where functional annotation tracks and individual *x*-axes are hidden. Instead, a scale line at the bottom is displayed to indicate region size.

distance matrix. The clusters and order of input sequences for visualization are determined using hierarchical clustering with average linkage applied to the distance matrix. This approach allows us to consider a set of genes for each cluster analogously to pangenome analyses, where we define 'conserved', 'intermediate', and 'variable' protein group classes that are roughly equivalent to 'core', 'shell', and 'cloud' terms of pangenomics [21]. Finally, the optional PyHMMER [22, 23] hmmscan search step for protein functional annotation can be applied. The list of available databases for search includes DefenseFinder and CasFinder (defence) [24, 25], PADLOC (defence) [26], dbAPIS_Acr (anti-defence) [27], VFDB (virulence factors) [28], and AMRFinderPlus (AMR genes) [29]. LoVis4u can read genomic signal data in bedGraph and bigWig file formats [30]. A detailed workflow description is available on the LoVis4u home page (art-egorov.github.io/lovis4u).

LoVis4u can be run in quick-start mode, with few required options, but also has a range of advanced customization options to give users full control over the output. Sequence order, sequence clusters, and protein group classes can be manually specified through feature and locus annotation tables, which can be provided as optional arguments. LoVis4u has a set of editable configuration files with advanced parameters already adjusted for different tasks and page layouts, including one- and two-column A4 page layouts. Users can also customize colours and labels with these files. A step-by-step guide on the home page demonstrates how to experiment with and optimize visualizations using these features. Finally, LoVis4u uses the ReportLab library API to generate the output vector image, which is saved in PDF format. This format allows further editing of all objects in vector image-editing programs like Adobe Illustrator or Inkscape if needed.

LoVis4u is implemented in Python3 and uses multiple Python3 libraries: biopython [31], bcbio-gff, scipy [32], configs, argparse, pandas [33], distinctipy [34], matplotlib [35], seaborn [36], reportlab, pyhmmer [22, 23], progress, and requests. LoVis4u also uses MMseqs2 [20] and bigWigToBed-

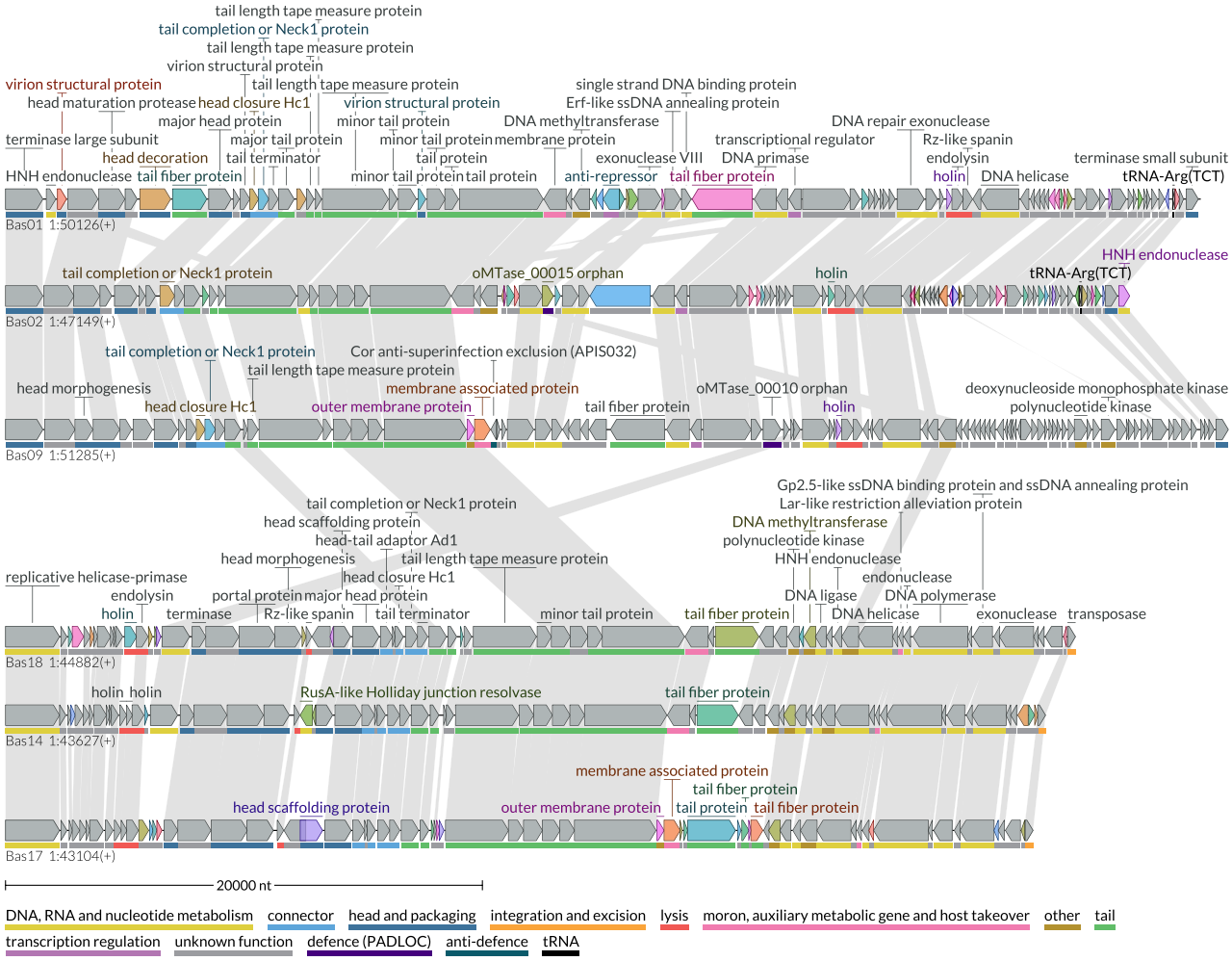


Figure 2. LoVis4u identifies regions of homology within the BASEL phage collection. Visualization of a set of BASEL [40] phage genomes. Colours and annotations are as per Fig. 1.

Graph [30] as non-Python dependencies, which are embedded in the library.

Results

Demonstration of LoVis4u key features in comparative genomics

To test and demonstrate a subset of the tool's key features, we visualized a set of P2-like phage genomes obtained from the RefSeq database [37] (Fig. 1B and C). LoVis4u allows users to specify multiple regions for each sequence, which are displayed on a single line. Additionally, if a sequence is circular (as is the case with many phage and plasmid sequences), the visualization will continue without a gap between sequence end and start coordinates. By default, LoVis4u highlights 'variable' proteins by automatically assigning different colours to each homologous protein group. Alternatively, conserved proteins can be highlighted. LoVis4u also automatically interprets PHROG [38] functional group annotations, as provided in GFF and GenBank files produced by Pharokka together with results of PyHMMER search, and can display these annotations with a functional category line beneath the ORFs. To better visualize long sequences, LoVis4u includes options for a more minimalistic and compact design as shown in panel Fig. 1C. A step-by-step guide for creating these figures can be found in the user guide on the tool's home page together with the gallery of more examples and command-line options to reproduce them: art-egorov.github.io/lovis4u/Gallery/gallery.

The capabilities and potential of the LoVis4u API was recently demonstrated in the systematic annotation of variable hotspots in phage genomes and plasmid sequences, where LoVis4u was integrated for high-throughput and fast visualization of millions of loci [39].

Exploration of the BASEL collection of phages with LoVis4u

To demonstrate full-length sequence visualization of a larger set of sequences, we used LoVis4u to create an overview of the complete BASEL collection of 78 phages [40] in a single figure, highlighting the variable genes (Supplementary File S1). The pipeline, which includes clustering of 13 630 protein sequences, hierarchical clustering of sequences, and generating the final graphic output, completed in only 50 seconds on an M1 MacBook Pro laptop. Beyond visualization, a key feature of LoVis4u is its ability to identify clusters of similar proteomes within the dataset and distinguish accessory and core proteins within each cluster. LoVis4u also generates a pairwise proteome similarity matrix, visualized for the BASEL col-

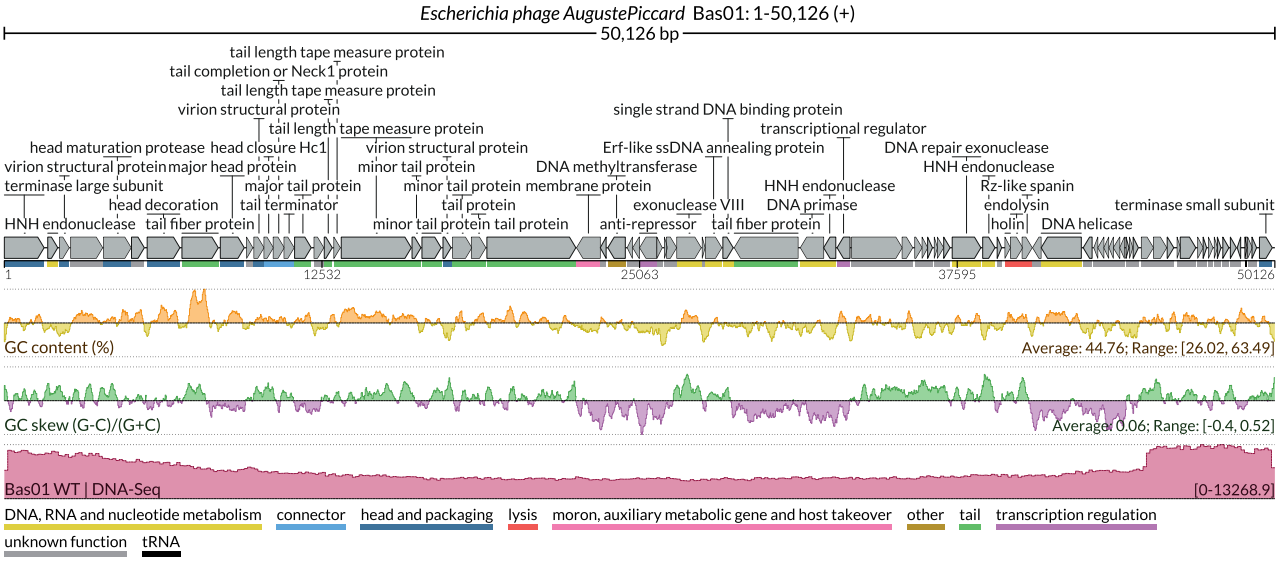


Figure 3. LoVis4u genome browser mode. Visualization of the *Bas01 Escherichia* phage Auguste Piccard genome. The functional annotation track and gene map are as shown in Fig. 1. GC content and GC skew tracks are centred at average values, with positive and negative relative values represented by different colours. The range and average values are annotated in the bottom right corner of each track. The genomic signal track is shown in the lowest track, in dark red. Multiple additional signal tracks, e.g. to compare wild type and mutants, can be added below.

lection in Supplementary Fig. S1. The dendrogram produced through clustering has been used to illustrate results from a phage screening experiment [41, 42]. Visualization of a subset of six BASEL phages is presented in Fig. 2, highlighting an additional capability of LoVis4u: its ability to process non-coding features, such as tRNAs.

Genome browser mode

In addition to comparative analysis and visualization of multiple related loci, another common task in microbiology is the analysis of a single genome and the visualization of coverage profiles from sequencing experiments (e.g. DNA-seq or RNA-seq), either for an entire genome or for user-defined windows. To address this, LoVis4u includes a genome browser mode specifically designed for prokaryotic genomes. In addition to visualization of multiple genomic signal tracks from sequencing experiments (e.g. DNA-seq, RNA-seq), it can generate output figures with sequence property tracks, such as GC content and GC skew profiles. LoVis4u also automatically applies moving-average smoothing to signal tracks when needed, taking into account the image width and the size of the visualized window.

To demonstrate this mode, we applied LoVis4u to Bas01 *Escherichia* phage Auguste Piccard from the BASEL collection [40], using a genomic signal track obtained from an example DNA-seq experiment (data distributed with the package sample data) (Fig. 3).

Discussion

Comparative genomic analysis in microbiological studies often requires specialized visualization tools for various tasks. Here, we present LoVis4u, a software tool designed to produce publication-quality figures of genomic loci with several automated analysis steps integrated into the pipeline. LoVis4u offers a much-needed compromise between tools with advanced R/Python APIs and those with user-friendly graphical interfaces and interactivity. It can be used for fast and automated generation of multiple figures in bioinformatics pipelines or other libraries, or can be used as a stand-alone tool for data exploration. By sensitively finding and annotating conserved and variable regions, LoVis4u can facilitate comparative evolutionary analyses of genomes or genomic regions, and the discovery of new biology. Finally, the LoVis4u genome browser mode provides a powerful tool for exploring results from sequencing experiments, such as DNA-seq or RNA-seq, facilitating both analysis and visualization to address different microbiological problems.

Acknowledgements

The tool was tested using resources provided by LUNARC, The Centre for Scientific and Technical Computing at Lund University.

Author contributions: Artyom A. Egorov (Conceptualization [equal], Visualization [lead], Software [lead], Writing [lead], Funding acquisition [supporting]), Gemma C. Atkinson (Conceptualization [equal], Writing [supporting], Funding acquisition [lead])

Supplementary data

Supplementary data is available at NAR Genomics & Bioinformatics online.

Conflict of interest

None declared.

Funding

The work was supported by grants from the Knut and Alice Wallenberg Foundation (number 2020-0037 to G.C.A.), Vetenskapsrådet (the Swedish Research Council; numbers

2023-02353 and 2022-01205 to G.C.A.), eSSENCE@LU (number 10:2 to G.C.A.), and The Royal Physiographic Society of Lund (Endowments for the Natural Sciences, Medicine and Technology, number 45379 to A.A.E.).

Data availability

The Python LoVis4u package is available in PyPI (*python3 -m pip install lovis4u*), and the source code is provided on GitHub (github.com/art-egorov/lovis4u) and Zenodo (https://doi.org/10.5281/zenodo.14800717). Detailed documentation with an installation guide and an example-driven manual are available on the LoVis4u home page (art-egorov.github.io/lovis4u).

References

1. Koonin EV, Makarova KS, Wolf YI. Evolution of microbial genomics: conceptual shifts over a quarter century. *Trends Microbiol* 2021;**29**:582–92. https://doi.org/10.1016/j.tim.2021.01.005
2. Saha CK, Sanches Pires R, Brolin H *et al.* FlaGs and webFlaGs: discovering novel biology through the analysis of gene neighbourhood conservation. *Bioinformatics* 2021;**37**:1312–4. https://doi.org/10.1093/bioinformatics/btaa788
3. Pereira J. GCsnap: interactive snapshots for the comparison of protein-coding genomic contexts. *J Mol Biol* 2021;**433**:166943. https://doi.org/10.1016/j.jmb.2021.166943
4. Andrews SJ, Rothnagel JA. Emerging evidence for functional peptides encoded by short open reading frames. *Nat Rev Genet* 2014;**15**:193–204. https://doi.org/10.1038/nrg3520
5. Egorov AA, Atkinson GC. uORF4u: a tool for annotation of conserved upstream open reading frames. *Bioinformatics* 2023;**39**:btad323. https://doi.org/10.1093/bioinformatics/btad323
6. Yutin N, Tolstoy I, Mutz P *et al.* Jumping DNA polymerases in bacteriophages. bioRxiv, https://doi.org/10.1101/2024.04.26.591309, 27 April 2024, preprint: not peer reviewed.
7. Rousset F, Depardieu F, Miele S *et al.* Phages and their satellites encode hotspots of antiviral systems. *Cell Host Microbe* 2022;**30**:740–53. https://doi.org/10.1016/j.chom.2022.02.018
8. Hochhauser D, Millman A, Sorek R. The defense island repertoire of the *Escherichia coli* pan-genome. *PLoS Genet* 2023;**19**:e1010694. https://doi.org/10.1371/journal.pgen.1010694
9. Carver TJ, Rutherford KM, Berriman M *et al.* ACT: the Artemis Comparison Tool. *Bioinformatics* 2005;**21**:3422–3. https://doi.org/10.1093/bioinformatics/bti553
10. Sullivan MJ, Petty NK, Beatson SA. Easyfig: a genome comparison visualizer. *Bioinformatics* 2011;**27**:1009–10. https://doi.org/10.1093/bioinformatics/btr039
11. Garcia PS, Jauffrit F, Grangeasse C *et al.* GeneSpy, a user-friendly and flexible genomic context visualizer. *Bioinformatics* 2019;**35**:329–31. https://doi.org/10.1093/bioinformatics/bty459
12. Harrison KJ, Crecy-Lagard V, Zallot R. Gene Graphics: a genomic neighborhood data visualization web application. *Bioinformatics* 2018;**34**:1406–8. https://doi.org/10.1093/bioinformatics/btx793
13. Guy L, Kultima JR, Andersson SG. genoPlotR: comparative gene and genome visualization in R. *Bioinformatics* 2010;**26**:2334–5. https://doi.org/10.1093/bioinformatics/btq413
14. Wilkins D. *gggenes: draw Gene Arrow Maps in 'ggplot2'*. R package version 0.5.0. 2023. https://wilkox.org/gggenes/ (21 January 2025, datelast accessed).
15. Pritchard L, White JA, Birch PR *et al.* GenomeDiagram: a python package for the visualization of large-scale genomic data. *Bioinformatics* 2006;**22**:616–7. https://doi.org/10.1093/bioinformatics/btk021
16. Ebmeyer S, Coertze RD, Berglund F *et al.* GEnView: a gene-centric, phylogeny-based comparative genomics pipeline for bacterial genomes and plasmids. *Bioinformatics* 2022;**38**:1727–8. https://doi.org/10.1093/bioinformatics/btab855
17. Gilchrist CLM, Chooi YH. clinker & clustermap.Js: automatic generation of gene cluster comparison figures. *Bioinformatics* 2021;**37**:2473–5.
18. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;**30**:2068–9. https://doi.org/10.1093/bioinformatics/btu153
19. Bouras G, Nepal R, Houtak G *et al.* Pharokka: a fast scalable bacteriophage annotation tool. *Bioinformatics* 2023;**39**:btac776. https://doi.org/10.1093/bioinformatics/btac776
20. Steinegger M, Soding J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol* 2017;**35**:1026–8. https://doi.org/10.1038/nbt.3988
21. Brockhurst MA, Harrison E, Hall JPJ *et al.* The ecology and evolution of pangenomes. *Curr Biol* 2019;**29**:R1094–103. https://doi.org/10.1016/j.cub.2019.08.012
22. Eddy SR. Accelerated profile HMM searches. *PLoS Comput Biol* 2011;**7**:e1002195. https://doi.org/10.1371/journal.pcbi.1002195
23. Larralde M, Zeller G. PyHMMER: a Python library binding to HMMER for efficient sequence analysis. *Bioinformatics* 2023;**39**:btad214. https://doi.org/10.1093/bioinformatics/btad214
24. Tesson F, Planel R, Egorov AA *et al.* A comprehensive resource for exploring antiphage defense: DefenseFinder webservice,wiki and databases. *Peer Community J* 2024;**4**:e91. https://doi.org/10.24072/pcjournal.470
25. Couvin D, Bernheim A, Toffano-Nioche C *et al.* CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res* 2018;**46**:W246–51. https://doi.org/10.1093/nar/gky425
26. Payne LJ, Todeschini TC, Wu Y *et al.* Identification and classification of antiviral defence systems in bacteria and archaea with PADLOC reveals new system types. *Nucleic Acids Res* 2021;**49**:10868–78. https://doi.org/10.1093/nar/gkab883
27. Yan Y, Zheng J, Zhang X *et al.* dbAPIS: a database of anti-prokaryotic immune system genes. *Nucleic Acids Res* 2024;**52**:D419–25. https://doi.org/10.1093/nar/gkad932
28. Liu B, Zheng D, Zhou S *et al.* VFDB 2022: a general classification scheme for bacterial virulence factors. *Nucleic Acids Res* 2022;**50**:D912–7. https://doi.org/10.1093/nar/gkab1107
29. Feldgarden M, Brover V, Gonzalez-Escalona N *et al.* AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. *Sci Rep* 2021;**11**:12728. https://doi.org/10.1038/s41598-021-91456-0
30. Kent WJ, Zweig AS, Barber G *et al.* BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics* 2010;**26**:2204–7. https://doi.org/10.1093/bioinformatics/btq351
31. Cock PJ, Antao T, Chang JT *et al.* Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 2009;**25**:1422–3. https://doi.org/10.1093/bioinformatics/btp163
32. Virtanen P, Gommers R, Oliphant TE *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* 2020;**17**:261–72. https://doi.org/10.1038/s41592-019-0686-2
33. The pandas development team. pandas-dev/pandas: Pandas. Version 2.2.2. Zenodo, 2024. https://doi.org/10.5281/zenodo.3509134
34. Roberts J, Crall J, Ang K-M, Brandt Y. alan-turing-institute/distinctipy: v1.3.4. Zenodo, 2024. https://doi.org/10.5281/zenodo.3985191
35. Hunter JD. Matplotlib: a 2D graphics environment. *Comput Sci Eng* 2007;**9**:90–5.
36. Waskom M. seaborn: statistical data visualization. *J Open Source Softw* 2021;**6**:3021. https://doi.org/10.21105/joss.03021

37. O'Leary NA, Wright MW, Brister JR *et al.* Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* 2016;**44**:D733–45. https://doi.org/10.1093/nar/gkv1189
38. Terzian P, Olo Ndela E, Galiez C *et al.* PHROG: families of prokaryotic virus proteins clustered using remote homology. *NAR Genom Bioinform* 2021;**3**:lqab067. https://doi.org/10.1093/nargab/lqab067
39. Egorov AA, Hauryliuk V, Atkinson GC. Systematic annotation of hyper-variability hotspots in phage genomes and plasmids. bioRxiv, https://doi.org/10.1101/2024.10.15.618418, 16 October 2024, preprint: not peer reviewed.
40. Maffei E, Shaidullina A, Burkolter M *et al.* Systematic exploration of *Escherichia coli* phage–host interactions with the BASEL phage collection. *PLoS Biol* 2021;**19**:e3001424. https://doi.org/10.1371/journal.pbio.3001424
41. Gapinska M, Zajko W, Skowronek K *et al.* Structure-functional characterization of *Lactococcus* AbiA phage defense system. *Nucleic Acids Res* 2024;**52**:4723–38. https://doi.org/10.1093/nar/gkae230
42. Mets T, Kurata T, Ernits K *et al.* Mechanism of phage sensing and restriction by toxin–antitoxin–chaperone systems. *Cell Host Microbe* 2024;**32**:1059–73. https://doi.org/10.1016/j.chom.2024.05.003

Received: November 24, 2024. **Revised:** January 16, 2025. **Editorial Decision:** February 3, 2025. **Accepted:** February 5, 2025

nature chemical biology

Article

https://doi.org/10.1038/s41589-024-01630-4

Mechanisms of neutralization of toxSAS from toxin–antitoxin modules

Received: 17 July 2023

Accepted: 22 April 2024

Published online: 4 June 2024

Check for updates

Lucia Dominguez-Molina[1,5], Tatsuaki Kurata[2,5], Albinas Cepauskas[1,5], Dannele Echemendia-Blanco[1], Safia Zedek[1], Ariel Talavera-Perez[1], Gemma C. Atkinson[2,6] ✉, Vasili Hauryliuk[2,3,4,6] ✉ & Abel Garcia-Pino[1,6] ✉

Toxic small alarmone synthetase (toxSAS) enzymes constitute a family of bacterial effectors present in toxin–antitoxin and secretion systems. toxSASs act through either translation inhibition mediated by pyrophosphorylation of transfer RNA (tRNA) CCA ends or synthesis of the toxic alarmone adenosine pentaphosphate ((pp)pApp) and adenosine triphosphate (ATP) depletion, exemplified by FaRel2 and FaRel, respectively. However, structural bases of toxSAS neutralization are missing. Here we show that the pseudo-Zn^{2+} finger domain (pZFD) of the ATfaRel2 antitoxin precludes access of ATP to the pyrophosphate donor site of the FaRel2 toxin, without affecting recruitment of the tRNA pyrophosphate acceptor. By contrast, (pp)pApp-producing toxSASs are inhibited by Tis1 antitoxin domains though occlusion of the pyrophosphate acceptor-binding site. Consequently, the auxiliary pZFD of AT2faRel is dispensable for FaRel neutralization. Collectively, our study establishes the general principles of toxSAS inhibition by structured antitoxin domains, with the control strategy directly coupled to toxSAS substrate specificity.

Small alarmone synthetases (SASs) are a diverse group of monofunctional RelA/SpoT Homolog (RSH) enzymes that catalyze the transfer of the adenosine triphosphate (ATP)-derived pyrophosphate moiety on the 3′ ribose position of the acceptor substrate[1]. The acceptor substrate is commonly a purine nucleotide, either guanosine triphosphate, diphosphate or monophosphate (GTP, GDP or GMP, yielding guanosine pentaphosphate ((pp)pGpp) alarmone nucleotides)[2–4] or their adenosine equivalents (ATP, ADP or AMP, yielding (pp)pApp)[5,6]. SASs also catalyze pyrophosphate transfer to the adenine moiety of the 3′ CCA end of transfer RNAs (tRNAs), yielding pyrophosphorylated tRNA (tRNA-PP)[7]. As such, the substrate specificity of SASs is broadly determined by their biological function; guanosine-specific SASs are bacterial stress response factors that modulate the intracellular levels of the (pp)pGpp alarmone[3,8,9], while adenine-specific toxic SASs (toxSASs) are potent inhibitors of bacterial growth[5,7,10,11]. The accumulation of toxSAS-produced (pp)pApp abrogates ATP production[5,6], while pyrophosphorylation of tRNA CCA abrogates protein synthesis by rendering the modified tRNA aminoacylation incompetent[7].

Due to their extreme toxicity, the enzymatic activity of toxSASs is tightly controlled. Acting in trans, dedicated immunity proteins and antitoxins sequester toxSASs into inactive complexes[5–7,11]. Acting in cis, the pseudo-Zn^{2+} finger domain (pZFD) autoinhibitory antitoxin element of the monomeric fused single-polypeptide toxin–antitoxin system (TA) $CapRel^{SJ46}$ inhibits its toxic enzymatic domain (toxSYNTH)[10]. The $pZFD^{CapRel}$ was predicted to block access of ATP, the pyrophosphate donor, through a conserved YXXY sequence motif that switches to a 3_{10}-helical conformation when anchored to the toxSYNTH of $CapRel^{SJ46}$ for neutralization[10]. Furthermore, $pZFD^{CapRel}$ also acts a phage sensor domain that mediates the activation of the toxin via the recognition of the phage SECΦ27 major capsid protein Gp57[10]. The proposed

[1]Cellular and Molecular Microbiology, Faculté des Sciences, Université libre de Bruxelles (ULB), Brussels, Belgium. [2]Department of Experimental Medical Science, Lund University, Lund, Sweden. [3]Faculty of Science and Technology, University of Tartu Institute of Technology, Tartu, Estonia. [4]Science for Life Laboratory, Lund, Sweden. [5]These authors contributed equally: Lucia Dominguez-Molina, Tatsuaki Kurata, Albinas Cepauskas. [6]These authors jointly supervised this work and are corresponding authors: Gemma C. Atkinson, Vasili Hauryliuk, Abel Garcia-Pino. ✉e-mail: gemma.atkinson@med.lu.se; vasili.hauryliuk@med.lu.se; abel.garcia.pino@ulb.be

disengagement of pZFD^{CapRel} from the toxSYNTH active site enables CapRel^{SJ46}-mediated tRNA pyrophosphorylation, ultimately resulting in translational shutoff, which quarantines the infected cell[10].

A different regulatory strategy is used to control *Pseudomonas aeruginosa* PAO1 Tas1 (type VI secretion effector (p)ppApp synthetase 1), which contains a divergent toxSAS-related domain[6]. Unlike the tRNA-targeting FaRel2 and CapRel^{SJ46}, Tas1 produces the toxic alarmone (pp)pApp, a potent inhibitor of purine biosynthesis, which depletes the cellular ATP pool[6]. Tas1 is neutralized by the immunity factor Tis1 (type VI secretion immunity to (p)ppApp synthetase 1). The formation of the Tas1:Tis1 heterodimer complex is proposed to block access of the pyrophosphate acceptor, not donor, ATP nucleotide to the toxSYNTH domain of Tas1 (ref. 6). The toxic TA effector *Cellulomonas marina* FaRel is a well-characterized (p)ppApp synthetase[5]. The *faRel* gene is encoded in an operon with a structure that is highly uncommon for typically bicistronic TAs; it is flanked by two antitoxin genes: *aTfaRel*, encoding a small alarmone hydrolase (SAH) enzyme that can promiscuously neutralize the toxic products of toxSASs, and *aT2faRel*, encoding a type II TA-like antitoxin specific to FaRel[5].

The *Coprobacillus* sp. D7 *faRel2–aTfaRel2* operon encodes a type II TA system, with the FaRel2 toxSAS toxin being specifically neutralized by the ATfaRel2 antitoxin[5] (Supplementary Fig. 1a–c). FaRel2 is a protein synthesis inhibitor that pyrophosphorylates uncharged tRNAs, similarly to CapRel^{SJ46} (ref. 7). As the ATfaRel2 antitoxin is a distant homolog of pZFD^{CapRel}, it is also likely to prevent the binding of ATP to the pyrophosphate donor site of the toxSYNTH domain[10]. However, direct structural evidence for the mode of neutralization of either fused (CapRel^{SJ46}) or bipartite (FaRel2:ATfaRel2) translation-targeting toxSAS TAs is lacking[5,7].

Here we uncover how tRNA-modifying toxSAS are neutralized. We determine the crystal structure of the individual FaRel2 toxin and ATfaRel2 antitoxin, as well as the heterotetrameric FaRel2$_2$:ATfaRel2$_2$ complex. Our structural, microbiological and biochemical evidence suggests that bipartite toxSAS TA operons use stable multimerization as an alternative to the colocalization of toxin and antitoxin domains of monomeric fused TAs, such as CapRel^{SJ46}, to ensure efficient toxin inhibition. We directly demonstrate that ATfaRel2 neutralizes FaRel2 by blocking access of ATP to the pyrophosphate donor site of toxSYNTH. Lastly, we propose a unifying conceptual framework that connects the mechanisms of allosteric control and substrate specificity in toxSASs.

Results

ATfaRel2 is a compact, well-structured antitoxin

We determined the crystal structure of ATfaRel2 (residues 1–73) to 1.2-Å resolution (Fig. 1a,b and Supplementary Table 1). The antitoxin protein is monomeric in the crystal (Supplementary Fig. 2a), in good agreement with molecular weight estimates by size-exclusion chromatography (SEC) (Supplementary Fig. 2b and Supplementary Table 2). At the core of its compact, well-folded structure is an antiparallel β-sheet (β-strands β1–β3), with β2 and β3 connected by the central α-helix α1. The C-terminal extension provides an additional β-strand to the β-sheet, β4, which folds parallel to β2, as well as a second α-helix, α2, which forms a part of the protein's hydrophobic core (Fig. 1b). Despite the lack of sequence similarity, ATfaRel2 is structurally similar to pZFD^{CapRel}, with the two proteins superimposing with a root-mean-square deviation (r.m.s.d.) of 2.8 Å, and they both contain conserved tyrosine residues at the C terminus (Fig. 1c).

Interestingly, the predicted topologies of pZFD^{CapRel} in the neutralizing state and ATfaRel2 are similar except for the order of structural elements at the termini. The C-terminal β/α extension of ATfaRel2 is structurally similar to the N-terminal β/α region of pZFD^{CapRel} that connects the antitoxin with toxSYNTH^{CapRel} via Anchor1 (Fig. 1c and Supplementary Fig. 2c–e). This is suggestive of circular permutation (compare Fig. 1b to Fig. 1d). However, as their similarities extend only to the topology level, these two small β/α elements may have evolved independently.

Changes in topology in the antitoxin elements likely reflect the in cis versus in trans neutralization of different toxSASs. Presumably, ATfaRel2 fully dissociates from FaRel2 when the toxin becomes active. Conversely, in the fused CapRels, dissociation of pZFD^{CapRel} is not possible and activation is believed to be mediated by a conformational change[10] (Supplementary Fig. 2d,e). Despite these topological variations, both domains retain strong structural similarities, including the YXXY sequence motif linked to toxSYNTH inhibition in CapRel^{SJ46}. This suggests that ATfaRel2 and pZFD^{CapRel} neutralize toxSYNTH domains by a common mechanism.

The YXXY motif is crucial for toxSYNTH inhibition

In CapRel^{SJ46}, the conserved YXXY motif is located in the switch region of the pZFD and is predicted to lock the enzyme in a catalytically inactive state with both Y residues blocking the donor pyrophosphate ATP-binding site of toxSYNTH (Supplementary Fig. 2c–e). Substitutions directly or indirectly targeting this motif render CapRel^{SJ46} constitutively active[10]. In the crystal structure of the catalytically active state of CapRel^{SJ46}, the YXXY motif assumes an extended conformation[10]. Conversely, in the AlphaFold2-generated neutralized state model, the YXXY motif is predicted to fold into a 3_{10}-helix that locks into the donor site (Supplementary Fig. 2d,e).

The structure of the free ATfaRel2 reveals that the ^{51}YXXY^{54} motif indeed folds into a 3_{10}-helix structure that was predicted to be key for toxSAS neutralization, scaffolded by β-strands β1 and β3 (Fig. 1a and Supplementary Fig. 1b). This suggests that, even in the absence of the toxin, ATfaRel2 is primed for efficient FaRel2 neutralization. Guided by the structural similarities between ATfaRel2 and pZFD^{CapRel}, we probed the role of the individual residues of the ^{51}YXXY^{54} motif in the neutralization of FaRel2 toxicity in vivo (Fig. 1e). While Y54A substitution fully ablates the neutralizing activity of ATfaRel2, I45A and Y51A result in modest but clearly detectable defects in FaRel2 neutralization. Our isothermal titration calorimetry (ITC) assays lend further support to the in vivo neutralization results. To overcome the challenges of producing an otherwise highly toxic FaRel2, we followed a well-established substitution strategy for RSH enzymes[10,12,13] and we used a catalytically impaired Y128F-substituted variant, FaRel2-Y128F. The substitution interferes with the accommodation of the acceptor nucleotide in the active site but is located far from the predicted pZFD^{CapRel}–toxSYNTH interface. We characterized complex formation between FaRel2-Y128F and wild-type (WT) ATfaRel2 or variants carrying substitutions in the ^{51}YXXY^{54} motif (Fig. 1f,g and Supplementary Table 3). V43A and A50M substitutions had a negligible effect on complex stability. These mutant antitoxins bound FaRel2-Y128F with K_D values of 78.8 nM and 38.0 nM, respectively, compared with a K_D value of 35.2 nM in the case of WT ATfaRel2. The destabilizing effect was more pronounced in the case of I45A and Y51A variants. While these substitutions resulted in 25-fold and 43-fold reductions in affinity, respectively, the remaining affinity was still sufficient for partial neutralization in vivo (Fig. 1e). Lastly, the Y54A variant, which was unable to neutralize FaRel2 in vivo, had a 250-fold lower affinity to the toxin than WT ATfaRel2.

FaRel2 binds ATP in a partially folded state

Next, we determined the structure of the catalytically impaired FaRel2-Y128F bound to the non-hydrolyzable ATP analog adenosine-5′-[(α,β)-methyleno]triphosphate (APCPP) at 2.6-Å resolution (Fig. 2a). The toxSYNTH domain of FaRel2 shares the overall topology of other nucleotide pyrophophotransferases, retaining a presumably ancestral fold composed of a central five-stranded β-sheet framed by two α-helices, α3 and α6 (refs. 13,14) (Fig. 2b). The core toxSYNTH domain is very similar to that of the (pp)pApp alarmone synthetase Tas1 (Protein Data Bank (PDB) 6OX6) and the tRNA-pyrophosphokinase CapRel^{SJ46}

https://doi.org/10.1038/s41589-024-01630-4

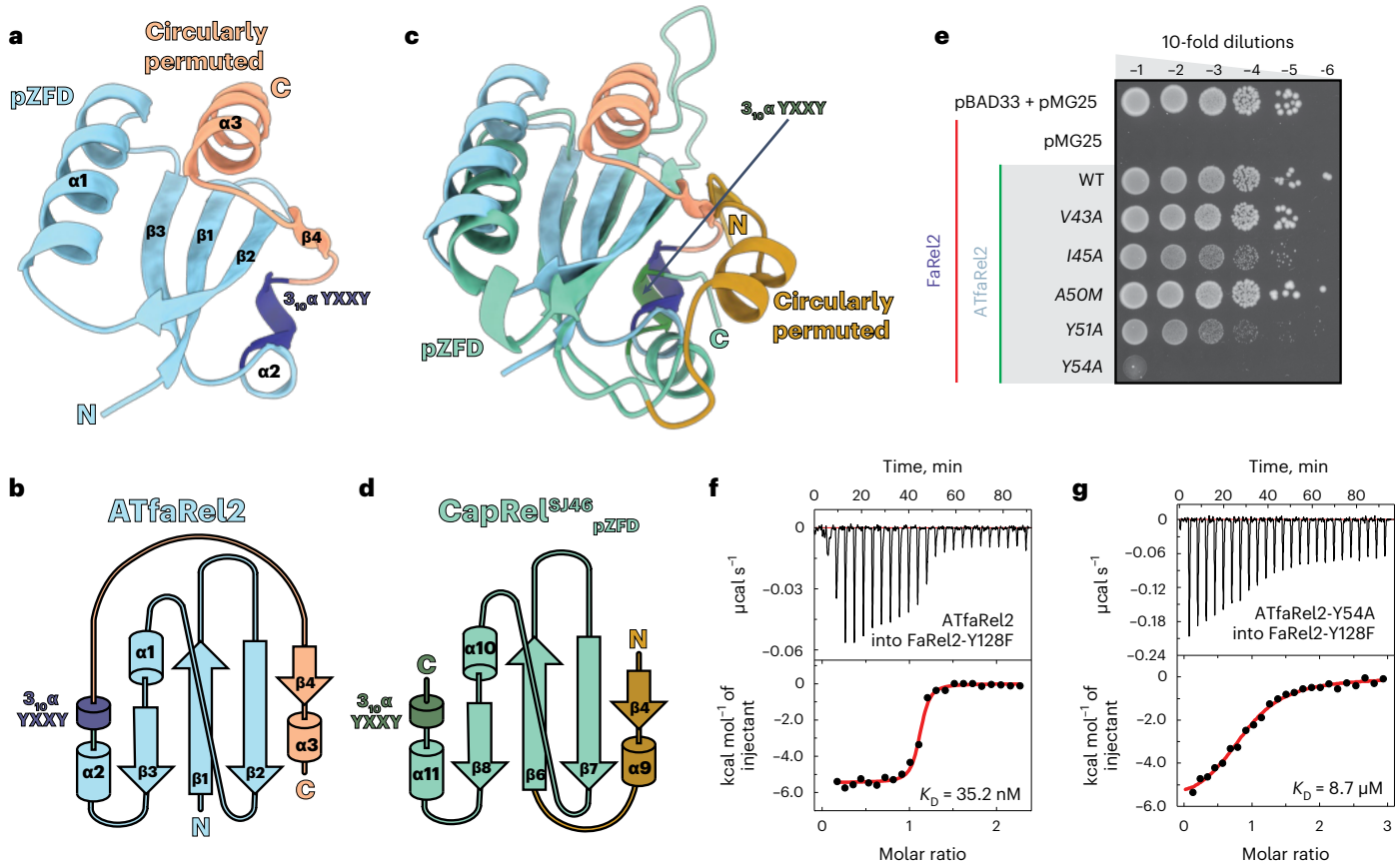


Fig. 1 | Structure of the ATfaRel2 antitoxin. a, Cartoon representation of *Coprobacillus* sp. D7 ATfaRel2 structure. The pZFD core is colored light blue, the C-terminal extension elements are colored salmon and the ^{51}YXXY^{54} motif is highlighted in dark blue. **b**, Topology representation of ATfaRel2, colored as in **a**. **c**, Superposition of the experimental ATfaRel2 structure (colored as in **a**) onto the AlphaFold2-predicted structure of the CapRel^{SJ46} antitoxin domain in the neutralizing state. pZFD^{CapRel} is colored light green and the N-terminal extension that is part of Anchor1 is colored ocher, with the conformational switch region that contains the YXXY neutralization motif colored dark green and labeled in the figure. **d**, Topology representation of the CapRel^{SJ46} antitoxin in the neutralizing state, highlighting the circular permutation of ATfaRel2. **e**, Toxicity neutralization assays probing the ^{51}YXXY^{54} motif of ATfaRel2 and its scaffolding structure from β1. Serial dilutions of *E. coli* strains expressing WT FaRel2 alone or coexpressed with ATfaRel2 (WT or V43A, I45A, A50M, Y51A and Y54A variants) were plated on solid LB medium and scored after 16 h at 37 °C. **f**,**g**, Binding of WT (**f**) and Y54A-substituted (**g**) ATfaRel2 to Y128F-substituted catalytically compromised FaRel2, monitored by ITC.

(PDB 7ZTB), superimposing with r.m.s.d. values of 1.0 Å and 0.8 Å, respectively (Supplementary Fig. 3a,b).

The catalytic core of RSH enzymes is typically decorated by regulatory elements that exert allosteric control on the enzymatic domains[12,15]. The FaRel2 toxSYNTH domain has a well-resolved N-terminal α-helical extension comprising helices α1 and α2, with residues K28 and R29 from the α2–α3 loop being crucial for tRNA binding[7] (Supplementary Fig. 1c). While the N terminus is unresolved (Fig. 2c) in the non-toxic (p)ppGpp alarmone synthetases RelP[15,16] and RelQ[17], the N terminus folds back toward the toxSYNTH core in FaRel2 and intercalates α1 and α2 between α3 and α5, providing an anchor point for tRNAs near the active site G-loop (Fig. 2a,c). The C-terminal region of SYNTH and toxSYNTH domains of pyrophosphotransferases have varied architectures composed predominantly of small α-helical domains[6,10,13,14,18,19]; the C-terminal bundle of four α-helices of SASs RelQ and RelP, which synthesizes (p)ppGpp, acts as an oligomerization interface[15,17], toxSYNTH of the monomeric Tas1 is followed by a small α-helical domain[6] and the SYNTH domain of long RHSs is followed by a flexible core domain with a high α-helical propensity[18]. In the case of FaRel2 complexed with APCPP, the C terminus is disordered and not visible in the electron density (Fig. 2d).

Despite these differences, the FaRel2-bound APCPP superimposes remarkably well with the APCPP bound in the donor site of other (p)ppGpp-synthesizing RSHs SAS RelP and RelQ, as well as the long RSH Rel[13,15,17] (Fig. 2d). As in the other alarmone synthetases, the adenosine group stacks with the conserved R64 and R95 from β1 and β2, with β5 E145 hydrogen bonding the adenosine NH_2 group (Supplementary Fig. 1c). The R64 residue has a crucial role in providing van der Waals contacts to accommodate the ribose while directly coordinating the 5′ α and β phosphates together with K66. The strongly basic α4 that follows β1 further stabilizes the triphosphates through S70, K74 and R77 (Fig. 2a). At the pyrophosphate acceptor side of the active site, the conformation of the G-loop and the orientation of the base-coordinating F128 (Y128 in the WT FaRel2) deviate from what was observed in precatalytic and postcatalytic complexes of (p)ppGpp synthetases[15,17] (Supplementary Fig. 3c,d). It is tempting to speculate that, while the ground state of FaRel2 is primed to bind ATP, efficient tRNA binding would likely involve further conformational rearrangements, such as folding of the C terminus and alignment of the active site.

ATfaRel2 binds FaRel2 by β-sheet extension

To uncover the mechanism of toxin neutralization, we determined the structure of the ATfaRel2:FaRel2 complex. The structure reveals a ATfaRel2$_2$:FaRel2$_2$ heterotetrametric arrangement (Fig. 3a–c). The 2:2 stoichiometry was confirmed in solution by SEC (62 kDa versus the

https://doi.org/10.1038/s41589-024-01630-4

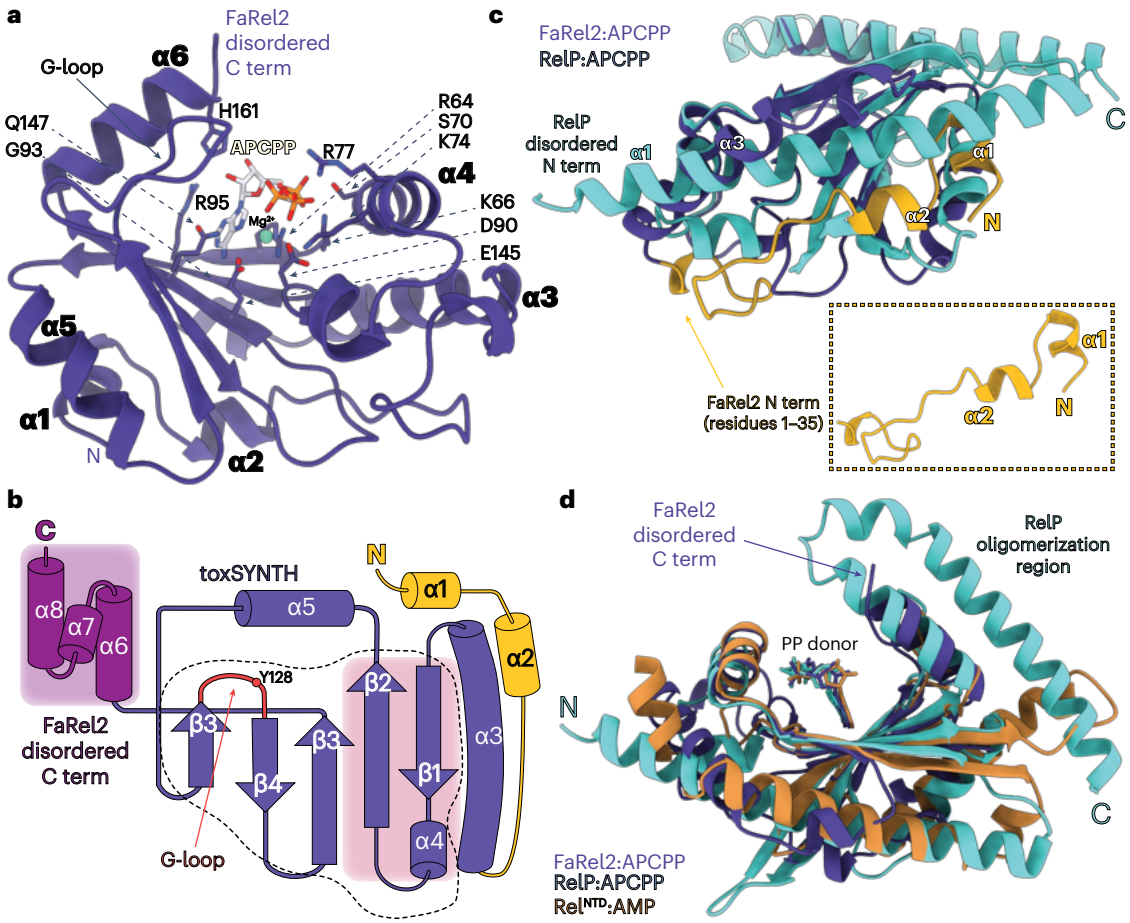


Fig. 2 | Structure of FaRel2 in complex with APCPP. a, The structure of *Coprobacillus* sp. D7 FaRel2 in complex with APCPP bound to the pyrophosphate donor site. The adenine base is coordinated by conserved residues R64, R95 and Q147, as well as residue G93, while E145 provides the specificity for adenosine nucleotides. R64 provides additional van der Waals contacts with the ribose moiety. The 5′ triphosphate moiety is stabilized by R64, K66, S70, K74 and R77, while the Mg^{2+} ion is held in place by the catalytic D90. **b**, Topology diagram of FaRel2, highlighting the N-terminal region involved in tRNA recognition (yellow), the toxSYNTH domain (dark blue, with the catalytic region outlined by a dashed line) and the disordered C-terminal region (dark magenta). The acceptor site is highlighted by the G-loop (β3–β4 loop) in red and the donor site involving β1, β2 and α3 is shaded in pink. **c**, Superposition of the FaRel2:APCPP complex on the structure of *S. aureus* RelP in complex with APCPP (PDB 6EWZ; turquoise). The ordered N terminus of FaRel2 that is lacking in RelP is highlighted in yellow. **d**, Superposition of the complexes of FaRel2:APCPP (dark blue), RelP:APCPP (turquoise) and *Thermus thermophilus* Rel^{NTD}:AMP (dark orange; PDB 6S2U) illustrates the conserved mode of pyrophosphage donor coordination in RSH pyrophosphotransferases.

theoretical 68 kDa; Fig. 3d) and was consistent with the ITC measurements (Fig. 1f).

The C-terminal region of FaRel2, disordered in the FaRel2:APCPP complex, folds into an α-helical subdomain upon complex formation and provides a dimerization interface mediated only by toxin–toxin interactions. By contrast, in the bound state, the conformation of ATfaRel2 matches that of the free antitoxin, with both structures superimposing with an r.m.s.d. of 0.4 Å (Supplementary Fig. 3e). The primary interface between ATfaRel2 and FaRel2 has an area of 1,190.0 $Å^2$, with ATfaRel2 sterically blocking access of the ATP substrate to the pyrophosphate donor binding site of the toxSYNTH domain (Fig. 3b). Through this large interface, ATfaRel2 contacts several key functional regions of FaRel2: (1) the long N-terminal α-helix α3 (a structural element that is often involved in allosteric crosstalk in many RSH enzymes and interacts with α1 from ATfaRel2); (2) the basic α-helix α4 (involved in the stabilization of the ATP triphosphate group, which is coordinated by ATfaRel2 through β3 and the β3–α1 loop); (3) the central β-sheet (through β1, β2 and β5, which harbor the catalytic center and adenine coordinating residues); and (4) the C-terminal cap of the α-helix α6 (which is disordered in the FaRel2:APCPP complex and folds into an α-helical dimerization region when bound to ATfaRel2). The only notable exception was the predicted tRNA recognition site that remains solvent accessible (Supplementary Fig. 3f).

The center of the neutralization interface is formed by the antiparallel β-strand interaction between FaRel2 β1 and ATfaRel2 β3 that connects the β-sheets of both proteins, extending the core of the complex. Residues I42–T46 of β3 form multiple van der Waals contacts with FaRel2 β1, further stabilizing the complex. Residues V43 and I45 serve as a scaffold, orienting the YXXY 3_{10}-helix that anchors ATfaRel2 (Fig. 3b). As predicted for $pZFD^{CapRel}$ (ref. 10), this hydrophobic tether projects Y51 and Y54 into the ATP-binding site through a *π*-stacking arrangement with R64 and R95, which precludes adenine coordination to the pyrophosphate donor site of FaRel2. These results suggest that, while this mechanism of neutralization is likely the same at the structural level to that proposed for $CapRel^{SJ46}$, the energetics of neutralization are certainly different, with the dynamic association of $pZFD^{CapRel}$ regulating $toxSYNTH^{CapRel}$ in cis, contrasting with the stable in trans neutralization in the ATfaRel2:FaRel2 complex. These differences could have important implications for triggering these systems.

https://doi.org/10.1038/s41589-024-01630-4

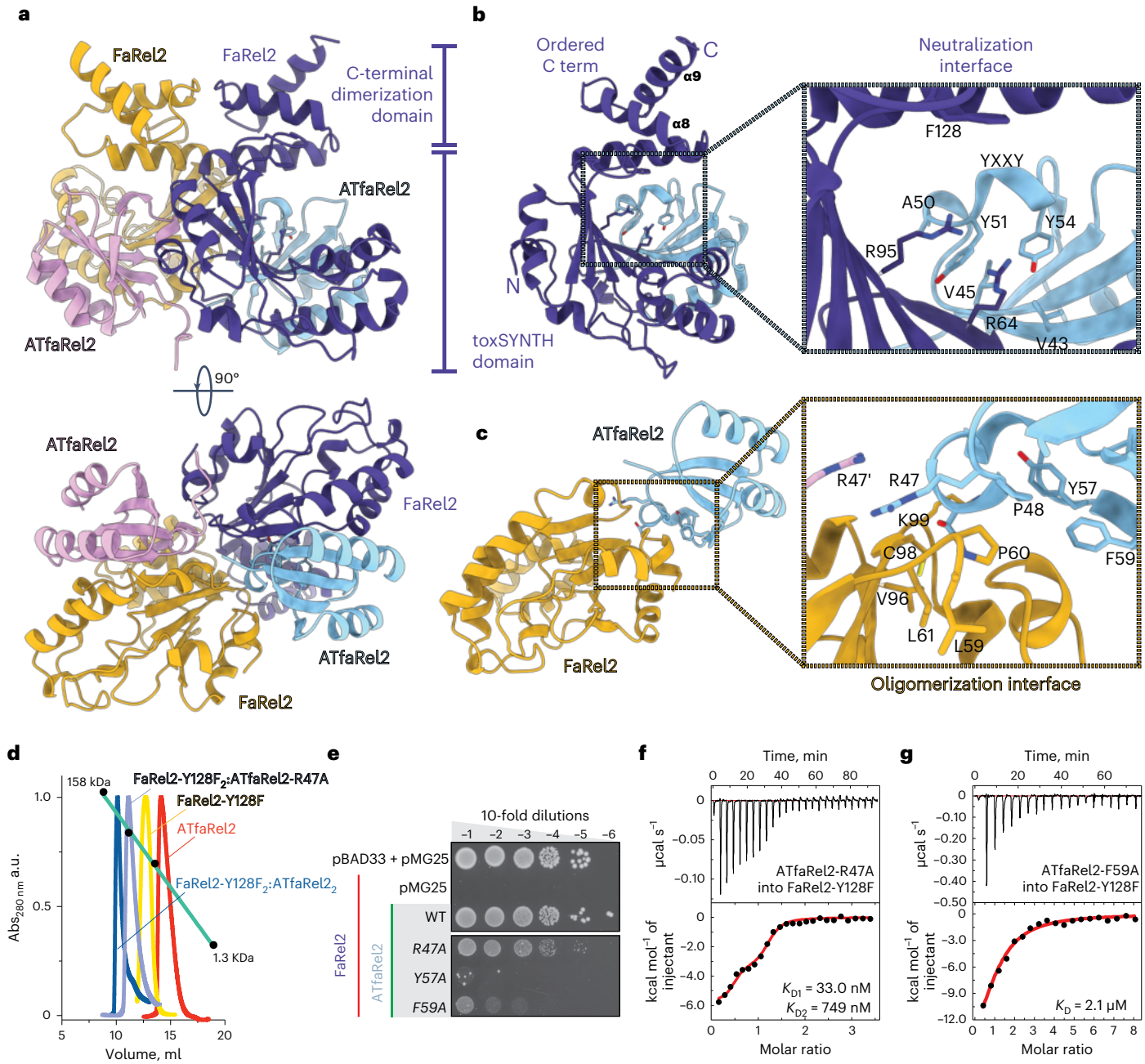


Fig. 3 | Structure of the ATfaRel2:FaRel2 complex. a, Structure of the heterotetrametic ATfaRel2$_2$:FaRel2$_2$ complex. Antitoxin units are colored light blue and pink, while the toxin units are colored dark blue and gold. The C-terminal region of FaRel2, disordered in the FaRel2:APCPP structure, in the TA complex folds into a small α-helical subdomain. **b**, Stabilization of the ATfaRel2:FaRel2 heterodimer by the primary complex interface. The ATP-binding catalytic sites have key structural roles in the interface. **c**, The secondary interface ATfaRel2:FaRel2 is formed at the two-fold symmetry axis of the complex. The interface comprises (1) cross-coordination of R47 from each antitoxin unit; (2) formation of an alternative hydrophobic interface between ATfaRel2 and FaRel2 (that is, different from the YXXY motif); and (3) interactions between the C-terminal α-helical regions of the two FaRel2 units. **d**, Analytical SEC of ATfaRel2$_2$:FaRel2-Y128F$_2$ (dark-blue trace), ATfaRel2-R47A$_2$:FaRel2-Y128F$_2$ (lilac), FaRel2-Y128F (yellow) and ATfaRel2 (red). **e**, Probing of the secondary FaRel2:ATfaRel2 interface through toxicity neutralization assays. Serial dilutions of *E. coli* strains expressing FaRel2 alone or coexpressed with ATfaRel2 (WT or R47A, Y57A and Y59A variants) were plated on solid LB medium and scored after 16 h at 37 °C. **f**,**g**, Binding of R47A-substituted (**f**) and F59A-substituted (**g**) ATfaRel2 to Y128F-substituted FaRel2, monitored by ITC.

Dimerization enhances toxin neutralization by ATfaRel2

Compared with pZFD^{CapRel}, ATfaRel2 is considerably more tolerant to substitutions in the toxin-binding interface (compare Fig. 1e with Extended Data Fig. 3j in ref. 10). The structure of the ATfaRel2$_2$:FaRel2$_2$ complex reveals that ATfaRel2 engages the neighboring FaRel2 in the heterotetramer through a secondary interface that is half the size of the primary one (~550.0 Å^2 versus 1,190.0 Å^2), thus effectively crosslinking the complex (Supplementary Fig. 4a–c). We hypothesized that the stable oligomeric nature of ATfaRel2$_2$:FaRel2$_2$ compensates for the lack of TA colocalization enforced in the monomeric CapRel^{SJ46} through fusion of the toxin and antitoxin domains into one polypeptide. On this basis, substitutions disrupting the ATfaRel2$_2$:FaRel2$_2$ oligomerization interface (and not affecting the primary neutralization interface at the active site) would compromise the TA recognition.

To test this hypothesis, we subjected the secondary interface to single-residue substitutions and assessed the complex stability in vivo through toxicity neutralization assays (Fig. 3e). The three targeted residues R47A, Y57A and F59A of ATfaRel2 are all distant from the main contact interface that blocks access of ATP to the active site (Fig. 3c). R47 is located on the two-fold symmetry axis of the complex and the side chains of R47 from each ATfaRel2 interlock through π–π interactions. Y57 and F59 are part of a small hydrophobic core that defines the secondary oligomerization interface. The R47A substitution resulted in a modest defect, whereas Y57A and F59A compromised the neutralization severely (Fig. 3d).

The direct interrogation of these interactions by ITC was in good agreement with the in vivo data. The R47A substitution efficiently perturbs the secondary interface and decouples the highly cooperative tetramer formation observed in the WT protein (Fig. 3f and Supplementary Table 3). The first high-affinity binding event (K_D = 35 nM with a stoichiometry of 0.4) followed a lower-affinity recognition event (K_D = 750 nM with a stoichiometry of 0.8). This likely represents the initial neutralization of FaRel2 (with a 2:1 TA ratio) followed by the formation of a less stable 2:2 tetramer. The impact of F59A on the affinity was even stronger, with a 60-fold decrease in affinity and a confirmed 1:1 binding molar ratio, indicating an interaction mediated by only the primary interface (Fig. 3g and Supplementary Table 3). These results were consistent with SEC experiments that revealed a decrease in size of the TA complex from the estimated 70.8 kDa of the WT (consistent with a 2:2 A:T stoichiometry) to 51.3 kDa suggestive of a ATfaRel2-R47A:FaRel2-Y128F$_2$ complex (1:2 A:T stoichiometry) (Fig. 3d and Supplementary Table 2). The observation of a stable ATfaRel2-R47A:FaRel2-Y128F$_2$ complex in SEC matched the high-affinity interaction observed by ITC with ATfaRel2-R47A (Supplementary Table 2). It is, thus, likely that the C-terminal region of FaRel2 that folds upon TA complex formation and provides a large FaRel2:FaRel2 interface in the complex is still capable of partially stabilizing the oligomer against the effect of mild substitutions such as R47A but not against F59A, which had a major effect on complex formation.

FaRel2 C terminus stabilizes the heterotetrameric complex

Upon formation of the ATfaRel2$_2$:FaRel2$_2$ complex, the C-terminal regions of the two individual FaRel2 toxin polypeptides fold into a dimerization region with four α-helices that contributes 670 Å^2 to the TA interface (Supplementary Fig. 4d). This folding upon binding interaction likely has an important role in the overall stability of the heterotetramer. Guided by the structure of FaRel2 bound to APCPP, we constructed a truncated version of FaRel2 (FaRel2-Δ166–206) lacking the C-terminal disordered part. FaRel2-Δ166–206 interacted with ATfaRel2 with a K_D of 3.3 μM, as measured by ITC (Supplementary Fig. 4e). This ~90-fold drop in affinity of FaRel2-Δ166–206 for ATfaRel2 underscores the strong contribution of oligomerization to the overall energetics of complex formation. In the case of the WT TA complex, the binding was both entropically and enthalpically driven. The entropic penalty from the folding of the FaRel2 C terminus was likely compensated for by the configurational entropy associated with the large hydrophobic surface buried upon binding (Supplementary Table 3). Together with the strong enthalpic component that accompanied the oligomerization through the C-terminal α-helical region, this resulted in very stable heterotetramerization. Because the FaRel2-Δ166–206 truncation removes the enthalpic contribution of the C terminus, binding to ATfaRel2 was, as expected, predominantly entropically driven, resulting in a less stable complex (Supplementary Table 3). Collectively, these results suggest that, while the main TA interface drives toxin neutralization, the oligomerization further stabilizes the interaction between ATfaRel2 and the toxSYNTH domain of FaRel2. Additional contacts of the main interface YXXY motif with the folded C-terminal α-helical FaRel2:FaRel2 interface link the oligomerization with toxin neutralization, hinting at a potential allosteric path for activation and toxin release.

tRNA-pyrophosphorylating toxins specifically bind tRNA

Given the low concentrations toxSAS toxins are typically found in the cell (below the detection levels of current techniques)[20,21], tRNA-phosphorylating activity would likely depend on a strong and specific association with tRNAs. We used ITC to examine the tRNA binding capacity of a representative toxSAS and housekeeping SASs: tRNA-phosphorylating *Coprobacillus* sp. D7 FaRel2 and *Mycobacterium* phage Phrann PhRel, (pp)pApp-synthesizing *C. marina* FaRel and (p)ppGpp-synthesizing *Staphylococcus aureus* SAS RelQ. FaRel2 and PhRel bound deacylated initiator tRNA$_i^{fMet}$ with similar affinities (K_D values of 483 nM and 825 nM, respectively; Fig. 4a,b and Supplementary Table 3). (pp)pApp synthetase FaRel had a 37-fold lower affinity to tRNA$_i^{fMet}$ (K_D value of 17.8 μM). No tRNA$_i^{fMet}$ binding was observed for (p)ppGpp-producing *Enterococcus faecalis* SAS RelQ (Fig. 4c,d and Supplementary Table 3). Lastly, *E. faecalis* RelQ was shown to interact with a short single-stranded model mRNA(MF) coding for MF dipeptide[22]. Our ITC experiments demonstrated that *S. aureus* RelQ similarly bound mRNA(MF) with submicromolar affinity (K_D value of 922 nM) (Fig. 4e and Supplementary Table 3).

ATfaRel2 interferes with APCPP binding but not tRNA recognition

While association with ATfaRel2 decreased the affinity (K_D) to APCPP ~20-fold, from 2.1 to 41.3 μM (Fig. 4f,g and Supplementary Table 3), the affinity to deacylated tRNA$_i^{fMet}$ was virtually the same for the free toxin and inactive TA complex (K_D value of 483 nM versus 460 nM) (Fig. 4h and Supplementary Table 3). In the cases of both free monomeric FaRel2 and the heterotetrametric ATfaRel2:FaRel2 complex, the tRNA binding had 1:1 stoichiometry with respect to FaRel2.

To further investigate the effect of AtfaRel2 binding on the interaction of FaRel2 with ATP, we determined the structure of the ATfaRel2:FaRel2 complex bound to APCPP (Fig. 4i). The high protein concentration intrinsic of the crystal lattice combined with high nucleotide concentration used for soaking facilitated the binding of APCPP to a partially blocked active site. As predicted from the structure of ATfaRel2:FaRel2, the coordination sites for the adenine, ribose and α phosphate groups at the pyrophosphate donor site are blocked by ATfaRel2. Thus, APCPP is bound in the pyrophosphate acceptor site in a conformation incompatible with pyrophosphate transfer (Fig. 4j). The adenine base is coordinated by Y128 and R95, resembling the expected coordination of the terminal adenine of the CCA tRNA moiety. The β and γ phosphates further anchor the nucleotide; however, they are observed in a reversed orientation compared with the FaRel2:APCPP complex (Fig. 4j–k).

It is instructive to compare the local charge distributions in the active sites of tRNA-modifying FaRel2 to those of (pp)pGpp and (p)ppApp alarmone synthetases Rel[13], RelP[15] and Tas1 (ref. 6). All alarmone synthetases have a large positive patch that accommodates diphosphate and triphosphate nucleotide substrates, located on the acceptor site close to the conserved Y residue that interacts with the acceptor base (Supplementary Fig. 4f–h). This positive patch is considerably smaller in FaRel2 (Supplementary Fig. 4i), which explains the misorientation of the β and γ phosphates of APCPP bound in the acceptor site of FaRel2 and the lack of alarmone synthetase activity of the tRNA-targeting toxSAS.

Collectively, our results demonstrate that FaRel2 is neutralized by the ATfaRel2 antitoxin by compromising the accommodation of ATP in the toxSYNTH active site without affecting the interaction with uncharged tRNAs. This suggests that the ATfaRel2$_2$:FaRel2$_2$:tRNA$_2$ ternary complex could be preformed in the cell, with the toxSAS neutralized in the complex until activation is triggered.

https://doi.org/10.1038/s41589-024-01630-4

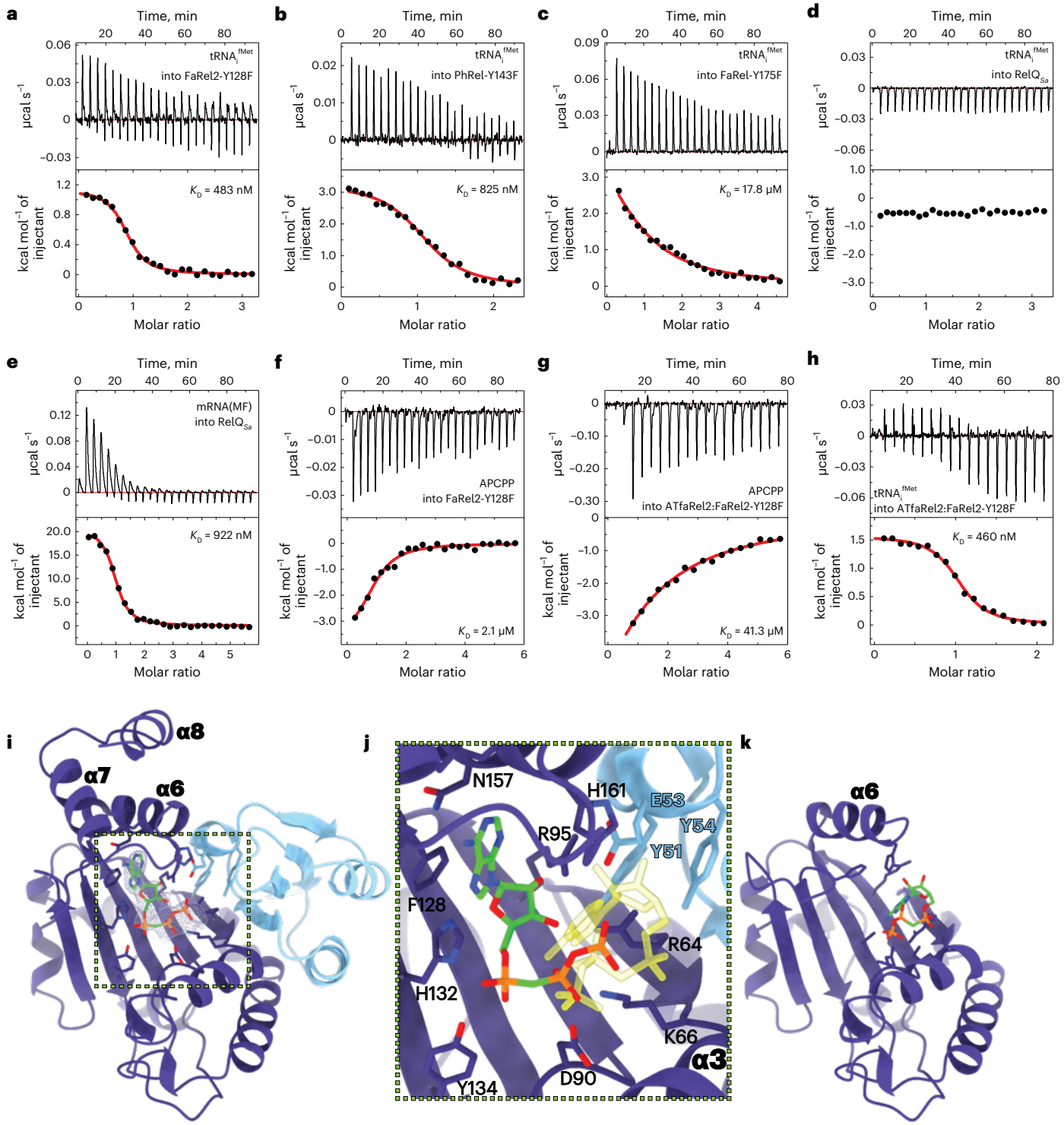


Fig. 4 | Energetic and structural basis of substrate recognition by toxSASs and non-toxSASs. a–**d**, Binding of FaRel2-Y128F (**a**), PhRel2-Y143F (**b**), FaRel-Y175F (**c**) and WT RelQ$_{Sa}$ (**d**) to deacylated initiator tRNA$_i^{fMet}$, monitored by ITC. **e**, Binding of mRNA(MF) to WT RelQ$_{Sa}$, monitored by ITC. **f**,**g**, Binding of APCPP to FaRel2-Y128F (**f**) and ATfaRel2$_2$:FaRel2-Y128F$_2$ complex (**g**), monitored by ITC. **h**, Binding of tRNA$_i^{fMet}$ to the ATfaRel2$_2$:FaRel2-Y128F$_2$ complex, monitored by ITC. **i**, Structure of ATfaRel2$_2$:FaRel2-Y128F$_2$ bound to APCPP (green). The unbiased mFo-DFc electron density map corresponding to the bound APCPP is shown in gray. **j**, Details of the coordination of APCPP (green) in the acceptor site when bound to ATfaRel2$_2$:FaRel2-Y128F$_2$. The adenosine base is coordinated by the G-loop F128 and the 5′ β and γ phosphates extend to the basic patch of α4. There, they bind in a reverse orientation compared with APCPP in the FaRel2:APCPP complex, underscoring that the nucleotide is bound in a state incompatible with phosphate transfer. For comparison, the APCPP in the orientation observed in the complex with FaRel2 is shown in light yellow. Active site residues of FaRel2 are labeled in black and the residues from the YXXY motif of ATfaRel2 are labeled in light blue. **k**, FaRel2-Y128F:APCPP complex with APCPP placed in the donor site in a catalytically compatible orientation, presented in the same pose as **i**. In the absence of ATfaRel2, α7 and α8 are not visible in the electron density.

https://doi.org/10.1038/s41589-024-01630-4

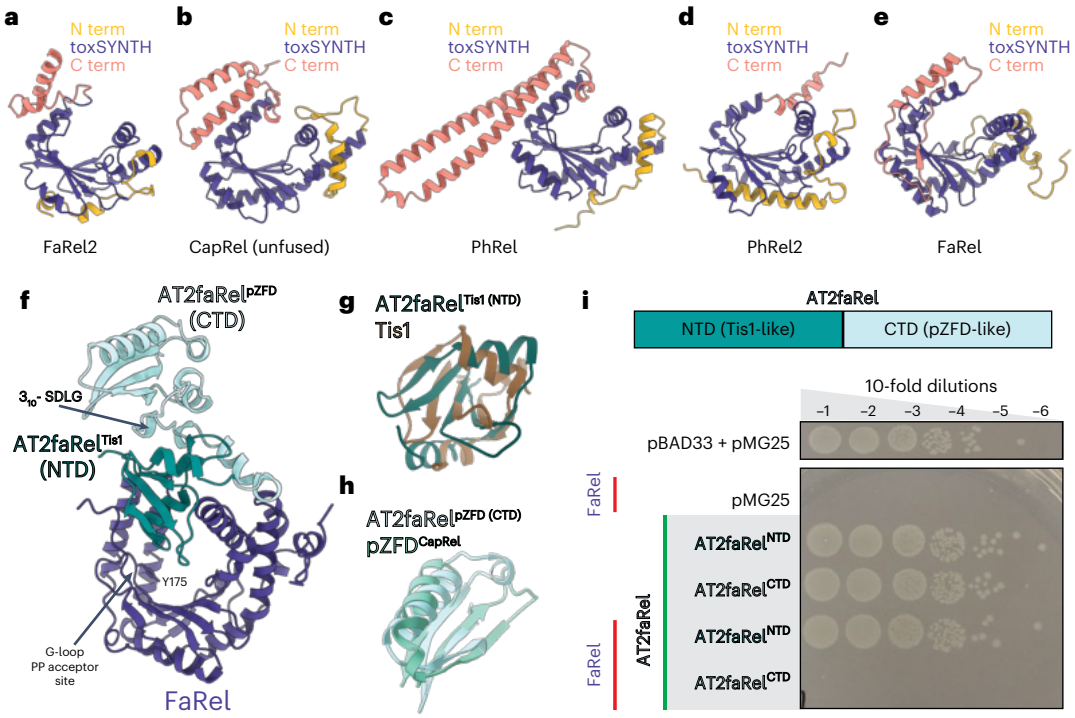


Fig. 5 | Neutralization of (pp)pApp-producing FaRel toxSAS. a–f, AlphaFold2-generated structural models of *Coprobacillus* sp. D7 FaRel2 (**a**), *M. tuberculosis* CapRel (unfused) (**b**), *M. tuberculosis* PhRel (**c**), bacteriophage Lily PhRel2 (**d**) and *C. marina* FaRel (**e**) and the AT2faRel:FaRel complex (**f**). The Tis1-like NTD of AT2faRel is colored dark green and the pZFD/ATfaRel2-like CTD is colored light cyan. AlphaFold2 predicts that FaRel is neutralized by the Tis1-like domain through the pyrophosphate acceptor site. **g**,**h**, Structural superposition of the Tis1-like NTD of AT2faRel (dark green) on Tis1 (PDB 6OX6), colored brown (**g**) and the CTD (light cyan) on pZFD^{CapRel} (PDB 7ZTB), colored light teal (**h**). **i**, Probing of the secondary FaRel2:ATfaRel2 interface through toxicity neutralization assays. Serial dilutions of *E. coli* strains expressing either AT2faRel^{NTD} or AT2faRel^{CTD} alone or together with FaRel were plated on solid LB medium and scored after 16 h at 37 °C.

toxSAS neutralization is defined by catalytic activity

Prompted by the conceptual differences between Tis1-mediated neutralization of Tas1 and ATfaRel2-mediated neutralization of FaRel2, we next used AlphaFold2[23] to explore the general principles underlying the mechanisms of toxin neutralization across known toxSAS functional diversity (Fig. 5 and Supplementary Fig. 5a–l). On the toxSAS toxin side, AlphaFold2 predicts a strong conservation of a core toxSYNTH fold decorated with a variety of insertions at the N and C termini (Fig. 5a–e). On the antitoxin side, the pZFD fold is found as either a standalone neutralizing domain or part of multidomain antitoxins combined with either Tis1 or Tis1-like (Fig. 5f–h) or PanA domains[11,24] (Supplementary Fig. 5a–j).

Structural predictions of the different neutralized complexes uncovered a general trend. Translation-targeting tRNA-pyrophosphorylating toxSASs such as fused and split CapRel, FaRel2, PhRel and PhRel2 are inhibited through the pyrophosphate donor site (Supplementary Fig. 5a–j). Conversely, metabolism-targeting (pp)pApp-producing toxSASs such as FaRel and Tas1 are neutralized through the pyrophosphate acceptor site (Fig. 5f and Supplementary Fig. 5k,l). The generality of this observation holds even in cases of multidomain antitoxins. In the case of the translation-targeting PhRel2 of bacteriophage Lily and *Bacillus subtilis* Ia1a. These toxins are neutralized by multidomain antitoxins that contain a pZFD fold subtype PAD1 (panacea-associated domain 1)[11]. However, ATphRel2 antitoxins neutralize the toxins analogously to the pZFD-mediated neutralization of CapRel and FaRel2 (Supplementary Fig. 5f,g)[24]. By contrast, the (pp)pApp-producing FaRel is inhibited by the Tis1-like domain of AT2faRel in a manner analogous to Tis1-mediated inhibition of Tas1 (Fig. 5f and Supplementary Fig. 5k,l).

We validated the structural predictions of Alphafold2 through mutagenesis and toxicity neutralization assays. In the case of PhRel2, as we showed previously[24], the isolated PAD1 domain of *B. subtilis* Ia1a ATphRel2 is sufficient to neutralize the toxin. In the case of FaRel:AT2faRel, the Tis1-like N-terminal domain (NTD) of AT2faRel (Fig. 5g) is sufficient to neutralize the (pp)pApp synthetase FaRel, while the pZFD C-terminal domain (CTD; Fig. 5h) has no neutralizing activity (Fig. 5i). Interestingly, this loss of neutralizing activity by the CTD of AT2faRel is accompanied by the loss of the YXXY recognition motif of the pZFD (Fig. 5f). Collectively, our results suggest a coupling between the substrate specificity of toxSASs and the mechanism of neutralization.

Discussion

Our current mechanistic understanding of toxSYNTH inhibition by immunity proteins and antitoxins is largely based on two landmark studies. The first study involves the structure of the neutralized complex between a monomeric (p)ppApp-producing toxSAS Tas1 and its immunity protein Tis1 (ref. 6). Ahmad and colleagues proposed that the enzymatic activity of Tas1 was suppressed by Tis1 through distortion of the acceptor nucleotide-binding site of toxSYNTH. The second study involves the structure of the tRNA-pyrophosphorylating toxSAS CapRel^{SJ46} (ref. 10), a fused monocistronic TA, with the antitoxin part comprising two anchor regions and a pZFD. In this structure, CapRel^{SJ46} appears in a catalytically competent state (that is not autoinhibited by the pZFD), as the antitoxin domain was in a conformation compatible with the enzymatic activity of toxSYNTH. Further exploration of the structural dynamics of CapRel^{SJ46} with AlphaFold2 (ref. 23) predicted a possible mechanism of autoinhibition mediated by the pZFD blocking the donor nucleotide-binding site of toxSYNTH (Fig. 6a,b).

Our study provides detailed mechanistic understanding of toxSAS regulation and allows generalizing the observations previously

https://doi.org/10.1038/s41589-024-01630-4

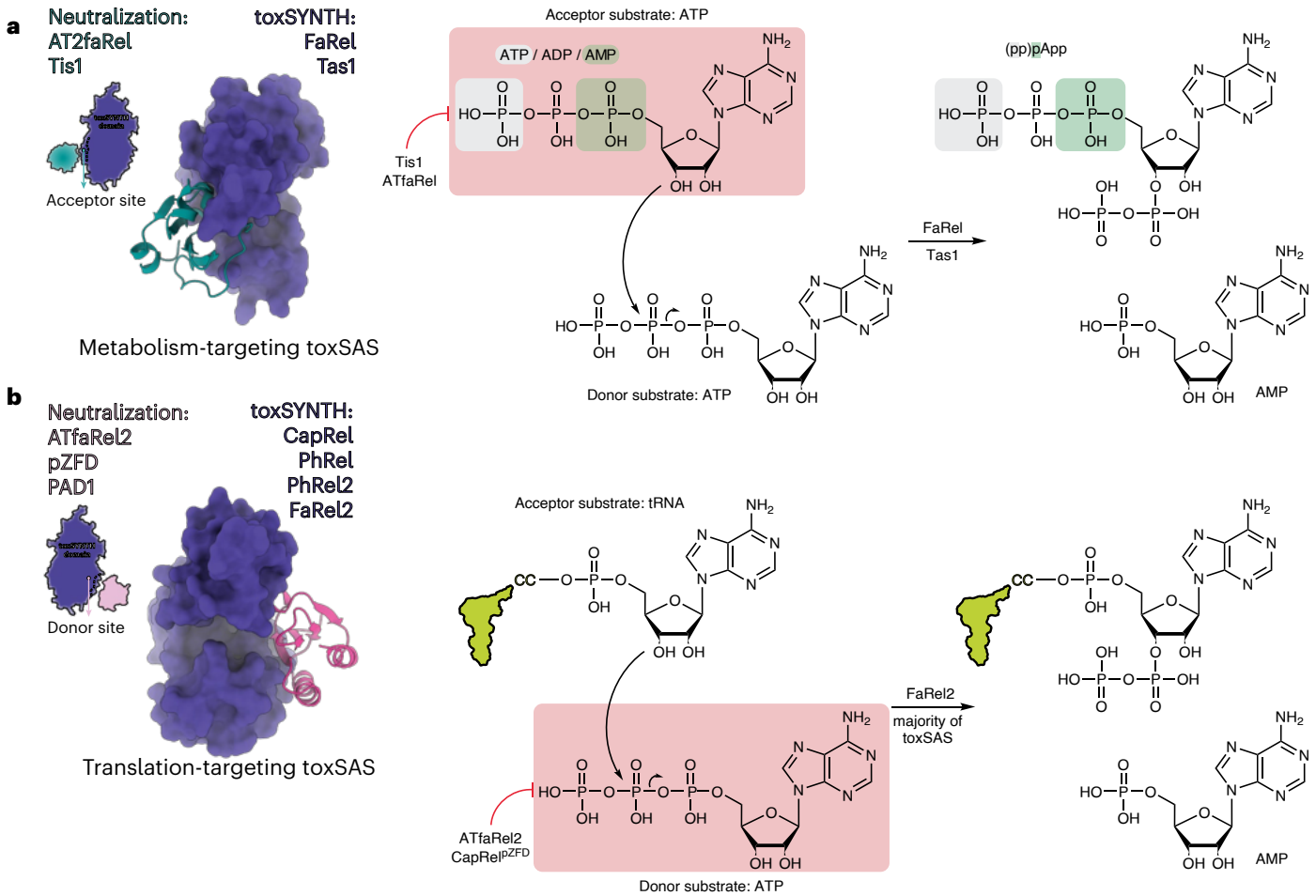


Fig. 6 | The substrate specificity of toxSASs is coupled to neutralization mechanism. Metabolism-targeting and translation-targeting toxSAS are inhibited by different strategies. **a**, Enzymatic activity of metabolism-targeting toxSASs is suppressed by inhibition of the recruitment of the pyrophosphate acceptor substrate. **b**, Translation-targeting toxSASs are neutralized by inhibition of the recruitment of the pyrophosphate donor, while binding of the pyrophosphate acceptor substrate, tRNA, is permitted. This strategy is analogous to that used for the regulation of long (p)ppGpp-producing RSHs Rel, RelA and SpoT; while the affinity to the GTP or GDP substrate nucleotide is constitutive, binding of the ATP nucleotide to the pyrophosphate donor site is under strict allosteric control.

made with CapRel^{SJ46} and Tas1:Tis1 to the whole toxSAS superfamily. We put forward a model that couples toxSAS substrate specificity to the inhibition strategies used by antitoxins and immunity proteins (Fig. 6a,b). Type II TA modules are notoriously evolutionarily promiscuous, with members of the same toxin family being neutralized by unrelated antitoxins through different mechanisms[24–26]. toxSAS TAs display a strikingly clear-cut dichotomy of neutralization mechanisms. Metabolism-targeting (pp)pApp synthetases such as Tas1 and FaRel are inhibited by Tis1-like antitoxins through the occlusion of the pyrophosphate acceptor nucleotide-binding site[6] (Fig. 6a). Conversely, translation-targeting tRNA pyrophosphotransferases such as FaRel2 and CapRel are inhibited by pZFD interfering with the binding site for pyrophosphate donor ATP (Fig. 6b), which echoes the regulatory strategy used in the autoinhibition of multidomain 'long' (p)ppGpp-synthesizing housekeeping RSHs[12,27]. While the antitoxins and immunity proteins mediating the two neutralization strategies have diverged in sequence, the neutralizing elements display strong structural conservation.

Type II TA antitoxins often rely on unstructured elements and disordered domains that fold upon binding for toxin neutralization[25,28–34]. In many cases, the structural plasticity of the disordered region of antitoxins allows them to couple toxin neutralization with transcriptional autoregulation[29,35,36] to balance the cellular T:A ratio. Interestingly, this feature of TA regulation seems so prevalent that, in the exception provided by the GraTA system (from the RelBE superfamily), which contains a well-folded and globular antitoxin, the disordered region leaps to the toxin GraT retaining a role as a crucial regulatory element[37]. The structurally defined lock-and-key neutralization specificity of toxSASs is, thus, uncommon and may underlie their 'switch' nature. In this sense, it is instructive to compare the structural–energetic interplay of the fused CapRel^{SJ46} TA with the bipartite ATfaRel2:FaRel2 TA. The forced colocalization of the toxin and antitoxin elements of CapRel^{SJ46} is compatible with a conformationally dynamic enzyme and offsets the entropic penalty associated with the antitoxin assuming a compact and structured neutralizing state (Supplementary Fig. 6a). These intrinsic dynamics facilitate the formation of the CapRel^{SJ46}:Gp57 complex that triggers the enzyme. By contrast, in the bipartite ATfaRel2:FaRel2 TA system, the free antitoxin naturally assumes the optimal conformation for neutralization. The entropic penalty associated with FaRel2 folding is compensated by the formation of the heterotetramer and a tight TA complex. Thus, the loss of colocalization is compensated by the stabilizing effect of oligomerization. Our findings suggest that offsetting this oligomeric structure could be the key to triggering these bipartite systems (Supplementary Fig. 6b).

ZFDs perform various molecular recognition functions[38] facilitated by a marked structural plasticity[39]. Structurally related pZFDs

mediate both phage recognition and toxin autoinhibition in fused tRNA-pyrophosphorylating TA CapRel^{SJ46} (ref. 10). PAD1 has the same fold as pZFD and directly mediates the neutralization of *B. subtilis* Ia1a and *Clostridium hylemonae* the tRNA-pyrophosphorylating PhRel2 toxSAS[11,24]. Conversely, the pZFD-like CTD of AT2faRel is not essential for the inhibition of the (pp)pApp synthetase FaRel, which is neutralized by its Tis1-like NTD (Fig. 5c,b). This suggests that the pZFD-like domain of AT2faRel performs a different sensory function. This evolutionary dynamic is reminiscent of PanA[11] and HigA[37,40] antitoxins. While, in the case of single-domain antitoxins, PanA directly mediates toxin neutralization, in multidomain antitoxins, PanA domains are not the direct neutralization element; instead, these domains were hypothesized to sense the TA-activating triggers[24]. Establishing the putative sensory functions of pZFDs in translation-targeting and metabolism-targeting toxSASs is one of the remaining challenges in the field.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41589-024-01630-4.

References

1. Atkinson, G. C., Tenson, T. & Hauryliuk, V. The RelA/SpoT homolog (RSH) superfamily: distribution and functional evolution of ppGpp synthetases and hydrolases across the tree of life. *PLoS ONE* **6**, e23479 (2011).
2. Lemos, J. A., Lin, V. K., Nascimento, M. M., Abranches, J. & Burne, R. A. Three gene products govern (p)ppGpp production by *Streptococcus mutans*. *Mol. Microbiol.* **65**, 1568–1581 (2007).
3. Nanamiya, H. et al. Identification and functional analysis of novel (p)ppGpp synthetase genes in *Bacillus subtilis*. *Mol. Microbiol.* **67**, 291–304 (2008).
4. Srivatsan, A. et al. High-precision, whole-genome sequencing of laboratory strains facilitates genetic studies. *PLoS Genet.* **4**, e1000139 (2008).
5. Jimmy, S. et al. A widespread toxin–antitoxin system exploiting growth control via alarmone signaling. *Proc. Natl Acad. Sci. USA* **117**, 10500–10510 (2020).
6. Ahmad, S. et al. An interbacterial toxin inhibits target cell growth by synthesizing (p)ppApp. *Nature* **575**, 674–678 (2019).
7. Kurata, T. et al. RelA–SpoT homolog toxins pyrophosphorylate the CCA end of tRNA to inhibit protein synthesis. *Mol. Cell* **81**, 3160–3170 (2021).
8. Geiger, T., Kastle, B., Gratani, F. L., Goerke, C. & Wolz, C. Two small (p)ppGpp synthases in *Staphylococcus aureus* mediate tolerance against cell envelope stress conditions. *J. Bacteriol.* **196**, 894–902 (2014).
9. Takada, H. et al. Ribosome association primes the stringent factor Rel for tRNA-dependent locking in the A-site and activation of (p)ppGpp synthesis. *Nucleic Acids Res.* **49**, 444–457 (2021).
10. Zhang, T. et al. Direct activation of a bacterial innate immune system by a viral capsid protein. *Nature* **612**, 132–140 (2022).
11. Kurata, T. et al. A hyperpromiscuous antitoxin protein domain for the neutralization of diverse toxin domains. *Proc. Natl Acad. Sci. USA* **119**, e2102212119 (2022).
12. Roghanian, M. et al. (p)ppGpp controls stringent factors by exploiting antagonistic allosteric coupling between catalytic domains. *Mol. Cell* **81**, 3310–3322 (2021).
13. Tamman, H. et al. A nucleotide-switch mechanism mediates opposing catalytic activities of Rel enzymes. *Nat. Chem. Biol.* **16**, 834–840 (2020).
14. Hogg, T., Mechold, U., Malke, H., Cashel, M. & Hilgenfeld, R. Conformational antagonism between opposing active sites in a bifunctional RelA/SpoT homolog modulates (p)ppGpp metabolism during the stringent response [corrected]. *Cell* **117**, 57–68 (2004).
15. Manav, M. C. et al. Structural basis for (p)ppGpp synthesis by the *Staphylococcus aureus* small alarmone synthetase RelP. *J. Biol. Chem.* **293**, 3254–3264 (2018).
16. Steinchen, W. et al. Structural and mechanistic divergence of the small (p)ppGpp synthetases RelP and RelQ. *Sci. Rep.* **8**, 2195 (2018).
17. Steinchen, W. et al. Catalytic mechanism and allosteric regulation of an oligomeric (p)ppGpp synthetase by an alarmone. *Proc. Natl Acad. Sci. USA* **112**, 13348–13353 (2015).
18. Tamman, H. et al. Structure of SpoT reveals evolutionary tuning of catalysis via conformational constraint. *Nat. Chem. Biol.* **19**, 334–345 (2022).
19. Pausch, P. et al. Structural basis for regulation of the opposing (p)ppGpp synthetase and hydrolase within the stringent response orchestrator Rel. *Cell Rep.* **32**, 108157 (2020).
20. LeRoux, M., Culviner, P. H., Liu, Y. J., Littlehale, M. L. & Laub, M. T. Stress can induce transcription of toxin–antitoxin systems without activating toxin. *Mol. Cell* **79**, 280–292 (2020).
21. Overgaard, M., Borch, J., Jorgensen, M. G. & Gerdes, K. Messenger RNA interferase RelE controls *relBE* transcription by conditional cooperativity. *Mol. Microbiol.* **69**, 841–857 (2008).
22. Beljantseva, J. et al. Negative allosteric regulation of *Enterococcus faecalis* small alarmone synthetase RelQ by single-stranded RNA. *Proc. Natl Acad. Sci. USA* **114**, 3726–3731 (2017).
23. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
24. & Ernits, K. et al. The structural basis of hyperpromiscuity in a core combinatorial network of type II toxin–antitoxin and related phage defense systems. *Proc. Natl Acad. Sci. USA* **120**, e2305393120 (2023).
25. Garcia-Pino, A., Zenkin, N. & Loris, R. The many faces of Fic: structural and functional aspects of Fic enzymes. *Trends Biochem. Sci.* **39**, 121–129 (2014).
26. Loris, R. & Garcia-Pino, A. Disorder- and dynamics-based regulatory mechanisms in toxin–antitoxin modules. *Chem. Rev.* **114**, 6933–6947 (2014).
27. Ainelo, A. et al. The structure of DarB in complex with Rel(NTD) reveals nonribosomal activation of Rel stringent factors. *Sci. Adv.* **9**, eade4077 (2023).
28. Dalton, K. M. & Crosson, S. A conserved mode of protein recognition and binding in a ParD–ParE toxin–antitoxin complex. *Biochemistry* **49**, 2205–2215 (2010).
29. De Jonge, N. et al. Rejuvenation of CcdB-poisoned gyrase by an intrinsically disordered protein domain. *Mol. Cell* **35**, 154–163 (2009).
30. Engel, P. et al. Adenylylation control by intra- or intermolecular active-site obstruction in Fic proteins. *Nature* **482**, 107–110 (2012).
31. Garcia-Pino, A. et al. Allostery and intrinsic disorder mediate transcription regulation by conditional cooperativity. *Cell* **142**, 101–111 (2010).
32. Garcia-Pino, A. et al. Doc of prophage P1 is inhibited by its antitoxin partner Phd through fold complementation. *J. Biol. Chem.* **283**, 30821–30827 (2008).
33. Kamada, K., Hanaoka, F. & Burley, S. K. Crystal structure of the MazE/MazF complex: molecular bases of antidote–toxin recognition. *Mol. Cell* **11**, 875–884 (2003).
34. Sterckx, Y. G. et al. A unique hetero-hexadecameric architecture displayed by the *Escherichia coli* O157 PaaA2–ParE2 antitoxin–toxin complex. *J. Mol. Biol.* **428**, 1589–1603 (2016).

35. Garcia-Pino, A. et al. An intrinsically disordered entropic switch determines allostery in Phd–Doc regulation. *Nat. Chem. Biol.* **12**, 490–496 (2016).
36. Jurenas, D., Van Melderen, L. & Garcia-Pino, A. Mechanism of regulation and neutralization of the AtaR–AtaT toxin–antitoxin system. *Nat. Chem. Biol.* **15**, 285–294 (2019).
37. Talavera, A. et al. A dual role in regulation and toxicity for the disordered N-terminus of the toxin GraT. *Nat. Commun.* **10**, 972 (2019).
38. Laity, J. H., Lee, B. M. & Wright, P. E. Zinc finger proteins: new insights into structural and functional diversity. *Curr. Opin. Struct. Biol.* **11**, 39–46 (2001).
39. Azarkan, M., Martinez-Rodriguez, S., Buts, L., Baeyens-Volant, D. & Garcia-Pino, A. The plasticity of the β-trefoil fold constitutes an evolutionary platform for protease inhibition. *J. Biol. Chem.* **286**, 43726–43734 (2011).
40. Hadzi, S. et al. Ribosome-dependent *Vibrio cholerae* mRNAse HigB2 is regulated by a β-strand sliding mechanism. *Nucleic Acids Res.* **45**, 4972–4983 (2017).

Methods

Plasmid construction

Fragments of WT *aTfaRel2* and its substituted variants (V43A, I45A, A50M, Y51A and Y54A) were PCR-amplified with primers VTK198 and VTK199 and templates VHp278 (WT), VHp1225 (V43A), VHp1226 (I45A), VHp1227 (A50M), VHp1236 (Y51A) or VHp1228 (Y54A). Using Gibson assembly, the resulting linear DNA fragment was inserted into linearized pMG25 using pMG HiFi For and pMG HiFi Rev primers.

Sequence analysis

Representative FaRel2 and cognate ATFaRel2 sequences were retrieved using webFlaGs[41], implementing the protein basic local alignment search tool (BLASTp) in the National Center for Biotechnology Information (NCBI) RefSeq Select database and otherwise default settings. Sequences were aligned using MAFFT version 7.490 with the L-INS-i strategy[42].

Toxicity neutralization assays

The experiments were performed as described previously[7]. The assays were performed on Luria–Bertani (LB) medium plates (BD). We used the *Escherichia coli* BW25113 strain cotransformed with two different plasmid systems for controllable expression of toxins and antitoxins. We used a pair of compatible plasmids: pMG25 for antitoxin expression (high copy number, ColE1 origin of replication (pUC), Amp^R, antitoxin expressed under the control of IPTG-inducible $P_{A1/04/03}$ promoter[43]] and pBAD33 for toxin expression (medium copy number, p15A origin of replication, Cml^R, toxins expressed under the control of arabinose-inducible P_{BAD} promoter[44]). The cells were grown in liquid LB medium (BD) supplemented with 0.2% glucose (repression conditions), 100 μg ml^{-1} ampicillin (AppliChem) and 20 μg ml^{-1} chloramphenicol (AppliChem). Serial dilutions were spotted on solid LB plates supplemented with 0.2% arabinose, as well as 100 μg ml^{-1} ampicillin (AppliChem) and 20 μg ml^{-1} chloramphenicol (AppliChem), and bacterial growth was scored after 16-h incubation at 37 °C.

Protein purification

ATfaRel2, Farel2-Y128F and the different variants of the proteins were expressed in *E. coli* BL21DE3. The proteins were produced with a His_6 tag at the N terminus, followed by a tobacco etch virus (TEV) protease cleavage site for ATfaRel2 and the different ATfaRel2 variants and a SUMO (small ubiquitin-like modifier) tag for FaRel2-Y128F. Cultures were grown in LB medium supplemented with kanamycin (50 μg ml^{-1}) at 37 °C with aeration. Expression was induced with 0.5 mM IPTG when the cells carrying the plasmid reached an optical density at 600 nm ($OD_{600\,nm}$) of ~0.5–0.8. After induction, the cells were harvested 16 h later by centrifugation and resuspended with buffer (25 mM HEPES pH 7.6, 1 M NaCl, 5 mM $MgCl_2$ and 1 mM TCEP) supplied with cOmplete protease inhibitor cocktail (Roche). The resuspended cells were flash-frozen in liquid nitrogen and stored at −80 °C.

The cell extracts were lysed using an Emulsiflex cell disruptor and the lysate was centrifuged to remove cell debris for 45 min at 25,000*g*. In both cases, the supernatant was loaded onto a 1-ml HiTrap Ni-NTA column (Cytiva) coupled to a fast protein liquid chromatography (FPLC) system (ÄKTA Explorer) equilibrated with buffer A (25 mM HEPES pH 7.6, 1 M NaCl, 5 mM $MgCl_2$, 1 mM TCEP and 20 mM imidazole). The column was washed with a linear gradient of buffer B (25 mM HEPES pH 7.6, 1 M NaCl, 5 mM $MgCl_2$, 1 mM TCEP and 500 mM imidazole). After tag removal, all individual proteins were further purified by SEC in a Superdex 75 Increase 10/30 (Cytiva) column using 25 mM HEPES pH 7.6, 300 mM NaCl, 2 mM $MgCl_2$ and 1 mM TCEP. Sample purity was confirmed by SDS–PAGE. The ATfaRel2:FaRel2-Y128F complex was obtained by mixing both proteins in a 1:1.2 molar ratio with an excess of antitoxin that was separated by SEC.

Analytical SEC

For the analytical SEC, 150 μl of each protein at a concentration of 1 mg ml^{-1} was loaded on a Superdex 300 Increase 1030 column (Cytiva) previously equilibrated in SEC buffer (25 mM HEPES pH 7.6, 200 mM NaCl, 2 mM $MgCl_2$ and 1 mM TCEP). The progress of the chromatography was monitored by the OD_{280} value.

Crystallization

Before crystallization, FaRel2-Y128F, ATfaRel2 and the ATfaRel2 variants were purified from the TEV cleavage reaction in the SEC buffer and concentrated to 8–10 mg ml^{-1}. Screening of the crystallization conditions was carried out using the sitting-drop vapor diffusion method. The drops were set up in Swiss (MRC) 96-well two-drop UVP sitting-drop plates using the Mosquito HTS system (TTP Labtech). Then, 0.1-μl drops of protein and precipitant solution were equilibrated to 80 μl of precipitant solution in the reservoir. Commercially available screens LMB and SG1 (Molecular Dimensions) were used to test the crystallization conditions. The conditions resulting in diffracting crystals are listed in Supplementary Table 1. The crystals used for data collection the 0.2-μl drops used in the screens were scaled to 2-μl drops.

Structure determination

All the data were processed with the XDS suite[45] and scaled with Aimless[46]. In all cases, the unit cell content was estimated with the program MATTHEW COEF from the CCP4 program suite[47]. The crystals of ATfaRel2 diffracted on average to ~1.3 Å. The analysis of the crystal anisotropy by the STARANISO server (http://staraniso.globalphasing.org/) proposed a resolution of 1.24 Å (with 1.24 Å in a^*, 1.33 Å in b^* and 1.43 Å in c^*). We used Arcimboldo_Lite[48] to perform ab initio phasing and solved the structure of ATfaRel2 in combination with Phaser[49] and SHELXE[50,51]. The solution contained 75 of the 99 residues; the final structure was completed by manual building using Coot[52] and refined with Buster/TNT[53] (R/R_{free} = 18.6/20.8).

In the case of the crystals of the FaRel2:APCPP complex, the analysis of the diffraction data suggested a resolution of 2.62 Å (with diffraction limits of 2.53 Å in a^*, 2.67 Å in b^* and 2.97 Å in c^*) based on which we selected 2.62 Å as the resolution cut-ff. We used the coordinates of $CapRel^{SJ46}$ (PDB 7ZTB) as the search model for the toxSYNTH domain of FaRel2 in complex with APCPP. The molecular replacement (MR) solution from Phaser[49] was used in combination with Rosetta as implemented in the MR-Rosetta suite from the Phenix package[54]. After several iterations of manual building with Coot[52] and maximum likelihood refinement as implemented in Buster/TNT[53], the model was extended to cover all the residues (R/R_{free} = 19.4/25.5).

The crystals of the ATfaRel2:FaRel2 complex were obtained in the $P2_12_12_1$ space group. The anisotropic analysis of the diffraction data suggested a resolution of 2.14 Å (with diffraction limits of 2.25 Å in a^*, 3.00 Å in b^* and 2.13 Å in c*). We used the refined coordinates of FaRel2 (PDB 8PU4, this work) and ATfaRel2 (PDB 8PU2, this work) as the search model for phasing and estimated the unit cell content with MATTHEW COEF from the CCP4 program suite[47]. The MR solution from Phaser[49] was completed by manual building with Coot[52] and refined with Buster/TNT[53] (R/R_{free} = 18.4/23.8). To obtain the structure of the ATfaRel2:FaRel2:APCPP complex, the $P2_12_12_1$ space group did not tolerate soaking; therefore, we grew crystals in a different condition ($F4_132$ space group). The coordinates of FaRel2 (PDB 8PU4, this work) and ATfaRel2 (PDB 8PU2, this work) were used for MR with Phaser[49] and modeling was completed by manual building with Coot[52] and refinement with Buster/TNT[53] (R/R_{free} = 21.8/23.2). Supplementary Table 1 details all the X-ray data collection and refinement statistics.

ITC

All titrations were performed with an Affinity ITC (TA instruments) at 25 °C. For antitoxin versus toxin titrations, ATfaRel2 and its substituted variants were loaded in the instrument syringe at 200–150 μM

and FaRel2-Y128F was used in the cell at 15–20 µM. In the case of the titrations of tRNA and mRNA versus toxSASs, 150 µM tRNA or mRNA was titrated into 15 µM FaRel2-Y128F, PhRel-Y143F, FaRel-Y175F, RelQ or the ATfaRel2:FaRel2-Y128F complex. For the titrations with nucleotides, APCPP was loaded in the instrument syringe at 180 µM and FaRel2-Y128F or the ATfaRel2:FaRel2-Y128F complex was used in the cell at 15 µM. All titrations were performed in 25 mM HEPES pH 7.6, 300 mM NaCl, 2 mM $MgCl_2$ and 1 mM TCEP. Final concentrations were verified by the OD_{280} value using a Nanodrop One (Thermo Fisher Scientific). All ITC measurements were performed by titrating a constant volume of 2 µl into the ITC cell using a constant stirring rate of 75 rpm. All data were processed, buffer-corrected and analyzed using the NanoAnalyse and Origin software packages. Supplementary Table 3 details all the thermodynamic parameters derived from the ITC titrations.

Reporting summary

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

Data availability

The data generated in this study are provided in the Supplementary Information provided with this paper. All coordinates were deposited in the PDB under accession numbers 8PU1, 8PU2, 8PU3 and 8PU4. Data are also available from the corresponding authors upon request.

References

41. Saha, C. K., Sanches Pires, R., Brolin, H., Delannoy, M. & Atkinson, G. C. FlaGs and webFlaGs: discovering novel biology through the analysis of gene neighbourhood conservation. *Bioinformatics* **37**, 1312–1314 (2021).
42. Katoh, K. & Standley, D. M. MAFFT: iterative refinement and additional methods. *Methods Mol. Biol.* **1079**, 131–146 (2014).
43. Jaskólska, M. & Gerdes, K. CRP-dependent positive autoregulation and proteolytic degradation regulate competence activator Sxy of *Escherichia coli*. *Mol. Microbiol.* **95**, 833–845 (2015).
44. Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.* **177**, 4121–4130 (1995).
45. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
46. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1204–1214 (2013).
47. Collaborative Computational Project, Number 4.The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760–763 (1994).
48. Sammito, M. et al. ARCIMBOLDO_LITE: single-workstation implementation and use. *Acta Crystallogr. D Biol. Crystallogr.* **71**, 1921–1930 (2015).
49. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
50. Thorn, A. Experimental phasing: substructure solution and density modification as implemented in SHELX. *Methods Mol. Biol.* **1607**, 357–376 (2017).
51. Uson, I. & Sheldrick, G. M. An introduction to experimental phasing of macromolecules illustrated by SHELX; new autotracing features. *Acta Crystallogr. D Struct. Biol.* **74**, 106–116 (2018).
52. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
53. Smart, O. S. et al. Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 368–380 (2012).
54. Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D Struct. Biol.* **75**, 861–877 (2019).

Acknowledgements

This work was supported by the Fonds National de Recherche Scientifique (FNRS; CDR J.0068.19 and J.0065.23F, EQP UN.025.19 and PDR T.0066.18 and T.0090.22 to A.G.P.), European Research Council (ERC; CoG DiStRes, no. 864311 to A.G.P.), Joint Programming Initiative on Antimicrobial Resistance (JPIAMR; JPI-EC-AMR-R.8004.18 to A.G.P.), Fonds Jean Brachet and the Fondation Van Buuren (A.G.P.). This work was supported by the Swedish Research Council (Vetenskapsrådet; grants 2019-01085 and 2022-01603 to G.C.A. and 2021-01146 to V.H.), Crafoord Foundation (project grant no. 20220562 to V.H.), Estonian Research Council (PRG335 to V.H.) and Cancerfonden (20 0872 Pj to V.H.). G.C.A. and V.H. were also supported by a project grant from the Knut and Alice Wallenberg Foundation (2020-0037 to G.C.A.). A.C. was supported by the Fund for Research in Industry and Agronomy (FRIA; FC31211 to A.C.). We acknowledge the use of beamtimes PROXIMA 1 and 2A at the Soleil synchrotron (Gif-sur-Yvette, France) and I24 at the Diamond Light Source (United Kingdom).

Author contributions

A.G.P. and V.H. drafted the paper with contributions from all authors and coordinated the study. V.H., A.G.P., G.C.A., A.T.P. and T.K. designed the experiments and analyzed the data. L.D.M., A.C., A.T.P. and T.K. performed the biochemical and microbiological experiments. D.E.B., S.Z. and A.C. performed the mutagenesis of FaRel2, ATfaRel2, FaRel and AT2faRel. L.D.M., A.C. and A.T.P. performed the biophysical measurements. L.D.M. crystallized the proteins. A.T.P. and A.G.P. determined the structures of FaRel2, ATfaRel2 and FaRel2:ATfaRel2. A.G.P. and G.C.A. performed the bioinformatic analyses. All authors approved the final revision of the paper.

Funding

Open access funding provided by Lund University.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41589-024-01630-4.

Correspondence and requests for materials should be addressed to Gemma C. Atkinson, Vasili Hauryliuk or Abel Garcia-Pino.

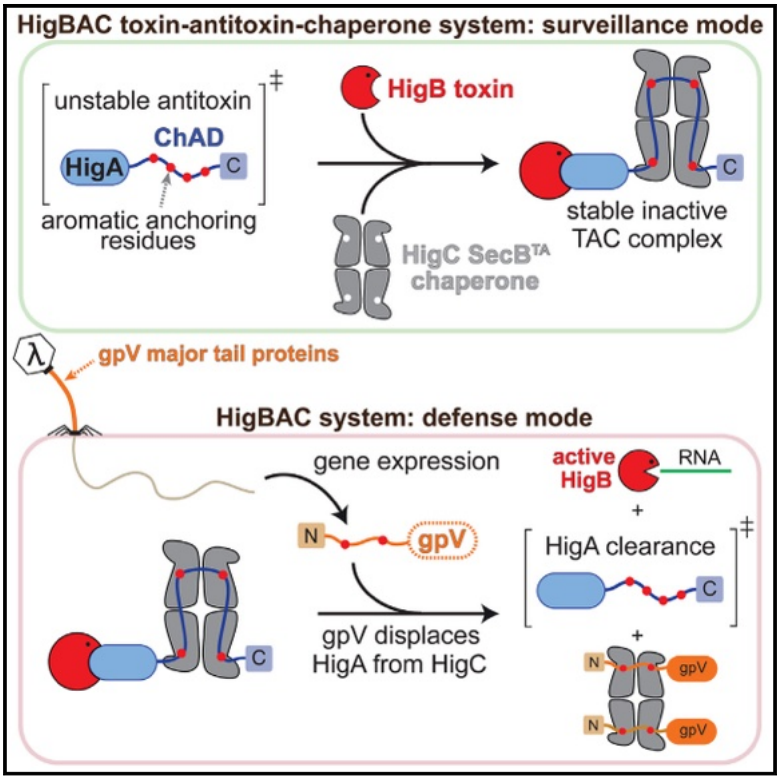
Peer review information *Nature Chemical Biology* thanks Kevin Forsberg, Mahavir Singh and the other, anonymous reviewer(s) for their contribution to the peer review of this work.

Article

Cell Host & Microbe

Mechanism of phage sensing and restriction by toxin-antitoxin-chaperone systems

Graphical abstract



Highlights

- *E. coli* HigBAC and CmdTAC are translation-targeting phage immunity TAC systems
- The HigC chaperone senses the phage λ major tail protein to trigger the HigB toxin
- The CmdT ADP-ribosyltransferase toxin inhibits translation by modifying mRNA
- HigC combined with CmdTA yields a hybrid broad-spectrum antiphage defense system

Authors

Toomas Mets, Tatsuaki Kurata, Karin Ernits, ..., Abel Garcia-Pino, Gemma C. Atkinson, Vasili Hauryliuk

Correspondence

abel.garcia.pino@ulb.be (A.G.-P.), gemma.atkinson@med.lu.se (G.C.A.), vasili.hauryliuk@med.lu.se (V.H.)

In brief

Toxin-antitoxin-chaperone (TAC) systems have been shown to mediate antiphage defense via as-yet unexplored mechanisms. Focusing on two TACs encoded by *Escherichia coli* prophages—HigBAC and CmdTAC—Mets et al. dissect the mechanisms of phage recognition by the SecB-like chaperone phage sensor and phage restriction by the TA warhead.

Mets et al., 2024, Cell Host & Microbe *32*, 1059–1073
July 10, 2024
https://doi.org/10.1016/j.chom.2024.05.003

CellPress

Cell Host & Microbe

CellPress
OPEN ACCESS

Article

Mechanism of phage sensing and restriction by toxin-antitoxin-chaperone systems

Toomas Mets,[1,2,11] Tatsuaki Kurata,[1,11] Karin Ernits,[1,11] Marcus J.O. Johansson,[1,11] Sophie Z. Craig,[3] Gabriel Medina Evora,[1,3] Jessica A. Buttress,[4] Roni Odai,[1] Kyo Coppieters't Wallant,[5] Jose A. Nakamoto,[1] Lena Shyrokova,[1] Artyom A. Egorov,[1] Christopher Ross Doering,[6] Tetiana Brodiazhenko,[2] Michael T. Laub,[6,7] Tanel Tenson,[2] Henrik Strahl,[4] Chloe Martens,[5] Alexander Harms,[8] Abel Garcia-Pino,[3,*] Gemma C. Atkinson,[1,9,*] and Vasili Hauryliuk[1,2,9,10,12,*]

[1]Department of Experimental Medical Science, Lund University, 221 00 Lund, Sweden
[2]University of Tartu, Institute of Technology, 50411 Tartu, Estonia
[3]Cellular and Molecular Microbiology (CM2), Faculté des Sciences, Université Libre de Bruxelles (ULB), Campus La Plaine, Building BC, Room 1C4203, Boulevard du Triomphe, 1050 Brussels, Belgium
[4]Centre for Bacterial Cell Biology, Biosciences Institute, Newcastle University, Newcastle upon Tyne NE2 4AX, UK
[5]Centre for Structural Biology and Bioinformatics, Université Libre de Bruxelles (ULB), Boulevard du Triomphe, Building BC, 1050 Bruxelles, Belgium
[6]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA
[7]Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA, USA
[8]ETH Zurich, Institute of Food, Nutrition and Health, 8092 Zürich, Switzerland
[9]Virus Centre, Lund University, Lund, Sweden
[10]Science for Life Laboratory, Lund, Sweden
[11]These authors contributed equally
[12]Lead contact
*Correspondence: abel.garcia.pino@ulb.be (A.G.-P.), gemma.atkinson@med.lu.se (G.C.A.), vasili.hauryliuk@med.lu.se (V.H.)
https://doi.org/10.1016/j.chom.2024.05.003

SUMMARY

Toxin-antitoxins (TAs) are prokaryotic two-gene systems composed of a toxin neutralized by an antitoxin. Toxin-antitoxin-chaperone (TAC) systems additionally include a SecB-like chaperone that stabilizes the antitoxin by recognizing its chaperone addiction (ChAD) element. TACs mediate antiphage defense, but the mechanisms of viral sensing and restriction are unexplored. We identify two *Escherichia coli* antiphage TAC systems containing host inhibition of growth (HigBA) and CmdTA TA modules, HigBAC and CmdTAC. HigBAC is triggered through recognition of the gpV major tail protein of phage λ. Chaperone HigC recognizes gpV and ChAD via analogous aromatic molecular patterns, with gpV outcompeting ChAD to trigger toxicity. For CmdTAC, the CmdT ADP-ribosyltransferase toxin modifies mRNA to halt protein synthesis and limit phage propagation. Finally, we establish the modularity of TACs by creating a hybrid broad-spectrum antiphage system combining the CmdTA TA warhead with a HigC chaperone phage sensor. Collectively, these findings reveal the potential of TAC systems in broad-spectrum antiphage defense.

INTRODUCTION

Toxin-antitoxin (TA) systems are diverse two-gene (bicistronic) elements that are ubiquitous in the genomes of archaea, bacteria, and temperate bacteriophages.[1] Although multiple functions have been demonstrated for TAs over the years, their role as abortive infection antiphage defense systems has become a particularly active topic of investigation in recent years.[2–4] Based on the nature of the antitoxin (protein or RNA) and the mechanism of toxin neutralization (such as the formation of an inactive TA complex, protection of the cellular target from the toxin, or degradation of the toxin mRNA), TAs are classified into eight groups, from type I to type VIII. Type II TAs employ proteinaceous antitoxins that neutralize cognate toxins through the formation of a tight, non-toxic TA complex.

Type II toxins have diverse mechanisms of toxicity, with the most common cellular targets being (1) protein synthesis machinery, (2) replication apparatus, (3) cell wall and cell skeleton, and (4) nucleotide metabolism.[1] One of the most well-known TA systems is host inhibition of growth (HigBA), originally identified in *Proteus vulgaris* as a locus that increases the stability of the Rts1 plasmid on which it is encoded.[5] The HigB RNase toxin is a ribosome-dependent mRNA interferase that is neutralized by the N-terminal intrinsically disordered regions (IDRs) of the dimeric HigA antitoxin.[6–8] DarTG is a recently discovered prophage-encoded antiphage defense TA system.[9,10] A member of the ADP-ribosyltransferase (ART) protein family,[11] the DarT

toxins link the ADP-ribose moiety of the NAD^+ (nicotinamide adenine dinucleotide) cofactor to the amino group of the guanine or thymine base of the DNA,[9,12] which results in inhibition of both RNA and DNA synthesis.[10,13] Detection of the invading phage by DarTG triggers the toxin, thus shutting down cellular transcription and replication in the infected cell to halt virus production.[10]

A variation on the type II TA theme is provided by tripartite TA-chaperone (TAC) systems. Although TAs are encoded by bicistronic operons, TACs are encoded by tricistronic operons: a "classical" TA gene arrangement followed by a third gene encoding a SecB-like ($SecB^{TA}$) chaperone.[14,15] SecB is a general housekeeping chaperone of Pseudomonadota (Proteobacteria) that both post- and co-translationally assists the folding of diverse proteins, both membrane-targeted and cytoplasmic.[16,17] $SecB^{TA}$ TAC chaperones are, on the other hand, highly specialized to ensure the stability of otherwise highly labile TAC antitoxins.[14] The strict dependence of the TAC TA units on $SecB^{TA}$-mediated stabilization is determined by the chaperone addiction (ChAD) element, an IDR located at the C terminus of the antitoxin.[18,19] Rather than mediating antitoxin neutralization, the ChAD region decreases the solubility of the antitoxin, promoting aggregation and degradation. This results in TA "addiction" to $SecB^{TA}$: the antitoxin and the chaperone act in concert to keep the toxin neutralized. Removal of the ChAD element converts the *Mycobacterium tuberculosis* HigBAC into a $SecB^{TA}$-independent TA system.[19] The interaction between the ChAD region of the *M. tuberculosis* HigA1 antitoxin and $SecB^{TA}$ chaperone HigC is strictly dependent on the aromatic Y114 residue of the ChAD.[19,20] ChAD elements are modular: grafting ChADs onto canonical type II TAs can convert TAs into chaperone-addicted TACs.[19] Finally, the destabilizing activity of the ChAD element is conditional on proteolysis of the antitoxin by the ClpXP protease, with *M. tuberculosis* HigBA being rendered non-toxic and $SecB^{TA}$-independent in Δ*clpX* or Δ*clpP* genetic backgrounds.[18]

A recent study by Vassalo and colleagues has established the biological function of two *Escherichia coli* (*E. coli*) TAC systems: MqsRAC from *E. coli* strain C496_10[19] and the ART toxin-containing system PD-T4-9 encoded in the hypervariable region of a P2-like prophage of *E. coli* ECOR22[21] Both of these TACs were shown to mediate antiphage defense, and the latter was renamed CmdTAC for chaperone-mediated defense TAC.[21] The CmdTAC PD-T4-9 protects *E. coli* from *Tevenvirinae* phage T4, with over-production of the CmdC $SecB^{TA}$ chaperone abrogating protection; the *E. coli* C496_10 MqsRAC protects from *Tevenvirinae* phage T2 but not T4.[21]

In this study, we address key outstanding questions in TAC biology: What is the mechanism of toxicity employed by CmdT toxins? How do TAC systems sense phage infection? Using our recent survey of type II TA diversity[22] as a starting point, we identify and validate two prophage-encoded TAC systems: HigBAC from *E. coli* strain NT1F31 and CmdTAC from *E. coli* O112ab:H26. We demonstrate HigBAC-mediated defense against siphoviruses (i.e., phages with long, noncontractile tails): *Queuovirinae* coliphage Bas25[23] and *Lambdavirus* λ_{vir}. We show that the toxicity of HigBAC is triggered through recognition of the gpV major tail protein of λ_{vir} through competition with the ChAD element of the antitoxin HigA. Binding of both HigA and gpV to $SecB^{TA}$ is strictly dependent on specific aromatic residues that compete for the same binding pockets of the chaperone. We uncover the molecular mechanism of anti-*Tevenvirinae* defense by CmdTAC: once the TAC system is activated upon phage infection, the CmdT ART toxin sequence-specifically ADP-ribosylates mRNA to shut down protein synthesis. Finally, by combining the CmdTA TA warhead with the HigC chaperone phage sensor, we create a hybrid broad-spectrum phage defense system, thus highlighting the evolutional malleability and modular nature of the TAC architecture. Collectively, our study establishes key principles for phage sensing and restriction by TAC defense systems.

RESULTS

E. coli NT1F31 HigBAC and *E. coli* O112ab:H26 CmdTAC are translation-targeting TAC systems that confer phage immunity

In our previous high-throughput analysis of type II TA systems, the NetFlax algorithm automatically identified TA systems based on their presence in conserved two-gene architectures.[22] This meant that toxin or antitoxin-like proteins encoded in longer conserved neighborhoods were not considered. However, when surveying "disregarded" genomic neighborhoods, we noticed that some toxins from the D2 MqsR-like toxin node of the TA network were associated with an antitoxin followed by a gene encoding a SecB-like protein, suggesting they are TACs, similar to MqsRAC.[19,21] To find homologs of these putative TACs that would be suitable for microbiological studies in the *E. coli* host, we searched for relatives of the SecB component in *Enterobacteriaceae* and carried out FlaGs gene neighborhood analysis.[24] This yielded a number of different TAC-like systems in *Enterobacteriaceae*, along with *yibN-grxC-secB-gpsA* operons encoding orthologues of the housekeeping *E. coli* SecB[25] (Figure S1A).

We focused on two representatives: HigBAC from *E. coli* strain NT1F31 and a tripartite CmdTAC system from *E. coli* O112ab:H26. Both of the systems are located on prophages: a lambdoid phage in the case of *E. coli* NT1F31 HigBAC and an unclassified prophage in the case of *E. coli* O112ab:H26 CmdTAC (Figures 1A and 1B). Importantly, HigBAC is encoded in a variable defense locus downstream of the *cI* repressor,[26] which in the classical λ phage encodes the RexAB phage exclusion system[27] (Figure S1B). Toxicity neutralization assays establish that both NT1F31 HigBAC and O112ab:H26 CmdTAC are, indeed, bona fide TACs. When expressed under the control of an arabinose-inducible P_{BAD} promoter from a pBAD33 plasmid vector, both *E. coli* NT1F31 HigBA (with a weak Shine-Dalgarno sequence) and *E. coli* O112ab:H26 CmdTA (with strong Shine-Dalgarno sequence) TA modules inhibit growth of the BW25113 test strain (Figures 1A and 1B). This demonstrates both the functionality of the toxins and the inability of the full-length antitoxins to neutralize the toxin in the absence of a chaperone. In the case of HigBA, the growth defect caused by TA expression is fully rescued when the cognate HigC chaperone is co-expressed in *trans* from a second plasmid (a pMG25 derivative) under the control of an IPTG (isopropyl β-D-1-thiogalactopyranoside)-inducible $P_{A1/O4/O3}$ promoter (Figure 1A). A partial neutralization of toxicity is observed for CmdTA co-expressed with the CmdC chaperone (Figure 1B). Importantly, although

CellPress
OPEN ACCESS

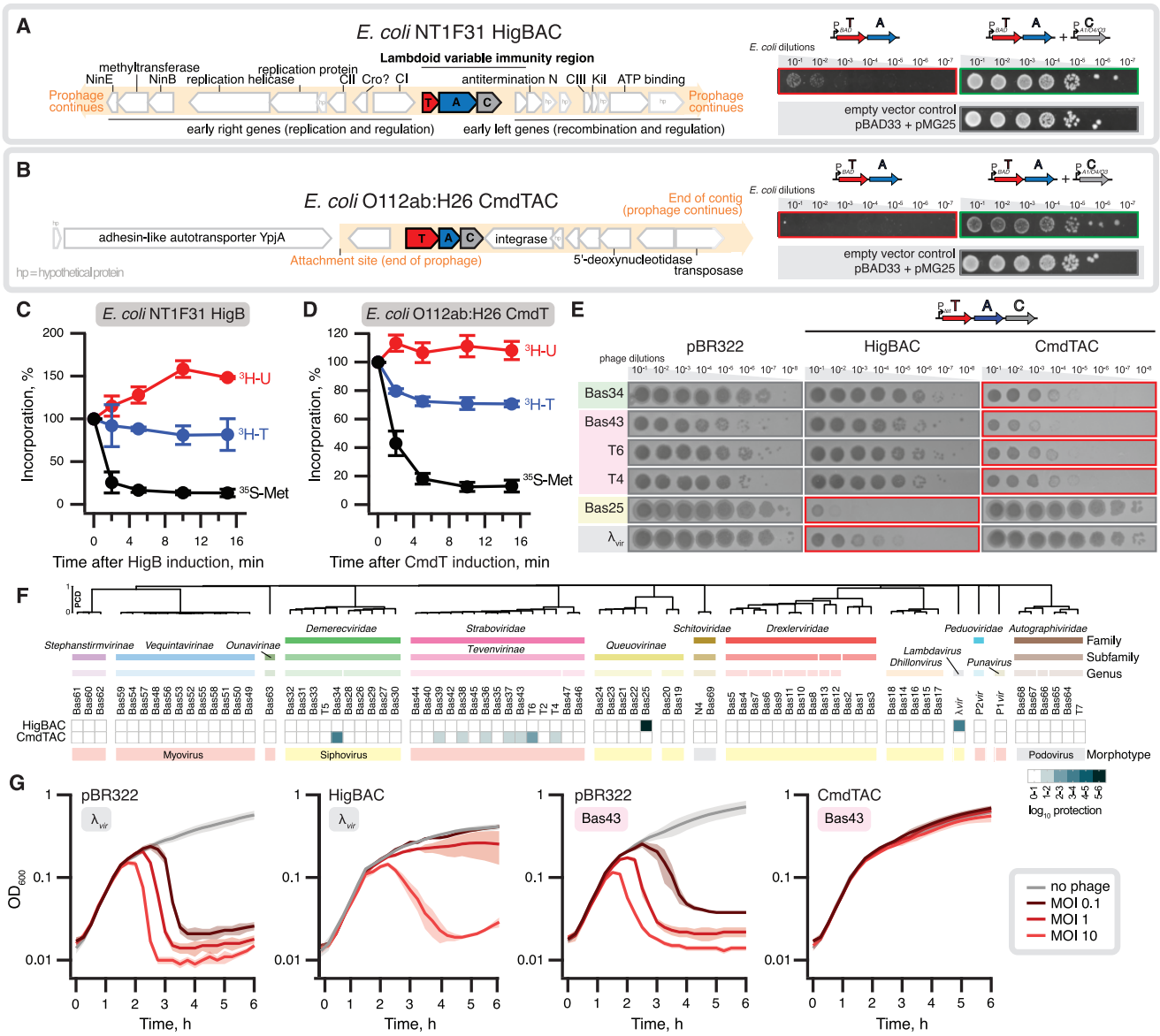


Figure 1. Phage defense by translation-targeting TAC systems: *E. coli* NT1F31 HigBAC and *E. coli* O112ab:H26 CmdTAC

(A and B) Gene neighborhoods of the validated TAC systems: (A) *E. coli* NT1F31 HigBAC and (B) *E. coli* O112ab:H26 CmdTAC. Both systems are encoded on prophage regions (left) and are functional in toxicity neutralization assays (right). To test TAC functionality, *E. coli* BW25113 strains were transformed with empty pBAD33 and pMG25 vectors or their derivatives expressing TAC toxin-antitoxin pairs (expression induced by 0.2% arabinose) and TAC chaperones (induced by 500 μM IPTG), respectively.

(C and D) Metabolic labeling assays with *E. coli* BW25113 expressing *E. coli* NT1F31 HigB RNase (D) or *E. coli* O112ab:H26 CmdT ART (C) toxins show specific inhibition of protein synthesis as manifested by a sharp decrease in ^{35}S-Met incorporation. The data is shown as geometric means with standard deviations.

(E and F) *E. coli* BW25113 cells transformed with either the pBR322 empty vector or either the two pBR322-based plasmids driving the expression of the TAC operons under the control of the constitutive P_{tet} promoter were challenged with 10-fold serial dilutions of BASEL[23] and common laboratory coliphages.

(E) Selected phages that were countered by one or other of the TAC systems tested.

(F) The results of the full screen shown as a heatmap of $\log_{10}$ protection values. The phage order and dendrogram are defined by hierarchical clustering applied to the proteome composition distance (PCD) matrix (see the STAR Methods section for details).

(G) Growth of *E. coli* BW25113 carrying the empty vector or the indicated plasmid-encoded TAC system in the presence of λ_{vir} or Bas43 phages at MOIs of 0, 0.1, 1, and 10. The shaded areas indicate the standard deviations. Additional liquid culture infection experiments with Bas25 and T4 are shown in Figures S3A–S3D.

P_{BAD}-driven expression of toxins abrogates *E. coli* growth, neither the antitoxin nor chaperone affect growth when expressed under the control of $P_{A1/O4/O3}$ (Figure S2A). Since the operon structure plays an important role in co-translational assembly of protein complexes,[28,29] we also tested the toxicity of CmdTAC and HigBAC TAC operons expressed from pBAD33. However, even with the chaperone expressed in *cis* (from the same plasmid), the CmdT toxicity is not fully neutralized

(Figure S2B); the likely explanation is excessive toxin expression in our constructs or/and the use of a non-native host strain.

We used metabolic labeling assays to assess the effects on translation (by following incorporation of ^{35}S-methionine in proteins), transcription (incorporation of ^{3}H-uridine in RNA), and replication (incorporation of ^{3}H-thymidine in DNA) to establish the mechanisms of toxicity employed by the newly identified TACs (Figures 1C and 1D). Given that the Rts1 plasmid-encoded HigB TA toxin is a ribosome-dependent RNase (mRNA interferase),[30] it is likely that the homologous *E. coli* NT1F31 HigB toxin also degrades mRNA. In support of this, *E. coli* NT1F31 HigB causes specific and potent inhibition of protein synthesis (Figure 1C). The outcome of the labeling experiments with *E. coli* O112ab:H26 ART CmdT toxin was more surprising. In stark contrast to the ART toxin DarT that inhibits replication and transcription,[10] the *E. coli* O112ab:H26 CmdT ART toxin is also a potent and specific inhibitor of protein synthesis, raising the question of the mechanism of CmdT toxicity (Figure 1D).

To test the functionality of the identified TACs in antiphage immunity, we performed a screen using the BASEL coliphage collection as well as a set of commonly used phages[23] (Figures 1E and 1F). Our collection of double-stranded DNA *Caudoviricetes* viruses included representatives of all the three morphologies: podovirus (short noncontractile tail), siphovirus (long noncontractile tail), and myovirus (long and flexible contractile tail). The tripartite TAC systems were expressed in the BW25113 *E. coli* K12 strain from a pBR322 derivative[31] under the control of a constitutive P_{tet} promoter. The *E. coli* NT1F31 HigBAC system provided protection from siphovirus phages such as *Lambdavirus* λ_{vir} and *Queuovirinae* Bas25, while *E. coli* O112ab:H26 CmdTAC granted protection against myoviruses from the *Tevenvirinae* subfamily such as Bas43, T4, and T6 as well as a siphovirus from the *Demerecviridae* family (Bas34). Liquid culture infection assays with increasing multiplicity of infection (MOI; 0.1, 1, and 10) support the defensive activity of HigBAC against λ_{vir} and Bas25 and CmdTAC against Bas43 and T4 (Figures 1G and S3A–S3D). The effects of TAC-mediated defense on phage-mediated culture collapse differ for different experimental systems. Even at MOI of 10, CmdTAC- and HigBAC-expressing *E. coli* strains are virtually immune to Bas43 and Bas25, respectively. This suggests that TACs do not necessarily act via abortive infection *sensu stricto*, that is by killing the infected cell. Rather, they compromise the virus more than the host, and for some phage-TAC pairs, this allows the infected cell to continue growth after defeating the invader. We have seen the same phenomenon in the case of *E. coli* expressing the *Lactococcus lactis* AbiA reverse transcriptase defense system, with bacterial growth being virtually unaffected by Bas20 and Bas25 even at an MOI of 10.[32] At an MOI of 10, λ_{vir} does eventually cause a collapse of HigBAC-expressing cultures, but more gradually than the rapid collapse in the absence of the defense system. Finally, in the case of CmdTAC-expressing T4-infected *E. coli* cultures, phage-induced lysis is abolished, and bacterial growth is arrested at the same time point at which culture collapse is observed in the cultures of *E. coli* that lack the system. As T4 rapidly degrades bacterial DNA,[33] this possibly reflects efficient abortion of the T4 lytic cycle at its later stages, that is, after host DNA degradation.

Collectively, our results establish these two TAC systems as translation-targeting antiphage immunity systems. Motivated by these results, we set out to answer the following questions: what is the mechanism of CmdT-mediated inhibition of translation? What is the mechanism of phage sensing and restriction by our TAC systems?

E. coli O112ab:H26 ART CmdT ADP-ribosylates mRNA to shut down translation

To validate the direct inhibition of translation by CmdT *in vitro*, we assayed the effect of CmdT on the production of the dihydrofolate reductase (DHFR) reporter protein in a reconstituted *E. coli* cell-free protein synthesis system (PURE).[34] The toxin was first produced *in situ* in the absence of the NAD^+ cofactor, and then DHFR production was assayed either in the presence or absence of NAD^+. The *E. coli* O112ab:H26 ART CmdT toxin potently inhibits DHFR synthesis in a strictly NAD^+-dependent manner, and, importantly, the 6-biotin-17-NAD^+ analog ($^{biot}NAD^+$) also supports the toxic activity (Figure 2A).

To identify the specific molecular target modified by CmdT, we used streptavidin-conjugated horseradish peroxidase (streptavidin-HRP) that detects $^{biot}NAD^+$. We first produced the CmdT toxin *in situ* in the PURE system in the presence of $^{biot}NAD^+$ and after a 60-min incubation at 37°C resolved the protein and RNA components on denaturing gels and probed with streptavidin-HRP (Figure 2B). Although no specific protein signal is detectable, a clear RNA signal is detectable with a size (about 0.8 kb) corresponding to CmdT-encoding mRNA. The signal is sensitive to treatment with single-stranded RNA-specific RNase A (Figure 2C). Next, we used a 24 nt-long model mRNA(MV) (5′-GGCAAGGAGGUAAAAAUGGUUAAA-3′) coding for the MV (Met-Val) dipeptide as an RNA substrate. This mRNA is commonly used to assemble ribosomal complexes for biochemical investigations.[35] mRNA(MV) but not the corresponding DNA oligo, mDNA(MV), is readily modified by the toxin (Figure 2D). In liquid culture infection assays, CmdTAC renders *E. coli* immune to Bas43 (Figure 1G). This suggests that the system preferentially targets Bas43 mRNAs, thus allowing the bacterial host to repel the infection and continue growing. We used a set of RNA oligos to test whether RNA modification by O112ab:H26 CmdT is sequence-specific. None of the homopolymeric 14 nt-long RNAs (poly(G), poly(C), poly(U), or poly(A)) are effectively modified (Figure S4). Furthermore, an RNA oligo that is reverse-complementary to mRNA(MV) was not modified, suggesting that the modification is sequence-specific (Figure 2E). Collectively, our results establish O112ab:H26 CmdT as a sequence-specific mRNA-modifying ART toxin.

A pattern of conserved aromatic residues of the HigA ChAD determines antitoxin solubility and mediates its recognition by SecB^{TA}

Due to more efficient suppression of TA toxicity by SecB^{TA} co-expression in the case *E. coli* NT1F31 HigBAC as compared with *E. coli* O112ab:H26 CmdTAC (Figure 1B), the HigBAC is more experimentally amenable for studying the molecular mechanism of ChAD recognition by SecB^{TA}. Therefore, we focused on this system. We used AlphaFold-multimer[36,37] to co-fold the tetrameric HigC SecB^{TA} in complex with one HigBA TA module (Figures 3A and S5A). At 96 amino acids (residues 122–218),

CellPress
OPEN ACCESS

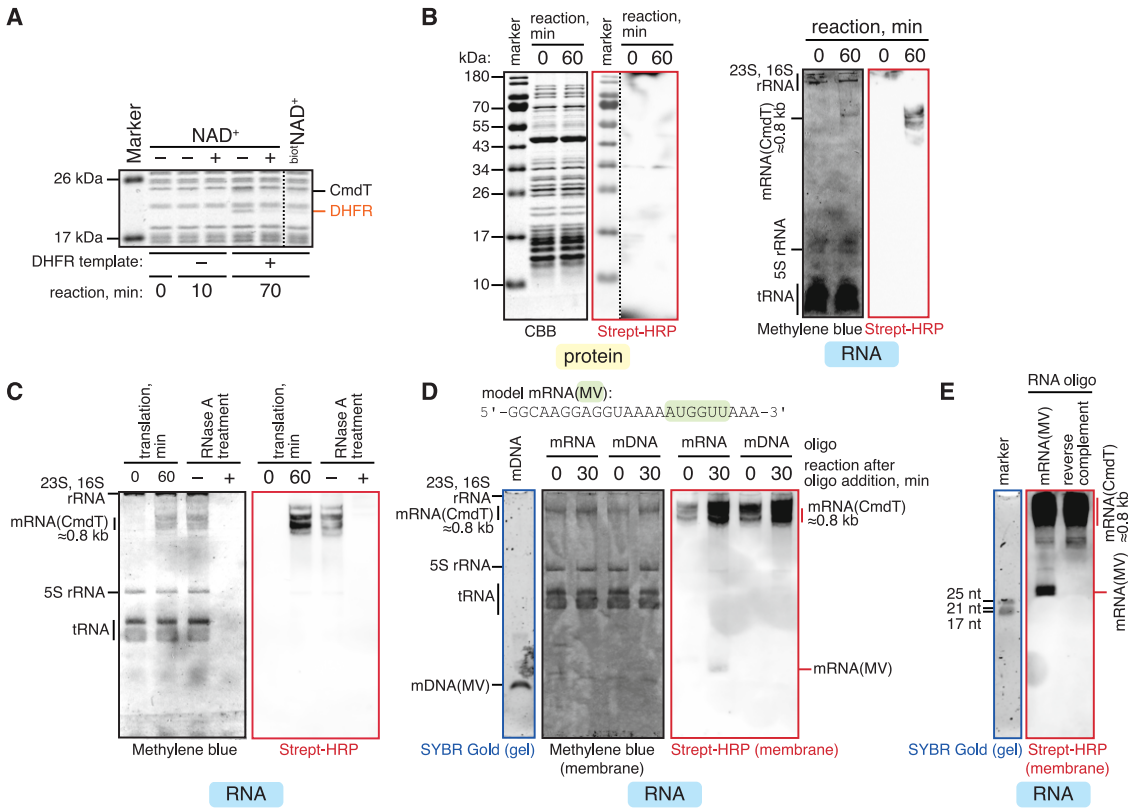


Figure 2. The *E. coli* O112ab:H26 CmdT ART toxin disrupts translation through sequence-specific ADP-ribosylation of mRNA

(A) CmdT NAD$^+$-dependently abrogates production of DHFR in cell-free expression assays.

(B and C) CmdT modifies the *cmdT*-encoding mRNA in the presence of the biotinylated NAD$^+$ (^{biot}NAD$^+$) substrate. The toxin modifies mRNA(CmdT) but not the protein components of the cell-free expression system (B). The modification signal is sensitive to the addition of RNase A (C).

(D) CmdT modifies the model mRNA(MV) oligonucleotide but not the corresponding DNA oligonucleotide, mDNA.

(E) The model RNA(MV) but not its reverse complement RNA is modified by CmdT in the presence of ^{biot}NAD$^+$.

the predicted unstructured HigA ChAD is more than 2-fold longer than that of the *M. tuberculosis* HigA1.[19] The ChAD element is predicted to wrap around the SecB^{TA} tetramer, contributing four additional antiparallel β-strands to the conserved β-sheet substructure of the SecB fold. The ChAD β-strands are predicted to occupy the peptide-binding channel of HigC that was previously defined by Xu and colleagues for *Haemophilus influenzae* SecB.[38] Four aromatic residues of the ChAD element—W128, F158, F188, and Y214—slot into four equivalent pockets of the chaperone subunits involving residues Y39, N41, R49, and D71 of HigC. Analogous pockets accommodating aromatic and hydrophobic ChAD residues were previously observed in the complex of *M. tuberculosis* SecB^{TA} bound to a short ChAD peptide fragment.[20]

The aromatic residues of HigA ChAD are conserved, supporting their functional importance (Figure S6A). IDRs often recognize their target proteins through recognition of short linear motifs (SLiMs) or/and shorter specific sequence features, such as clusters of aromatic of charged residues.[39] Closer inspection of the four aromatic residues reveals that (1) there is a loose SLiM consensus associated with the residues: with the exception of F188, aromatic residues are preceded by polar residue and followed by an amino acid with a small side chain; (2) the three potential SLiMs are regularly interspaced across the ChAD with a step of ≈30 amino acids; and (3) as predicted by CamSol,[40] F188 is located in a region with a high propensity for aggregation (Figure 3B), thus suggesting a key role in the destabilizing activity of the ChAD element.

Removal of the predicted ChAD region (Δ122-218) renders the *E. coli* NT1F31 HigA antitoxin able to efficiently neutralize the HigB toxin in the absence of the chaperone, behaving like a classical *higBA* TA pair (Figure 3C). To validate our AlphaFold predictions of specific interactions, we next targeted the four aromatic residues of the HigA ChAD element. First, we substituted each individually for alanine and tested the toxicity of the resultant mutant *higBA* TA units in the presence or absence of the chaperone (Figure 3D). The only mutant *higBA* TA variant that had a (very mild) phenotype was the F188A substitution that is predicted to be a part of the CamSol-predicted aggregation-prone patch. This TA variant is slightly less toxic than the wild type (Figure 3D, left), while the HigBA toxicity is still efficiently neutralized by HigC (Figure 3D, right). Next, we substituted W128, F158, and Y214 in combinations, leaving out F188. Although none of the tested pairwise substitutions abolished the HigC-mediated suppression of toxicity, substitution of the three residues for alanine completely abrogated the effect of HigC co-expression (Figure 3D, right). This suggests that although they are not individually essential for SecB^{TA} addiction, these residues collectively

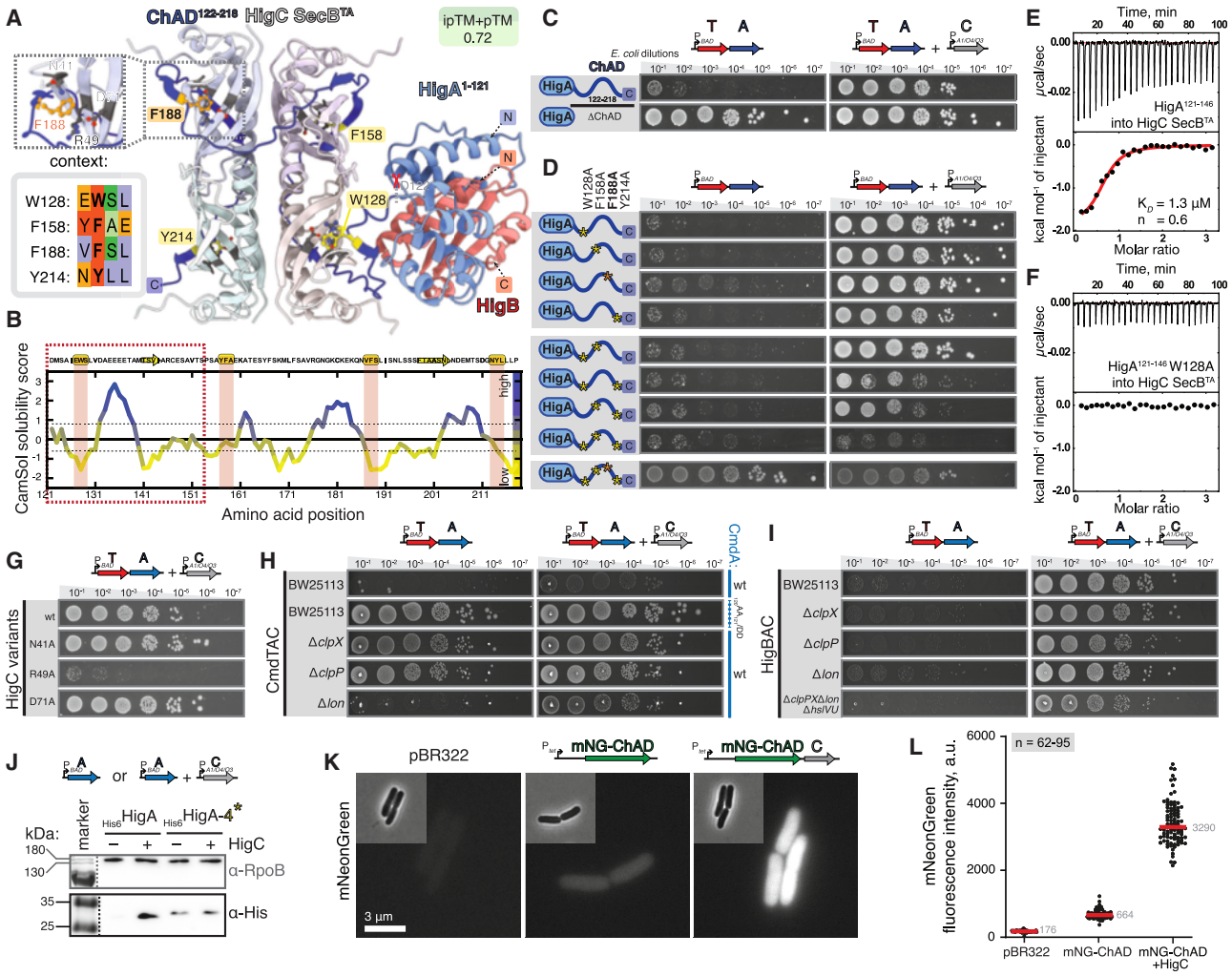


Figure 3. The aromatic residues of the ChAD element play key roles in HigC-dependent neutralization of the HigB toxin by HigA

(A) An AlphaFold model of HigBA TA unit in complex with the HigC tetramer. The four aromatic residues of the HigA ChAD element that are predicted to interact with the four equivalent pockets of the $HigC_4$ tetramer are highlighted.

(B) CamSol solubility analysis of the HigA ChAD region, residues 122–218. The four putative SLiM motifs are highlighted, while the ChAD section used for ITC experiments is marked with a dashed box.

(C and D) Toxicity assays with wild-type and ChAD-mutated HigA antitoxin variants. Removal of the ChAD element renders the HigBA TA unit non-toxic, indicative of efficient HigA stabilization (C). Effects of individual substitutions of the four aromatic residues of the ChAD element on HigBA toxicity (D, left) and its suppression through co-expression of the HigC (D, right). HigBA and HigC expression was induced by 0.2% arabinose and 50 μM IPTG, respectively.

(E and F) Binding of the wild-type (E) and W128A-substituted (F) $HigA^{121–146}$ peptide to HigC as monitored by ITC. The stoichiometry is calculated per HigC protein chain.

(G) Suppression of HigBA toxicity through IPTG-induced co-expression of HigC variants carrying amino acid substitutions in the pocket that is predicted to recognize the aromatic residues of the HigA ChAD. The R49A HigC variant is unable to suppress the HigBA toxicity.

(H and I) TA and TAC toxicity assays in *E. coli* BW25113 protease-deficient strains. Indicative of antitoxin degradation by ClpXP, the CmdTA TA unit is non-toxic in Δ*clpP* and Δ*clpX* but not Δ*lon* strains (H, left). The CmdTA TA toxicity is abrogated upon the Asp-Asp substitution of the C-terminal Ala-Ala motif of CmdA that is recognized by ClpX (H). The HigBA TA is toxic in all of the tested backgrounds, suggesting that proteolytic degradation is not essential for ChAD-mediated destabilization of the antitoxin (I, left). TA and chaperone expression was induced by 0.2% arabinose and 500 μM IPTG, respectively.

(J) Immunoblotting analysis of *E. coli* BW25113 cells expressing either wild-type or substituted (4*) N-terminally His_6-tagged HigA, either in the absence or in the presence of HigC.

(K and L) Fluorescence and phase contrast microscopy of *E. coli* BW25113 cells transformed with an empty pBR322 vector or pBR322 derivatives expressing either mNeonGreen-ChAD alone or mNeonGreen-ChAD together with HigC (K). Quantification of mNeonGreen fluorescence for individual cells from the same imaging dataset (n = 62–95 cells, the red line represents the median value) (L).

participate in the interaction through avidity effects. Finally, the tetra-substituted variant has a dramatically reduced toxicity, which does not need to be suppressed by HigC. This suggests that the four residues are collectively essential for rendering the ChAD-tagged antitoxin aggregation-prone and chaperone-addicted. Next, we used isothermal titration calorimetry (ITC) to

Table 1. Binding parameters of gpV and HigA fragments to HigC

Titrations	KD (μM)	ΔH (kcal/mol)	−TΔS (kcal/mol)	ΔG (kcal/mol)	Molar ratio (n)
$gpV^{1–160}$ into HigC	N.D.	N.D.	N.D.	N.D.	N.D.
$gpV^{48–78}$ into HigC	0.25	−3.8	−5.3	−9.1	0.47
$gpV^{48–78}{}_{W67A}$ into HigC	N.D.	N.D.	N.D.	N.D.	N.D.
$gpV^{48–78}{}_{61DDED64/AAAA}$ into HigC	7.0	−4.5	−2.6	−7.1	0.49
$HigA^{121–146}$ into HigC	1.3	−1.9	−6.3	−8.2	0.6
$HigA^{121–146}{}_{W128A}$ into HigC	N.D.	N.D.	N.D.	N.D.	N.D.
($gpV^{48–78}$ + $HigA^{121–146}$) into (HigC + $HigA^{121–146}$)	6.7	−5.6	−1.6	−7.2	0.5

Experiments were performed with either 30 μM HigC (titrated with 350 μM $gpV^{1–160}$) or 16 μM HigC (titrated with 160 μM $gpV^{48–78}$ or $HigA^{121–146}$, either wild-type or substituted variants). In the competition experiment, 12 μM HigC supplemented with 18 μM $HigA^{121–146}$ was titrated with 160 μM $gpV^{48–78}$ similarly supplemented with 18 μM $HigA^{121–146}$ to prevent the dilution of the ChAD-mimicking peptide. The binding parameters were determined by fitting the ITC data to a single interaction model. Data represent mean values ± SD., and N.D. stands for "not detectable." The presented titrations are background-subtracted.

test the HigC-mediated recognition of a peptide mimicking the first of the four repetitive SLiM elements of the ChAD region (Figures 3E and 3F; Table 1). The peptide corresponding to residues 121–148 of HigA binds HigC with low-μM affinity (K_D = 1.3 μM, n = 0.6; n is calculated per HigC monomer). The W128A substitution targeting the aromatic core of the SLiM abrogates the interaction, directly validating specificity of the interaction.

Finally, we probed the pockets of the HigC chaperone that are predicted to recognize the aromatic residues of the ChAD element. Although alanine substitution of N41 and D71 do not affect HigC-mediated neutralization of HigBA, the R49A variant is unable to rescue the growth defect of HigBA-expressing cells (Figure 3G).

Inactivation of the *E. coli* NT1F31 HigA antitoxin does not rely on the ClpXP protease

The protease ClpXP recognizes specific motifs, or degrons, that mark a protein for degradation. The most well-characterized ClpXP degron is the SsrA-tag. The key element of this tag is the C-terminal Ala-Ala-COO^- moiety.[41,42] The ChAD of *M. tuberculosis* HigA1 terminates with the similarly aliphatic Val-Ala dipeptide degron, marking the antitoxin for ClpXP-mediated degradation.[18] Substitution of the C-terminal Val-Ala motif of *M. tuberculosis* HigA1 for Asp-Asp abolishes antitoxin degradation by ClpXP, rendering the well-neutralized HigBA1 TA unit non-toxic.[18]

The ChAD element of CmdA terminates with Ala-Ala, suggesting targeting by ClpXP. The *E. coli* NT1F31 HigA, on the other hand, terminates with Leu-Pro, which makes it an unlikely substrate for ClpXP. To test the role of ClpXP in the function of the two TAC systems, we substituted the terminal AA motif of CmdA for DD and tested the toxicity of the resultant CmdTA construct in wild-type BW25113 *E. coli* (Figure 3H). Consistent with ClpXP being responsible for the instability of wild-type CmdA for degradation, the $_{120}AA_{121}$/DD mutant CmdTA variant is non-toxic. Next, we tested the toxicity of the two TA modules in BW25113 *E. coli* Δ*clpP*, Δ*clpX*, and Δ*lon* strains; the latter strain was used as a specificity control. As expected, the CmdTA unit is not toxic in Δ*clpP* and Δ*clpX*—but is toxic in Δ*lon*—*E. coli* strains, supporting the direct role of ClpXP in CmdTAC triggering through antitoxin degradation (Figure 3H). As expected from its likely lack of a degron in the ChAD element, HigAB is toxic in all tested protease-deficient backgrounds, including Δ*clpPX* Δ*lon* Δ*hslVU* (Figure 3I). This suggests that HigBAC activation does not rely on the proteolytic activity of ClpX, Lon, or HslVU. Rather, it may rely on antitoxin aggregation or/and clearance by an as yet undetermined protease(s).

To directly probe the effects of HigC on HigA abundance and solubility, we used immunoblotting (Figure 3J) and fluorescence microscopy (Figures 3K and 3L). First, we tested the effects of HigC co-expression as well as simultaneous alanine substitutions of the four aromatic residues of the HigA ChAD element (W128A, F158A, Y214A, and F188A; referred to as the 4*) on the expression levels of N-terminally hexahistidine (His_6)-tagged HigA (Figure 3J). In the absence of HigC, wild-type HigA is not detectable by α-His_6 immunoblotting; co-expression of HigC results in a strong and specific α-His_6 signal. Conversely, the 4* ChAD-mutated HigA is stable regardless of the presence or absence of HigC. These results demonstrate that the HigC:ChAD interaction is essential for stabilizing the otherwise highly labile HigA and, second, that the aromatic residues of the ChAD element are crucial for its degron function. Next, we imaged *E. coli* cells expressing C-terminally ChAD-tagged monomeric yellow-green fluorescent protein mNeonGreen[43] with or without co-expression of HigC; *E. coli* transformed with empty pBR322 vector were used as a control (Figures 3K and 3L). In good agreement with immunoblotting experiments, HigC co-expression increased the fluorescent signal 5-fold. In the absence of HigC, the mNeonGreen-ChAD fluorescent signal is weak and evenly distributed throughout the cell with no aggregate formation detected, indicating that the association of HigA with HigC protects the antitoxin from proteolytic destruction rather than aggregation.

Toxic activity of the *E. coli* NT1F31 HigBAC system is triggered by direct sensing of the λ phage major tail protein gpV

To discover the nature of the phage triggers that activate the *E. coli* NT1F31 HigBAC defense system, we isolated spontaneous escape mutants of the λ_{vir} phage that can overcome HigBAC-mediated immunity. We sequenced four $\lambda_{vir}{}^{escape}$ mutant phages immune to HigBAC, and all shared the same S54P substitution in the major tail protein gpV (Dataset S1). In the absence of HigBAC, wild-type λ_{vir} and $\lambda_{vir}{}^{escape\_1}$ are similarly

active against *E. coli*; in the presence of HigBAC only the wild-type λ_{vir} is restricted (Figures 4A, S3E, and S3F). Given that structural proteins are well-established triggers of antiphage defense systems,[44,45] we focused our attention on gpV as a candidate HigBAC trigger. S54 is located in a highly flexible region of the β2-β3 loop (residues 50–78) of the N-terminal domain of gpV (gpV_N, residues 1–160) that mediates polymerization of the phage tail.[46,47] Thus, the S54P substitution would likely not significantly alter the structural properties of gpV given the intrinsic flexibility of the region.

To probe the possible physical interaction between gpV and HigBAC in the context of infection, we N-terminally FLAG-tagged HigC in the context of the full HigBAC system and performed immunoprecipitation followed by tandem mass spectrometry (IP-MS/MS) during infection with both wild-type λ_{vir} and the $\lambda_{vir}^{escape\_1}$ mutant. We sampled at 30 min post-infection, by which time late genes such as structural proteins are expected to be expressed.[48] gpV is dramatically enriched in HigC pull-down samples in the case of wild-type λ_{vir} as compared with samples generated using $\lambda_{vir}^{escape\_1}$ (Figure 4C). This suggests that the *gpV S54P* substitution allows λ_{vir}^{escape} to circumvent HigBAC defense by preventing gpV recognition by HigC and consequent HigBAC activation.

Prediction of the complex of the gpV_N region and the $SecB^{TA}$ tetramer by AlphaFold-Multimer yielded a high-confidence model, providing structural insight into how HigBAC could be triggered through direct recognition of the tail protein by the HigC chaperone (Figures 4D and S5B). The structure suggests that HigC prevents the folding of the N-terminal amino acid region of gpV_N (aa 1–78), which, reminiscent of the ChAD:$SecB^{TA}$ interaction, becomes linear and largely disordered. The residues constituting the β2-β3 loop as well as the flanking β-strands β2 and β3 are predicted to wrap around the chaperone, with aromatic residues W36 and W67 slotting into the same pockets of the chaperone that mediate ChAD recognition. The two aromatic residues are spaced out by ≈30 residues and are located in a context similar to that of the ChAD SLiMs, i.e., preceded by an aspartic acid and followed by a polar residue, serine, or threonine. Finally, the acidic patch $D_{61}DED_{64}$ is predicted to make extensive electrostatic interactions with the chaperone.

Co-expression of wild-type gpV—but not the S54P variant—induces HigBAC toxicity (Figure 4E). We used this reductionist system to probe our structural model through mutagenesis. Co-expression with gpV variants that are expected to compromise the interaction with HigC reveal differences in importance for activation. Activation is still seen with W36A and—to some extent—disruption of the acid patch $D_{61}DED_{64}$. S54P and W67A variants, on the other hand, do not activate HigBAC toxicity. Importantly, the N-terminal residues 1–78 of gpV_N fused with a stabilizing C-terminal SUMO (small ubiquitin-like modifier) tag activates the TAC toxicity more efficiently than the full-length gpV, with the S54P substitution in $gpV^{1–78}$-SUMO merely decreasing its activity but not abolishing it fully as observed in the case of full-length gpV.

gpV co-translationally competes with HigA ChAD for binding to HigC

Our structural modeling suggests the topology of the fully folded gpV is incompatible with the N-terminal region forming a complex with HigC. This is due to the β strands β2 and β3 that flank the disordered loop β2-β3 coming together to form part of a twisted β sheet[46] (compare Figures 4B and 4D). Supporting this, the C-terminally truncated and SUMO-tagged gpV_N fragment (1–78) lacking the β-sheet element is a more potent activator of HigBAC than the full-length gpV (Figure 4E). In the context of infection, where it is assisted by phage-encoded chaperones gpG and gpGT, gpV readily folds and assembles into a tail tube superstructure.[49] Therefore, we reasoned that there is only a short window of opportunity for HigC to recognize the nascent N-terminal fraction of gpV as the protein is still being synthesized and not yet folded.

To probe the interaction between unfolded gpV and HigC as would occur in a co-translational context, we designed an unstructured gpV peptide fragment comprising residues 48–78 that contains the W67 and $D_{61}DED_{64}$ elements predicted to be recognized by HigC. ITC experiments showed that the two partners form a tight complex (K_D = 250 nM, n = 0.47) (Figure 4F; Table 1). The W67A substitution disrupts the interaction between the $gpV^{48–78}$ peptide, while the poly-alanine substitution of the $D_{61}DED_{64}$ element decreases the affinity more than 20-fold (K_D = 7 μM, n = 0.5; n is calculated per HigC monomer) (Figures 4G and 4H). By contrast, our ITC experiments detect no interaction between gpV_N (1–160) and HigC, even when the phage protein is used at 350 μM concentration and HigC at 30 μM, a 5-fold increase compared with the concentrations used for the peptide-HigC interactions (Figure 4I). The lack of interaction further strengthens the idea that gpV needs to be partially folded for gpV:HigC complex formation to take place. To mimic HigBAC triggering by gpV, we performed a competition experiment with the ChAD of HigA. HigC was saturated with 1.5-fold molar excess of the $HigA^{121–148}$ ChAD-mimicking peptide, followed by titration with the $gpV^{48–78}$ peptide while keeping the ChAD peptide concentration stable. Under these conditions, $gpV^{48–78}$ binds HigC with an effective K_D of 6.6 μM, an about 25-fold drop in affinity (Figure 4J), which is strongly suggestive of direct competition between the two peptide ligands for HigC.

Finally, we probed the HigC:gpV interaction through hydrogen deuterium exchange mass spectrometry (HDX-MS) (Figures 4K–4M and S5E). To map the binding interface, we compared the deuterium exchange—a proxy for solvent accessibility—for individual HigC and $gpV^{48–78}$ with that for the HigC:$gpV^{48–78}$ complex (Figures 4K and 4L). A decrease in deuterium exchange (ΔHDX) in the HigC:$gpV^{48–78}$ complex as compared with unbound HigC was localized to β1, β2, β3, and α1 elements of HigC that constitute the peptide-binding groove as well as the pockets that accommodate the aromatic residues of the ChAD (Figure 4M). On the $gpV^{48–78}$ side, the entire peptide was strongly protected from deuterium exchange (Figure 4M). These results are in agreement with AlphaFold modeling and functional assays.

A hybrid TAC system comprised of the HigC chaperone and the CmdTA TA unit has an expanded antiphage defense spectrum and increased potency

Earlier studies have shown that $SecB^{TA}$ chaperones are specific and preferentially recognize the ChAD elements of cognate TA modules.[19] To establish whether this is also the case for our

CellPress
OPEN ACCESS

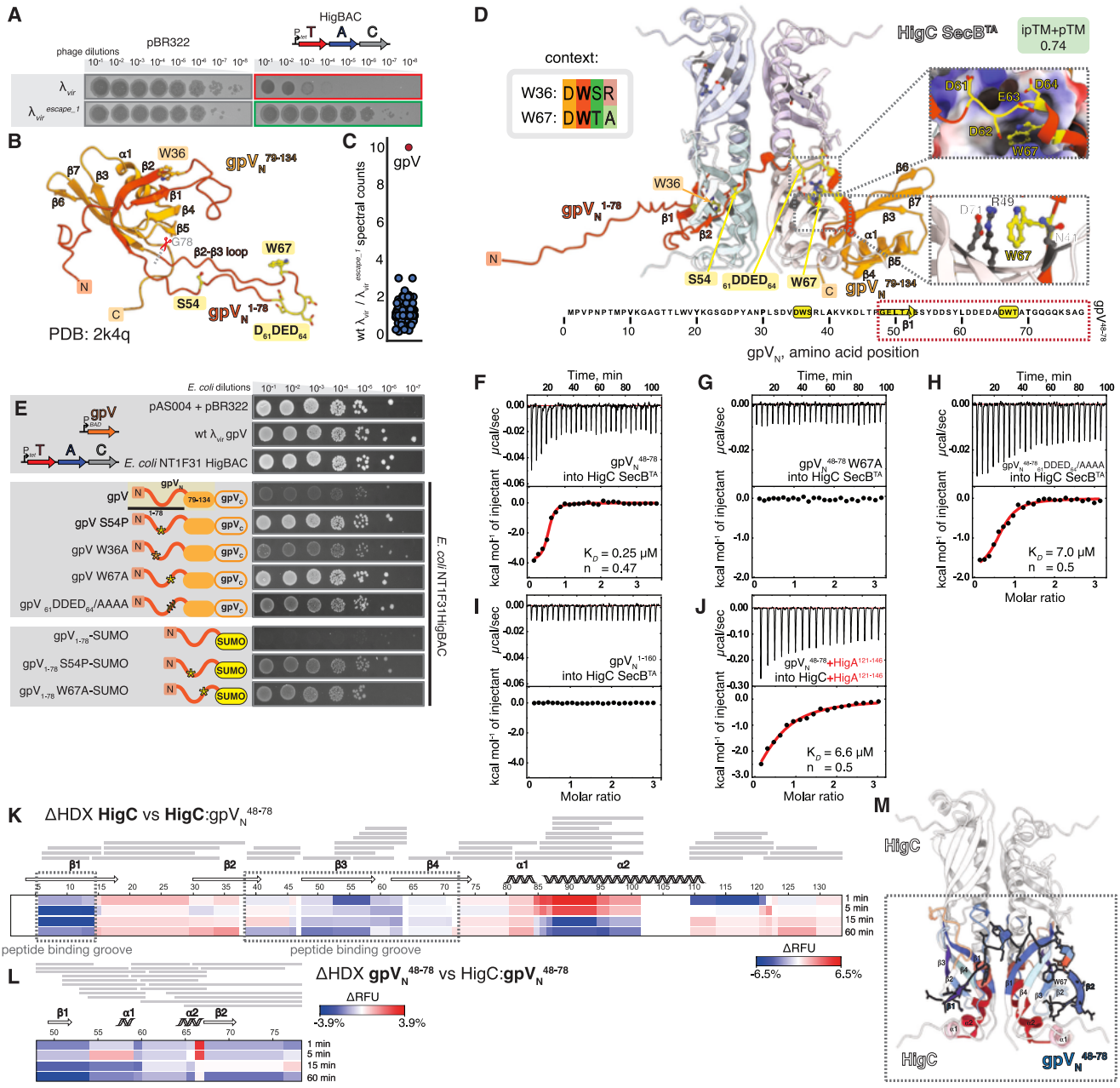


Figure 4. The λ phage major tail protein gpV triggers the *E. coli* NT1F31 HigBAC system

(A) Constitutive P_{tet}-driven expression of HigBAC provides protection against λ_{vir} but not the $\lambda_{vir}^{escape\_1}$ escape mutant carrying an *S54P* substitution in *gpV*. Complementary liquid culture infection experiments are shown in Figures S3E and S3F.

(B) Experimentally determined structure of gpV_N (PDB 2k4q).[46]

(C) Comparative IP-MS/MS analysis of HigBA_FLAG-HigC pull-down samples purified from *E. coli* culture infected with λ_{vir} and $\lambda_{vir}^{escape\_1}$ mutant reveals relative enrichment of gpV in the wild-type sample.

(D) An AlphaFold model of the λ gpV_N domain in complex with the HigC tetramer. The gpV_N^{48-78} section used for ITC experiments is highlighted with a dashed box.

(E) Induction of HigBAC toxicity through co-expression of wild-type and engineered gpV variants. Although expression of either HigBAC or gpV alone has no effect, their co-expression results in synthetic toxicity. Co-expression of HigBAC with the escape S54P mutant variant of gpV fails to trigger TAC-mediated toxicity. No triggering is observed for W67A-substituted gpV while decreased triggering is observed for the $_{61}DDED_{64}$/AAAA variant. Expression of gpV and its derivatives was induced by 0.2% arabinose.

(F–J) Binding of gpV_N^{1-160} protein (I) as well as of gpV^{48-78} peptide variants (F–H and J) to HigC as monitored by ITC. The stoichiometry is calculated per HigC protein chain.

(K and L) ΔHDX between HigC vs. HigC:gpV^{48-78} (K) as well as gpV^{48-78} vs. HigC:gpV^{48-78} (L) plotted as a heatmap.

(M) $HigC_4$:gpV^{48-78} AlphaFold model colored as a function of the ΔHDX. The gpV^{48-78} fragment is outlined in bold. Only two subunits of the HigC tetramer were colored for clarity.

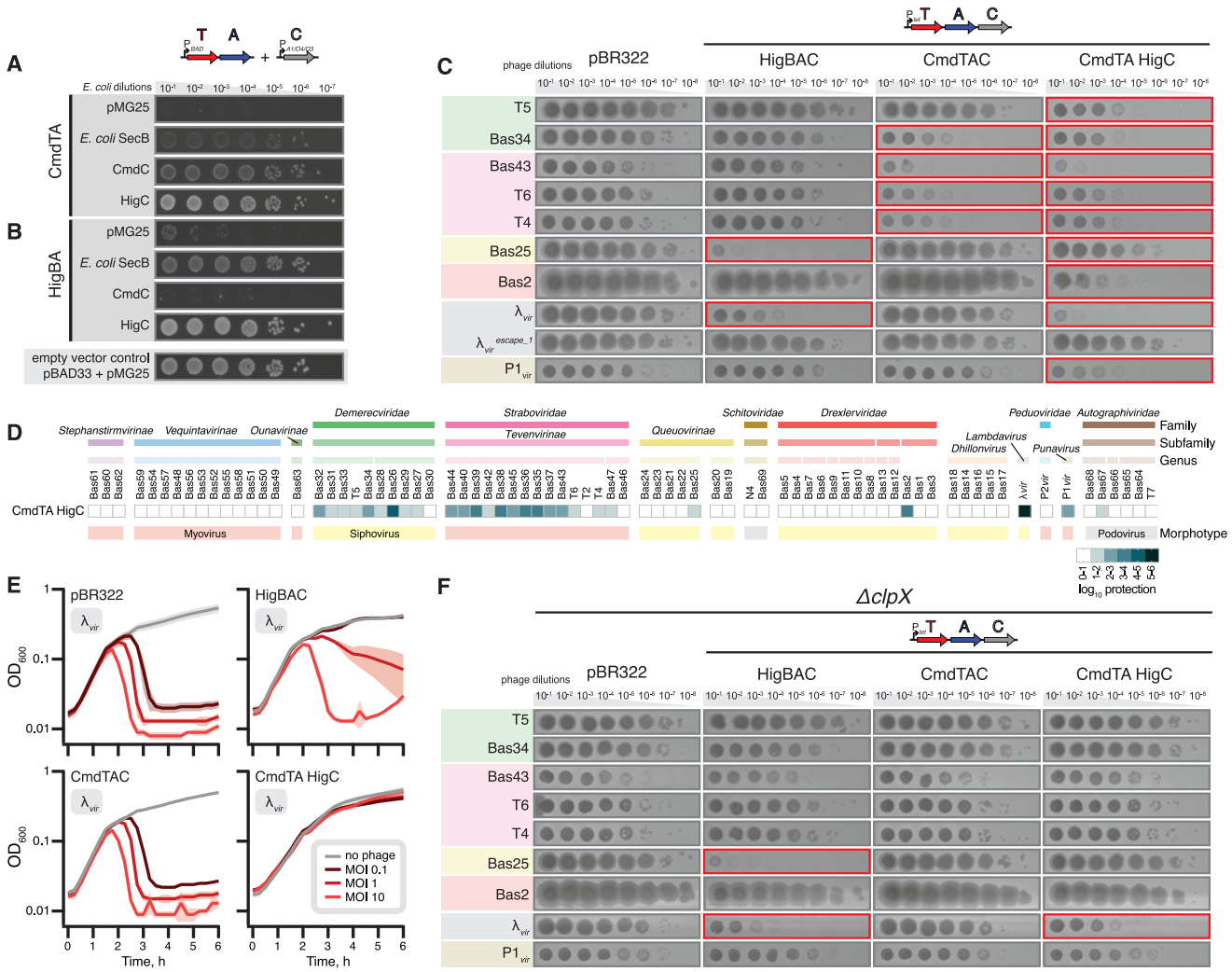


Figure 5. Engineering of a synthetic antiphage defense system through a combination of the CmdTA toxin-antitoxin unit and the HigC phage-sensing chaperone

(A and B) Suppression of CmdTA (A) or HigBA (B) toxicity through IPTG-induced co-expression of SecB chaperones: *E. coli* housekeeping SecB, CmdC, and HigC; pMG25 was used as an empty vector control. Expression TA and chaperone were induced by 0.2% arabinose and 500 μM IPTG, respectively.

(C and D) *E. coli* BW25113 strains expressing HigBAC, CmdTAC, or the hybrid CmdTA-HigC TAC operon composed of CmdTA toxin-antitoxin unit and HigC chaperone were challenged with 10-fold serial dilutions of BASEL[23] and common lab coliphages, including the $\lambda_{vir}^{escape\_1}$ escape mutant variant. (C) depicts results with select phages, and the results of the full screen are shown in (D).

(E) Growth of *E. coli* BW25113 carrying the empty vector or the indicated plasmid-encoded TAC systems in the presence of λ_{vir} at MOIs of 0, 0.1, 1, and 10. The shaded areas indicate the standard deviations.

(F) Activity of HigBAC, CmdTAC, and the hybrid TAC in antiphage immunity tested in Δ*clpX E. coli*.

two TAC systems, we tested the ability of non-cognate SecB^{TA}s—as well as that of housekeeping *E. coli* SecB—to neutralize the toxicity of HigBA and CmdTA when expressed in *trans* (Figure 5A). As expected, the non-cognate CmdC SecB^{TA} fails to counter the HigBA toxicity. Surprisingly, however, the HigC chaperone neutralizes the CmdTA unit efficiently, displaying better neutralizing activity than the cognate chaperone CmdC. Overexpression of the housekeeping *E. coli* SecB chaperone partially suppresses the toxicity of both HigBA and CmdTA. In the case of HigBA, the housekeeping *E. coli* SecB is more efficient than the non-cognate CmdC SecB^{TA} chaperone. We have observed analogous results when hybrid TAC systems were cloned as one operon (Figure S2B), suggesting the failure of CmdC to neutralize HigBA is not due to the disruption of co-translational folding of the TAC complex but reflects the intrinsic differences in the two SecB^{TA}s. Although AlphaFold predictions of CmdT in complex with either the cognate chaperone CmdC or the non-cognate HigC yielded models with good quality scores, the independent AlphaFold runs unfortunately failed to reveal a structural consensus that could help us rationalize the observed promiscuity (Figures S5C and S5D).

The ability of HigC to recognize and neutralize the non-cognate CmdTA makes it possible to directly test the

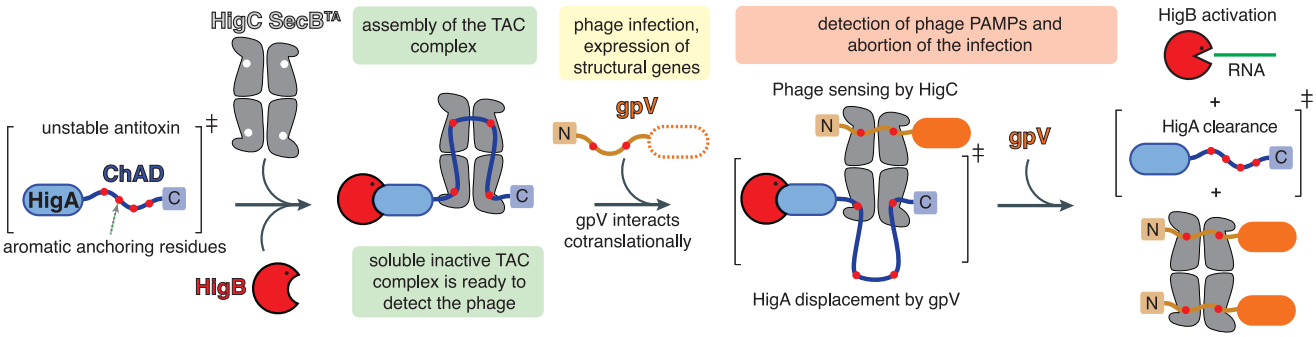


Figure 6. Sensing and abortion of phage infection by HigBAC defense system

In the absence of phage infection, the HigB RNase toxin, HigA antitoxin, and tetrameric $HigC_4$ chaperone form an inert complex. TAC complex formation stabilizes otherwise unstable HigA. Both HigA destabilization and HigA recognition are driven by the aromatic residues located in the ChAD region. Expression of the λ gpV protein is co-translationally recognized by the HigC chaperone, resulting in direct competition between the HigA and the major tail protein. Destabilization of HigA triggers HigB toxicity, which, in turn, restricts the phage propagation in the infected cell.

functions of the TA unit and the chaperone in sensing the phage infection. We cloned the hybrid TAC system comprised of CmdTA followed by the HigC chaperone for constitutive expression from a pBR322 derivative plasmid and performed an immunity screen with the BASEL phage collection. Our initial naive expectation was that the nature of the $SecB^{TA}$ chaperone (HigC in this case) would determine the spectrum of conferred defense, as the spectrum of the recognized phage triggers is defined by the nature of the chaperone. However, the result was strikingly different (Figures 5C and 5D). The hybrid TAC has a defense spectrum that (1) almost fully combines that of the parental systems, affording stronger protection than either of the parental systems for some phages (such as λ_{vir} and phages in the *Tevenvirinae* subfamily), (2) expands the spectrum of defense to all the three phage morphotypes, defending against viruses that were not protected against by either of the parental systems, such as additional myo- and siphoviruses (*Punavirus* $P1_{vir}$, Bas2 in the *Drexleviridae* family, and multiple phages in the *Demerecviridae* family, including T5) as well as podoviruses from the *Autographviridae* family (Bas66 and Bas67, weak protection). Importantly, the $\lambda_{vir}^{escape\_1}$ mutant variant efficiently overcomes the immunity mediated by the CmdTA_HigC system, suggesting that HigC-mediated phage-sensing operates similarly in hybrid TAC as it does in the native HigBAC (Figure 5C). Liquid culture λ_{vir} infection assays with increasing MOI (0.1, 1, and 10) are in good agreement with the plaquing experiments (Figure 5E). Although CmdTAC fails to provide any protection against the phage, the hybrid CmdTA-HigC system renders *E. coli* growth insensitive to λ_{vir}, even at an MOI of 10. The effect is reminiscent of the full protection granted by CmdTAC against Bas43 in liquid culture experiments (Figure 1G). The fact that, while sharing the chaperone, the hybrid system has a broader defense spectrum as compared with HigBAC suggests that either CmdTA has a broader spectrum of phages it is active against once triggered or that CmdTA could play a direct role in sensing the viral infection.

Finally, we tested the ability of HigBAC, CmdTAC, and the hybrid TAC to confer antiphage immunity in Δ*clpX E. coli* (Figure 5F). In good agreement with toxicity assays (Figures 3H and 3I), CmdTAC loses its defensive activity in Δ*clpX* background, while HigBAC is as active as in the wild-type *E. coli*. Similarly, the hybrid CmdTA-HigC system is functionally compromised in Δ*clpX E. coli*, which is to be expected as it shares the ChAD element with the ClpPX-dependent CmdTAC.

DISCUSSION

The CmdT TAC toxin from *E. coli* O112ab:H26 provides an example of an ART enzyme targeting mRNA to inhibit protein synthesis. Although ART effectors of type VI secretion systems are known to target RNA, these systems specifically modify structured RNA: *Photorhabdus laumondii* Rhs ADP-ribosylates the 23S rRNA Sarcin-Ricin Loop[50] and *Pseudomonas aeruginosa* RhsP2 targets diverse structured non-coding RNAs such as tRNA, tmRNA, 4.5S rRNA, and RNase P.[51] The first CmdTAC system to be discovered, PD-T4-9 from *E. coli* ECOR22,[21] was also recently shown to target mRNA, specifically modifying adenines in GA dinucleotide motifs within single-stranded RNAs (see the bioRxiv preprint by Vassalo and colleagues[52]), thus suggesting the generality of this mechanism.

This study has shed light on how the chaperone-mediated recognition of viral structures has been adopted by TAC defense systems to sense phage infection. Phage structural proteins provide pathogen-associated molecular patterns (PAMPs) that are recognized by diverse defense systems such as the $CapRel^{SJ46}$ toxSAS fused TA system, Avs STAND NTPases, HEPN domain-containing systems, the DSR2 NADase system, and the AbiT abortive infection system.[44,45,53–58] Here we show that HigBAC from *E. coli* strain NT1F31 senses phage λ through recognition of the major tail protein gpV by the HigC chaperone. We propose the following model for HigBAC triggering (Figure 6). In the absence of infection, the inactive TAC complex is co-translationally assembled, with the aromatic residues of the HigA ChAD nucleating ChAD recognition by the HigC $SecB^{TA}$ chaperone. Expression of gpV diverts the holdase activity of HigC, which leads to inactivation of the HigA antitoxin, unleashing the HigB RNase toxin that restricts the productive phage replication.

Phage replication and assembly commonly relies on both host and phage-encoded chaperones. GroEL/GroES and DnaK/DnaJ/GrpE chaperone systems were discovered as essential components for λ replication in *E. coli* through genetic screens selecting for the *gro* (inability to *gro*w the phage) phenotype.[59] GroEL is essential for folding and assembly of phage structural elements.[60] The dedicated phage-encoded chaperones gpG and gpGT are essential for the assembly of the λ phage tail from gpV subunits.[49,61] In the phage T4, a dedicated co-chaperonin gp31 substitutes for GroES to facilitate the folding of the major capsid protein gp23.[62,63] Thus, chaperones and their effects on modulating protein folding and aggregation are important facets of phage infection cycles, host:virus interactions, and—in the case of prophage-encoded defense systems such as TACs—virus:virus interactions.

Limitations of the study

This study lays the foundation for detailed studies of TAC-mediated antiphage defense. To uncover the molecular details, dedicated experimental structural studies are essential. Without these, understanding the molecular basis of $SecB^{TA}$ specificity toward phage triggers and TAC antitoxin ChAD elements will remain inherently limited. This study relies heavily on reconstitution of TAC sensing through co-expression of the identified phage trigger with the TAC system. Characterization of HigBAC TAC expression and assembly dynamics, along with its triggering by gpV during phage infection is essential for uncovering the co-translational nature of the $SecB^{TA}$-based sensory system. Dedicated substrate-specificity studies are needed to uncover the specific RNA species and RNA motifs targeted by HigB and CmdC TAC toxins. Finally, it is important to note that the phage defense experiments presented in this work relied on plasmid-based expression, warranting further studies of prophage-encoded TAC systems in their native genomic context.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Bacterial strains, plasmids, and phages
- METHOD DETAILS
  - Figure preparation
  - Bioinformatics
  - Construction of plasmids
  - Microbiological assays
  - *In vivo* functional assays
  - Biochemical assays
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.chom.2024.05.003.

ACKNOWLEDGMENTS

The AlphaFold2 computations were enabled by the supercomputing resource Berzelius provided by the National Supercomputer Centre (NSC) at Linköping University and the Knut and Alice Wallenberg foundation. Additional computational resources used were provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS) and the Swedish National Infrastructure for Computing at NSC, Chalmers University Centre for Computational Science and Engineering (C3SE), and PDC Centre for High Performance Computing, KTH Royal Institute of Technology, partially funded by the Swedish Research Council through grants 2018-05973 and 2022-06725. This work was supported by the Knut and Alice Wallenberg Foundation (project grant 2020-0037 to G.C.A. and V.H.), the Swedish Research Council (Vetenskapsrådet) grants (2019-01085, 2022-01603, and 2023-02353 to G.C.A. and 2021-01146 to V.H.), Crafoord Foundation (project grant Nr 20220562 to V.H.), the Estonian Research Council (PRG335 to V.H. and T.T.), Cancerfonden (20 0872 Pj to V.H.), the Fonds National de Recherche Scientifique (FNRS CDR J.0068.19 and J.0065.23F; FNRS-EQP UN.025.19; FNRS PDR T.0066.18 and PDR T.0090.22 to A.G.-P.), ERC (CoG DiStRes, n 864311 to A.G.-P.), Fonds Jean Brachet and the Fondation Van Buuren (A.G.-P.), FNRS-ASP (S.Z.C.), and Ambizione Fellowship PZ00P3_180085 and starting grant TMSGI3_211369 of the Swiss National Science Foundation (SNSF) (both A.H.). M.T.L. is an investigator of the Howard Hughes Medical Institute. J.A.B. and H.S. were funded by BBSRC grants BB/X003035/1 and BB/T017570/1. K.C.W. is a FRIA fellow. C.M. is a Research Associate of the FRS—FNRS supported by grant F.4532.22 from the FRS—FNRS.

AUTHOR CONTRIBUTIONS

V.H. coordinated the study and drafted the manuscript with contributions from all authors. T.M., T.K., K.E., M.J.O.J., S.Z.C., G.M.E., C.R.D., M.T.L., T.T., A.H., C.M., H.S., A.G.-P., G.C.A., and V.H. designed experiments and analyzed the data. T.M., T.K., K.E., M.J.O.J., S.Z.C., G.M.E., L.S., K.C.W., C.M., J.A.B., C.R.D., and T.B. performed experiments. R.O., J.A.N., A.A.E., A.G.-P., and G.C.A. performed bioinformatic analyses.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: February 28, 2024
Revised: April 10, 2024
Accepted: May 7, 2024
Published: May 30, 2024

REFERENCES

1. Jurėnas, D., Fraikin, N., Goormaghtigh, F., and Van Melderen, L. (2022). Biology and evolution of bacterial toxin-antitoxin systems. Nat. Rev. Microbiol. *20*, 335–350. https://doi.org/10.1038/s41579-021-00661-1.
2. LeRoux, M., and Laub, M.T. (2022). Toxin-Antitoxin Systems as Phage Defense Elements. Annu. Rev. Microbiol. *76*, 21–43. https://doi.org/10.1146/annurev-micro-020722-013730.
3. Song, S., and Wood, T.K. (2020). A Primary Physiological Role of Toxin/Antitoxin Systems Is Phage Inhibition. Front. Microbiol. *11*, 1895. https://doi.org/10.3389/fmicb.2020.01895.
4. Kelly, A., Arrowsmith, T.J., Went, S.C., and Blower, T.R. (2023). Toxin-antitoxin systems as mediators of phage defence and the implications for abortive infection. Curr. Opin. Microbiol. *73*, 102293. https://doi.org/10.1016/j.mib.2023.102293.
5. Tian, Q.B., Ohnishi, M., Tabuchi, A., and Terawaki, Y. (1996). A new plasmid-encoded proteic killer gene system: cloning, sequencing, and analyzing hig locus of plasmid Rts1. Biochem. Biophys. Res. Commun. *220*, 280–284. https://doi.org/10.1006/bbrc.1996.0396.
6. Schureck, M.A., Maehigashi, T., Miles, S.J., Marquez, J., Cho, S.E., Erdman, R., and Dunham, C.M. (2014). Structure of the *Proteus vulgaris*

HigB-(HigA)2-HigB toxin-antitoxin complex. J. Biol. Chem. *289*, 1060–1070. https://doi.org/10.1074/jbc.M113.512095.

7. Christensen-Dalsgaard, M., and Gerdes, K. (2006). Two *higBA* loci in the *Vibrio cholerae* superintegron encode mRNA cleaving enzymes and can stabilize plasmids. Mol. Microbiol. *62*, 397–411. https://doi.org/10.1111/j.1365-2958.2006.05385.x.

8. Hadži, S., Garcia-Pino, A., Haesaerts, S., Jurenas, D., Gerdes, K., Lah, J., and Loris, R. (2017). Ribosome-dependent *Vibrio cholerae* mRNAse HigB2 is regulated by a beta-strand sliding mechanism. Nucleic Acids Res. *45*, 4972–4983. https://doi.org/10.1093/nar/gkx138.

9. Jankevicius, G., Ariza, A., Ahel, M., and Ahel, I. (2016). The Toxin-Antitoxin System DarTG Catalyzes Reversible ADP-Ribosylation of DNA. Mol. Cell *64*, 1109–1116. https://doi.org/10.1016/j.molcel.2016.11.014.

10. LeRoux, M., Srikant, S., Teodoro, G.I.C., Zhang, T., Littlehale, M.L., Doron, S., Badiee, M., Leung, A.K.L., Sorek, R., and Laub, M.T. (2022). The DarTG toxin-antitoxin system provides phage defence by ADP-ribosylating viral DNA. Nat. Microbiol. *7*, 1028–1040. https://doi.org/10.1038/s41564-022-01153-5.

11. Suskiewicz, M.J., Prokhorova, E., Rack, J.G.M., and Ahel, I. (2023). ADP-ribosylation from molecular mechanisms to therapeutic implications. Cell *186*, 4475–4495. https://doi.org/10.1016/j.cell.2023.08.030.

12. Schuller, M., Raggiaschi, R., Mikolcevic, P., Rack, J.G.M., Ariza, A., Zhang, Y., Ledermann, R., Tang, C., Mikoc, A., and Ahel, I. (2023). Molecular basis for the reversible ADP-ribosylation of guanosine bases. Mol. Cell *83*, 2303–2315.e6. https://doi.org/10.1016/j.molcel.2023.06.013.

13. Schuller, M., Butler, R.E., Ariza, A., Tromans-Coia, C., Jankevicius, G., Claridge, T.D.W., Kendall, S.L., Goh, S., Stewart, G.R., and Ahel, I. (2021). Molecular basis for DarT ADP-ribosylation of a DNA base. Nature *596*, 597–602. https://doi.org/10.1038/s41586-021-03825-4.

14. Bordes, P., Cirinesi, A.M., Ummels, R., Sala, A., Sakr, S., Bitter, W., and Genevaux, P. (2011). SecB-like chaperone controls a toxin-antitoxin stress-responsive system in *Mycobacterium tuberculosis*. Proc. Natl. Acad. Sci. USA *108*, 8438–8443. https://doi.org/10.1073/pnas.1101189108.

15. Sala, A., Calderon, V., Bordes, P., and Genevaux, P. (2013). TAC from *Mycobacterium tuberculosis*: a paradigm for stress-responsive toxin-antitoxin systems controlled by SecB-like chaperones. Cell Stress Chaperones *18*, 129–135. https://doi.org/10.1007/s12192-012-0396-5.

16. Eismann, L., Fijalkowski, I., Galmozzi, C.V., Koubek, J., Tippmann, F., Van Damme, P., and Kramer, G. (2022). Selective ribosome profiling reveals a role for SecB in the co-translational inner membrane protein biogenesis. Cell Rep. *41*, 111776. https://doi.org/10.1016/j.celrep.2022.111776.

17. Hartl, F.U., Lecker, S., Schiebel, E., Hendrick, J.P., and Wickner, W. (1990). The binding cascade of SecB to SecA to SecY/E mediates preprotein targeting to the *E. coli* plasma membrane. Cell *63*, 269–279. https://doi.org/10.1016/0092-8674(90)90160-g.

18. Texier, P., Bordes, P., Nagpal, J., Sala, A.J., Mansour, M., Cirinesi, A.M., Xu, X., Dougan, D.A., and Genevaux, P. (2021). ClpXP-mediated Degradation of the TAC Antitoxin is Neutralized by the SecB-like Chaperone in *Mycobacterium tuberculosis*. J. Mol. Biol. *433*, 166815. https://doi.org/10.1016/j.jmb.2021.166815.

19. Bordes, P., Sala, A.J., Ayala, S., Texier, P., Slama, N., Cirinesi, A.M., Guillet, V., Mourey, L., and Genevaux, P. (2016). Chaperone addiction of toxin-antitoxin systems. Nat. Commun. *7*, 13339. https://doi.org/10.1038/ncomms13339.

20. Guillet, V., Bordes, P., Bon, C., Marcoux, J., Gervais, V., Sala, A.J., Dos Reis, S., Slama, N., Mares-Mejía, I., Cirinesi, A.M., et al. (2019). Structural insights into chaperone addiction of toxin-antitoxin systems. Nat. Commun. *10*, 782. https://doi.org/10.1038/s41467-019-08747-4.

21. Vassallo, C.N., Doering, C.R., Littlehale, M.L., Teodoro, G.I.C., and Laub, M.T. (2022). A functional selection reveals previously undetected anti-phage defence systems in the *E. coli* pangenome. Nat. Microbiol. *7*, 1568–1579. https://doi.org/10.1038/s41564-022-01219-4.

22. Ernits, K., Saha, C.K., Brodiazhenko, T., Chouhan, B., Shenoy, A., Buttress, J.A., Duque-Pedraza, J.J., Bojar, V., Nakamoto, J.A., Kurata, T., et al. (2023). The structural basis of hyperpromiscuity in a core combinatorial network of type II toxin-antitoxin and related phage defense systems. Proc. Natl. Acad. Sci. USA *120*, e2305393120. https://doi.org/10.1073/pnas.2305393120.

23. Maffei, E., Shaidullina, A., Burkolter, M., Heyer, Y., Estermann, F., Druelle, V., Sauer, P., Willi, L., Michaelis, S., Hilbi, H., et al. (2021). Systematic exploration of *Escherichia coli* phage-host interactions with the BASEL phage collection. PLoS Biol. *19*, e3001424. https://doi.org/10.1371/journal.pbio.3001424.

24. Saha, C.K., Sanches Pires, R., Brolin, H., Delannoy, M., and Atkinson, G.C. (2021). FlaGs and webFlaGs: discovering novel biology through the analysis of gene neighbourhood conservation. Bioinformatics *37*, 1312–1314. https://doi.org/10.1093/bioinformatics/btaa788.

25. Salgado, H., Gama-Castro, S., Peralta-Gil, M., Diaz-Peredo, E., Sanchez-Solano, F., Santos-Zavaleta, A., Martinez-Flores, I., Jimenez-Jacinto, V., Bonavides-Martinez, C., Segura-Salazar, J., et al. (2006). RegulonDB (version 5.0): *Escherichia coli* K-12 transcriptional regulatory network, operon organization, and growth conditions. Nucleic Acids Res. *34*, D394–D397. https://doi.org/10.1093/nar/gkj156.

26. Degnan, P.H., Michalowski, C.B., Babić, A.C., Cordes, M.H.J., and Little, J.W. (2007). Conservation and diversity in the immunity regions of wild phages with the immunity specificity of phage lambda. Mol. Microbiol. *64*, 232–244. https://doi.org/10.1111/j.1365-2958.2007.05650.x.

27. Snyder, L. (1995). Phage-exclusion enzymes: a bonanza of biochemical and cell biology reagents? Mol. Microbiol. *15*, 415–420. https://doi.org/10.1111/j.1365-2958.1995.tb02255.x.

28. Koubek, J., Schmitt, J., Galmozzi, C.V., and Kramer, G. (2021). Mechanisms of Cotranslational Protein Maturation in Bacteria. Front. Mol. Biosci. *8*, 689755. https://doi.org/10.3389/fmolb.2021.689755.

29. Wells, J.N., Bergendahl, L.T., and Marsh, J.A. (2016). Operon Gene Order Is Optimized for Ordered Protein Complex Assembly. Cell Rep. *14*, 679–685. https://doi.org/10.1016/j.celrep.2015.12.085.

30. Hurley, J.M., and Woychik, N.A. (2009). Bacterial toxin HigB associates with ribosomes and mediates translation-dependent mRNA cleavage at A-rich sites. J. Biol. Chem. *284*, 18605–18613. https://doi.org/10.1074/jbc.M109.008763.

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

32. Gapińska, M., Zajko, W., Skowronek, K., Figiel, M., Krawczyk, P.S., Egorov, A.A., Dziembowski, A., Johansson, M.J.O., and Nowotny, M. (2024). Structure-functional characterization of Lactococcus AbiA phage defense system. Nucleic Acids Res. *52*, 4723–4738. https://doi.org/10.1093/nar/gkae230.

33. Kutter, E., Bryan, D., Ray, G., Brewster, E., Blasdel, B., and Guttman, B. (2018). From Host to Phage Metabolism: Hot Tales of Phage T4's Takeover of *E. coli*. Viruses *10*, 387. https://doi.org/10.3390/v10070387.

34. Shimizu, Y., Kuruma, Y., Kanamori, T., and Ueda, T. (2014). The PURE system for protein production. Methods Mol. Biol. *1118*, 275–284. https://doi.org/10.1007/978-1-62703-782-2_19.

35. Takada, H., Roghanian, M., Murina, V., Dzhygyr, I., Murayama, R., Akanuma, G., Atkinson, G.C., Garcia-Pino, A., and Hauryliuk, V. (2020). The C-terminal RRM/ACT domain is crucial for fine-tuning the activation of 'long' RelA-SpoT Homolog enzymes by ribosomal complexes. Front. Microbiol. *11*, 277. https://doi.org/10.3389/fmicb.2020.00277.

36. Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Žídek, A., Bates, R., Blackwell, S., Yim, J., et al. (2022). Protein complex prediction with AlphaFold-Multimer. Preprint at bioRxiv. https://doi.org/10.1101/2021.10.04.463034.

37. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2.

38. Xu, Z., Knafels, J.D., and Yoshino, K. (2000). Crystal structure of the bacterial protein export chaperone secB. Nat. Struct. Biol. *7*, 1172–1177. https://doi.org/10.1038/82040.

39. Holehouse, A.S., and Kragelund, B.B. (2024). The molecular basis for cellular function of intrinsically disordered protein regions. Nat. Rev. Mol. Cell Biol. *25*, 187–211. https://doi.org/10.1038/s41580-023-00673-0.

40. Sormanni, P., and Vendruscolo, M. (2019). Protein Solubility Predictions Using the CamSol Method in the Study of Protein Homeostasis. Cold Spring Harb. Perspect. Biol. *11*, a033845. https://doi.org/10.1101/cshperspect.a033845.

41. Flynn, J.M., Neher, S.B., Kim, Y.I., Sauer, R.T., and Baker, T.A. (2003). Proteomic discovery of cellular substrates of the ClpXP protease reveals five classes of ClpX-recognition signals. Mol. Cell *11*, 671–683. https://doi.org/10.1016/s1097-2765(03)00060-1.

42. Fei, X., Bell, T.A., Barkow, S.R., Baker, T.A., and Sauer, R.T. (2020). Structural basis of ClpXP recognition and unfolding of *ssrA*-tagged substrates. eLife *9*, e61496. https://doi.org/10.7554/eLife.61496.

43. Shaner, N.C., Lambert, G.G., Chammas, A., Ni, Y., Cranfill, P.J., Baird, M.A., Sell, B.R., Allen, J.R., Day, R.N., Israelsson, M., et al. (2013). A bright monomeric green fluorescent protein derived from *Branchiostoma lanceolatum*. Nat. Methods *10*, 407–409. https://doi.org/10.1038/nmeth.2413.

44. Zhang, T., Tamman, H., Coppieters 't Wallant, K., Kurata, T., LeRoux, M., Srikant, S., Brodiazhenko, T., Cepauskas, A., Talavera, A., Martens, C., et al. (2022). Direct activation of a bacterial innate immune system by a viral capsid protein. Nature *612*, 132–140. https://doi.org/10.1038/s41586-022-05444-z.

45. Gao, L.A., Wilkinson, M.E., Strecker, J., Makarova, K.S., Macrae, R.K., Koonin, E.V., and Zhang, F. (2022). Prokaryotic innate immunity through pattern recognition of conserved viral proteins. Science *377*, eabm4096. https://doi.org/10.1126/science.abm4096.

46. Pell, L.G., Kanelis, V., Donaldson, L.W., Howell, P.L., and Davidson, A.R. (2009). The phage lambda major tail protein structure reveals a common evolution for long-tailed phages and the type VI bacterial secretion system. Proc. Natl. Acad. Sci. USA *106*, 4160–4165. https://doi.org/10.1073/pnas.0900044106.

47. Campbell, P.L., Duda, R.L., Nassur, J., Conway, J.F., and Huet, A. (2020). Mobile Loops and Electrostatic Interactions Maintain the Flexible Tail Tube of Bacteriophage Lambda. J. Mol. Biol. *432*, 384–395. https://doi.org/10.1016/j.jmb.2019.10.031.

48. Liu, X., Jiang, H., Gu, Z., and Roberts, J.W. (2013). High-resolution view of bacteriophage lambda gene expression by ribosome profiling. Proc. Natl. Acad. Sci. USA *110*, 11928–11933. https://doi.org/10.1073/pnas.1309739110.

49. Xu, J., Hendrix, R.W., and Duda, R.L. (2014). Chaperone-protein interactions that mediate assembly of the bacteriophage lambda tail to the correct length. J. Mol. Biol. *426*, 1004–1018. https://doi.org/10.1016/j.jmb.2013.06.040.

50. Jurėnas, D., Payelleville, A., Roghanian, M., Turnbull, K.J., Givaudan, A., Brillard, J., Hauryliuk, V., and Cascales, E. (2021). *Photorhabdus* antibacterial Rhs polymorphic toxin inhibits translation through ADP-ribosylation of 23S ribosomal RNA. Nucleic Acids Res. *49*, 8384–8395. https://doi.org/10.1093/nar/gkab608.

51. Bullen, N.P., Sychantha, D., Thang, S.S., Culviner, P.H., Rudzite, M., Ahmad, S., Shah, V.S., Filloux, A., Prehna, G., and Whitney, J.C. (2022). An ADP-ribosyltransferase toxin kills bacterial cells by modifying structured non-coding RNAs. Mol. Cell *82*, 3484–3498.e11. https://doi.org/10.1016/j.molcel.2022.08.015.

52. Vassallo, C.N., Doering, C.R., and Laub, M.T. (2024). Anti-viral defense by an ADP-ribosyltransferase that targets mRNA to block translation. Preprint at bioRxiv. https://doi.org/10.1101/2024.02.24.581662.

53. Stokar-Avihail, A., Fedorenko, T., Hör, J., Garb, J., Leavitt, A., Millman, A., Shulman, G., Wojtania, N., Melamed, S., Amitai, G., and Sorek, R. (2023). Discovery of phage determinants that confer sensitivity to bacterial immune systems. Cell *186*, 1863–1876.e16. https://doi.org/10.1016/j.cell.2023.02.029.

54. Schmitt, C.K., Kemp, P., and Molineux, I.J. (1991). Genes 1.2 and 10 of bacteriophages T3 and T7 determine the permeability lesions observed in infected cells of *Escherichia coli* expressing the F plasmid gene *pifA*. J. Bacteriol. *173*, 6507–6514. https://doi.org/10.1128/jb.173.20.6507-6514.1991.

55. Labrie, S.J., Tremblay, D.M., Moisan, M., Villion, M., Magadán, A.H., Campanacci, V., Cambillau, C., and Moineau, S. (2012). Involvement of the major capsid protein and two early-expressed phage genes in the activity of the lactococcal abortive infection mechanism AbiT. Appl. Environ. Microbiol. *78*, 6890–6899. https://doi.org/10.1128/AEM.01755-12.

56. Huiting, E., Cao, X., Ren, J., Athukoralage, J.S., Luo, Z., Silas, S., An, N., Carion, H., Zhou, Y., Fraser, J.S., et al. (2023). Bacteriophages inhibit and evade cGAS-like immune function in bacteria. Cell *186*, 864–876.e21. https://doi.org/10.1016/j.cell.2022.12.041.

57. Yin, H., Li, X., Wang, X., Zhang, C., Gao, J., Yu, G., He, Q., Yang, J., Liu, X., Wei, Y., et al. (2024). Insights into the modulation of bacterial NADase activity by phage proteins. Nat. Commun. *15*, 2692. https://doi.org/10.1038/s41467-024-47030-z.

58. Rostøl, J.T., Quiles-Puchalt, N., Iturbe-Sanz, P., Lasa, Í., and Penadés, J.R. (2024). Bacteriophages avoid autoimmunity from cognate immune systems as an intrinsic part of their life cycles. Nat. Microbiol. *9*, 1312–1324. https://doi.org/10.1038/s41564-024-01661-6.

59. Georgopoulos, C. (2006). Toothpicks, serendipity and the emergence of the *Escherichia coli* DnaK (Hsp70) and GroEL (Hsp60) chaperone machines. Genetics *174*, 1699–1707. https://doi.org/10.1534/genetics.104.68262.

60. Zeilstra-Ryalls, J., Fayet, O., and Georgopoulos, C. (1991). The universally conserved GroE (Hsp60) chaperonins. Annu. Rev. Microbiol. *45*, 301–325. https://doi.org/10.1146/annurev.mi.45.100191.001505.

61. Xu, J., Hendrix, R.W., and Duda, R.L. (2004). Conserved translational frameshift in dsDNA bacteriophage tail assembly genes. Mol. Cell *16*, 11–21. https://doi.org/10.1016/j.molcel.2004.09.006.

62. van der Vies, S.M., Gatenby, A.A., and Georgopoulos, C. (1994). Bacteriophage T4 encodes a co-chaperonin that can substitute for *Escherichia coli* GroES in protein folding. Nature *368*, 654–656. https://doi.org/10.1038/368654a0.

63. Takano, T., and Kakefuda, T. (1972). Involvement of a bacterial factor in morphogenesis of bacteriophage capsid. Nat. New Biol. *239*, 34–37. https://doi.org/10.1038/newbio239034a0.

64. Grenier, F., Matteau, D., Baby, V., and Rodrigue, S. (2014). Complete Genome Sequence of Escherichia coli BW25113. Genome Announc. *2*, e01038-14. https://doi.org/10.1128/genomeA.01038-14.

65. Song, Y., DiMaio, F., Wang, R.Y.R., Kim, D., Miles, C., Brunette, T., Thompson, J., and Baker, D. (2013). High-resolution comparative modeling with RosettaCM. Structure *21*, 1735–1742. https://doi.org/10.1016/j.str.2013.08.005.

66. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780. https://doi.org/10.1093/molbev/mst010.

67. Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M., and Barton, G.J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics *25*, 1189–1191. https://doi.org/10.1093/bioinformatics/btp033.

68. Larsson, A. (2014). AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics *30*, 3276–3278. https://doi.org/10.1093/bioinformatics/btu531.

69. Goddard, T.D., Huang, C.C., Meng, E.C., Pettersen, E.F., Couch, G.S., Morris, J.H., and Ferrin, T.E. (2018). UCSF ChimeraX: Meeting modern challenges in visualization and analysis. Protein Sci. *27*, 14–25. https://doi.org/10.1002/pro.3235.

70. Arndt, D., Grant, J.R., Marcu, A., Sajed, T., Pon, A., Liang, Y., and Wishart, D.S. (2016). PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res. *44*, W16–W21. https://doi.org/10.1093/nar/gkw387.

71. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676–682. https://doi.org/10.1038/nmeth.2019.

72. Sormanni, P., Aprile, F.A., and Vendruscolo, M. (2015). The CamSol method of rational design of protein mutants with enhanced solubility. J. Mol. Biol. *427*, 478–490. https://doi.org/10.1016/j.jmb.2014.09.026.

73. Perez-Riverol, Y., Bai, J., Bandla, C., García-Seisdedos, D., Hewapathirana, S., Kamatchinathan, S., Kundu, D.J., Prakash, A., Frericks-Zipper, A., Eisenacher, M., et al. (2022). The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. Nucleic Acids Res. *50*, D543–D552. https://doi.org/10.1093/nar/gkab1038.

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

75. Xu, J., and Zhang, Y. (2010). How significant is a protein structure similarity with TM-score = 0.5? Bioinformatics *26*, 889–895. https://doi.org/10.1093/bioinformatics/btq066.

76. Hauser, M., Steinegger, M., and Söding, J. (2016). MMseqs software suite for fast and deep clustering and searching of large protein sequence sets. Bioinformatics *32*, 1323–1330. https://doi.org/10.1093/bioinformatics/btw006.

77. Gu, Z. (2022). Complex heatmap visualization. iMeta *1*, e43. https://doi.org/10.1002/imt2.43.

78. Schoch, C.L., Ciufo, S., Domrachev, M., Hotton, C.L., Kannan, S., Khovanskaya, R., Leipe, D., McVeigh, R., O'Neill, K., Robbertse, B., et al. (2020). NCBI Taxonomy: a comprehensive update on curation, resources and tools. Database (Oxford) *2020*, baaa062. https://doi.org/10.1093/database/baaa062.

79. Turner, D., Shkoporov, A.N., Lood, C., Millard, A.D., Dutilh, B.E., Alfenas-Zerbini, P., van Zyl, L.J., Aziz, R.K., Oksanen, H.M., Poranen, M.M., et al. (2023). Abolishment of morphology-based taxa and change to binomial species names: 2022 taxonomy update of the ICTV bacterial viruses subcommittee. Arch. Virol. *168*, 74. https://doi.org/10.1007/s00705-022-05694-2.

80. Quan, J., and Tian, J. (2009). Circular polymerase extension cloning of complex gene libraries and pathways. PLoS One *4*, e6441. https://doi.org/10.1371/journal.pone.0006441.

81. Quan, J., and Tian, J. (2011). Circular polymerase extension cloning for high-throughput cloning of complex and combinatorial DNA libraries. Nat. Protoc. *6*, 242–251. https://doi.org/10.1038/nprot.2010.181.

82. Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., 3rd, and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods *6*, 343–345. https://doi.org/10.1038/nmeth.1318.

83. Naville, M., Ghuillot-Gaudeffroy, A., Marchais, A., and Gautheret, D. (2011). ARNold: a web tool for the prediction of Rho-independent transcription terminators. RNA Biol. *8*, 11–13. https://doi.org/10.4161/rna.8.1.13346.

84. Neidhardt, F.C., Bloch, P.L., and Smith, D.F. (1974). Culture medium for enterobacteria. J. Bacteriol. *119*, 736–747. https://doi.org/10.1128/jb.119.3.736-747.1974.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial strains		
E. coli K-12 BW25113	Grenier et al.[64]	N/A
E. coli DH5α	Laboratory stock	N/A
For other *E. coli* strains see Dataset S1	N/A	N/A
Bacteriophages		
For bacteriophages used see Dataset S1	N/A	N/A
Chemicals, Peptides, and Recombinant Proteins		
Phusion High-Fidelity PCR Master Mix with HF Buffer	Thermo Scientific	Cat#F531L
Phusion High-Fidelity DNA Polymerase	Thermo Scientific	Cat#F530L
T4 PNK	Thermo Scientific	Cat#EK0031
Dpn I	Thermo Scientific	Cat#FD1703
T4 DNA Ligase	Thermo Scientific	Cat#EL0011
Zymoclean Gel DNA Recovery Kit	Zymo Research	Cat#D4008
DNA Clean & Concentrator-5	Zymo Research	Cat#D4014
FavorPrep Plasmid Extraction Mini Kit	Favorgen	Cat#FAPDE 300
Phage DNA Isolation Kit	Norgen Biotek Corp.	Cat#46850
NEBuilder® HiFi DNA Assembly Master Mix	New England Biolabs	Cat#E2621L
L-[^{35}S]-Methionine	PerkinElmer	Cat#NEG009C005MC
[5,6-^{3}H]-Uridine	PerkinElmer	Cat#NET367250UC
[Methyl-^{3}H]-Thymidine	PerkinElmer	Cat#NET027W001MC
PURExpress In Vitro Protein Synthesis Kit	New England Biolabs	Cat#E6800
RNase Inhibitor Murine	New England Biolabs	Cat#M0314S
Anti-His Tag Antibody (Monoclonal, 9C11)	Boster Biological Technology	Cat#M30975; RRID: AB_3099525
Anti-RNA polymerase beta antibody [EPR18704]	abcam	Cat#ab191598; RRID: AB_257896
Goat anti-Mouse IgG (H&L), HRP conjugated	Agrisera AB	Cat#AS11 1772; RRID: AB_3099526
Anti-Rabbit IgG (whole molecule)–Peroxidase antibody produced in goat	Sigma-Aldrich	Cat#A0545; RRID; AB_257896
HRP-conjugated streptavidin	Thermo Scientific	Cat#N100
Carbenicillin	Gold Bio	Cat#C-103
Pierce Anti-DYKDDDDK Magnetic Agarose	Thermo Fisher	Cat#A36797
cOmplete, Mini Protease Inhibitor Cocktail	Millipore Sigma	Cat#11836153001
Triton X-100	Millipore Sigma	Cat#X100-1L
Benzonase nuclease, Purity >90%	Millipore Sigma	Cat#70746-3
Ready-Lyse lysozyme	Biosearch Technologies	Cat#R1810M
10X PBS Buffer	Thermo Fisher	Cat#AM9625
Sodium Chloride	Millipore Sigma	Cat#S5886-10KG
0.5M EDTA, pH 8.0	Boston BioProducts	Cat#BM-150
Trizma Base	Millipore Sigma	Cat#T1503-1KG
WesternBright Quantum	Advansta	Cat#K-12042-D10
Oligonucleotides		
For primers used for cloning of *E. coli* plasmids see Dataset S1	N/A	N/A
5'rGrGrCrArArGrGrArGrGrUrArArArArArUrGrGrUrUrArArA3'	Sigma	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
5'rUrUrUrArArCrCrArUrUrUrUrUr ArCrCrUrCrCrUrUrGrCrC3'	Integrated DNA Technologies	N/A
5'dGdGdCdAdAdGdGdAdGdGdT dAdAdAdAdAdTdGdGdTdTdAdAdA3'	Sigma	N/A
5'rArArArArArArArArArArArArA3'	Integrated DNA Technologies	N/A
5'rUrUrUrUrUrUrUrUrUrUrUrU3'	Integrated DNA Technologies	N/A
5'rCrCrCrCrCrCrCrCrCrCrCrC3'	Integrated DNA Technologies	N/A
5'rGrCrGrCrCrGrCrGrCrGrGrCrG3'	Integrated DNA Technologies	N/A
5'rCrGrCrCrCrGrCrGrCrGrGrCrC3'	Integrated DNA Technologies	N/A
5'rArUrArUrUrArUrArUrArArArUrArArUr ArUrUrArUrArUrArArArUrArUrArUr UrUrArUrArUrArArArUrA3'	Integrated DNA Technologies	N/A
5'rUrArUrUrUrArUrArUrArArUrArUrUrAr UrUrUrArUrArUrArArUrArUrUrUr UrArUrArUrArArUrArU3'	Integrated DNA Technologies	N/A
5'dAdAdAdAdAdAdAdAdAdAdAdAdA3'	Integrated DNA Technologies	N/A
5'dTdTdTdTdTdTdTdTdTdTdTdT3'	Integrated DNA Technologies	N/A
5'dCdCdCdCdCdCdCdCdCdCdCdC3'	Integrated DNA Technologies	N/A
5'dGdCdGdCdCdGdCdGdCdGdGdCdG3'	Integrated DNA Technologies	N/A
5'dCdGdCdCdCdGdCdGdCdGdGdCdC3'	Integrated DNA Technologies	N/A
5'dAdTdAdTdTdAdTdAdTdAdAdTdAdAdT dAdTdTdAdTdAdTdAdAdAdTdAdAdTdAdT dTdAdTdAdTdAdAdAdTdA3'	Integrated DNA Technologies	N/A
5'dTdAdTdTdTdAdTdAdTdAdAdTdAdTdTd AdTdTdTdAdTdAdTdAdAdTdAdTdTdAd TdTdTdAdTdAdTdAdAdTdAdT3'	Integrated DNA Technologies	N/A
Deposited data		
IP-MS mass spectrometry	This study	MassIVE: doi: https://doi.org/10.25345/C5T14V10M
HDX-MS proteomics	This study	PRIDE: PXD050182
AlphaFold2 predicted structures	This study	Github: https://github.com/GCA-VH-lab/tac_ms_af
Phage annotation table, phage clustering results and the R code	This study	Figshare: https://doi.org/10.6084/m9.figshare.24968247
Recombinant DNA		
For *E. coli* plasmids see Dataset S1	This work	N/A
Software and Algorithms		
SnapGene	GSL Biotech LLC	https://www.snapgene.com/
Geneious Prime	Biomatters	https://www.geneious.com/
AlphaFold2	Song et al.[65]	N/A
MAFFT	Katoh et al.[66]	RRID:SCR_011811
Jalview	Waterhouse et al.[67]	RRID:SCR_006459
Aliview	Larsson[68]	N/A
UCSF ChimeraX	Goddard et al.[69]	RRID:SCR_015872
PHASTER	Arndt et al.[70]	N/A
FlaGs2	Saha et al.[24]	N/A
GraphPad Prism 10.1.1	GraphPad Software	https://www.graphpad.com/features
Igor Pro 7.07	WaveMetrics, Inc.	https://www.wavemetrics.com
Adobe Illustrator 28.0	Adobe Inc.	https://www.adobe.com/products/illustrator.html
Adobe Photoshop 28.0	Adobe Inc.	https://www.adobe.com/products/photoshop.html

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
NanoAnalyze	TA Instruments	https://www.tainstruments.com/itcrun-dscrun-nanoanalyze-software/
NIS Elements AR	Nikon	https://www.microscope.healthcare.nikon.com/products/software/nis-elements/nis-elements-advanced-research
Origin	OriginLab	https://www.originlab.com/
Fiji	Schindelin et al.[71]	RRID:SCR_002285
CamSol	Sormanni et al.[72]	N/A
Other		
Zeta-Probe® Blotting Membranes	Bio-Rad	Cat#1620153
Trans-Blot® Turbo™ Transfer System	Bio-Rad	Cat#1704150
Amersham™ ImageQuant 800	Cytiva	Cat#29399484
5910 Ri centrifuge	Eppendorf	Cat#5943000061
S-4x universal rotor	Eppendorf	Cat#5895200001
5418 R Centrifuge	Eppendorf	Cat#5401000064
FA-45-18-11 rotor	Eppendorf	Cat#5418707005
Optima XPN-80 Ultracentrifuge	Beckman Coulter	Cat#A95765
SW-41Ti rotor	Beckman Coulter	Cat#331336

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to the lead contact, Vasili Hauryliuk (vasili.hauryliuk@med.lu.se).

Materials availability

Plasmids, bacterial strains, and phages used in this study will be provided upon request.

Data and code availability

- Raw data IP-MS mass spectrometry data to are deposited at the MassIVE database and the HDX-MS proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE[73] partner repository. AlphaFold2 predicted structures together with the accompanying quality metric scores were deposited at Github. The phage annotation table, results of phage protein clustering including heatmaps, and the R script used for clustering are available at figshare. The accession numbers for the datasets are listed in the key resources table.
- The study does not make use of unpublished data or software.
- Any additional information required to reanalyze the data reported in this work paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Bacterial strains, plasmids, and phages

Bacterial strains, bacteriophages, plasmids as well as oligonucleotide primers used in the study are listed in Dataset S1.

Bacteria

E. coli strains were grown at 37 °C on solid (1.5% w/v agar) or in liquid LB medium. To select for plasmids, the medium was when needed supplemented with 25 μg/mL chloramphenicol, 100 μg/mL carbenicillin/ampicillin, 50 μg/ml kanamycin, and/or 20 μg/mL gentamicin. The expression from inducible promoters were controlled by supplementing the medium with appropriate concentrations of glucose/arabinose and/or IPTG. For phage infection of liquid cultures, the medium was supplemented with 10 mM $MgSO_4$ and 2.5 mM $CaCl_2$.

Bacteriophages

Phages were propagated using the double-agar overlay method using *E. coli* BW25113 strain as the host. Top agar was prepared as LB agar (0.5% w/v agar) supplemented with 20 mM $MgSO_4$, and 5 mM $CaCl_2$. Phage stocks were prepared from the top agar layer and stored in SM buffer (0.1 M NaCl, 10 mM $MgSO_4$, and 0.05 M Tris-HCl pH 7.5) at 4 °C.

METHOD DETAILS

Figure preparation

Figures were prepared using UCSF ChimeraX 1.6.1, GraphPad Prism 10.1.1 (GraphPad Software), Igor Pro 7.07 (WaveMetrics, Inc.), Adobe Illustrator 28.0 (Adobe Inc.) and Adobe Photoshop 28.0 (Adobe Inc.), CamSol[40] and Fiji.[71]

Bioinformatics

Genomic neighbourhood search and multiple sequence alignment

Sequence searching for SecB homologues was carried with PSI-BLAST[74] against the RefSeq database, limiting by taxonomy to *Enterobacteriaceae* and an E value threshold of 0.01. The query was *Lactiplantibacillus garii* accession WP_125072952, identified with NetFlax as a SecB-like protein associated with a MqsRA-like TA.[22] Eight PSI-Blast iterations were carried out, with all identified hits going through to each round, after which no new hits were identified. All 558 hits were used as queries for FlaGs2[22] (https://github.com/GCA-VH-lab/FlaGs2) with default settings to produce (Dataset S1). Chaperone and Antitoxin sequences were aligned using MAFFT L-INS-i.[66] Alignments were visualized with AliView[68] and Jalview.[67] Prophage boundaries were predicted with the PHASTER[70] server.

Structural modeling

Protein structure predictions were made with AlphaFold2 v2.3.1[37] and complexes were predicted with the AlphaFold-Multimer[36] protocol. The structural template cut-off date was set to May 14, 2020 (–max_template_date=2020-05-14). As per Evans and colleagues,[36] the quality of multimeric predictions was assessed using a weighted sum of predicted template-modelling (pTM) scores [pTM (scoring all residues of the model) and ipTM (scoring only the interfacing residues)]: 0.2*pTM+0.8*ipTM, referred to as simply pTM+ipTM. Models with scores $\geq$0.5 are likely have the correct overall fold,[37,75] and we considered models with scores $\geq$0.7 to be of good quality.

Phage clustering based on proteome similarity

BASEL phages were clustered through hierarchical clustering using pairwise proteome similarity scores as described previously.[32] All proteins from the BASEL collection of phages were clustered using MMseqs[76] with optional parameters (*–cluster-mode 1 –cov-mode 0 -c 0.7 –min-seq-id 0.3*), with each protein being assigned to a cluster corresponding to a set of homologues. To calculate the similarity scores between the i^{th} and j^{th} phages (*sim(i,j)*) the number of overlapped homologues was normalized to the size of the i^{th} proteome. A symmetric proteome composition distance (PCD, *dist(i,j)*) matrix was calculated with the formula *dist(i,j) = 1 – (average (sim(i,j), sim(j,i)))*. The clustering was performed in R using the average-linkage method and data were visualized using the *ComplexHeatmap* library.[77] Phages were annotated with family, subfamily and genus using taxonomic information extracted from the NCBI Taxonomy database[78] (January 2024 version) and ICTV[79] (2022 update).

Construction of plasmids

SnapGene (GSL Biotech LLC) and Geneious Prime (Biomatters) software were employed for primer and plasmid map design. All plasmids were constructed with either circular polymerase extension cloning (CPEC)[80,81] with Phusion polymerase (Thermo Scientific), Gibson assembly,[82] or ligation of DNA fragments with T4 DNA Ligase (Thermo Scientific). The following TAC genes were used: CmdT WP_097333619.1, CmdA WP_053886482.1, CmdC WP_053886483.1, HigB WP_112844290.1, HigA WP_112844289.1 and HigC WP_112844288.1. pBAD33 plasmid (arabinose inducible P_{BAD} promoter) was used to express toxins, TAs and TACs and their variants, pAS004, a pBAD33 derivative, was used to express λ_{vir} gpV and its variants, and pMG25 (IPTG inducible $P_{A1/O4/O3}$ promoter) vector was used for expression of chaperones and antitoxins. The HigBA, HigBAC, and their mutated version fragments were inserted into the pBAD33 plasmid with a weak Shine-Dalgarno motif (ATTAGAAGAATAAG) from CmdTAC. pAS004, pMG25, CmdTAC pBAD33 derivates as well as pBAD33-SD-HigB were created with a strong SD (AGGAGGAATTAA). For immunity assays, TAC regions (coding sequences with 150 bp upstream for HigBAC region, and 118 bp upstream for CmdTAC region and native terminators for both predicted by ARNold[83]) were cloned into the pBR322 vector under the control of the constitutive P_{tet} promoter. For metabolic labelling assays the TAC toxin genes were cloned into pBAD33, either without the Shine-Dalgarno sequence (the more toxic HigB, VHp1718) or with a strong Shine-Dalgarno sequence AGGAGGAATTAA (the less toxic CmdT, VHp1140). pBR322-HigBA/FLAG-C construct used for pulldown experiments was created by site-directed mutagenesis using pBR322-HigBAC as a template. Detailed cloning schemes can be found in Dataset S1. All constructs were verified by sequencing.

Microbiological assays

TA and TAC toxicity assays

The TA regions of *E. coli* NT1F31 HigBAC system and *E. coli* O112ab:H26 CmdTAC system were expressed from pBAD33. HigBA and its derivatives were expressed with a weak Shine-Dalgarno motif from CmdTAC (ATTAGAAGAATAAG), while CmdTA with a strong Shine-Dalgarno sequence (AGGAGGAATTAA). Chaperones HigC, CmdC, and *E. coli* BW25113 housekeeping chaperone SecB were expressed with a strong SD from pMG25 plasmid. Plasmids were transformed into wild-type *E. coli* BW25113 or protease deletion strains. Bacterial cultures from single colonies were grown for five to six hours, adjusted to OD_{600} 1.0, serially diluted (10^1- to 10^8-fold) and spotted on LB medium supplemented with 100 μg/mL carbenicillin (Fisher Bioreagents), 25 μg/mL chloramphenicol (AppliChem) and inducers (0.2% arabinose for TA induction, 50 μM or 500 μM IPTG for chaperone induction). Plates were scored after overnight incubation at 37 °C. For *in cis* TAC neutralization assays the CDSs of HigBAC or CmdTAC were expressed from

pBAD33, with a weak SD for HigBAC and a strong SD for CmdTAC. Plasmids expressing the TA or TAC region were transformed into *E. coli* BW25113. Bacterial cultures were started from single colonies, grown for five to six hours, adjusted to OD_{600} 1.0, serially diluted (from 10^1- to 10^8-fold) and spotted on LB medium supplemented with 25 μg/mL chloramphenicol (AppliChem) and 0.2% arabinose for induction. Plates were incubated overnight at 37 °C and scored. All experiments were performed at least three times, representative images are shown.

TAC triggering assays

The effects of co-expression of *E. coli* NT1F31 HigBAC (pBR322 with constitutive P_{tet}-driven expression) and λ_{vir} gpV variants (pAS004 with arabinose-inducible P_{BAD}-driven expression) were tested on a spot assay. The plasmids were transformed into *E. coli* BW25113 and bacterial cultures from single colonies were grown for five to six hours. Optical density was adjusted to OD_{600} 1.0, serially diluted, and spotted on LB medium supplemented with 100 μg/mL carbenicillin (Fisher Bioreagents), 20 μg/mL gentamicin (Sigma-Aldrich) and 0.2% arabinose for gpV induction. Plates were scored after an overnight incubation at 37 °C. All experiments were performed at least three times, representative images are shown.

Experimental phage infections

To assess the activity of the TAC systems in phage defence, we performed efficiency of plating assays, essentially as described previously,[22] using the BASEL collection[23] and a subset of common laboratory phages (Dataset S1). Briefly, overnight cultures of *E. coli* BW25113 cells carrying either an empty vector (pBR322 derivative lacking the tetracycline resistance cassette[31]) or a TAC system (pBR322-P_{tet}-HigBAC (VHp1259), pBR322-P_{tet}-CmdTAC (VHp1256), or pBR322-P_{tet}-CmdTA-HigC (VHp1608)) were mixed with top agar (LB with 0.5% agar, 20 mM $MgSO_4$, and 5mM $CaCl_2$), to a final concentration of 0.075 OD_{600} units/ml, and overlayed on LB-agar plates (1.5% agar). Individual phage stocks were 10-fold serially diluted in SM buffer (0.1 M NaCl, 10 mM $MgSO_4$, and 0.05 M Tris-HCl pH 7.5) and 2.5 μL of each of eight dilutions spotted on solidified top agar plates. Plaque formation was monitored after 6 and 24h of incubation at 37 °C. The phages that showed sensitivity to at least one system were re-tested towards all three systems using three different transformants for each plasmid. Plaques were counted after 24 h and the efficiency of plaquing (EOP) was determined for each repeat by dividing the plaque forming units for a given TAC system by that for the vector control. The average EOP value from the three repeats was -$\log_{10}$ transformed, yielding the $\log_{10}$ protection value used in heatmaps. All experiments were performed at least three times, representative images are shown.

Phage infection of liquid cultures were performed essentially as described previously.[32] Three different transformants of *E. coli* BW25113 harbouring either an empty vector or a TAC-containing plasmid were grown overnight in LB medium supplemented with ampicillin, 10 mM $MgSO_4$, and 2.5 mM $CaCl_2$. The cells were diluted to $OD_{600} \approx 0.075$ in the same medium and 100 μL were added to wells of a 96-well plate. The relevant phages were diluted in SM buffer to generate final MOI values of 10, 1, and 0.1 when 10 μL of the dilution was added to a well. Ten μL of SM buffer were added to control wells. The growth was monitored at 37 °C in a Synergy H1 (BioTek) plate reader measuring OD_{600} every 15 min. The growth curves represent the mean of three replicates (n = 3), using different transformants in each replicate, and the shaded areas indicate the standard deviation.

Isolation of bacteriophage escape mutants

To identify bacteriophage escape mutants, 200 μl of overnight culture of *E. coli* BW25113 strain carrying pBR322-P_{tet}-HigBAC (VHp1259) plasmid was infected with 20 μl of different λ_{vir} stock dilutions, mixed with top agar and poured on LB agar plate. After incubation at 37 °C for 24h, plaques with normal morphology (as seen on control culture without the immunity system) were isolated with a sterile toothpick and re-streaked three times using BW25113 carrying pBR322-P_{tet}-HigBAC plasmid as the host. High-titer stocks were prepared from potential escape mutants as described previously.[23] Norgen Biotek Phage DNA Isolation Kit (Norgen Biotek Corp.) was used to isolate genomic DNA of bacteriophages. The DNA was sequenced at the Microbial Genome Sequencing Center and results analyzed using Geneious Prime (Biomatters).

In vivo functional assays

Metabolic labeling with ^{35}S-methionine, ^{3}H-uridine, and ^{3}H-thymidine

For metabolic labelling experiments the *E. coli* BW25113 strain was co-transformed with the pBAD33 plasmid carrying the toxin gene of interest (CmdT:VHp1140 and HigB:VHp1718) for L-arabinose-inducible expression and the empty pMG25 vector. Transformed cells were initially plated on LB plates supplemented with 100 μg/mL carbenicillin, 25 μg/mL chloramphenicol, and 0.2% glucose (to suppress leaky toxin expression). Using individual *E. coli* colonies for inoculation, 2 mL liquid cultures were prepared in defined Neidhardt MOPS minimal media,[84] supplemented with 100 μg/mL carbenicillin, 25 μg/mL chloramphenicol, 0.1% casamino acids, and 0.2% glucose, and grown overnight at 37 °C with shaking. Subsequently, experimental 15 mL cultures were prepared in 125 mL conical flasks in MOPS medium, supplemented with 0.5% glycerol, 100 μg/mL carbenicillin, 25 μg/mL chloramphenicol, as well as a set of 19 amino acids (lacking methionine), each at a final concentration of 25 μg/mL. These cultures were grown overnight, back-diluted to a final OD_{600} of 0.05, and grown at 37 °C with shaking until the OD_{600} reached 0.2. At this point, 1 mL aliquots (designated as the pre-induction zero time-point) were transferred to 1.5 mL Eppendorf tubes containing 10 μL of the respective radioisotope (^{35}S methionine - 4.35 μCi, Hartman; ^{3}H uridine - 0.65 μCi, Hartman; or ^{3}H thymidine - 2 μCi, Hartman) and placed in a heat block at 37 °C. Toxin expression in the remaining 14 mL culture was induced by adding L-arabinose to a final concentration of 0.2%. Throughout the toxin induction time course, 1 mL aliquots were taken from the 15 mL culture and transferred to 1.5 mL Eppendorf tubes containing 10 μL of the appropriate radioisotope (^{35}S methionine, ^{3}H uridine, or ^{3}H thymidine). Radioisotope incorporation was halted after 8 minutes of incubation at 37 °C by adding 200 μL of ice-cold 50% trichloroacetic acid (TCA) to the 1 mL cultures. Additionally, 1 mL aliquots were periodically sampled for OD_{600} measurements. The resultant 1.2 mL culture/TCA samples were loaded onto GF/C

filters (Whatman) prewashed with 5% TCA and unincorporated label was removed by washing the filter twice with 5 mL of ice-cold TCA followed by a 5 mL wash with 95% EtOH (twice). The filters were placed in scintillation vials, dried for at least two hours at room temperature, followed by the addition of EcoLite™-scintillation cocktail (5 mL per vial; MP Biomedicals). After shaking for 15 minutes, radioactivity was quantified using Tri-Carb 4910TR-scintillation counter (Perkin Elmer). Isotope incorporation was quantified by normalizing radioactivity counts (CPM) to OD_{600}, with the pre-induction zero time-point serving as the reference (set to 100%). All experiments were conducted in triplicates (n = 3), using three independent cultures initiated from distinct colonies, the data is shown as geometric means with standard deviations.

Immunoblotting

E. coli BW25113 strain co-transformed with pBAD33 plasmid carrying either the N-terminally His_6 tagged *higA* gene (VHp1720) or its tetra-substituted version [4*: *W128A F158A F188A Y214A*] (VHp1722) for L-arabinose-inducible expression, together with either empty pMG25 vector (VHp1069) or its derivative for IPTG-inducible expression of HigC (VHp1618). Experimental 20 mL cultures (LB media supplemented with 100 μg/mL carbenicillin, 25 μg/mL chloramphenicol and 50 μM IPTG) were inoculated to an OD_{600} of 0.05 and grown at 37 °C with shaking until reaching the OD_{600} of 0.5. Expression of $_{His6}$HigA was induced by addition of L-arabinose to a final concentration of 0.2%, 1 mL samples were collected after 1 hour, the cells pelleted by centrifugation, dissolved in 1x SDS-PAGE sample buffer (200 μL for 1 mL of OD_{600} 1.0 culture), denatured at 95 °C for 5 min and 10 μL of the final sample were resolved on SDS-PAGE (12% acrylamide/bis-acrylamide 37.5:1). Proteins were transferred to BioTrace™ NT nitrocellulose membranes (Pall Life Sciences) using the Trans-Blot® Turbo™ Transfer System (Bio-Rad), and the membranes were blocked for 1 hour in PBS-T with 5% skimmed milk at room temperature. Blocked membranes were incubated with either primary anti-His-tag antibodies (Boster Biological Technology, M30975; 1:1000) or primary anti-RpoB antibodies (Abcam, ab191598; 1:2000) at 4 °C overnight in PBS-T with 1% milk. After three 5-minute washes with fresh PBS-T, HRP-conjugated secondary antibodies were added (goat anti-mouse IgG, Argisera, AS11 1772; 1:5000 or anti-rabbit IgG, Sigma-Aldrich, A0545; 1:5000, respectively; both diluted in PBS-T) and the membranes were incubated for 1 hour at room temperature. The membranes were washed twice with PBS-T for 5 minutes followed by one 5-minute wash with PBS. The WesternBright Quantum HRP substrate (Advansta) signal was visualized with Amersham™ ImageQuant 800 (Cytiva) imaging system. All experiments were performed at least three times, representative blots are shown.

Fluorescence microscopy

Fluorescence microscopy was carried out with early-mid logarithmic growth phase *E. coli* BW25113 cells grown in LB media (Miller) (10 g l^{-1} tryptone, 5 g l^{-1}yeast extract, 10 g l^{-1} NaCl) supplemented with 100 μg/ml ampicillin for maintaining plasmids at 37 °C. Samples were immobilized on Teflon-coated multi-spot microscope slides (Thermo Fisher) covered with a thin layer of H_2O/1.2 % agarose and imaged immediately. Microscopy was performed using a Nikon TI2 equipped with Nikon CFI Plan Apo DM Lambda 100X Oil objective, CoolLED pE-4000 light source, and Photometrics Kinetix sCMOS camera. Images were acquired with Nikon NIS-Elements AR software and analyzed with Fiji.[71]

Immunoprecipitation followed by tandem mass spectrometry (IP-MS/MS)

Overnight cultures containing plasmid-based anhydrous tetracycline (aTc)-inducible HigBAC or HigBA/FLAG-C cells were back-diluted in 250 mL LB with 0.05 mg/mL carbenicillin and 100 ng/mL aTc and grown at 37 °C to an OD_{600} = 0.2. Cultures were infected with either λ_{vir} or $\lambda_{vir}^{escape\_1}$ mutant at an MOI 10 and samples collected at 0-, 15-, and 30-minutes post-infection. To collect samples, the culture was pelleted at 7,500g for 5 minutes, the pellet decanted and then resuspended in lysis buffer (25 mM Tris-HCL, 150 mM NaCl, 1 mM EDTA, 5% glycerol, 1% Triton X100) supplemented with 1 μL/mL Ready-Lyse™ Lysozyme (Fischer Scientific), 1 μL/mL benzonase (Sigma), and cOmplete™ Protease Inhibitor Cocktail (Roche) and then flash frozen in liquid nitrogen. Samples were thawed and additional lysis buffer added as necessary to normalize sample concentration by OD_{600} value taken concurrent with sample collection. Samples were refrozen in liquid nitrogen and thawed to ensure complete cell lysis. Samples were spun at 20,000g for 10 minutes at 4 °C to pellet any debris. For each sample, 50 μL of Pierce™ Anti-DYKDDDDK magnetic agarose beads was mixed with 450 μL of lysis buffer and then collected to the side of the tube using a magnetic rack. Beads were then washed twice with 500 μL of lysis buffer. After the final wash, beads were mixed with 1 mL of sample and incubated for 20 minutes at room temperature on an end-to-end rotor. After incubation, beads were washed with wash buffer (1X PBS, 150 mM NaCl) twice and then once with MilliQ H_2O. On-bead reduction, trypsin digest, and LC-MS/MS were done as previously.[44] Detected peptides were mapped to MG1655 and λ protein sequences and the abundance of proteins were estimated by number of spectrum counts/molecular mass to normalize for protein sizes.

Biochemical assays

In vitro translation assays

PURExpress *in vitro* protein synthesis kit (NEB, E6800) supplemented with 0.8 U/μL RNase Inhibitor Murine (NEB, M0314S) was used for reactions as per the manufacturer instructions. All reactions contained each ART template plasmids (10 ng/μL, VHp1221) with or without 0.1 mM NAD^+ or 6-biotin-17-NAD^+. After a 10-minute incubation at 37 °C, a 1.3 μL aliquot of the reaction mixture was taken and quenched by addition of 13.7 μL of 2x SDS-PAGE sample buffer (100 mM Tris:HCl pH = 6.8, 4% SDS, 0.02% bromophenol blue, 20% glycerol, 20 mM DTT and 4% β-mercaptoethanol), and DHFR template plasmid was added to the remaining reaction mixture at a final concentration of 20 ng/μL. After further incubation at 37 °C for 1 hour, the reaction mixture was mixed with 9-fold volume of 2x sample buffer, denatured at 95 °C for 5 min and resolved on SDS-PAGE gel (18% acrylamide/bis-acrylamide = 37.5:1). The SDS-PAGE gel was fixed by incubating for 5 min at room temperature in 50% ethanol solution supplemented with 2% phosphoric

acid, washed three times with water for 20 min at room temperature, and stained with "blue silver" solution (0.12% Brilliant Blue G250 (Sigma-Aldrich, 27815), 10% ammonium sulfate, 10% phosphoric acid, and 20% methanol) overnight at room temperature. After washing with water for 3 hours at room temperature, the gel was imaged on an Amersham™ ImageQuant 800 (Cytiva) imaging system.

Immunoblotting

For streptavidin blotting, the PURExpress reaction was incubated with 0.1 mM 6-biotin-17-NAD^+ and 5 nM PCR fragment amplified with VTK53 and VTK54 primers and VHp1221 template at 37 °C for 2 hours. In the case of model-substrate modification, 0.1 volume of 100 μM oligo RNA or DNA was added to the reaction at 30 min after starting the reaction then the reaction was incubated at 37 °C for a further 30 min. The sample was denatured in 1x SDS-PAGE sample buffer (for protein) at 95 °C for 5 min or in Urea-PAGE sample buffer (98% formamide, 10 mM EDTA, 0.3% BPB and 0.3% Xylene cyanol, for nucleic acids) at 65 °C for 5 min, and resolved on SDS-PAGE (12% acrylamide/bis-acrylamide = 37.5:1) or 8 M Urea-PAGE (6% or 8% acrylamide/bis-acrylamide = 19:1). When using double-strand model substrates, the sample was mixed with 0.2 volume of TriTrack DNA Loading Dye (6X) (Thermo Scientific, R1161) and resolved on native PAGE (8% acrylamide/bis-acrylamide = 19:1). Resolved samples were transferred to Zeta-Probe® Blotting Membranes (Bio-Rad) using Trans-Blot® Turbo™ Transfer System (Bio-Rad). The membrane was blocked in PBS with 3% BSA at room temperature for 30 min, and the first antibody incubation was performed at 4 °C for one hour in PBS with 3% BSA and HRP-conjugated streptavidin (Thermo scientific, N100; 1:5,000 dilution). After two 5-minute washes with fresh PBS-T and one 5-minute wash with PBS, biotinylated ADP signal was imaged on an Amersham™ ImageQuant 800 (Cytiva) imaging system using WesternBright Quantum HRP substrate (Advansta).

Purification of HigC

N-terminally His_6-TEV-tagged HigC was cloned in pET28b vector and expressed in *E. coli* BL21(DE3). Cultures were grown in LB medium supplemented with kanamycin (50 μg/ml) at 37 °C with aeration. Expression was induced with 0.1 mM IPTG when the cells carrying the plasmid reached an $OD_{600\ nm}$ of ∼ 0.5-0.8 at 25 °C. After induction, the cells were harvested 16h later by centrifugation and resuspended with buffer (25 mM HEPES, pH 7.6, 300 mM NaCl, 300 mM KCL and 2 mM $MgCl_2$, 1mM TCEP) supplied with complete protease-inhibitor cocktail (Roche). The resuspended cells were flash-frozen in liquid nitrogen and stored at – 80 °C. The cell extracts were lysed using an Emulsiflex cell disruptor and the lysate was centrifuged to remove cell debris for 45 min at 25 000g. Prior to purification the extract was filter through a 0.45 μm membrane and loaded onto 1 ml HiTrap Ni NTA column (Cytiva) coupled to an FPLC (ÄKTA Explorer) equilibrated with buffer A (25 mM HEPES, pH 7.6, 300 mM NaCl, 300 mM KCL and 2 mM $MgCl_2$, 1mM TCEP). The column was washed with buffer A before proceeding with the elution of HigC with buffer B (25 mM HEPES, pH 7.6, 300 mM NaCl, 300 mM KCl and 2 mM $MgCl_2$, 1mM TCEP and 500 mM Imidazole). After tag removal by incubating the protein sample for 10h with TEV (1:5 molar ratio) at 10 °C the solution was passed through an HiTrap Ni NTA column to trap the protease and the His-tag and collect the tagless HigC. A final step of size exclusion chromatography was performed with the pooled samples in 25 mM HEPES, pH 7.6, 300 mM NaCl, 300 mM KCl and 1mM TCEP.

Purification of gpV_N

N-terminally His_6-TEV-tagged gpV_N was cloned in pET28b vector and expressed in *E. coli* BL21(DE3). Cultures were grown in LB medium supplemented with kanamycin (50 μg/ml) at 37 °C with aeration. Expression was induced with 0.5 mM IPTG when the cells carrying the plasmid reached an $OD_{600\ nm}$ of ∼ 0.5-0.8 at 25 °C. Cells were harvested by centrifugation 4h after induction and resuspended in 25 mM HEPES, pH 7.5, 300 mM NaCl, 300 mM KCl and 10 mM $MgCl_2$, 1mM TCEP supplied with complete protease-inhibitor cocktail (Roche). DNase (10 μg/ml) was added to the resuspended cells before they were stored at –80 °C. The cell extracts were lysed using an Emulsiflex cell disruptor and the lysate was centrifuged to remove cell debris for 45 min at 42 000 g. The extract was filtered through a 0.45 μm membrane, then loaded onto a gravity-flow column (Cytiva) packed with Co^{2+}-affinity resin equilibrated with purification buffer (25 mM HEPES, pH 7.5, 300 mM NaCl, 300 mM KCl 1mM TCEP). The column was washed with 10 ml purification buffer, then the protein was eluted stepwise with purification buffer containing 500 mM imidazole. The eluted fractions were transferred to a size-exclusion chromatography (SEC) column Superdex 75 pg (Cytiva), equilibrated in the SEC buffer (25 mM HEPES, pH 7.5, 150 mM NaCl, 150 mM KCl, 1mM TCEP). The His-TEV tag was removed by incubating with TEV enzyme (1:50 molar ratio) at 10 °C overnight. The cleaved protein was recovered by passing the sample over a gravity-flow column (Cytiva) packed with Co^{2+}-affinity resin (Thermo Fisher Scientific).

Isothermal titration calorimetry (ITC)

All titrations were performed with an Affinity ITC (TA instruments) at 30 °C in 25 mM HEPES, 150 mM KCl, 150 mM NaCl, 1 mM TCEP at pH 7.5 (ITC buffer). The ChAD fragments and the gpV synthetic truncates were directly resuspended in the ITC buffer. The final concentrations were verified by the OD_{280} absorption using a Nanodrop One (ThermoScientific). All ITC measurements were performed by titrating a constant volume of 2 μl of peptide or gpV_N into the ITC cell (containing HigC at 15 μM) using a constant stirring rate of 75 rpm. All data were processed, buffer-corrected and analyzed using the NanoAnalyze and Origin software packages. Sequences of synthetic polypeptides used in the study are provided in the (Dataset S1).

HDX-MS sample preparation

The HigC protein was concentrated to 100 μM. To prepare the complex of HigC with gpV^{48-78}, the peptide was directly dissolved in the HigC sample in a 1:4 molar ratio. For each experiment, 10 μL of protein sample (HigC or HigC:gpV^{48-78} were incubated for 1 min, 5 min, 15 min or 60 min in 50 μL of labelling buffer (25 mM HEPES pH 7.5, 400 mM KCl, 400 mM NaCl, 1 mM TCEP) at 20 °C. The non-deuterated reference points were prepared by replacing the labelling buffer with the equilibration buffer (25 mM HEPES pH 7.5,

400 mM KCl, 400 mM NaCl, 1 mM TCEP). After labelling, the samples were quenched by mixing 60 μL of pre-chilled quench buffer (1.2 % formic acid, pH 2.4) with the labelled sample, flash-frozen in liquid N_2 and stored at -80 °C.

HDX-MS data collection

Prior to each injection, samples were thaw at room temperature and 150 μL of the quench samples were directly transferred to the Enzymate BEH Pepsin Column (Waters Corporation) at 200 μL per min and at 20 °C with a pressure of 3 kPSI. Peptic peptides were trapped for 3 min on an Acquity UPLC BEH C18 VanGuard Pre-column (Waters Corporation) at a 200 μL per min flow rate in water (0.1% formic acid in HPLC-grade water, pH 2.5) before elution to an Acquity UPLC BEH C18 Column for chromatographic separation. Separation was done with a linear gradient buffer (3–45% gradient of 0.1% formic acid in acetonitrile) at a flow rate of 40 μL per min. Peptides identification and deuteration uptake analysis was performed on the Synapt G2 in ESI ± MS^E mode (Waters Corporation). Leucine Enkephalin was applied for mass accuracy correction and sodium formate was used as calibration for the mass spectrometer. MS^E data were collected by a 20-30 V transfer collision energy ramp. The pepsin column was washed between injections using pepsin wash buffer (1.5 M Guanidinium HCl, 4% (v/v) acetonitrile, 0.8% (v/v) formic acid). A cleaning run was performed before loading the samples to prevent peptide carry-over. Optimized peptide identification and peptide coverage for all samples was performed from undeuterated controls (five replicates). All deuterium time points were performed in triplicates.

QUANTIFICATION AND STATISTICAL ANALYSIS

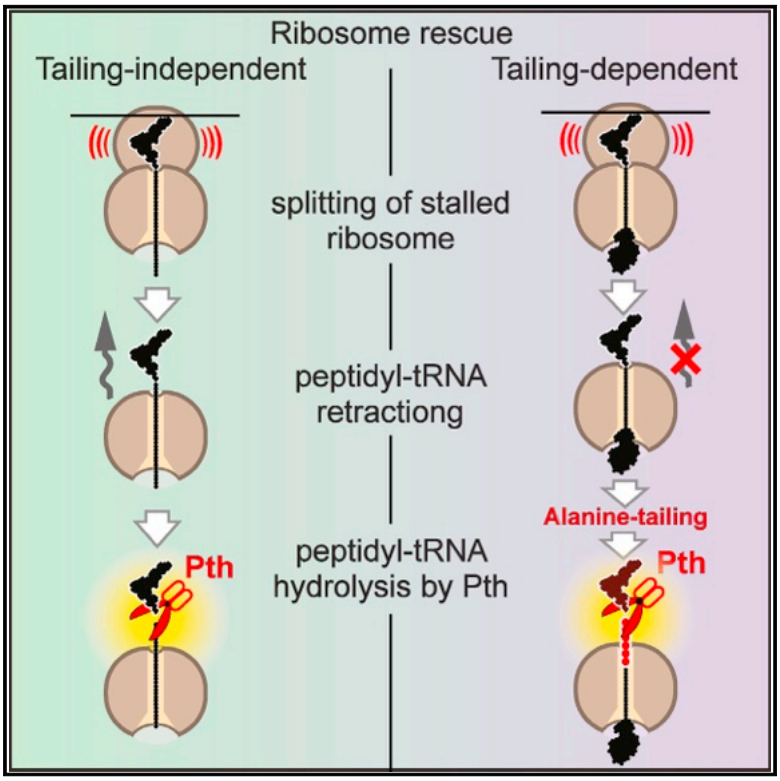
The statistical details of experiments can be found in the figure legends and in the STAR Methods. Statistical analyses were performed using GraphPad Prism 10.1.1, Igor Pro 7.07, Origin and NanoAnalyze. Metabolic labelling experiments were performed triplicates (n = 3), using three independent cultures initiated from distinct colonies, the data are shown as geometric means with standard deviations. Median values of mNeonGreen fluorescence were quantified using data from individual cells from the same imaging dataset (n = 62 to 95 cells) (Figures 3K and 3L). Spectral counts for each protein in the *E. coli* MG1655 and λ_{vir} proteomes and the HigBAC system were calculated, and high confidence hits assessed. The ratio of spectral counts between the λ_{vir} and $\lambda_{vir}^{escape\_1}$ escape mutant infected samples at 30 minutes post-infection with a pseudocount added to each count was used to generate Figure 4C. The growth curves in Figure 1G, 5G, and S3 represent the mean of three replicates (n = 3), using different transformants in each replicate, and the shaded areas indicate the standard deviation.

Molecular Cell

Article

Peptidyl-tRNA hydrolase is the nascent chain release factor in bacterial ribosome-associated quality control

Graphical abstract



Highlights

- Peptidyl-tRNA stuck inside the bacterial large ribosomal subunit is released by Pth
- Pth action requires the retraction of peptidyl-tRNA from the ribosomal catalytic center
- Nascent polypeptide folding may prevent peptidyl-tRNA retraction and release by Pth
- RQC machinery-mediated alanine tailing facilitates peptidyl-tRNA release by Pth

Authors

Maxim S. Svetlov, Clémence F. Dunand, Jose A. Nakamoto, ..., Daniel N. Wilson, Nora Vázquez-Laslop, Alexander S. Mankin

Correspondence

msvet2@uic.edu (M.S.S.), shura@uic.edu (A.S.M.)

In brief

Svetlov et al. show that, following ribosome splitting, peptidyl-tRNA that remains stuck inside the large ribosomal subunit is released by peptidyl-tRNA hydrolase (Pth). This release requires the retraction of peptidyl-tRNA from the subunit to expose its ester bond to Pth. The release is promoted by the C-terminal oligoalanine tailing of the nascent polypeptide.

Svetlov et al., 2024, Molecular Cell *84*, 715–726
February 15, 2024 Published by Elsevier Inc.
https://doi.org/10.1016/j.molcel.2023.12.002

CellPress

Molecular Cell

CellPress

Article

Peptidyl-tRNA hydrolase is the nascent chain release factor in bacterial ribosome-associated quality control

Maxim S. Svetlov,[1,2,5,*] Clémence F. Dunand,[1,2] Jose A. Nakamoto,[3] Gemma C. Atkinson,[3] Haaris A. Safdari,[4] Daniel N. Wilson,[4] Nora Vázquez-Laslop,[1,2] and Alexander S. Mankin[1,2,*]

[1]Center for Biomolecular Sciences, University of Illinois at Chicago, Chicago, IL 60607, USA

[2]Department of Pharmaceutical Sciences, University of Illinois at Chicago, Chicago, IL 60607, USA

[3]Department of Experimental Medicine, University of Lund, 221 00 Lund, Sweden

[4]Institute for Biochemistry and Molecular Biology, University of Hamburg, 20146 Hamburg, Germany

[5]Lead contact

*Correspondence: msvet2@uic.edu (M.S.S.), shura@uic.edu (A.S.M.)

https://doi.org/10.1016/j.molcel.2023.12.002

SUMMARY

Rescuing stalled ribosomes often involves their splitting into subunits. In many bacteria, the resultant large subunits bearing peptidyl-tRNAs are processed by the ribosome-associated quality control (RQC) apparatus that extends the C termini of the incomplete nascent polypeptides with polyalanine tails to facilitate their degradation. Although the tailing mechanism is well established, it is unclear how the nascent polypeptides are cleaved off the tRNAs. We show that peptidyl-tRNA hydrolase (Pth), the known role of which has been to hydrolyze ribosome-free peptidyl-tRNA, acts in concert with RQC factors to release nascent polypeptides from large ribosomal subunits. Dislodging from the ribosomal catalytic center is required for peptidyl-tRNA hydrolysis by Pth. Nascent protein folding may prevent peptidyl-tRNA retraction and interfere with the peptide release. However, oligoalanine tailing makes the peptidyl-tRNA ester bond accessible for Pth-catalyzed hydrolysis. Therefore, the oligoalanine tail serves not only as a degron but also as a facilitator of Pth-catalyzed peptidyl-tRNA hydrolysis.

INTRODUCTION

Cell growth and proliferation critically depend on efficient protein synthesis. Although most mRNAs are smoothly translated by ribosomes, cleavage or chemical damage to mRNAs, constellations of rare codons, or the polymerization of problematic amino acid sequences can cause prolonged translational pausing (reviewed in Filbeck et al.,[1], Müller et al.,[2], and Howard and Frost[3]). Such ribosome stalling is detrimental to cells because it sequesters ribosomes together with the associated tRNAs in inactive states, potentially leading to diminished global translation. In addition, the incomplete proteins trapped in the stalled ribosomes could be toxic for the cell. Therefore, organisms from all domains of life have evolved sophisticated mechanisms dedicated to rescuing and recycling stalled ribosomes and tRNAs and to targeting the associated truncated nascent peptides for degradation.

In eukaryotes, stalled ribosomes are recognized and split into subunits by the combined action of Hbs1/HBS1L, Dom34/Pelota(PELO), and Rli1/ABCE1 proteins (in yeast/mammals, respectively),[4,5] as reviewed in Franckenberg et al.,[6] Buskirk and Green,[7] and Joazeiro.[8] Although the resultant 40S subunits can readily reinitiate new rounds of translation, the split 60S subunits remain associated with peptidyl-tRNA (pept-tRNA). Such 60S-nascent chain complexes (60S-NCCs) are targeted by the ribosome-associated quality control (RQC) apparatus, which recycles ribosomal subunits and tRNAs by facilitating the release of the nascent polypeptides while tagging them for degradation.[9] These goals are achieved by the non-programmed addition of Ala (in mammals) or Ala/Thr C-terminal tails ("CAT-tails" in yeast) to the nascent chains in a reaction catalyzed by the Rqc2/NEMF proteins.[10] Tailing leads to the extrusion of the nascent protein's lysine residues out of the nascent peptide exit tunnel of the ribosome, allowing for their ubiquitination by the 60S-NCC-associated Ltn1/Listerin ubiquitin E3 ligase.[11,12] The nascent chain ubiquitination promotes the extraction and subsequent proteasomal degradation of the truncated proteins.[11,13–15] Additionally, in mammals, the Ala tails themselves serve as a degron recruiting other E3 ligases upon the release of the nascent protein.[16,17]

Recently, an RQC-like mechanism was discovered in bacteria.[18] In *Bacillus subtilis* and a number of other bacterial species (Figure 1A), the RqcU (MutS2) protein senses ribosomes that have collided due to translational arrest and promotes the splitting of the leading stalled ribosome into the small (30S) and large

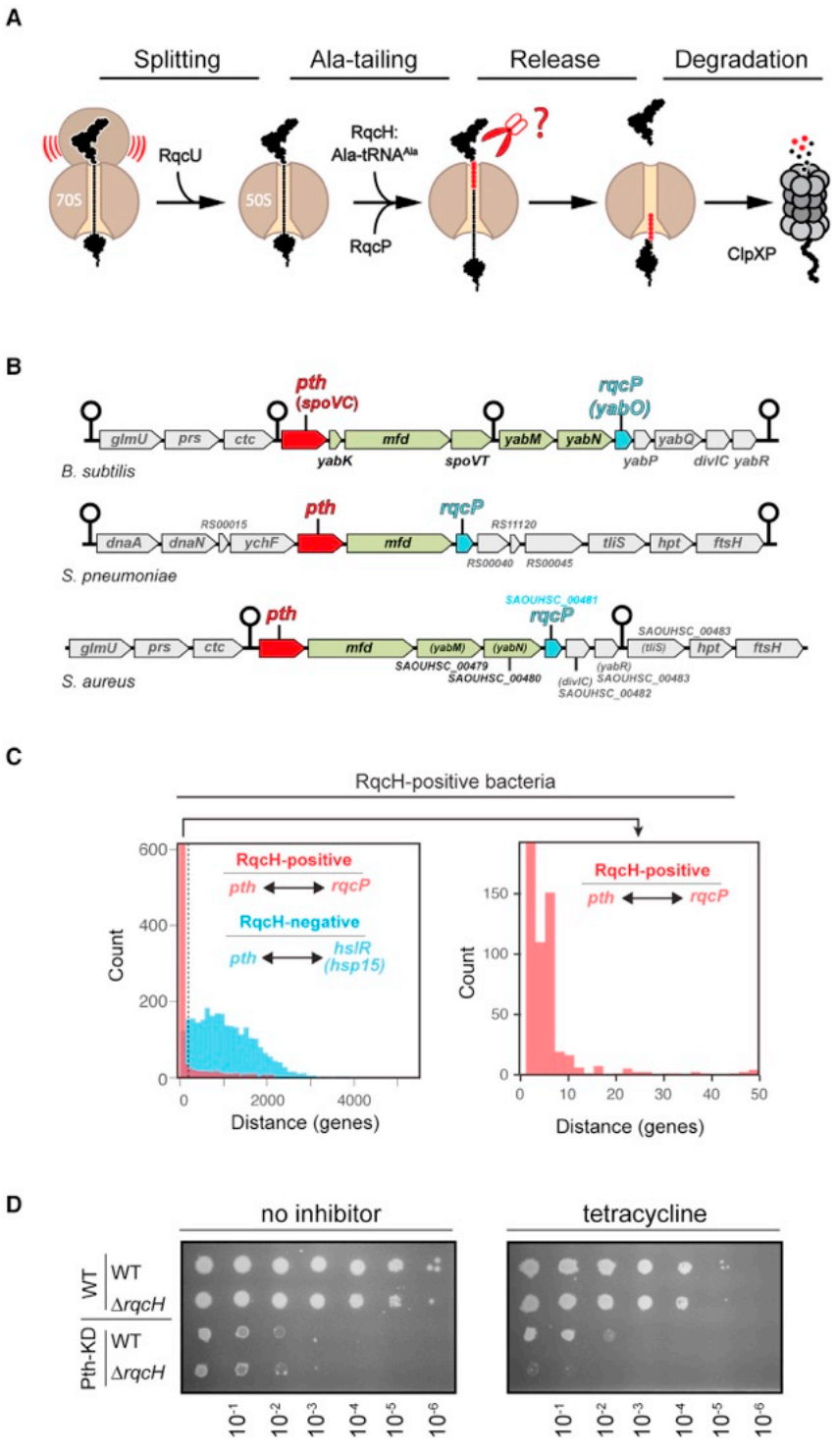


Figure 1. Pth is genetically and functionally linked to the bacterial RQC pathway

(A) Bacterial ribosome-associated quality control (RQC) pathway.[8,19] The factor responsible for disengaging the tRNA and the nascent peptide in the 50S-NCC remains unknown.

(B) Location of the *pth* gene in the neighborhood of *rqcP* (*yabO*) in the genomes of *rqcH*-positive bacterial species *B. subtilis*, *Streptococcus pneumoniae*, and *Staphylococcus aureus*. Location of the putative transcription termination sites[30–32] is indicated by black loops.

(C) Left: distance (number of genes) separating *pth* and *rqcP* genes in the genomes of 767 RqcH-positive bacteria (light red bars) or separating *pth* and the rqcP ortholog *hslR* (*hsp15*) in 2,616 RqcH-negative bacteria (blue bars). Bin size is 100 genes. Right: zoomed-in section of the histogram of RqcH-positive bacteria shown on the left.. Bin size is 2 genes. See also Table S1.

(D) Spot test showing the sensitivity of *B. subtilis* *ΔrqcH* cells with decreased expression of Pth to tetracycline. Both WT and Pth-KD cells express dCas9; Pth-KD strain also expresses *pth*-specific sgRNA.[33]

(50S) subunits.[19,20] The bacterial 50S-NCC is recognized by the NEMF ortholog RqcH, which, in concert with RqcP, recruits Ala-tRNA^{Ala} to extend the C terminus of the stalled polypeptide with an oligo-alanine tail.[18,21,22] Following the release of the nascent chain from the ribosome, the oligo-Ala tail serves as a degron facilitating the hydrolysis of the incomplete protein by the ClpXP protease.[18] The 50S-NCC can also accumulate in bacteria that lack the RqcU/RqcH/RqcP system due to the translation of specific ribosome-destabilizing nascent peptide sequences,[23,24] antibiotic treatment,[25] heat shock,[26,27] and possibly other stresses.[28,29]

A critical step in the recycling of 60S- or 50S-NCCs is the eventual uncoupling of the nascent peptide (Ala/CAT-tailed or not) from the associated tRNA. Only after disengaging from tRNA, the nascent chain can be extracted from the exit tunnel, the tRNA released, and the large subunit made available for new rounds of translation. In eukaryotes, the 60S-NCC-associated pept-tRNA, which carries a ubiquitinated nascent chain, is targeted by Vms1/ANKZF1 nuclease. This enzyme disengages the protein from tRNA by endonucleolytically cleaving off the tRNA's CCA end[34–37] rather than by hydrolyzing the pept-tRNA ester bond, as it occurs during conventional translation termination promoted by release factors. In contrast, in bacteria, it has remained unknown which cellular factor is responsible for the release of the nascent peptide from the 50S-NCC and how the release is coordinated with the RQC-catalyzed Ala tailing. Several bacterial proteins have been proposed to carry out this task, including the canonical (RF1 and RF2) and non-canonical (ArfB and PrfH) release factors; some yet-to-be-discovered endonucleases analogous to the eukaryotic Vms1/ANKZF1; or pept-tRNA

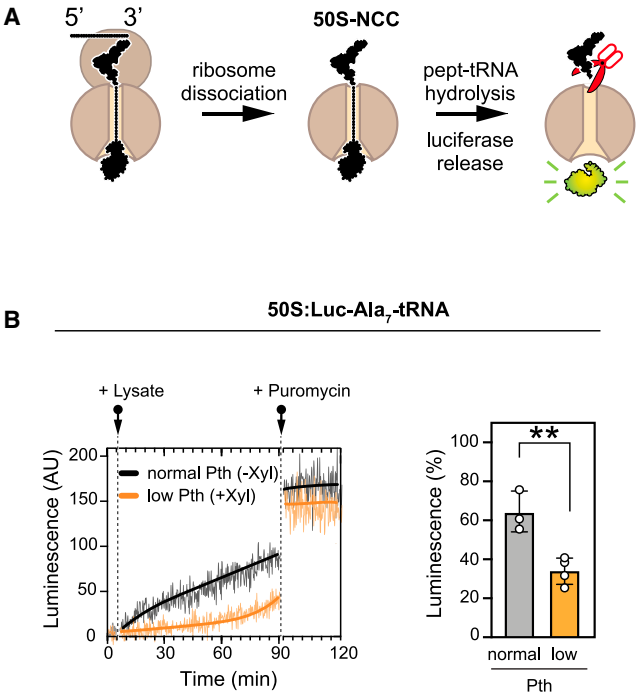


Figure 2. Pth in *B. subtilis* lysates hydrolyzes Ala-tailed peptidyl-tRNAs on split 50S subunits

(A) Principle of the luciferase-based nascent peptide release assay. Luciferase gains activity only after its release from the ribosome.[43]

(B) Release of Luc-A_7 polypeptide from the 50S-NCC in lysates of Pth-KD *B. subtilis* cells. Decreased expression of Pth in Pth-KD cells is induced by the addition of xylose. Puromycin addition (arrow) releases residual luciferase from the 50S-NCC. The bar graph shows the efficiency of Pth-mediated release in lysates from cells with normal (uninduced, gray trace) or low (xylose-induced, orange trace) Pth expression, estimated relative to the luciferase activity achieved after puromycin treatment. The error bars show the SD in four independent experiments. See also Figure S1A.

hydrolase (Pth), an enzyme known to hydrolyze ribosome-free pept-tRNAs.[1,21,26,27,38] However, in the absence of experimental evidence, none of these proteins so far stand out as strong candidates. Indeed, the association of release factors with the ribosome is driven to a large extent by their contacts with the small ribosomal subunit, and they are unlikely to bind to the isolated large ribosomal subunit; bacterial endonucleases with activity akin to that of Vms1/ANKZF1 are unknown; the remaining candidate, Pth, was described as an enzyme acting upon pept-tRNAs that drop off the ribosome primarily at the early rounds of translation.[39,40] Furthermore, if Pth were to act upon the 50S-NCC, how would it access the ester bond of the pept-tRNA enclosed within the active site of the peptidyl transferase center (PTC) of the large subunit?

In this paper, we reveal that the stalled pept-tRNA in bacterial 50S-NCCs is in fact released by the action of Pth. We show that pept-tRNA needs to be partially extracted (pulled back) from the 50S-NCC to render it accessible for Pth-catalyzed hydrolysis. We further show that Ala tailing facilitates the exposure of the pept-tRNA ester bond when the folding of the nascent protein prevents its extraction from the 50S-NCC. Our findings establish Pth as an integral component of RQC irrespective of whether it acts in concert with the Ala-tailing system or as a part of alternative ribosome rescue mechanism(s).

RESULTS

Pth is genetically linked to the RQC pathway

Genes with related functions tend to be closely clustered in bacterial genomes.[41] Therefore, we reasoned that the gene encoding the factor responsible for hydrolyzing pept-tRNA in 50S-NCCs may be located in the neighborhood of the genes encoding the Ala-tailing machinery. Examining the *B. subtilis* genome, we found the Pth-encoding gene *pth* in close vicinity of *rqcP*, which encodes the RqcP protein, the key accessory factor that facilitates the RqcH-catalyzed addition of the oligo-Ala tails (Figure 1B). In two other *rqcH*-positive bacterial species, *Streptococcus pneumoniae* and *Staphylococcus aureus*, *pth* and *rqcP* even reside in the same operon (Figure 1B). These initial observations prompted us to systematically analyze the proximity of *pth* and *rqcP* across fully sequenced reference bacterial genomes. The results of this analysis (Figure 1C) showed strong genomic association of the *pth* and *rqcP* genes specifically in *rqcH*-positive bacteria (Figure 1C; Table S1), suggesting that *pth* may be functionally related to the RQC pathway.

We then asked whether Pth is genetically linked to the RQC mechanism. Because Pth is essential for bacterial growth and its gene cannot be deleted, we utilized the *B. subtilis pth* knockdown (*pth*-KD) strain in which *pth* expression is downregulated through the inhibition of its promoter by catalytically dead dCas9 paired with $sgRNA^{pth}$.[33] As expected, even though *pth*-KD cells showed diminished growth, they could nevertheless readily form colonies on agar plates. Both *pth*-KD and Δ*rqcH* strains could also grow on plates with subinhibitory concentrations of tetracycline (Figure 1D), an antibiotic that causes genome-wide collisions of translating ribosomes and the accumulation of substrates for the RQC pathway.[18,42] However, Δ*rqcH* cells where Pth expression was repressed became hypersusceptible to this antibiotic (Figure 1D). Synthetic hypersensitivity to tetracycline showed that Pth is not only genomically associated with one of the key RQC components (Figure 1C) but also potentially functionally involved in the RQC pathway (Figure 1D), possibly playing a role in the release of nascent peptides from 50S-NCCs.

Pth can release Ala-tailed nascent peptides from the 50S-NCC

To explore a potential mechanistic role of Pth in the RQC pathway, we established an *in vitro* assay for monitoring the disengagement of an Ala-tailed nascent polypeptide from tRNA in a 50S-NCC (Figure 2A). The assay exploits the observation that the newly translated firefly luciferase (Luc) remains inactive until it is released from the tRNA and its C terminus is fully extracted from the ribosomal nascent peptide exit tunnel.[43,44] To prepare 50S-NCCs, a non-stop mRNA encoding Luc followed by a seven-alanine-residue-long tail (Luc-Ala_7) was translated *in vitro* by *B. subtilis* or *Escherichia coli* ribosomes. Following translation, ribosomes stalled at the 3′ end of the mRNA were dissociated into subunits, and 50S-NCCs were isolated by

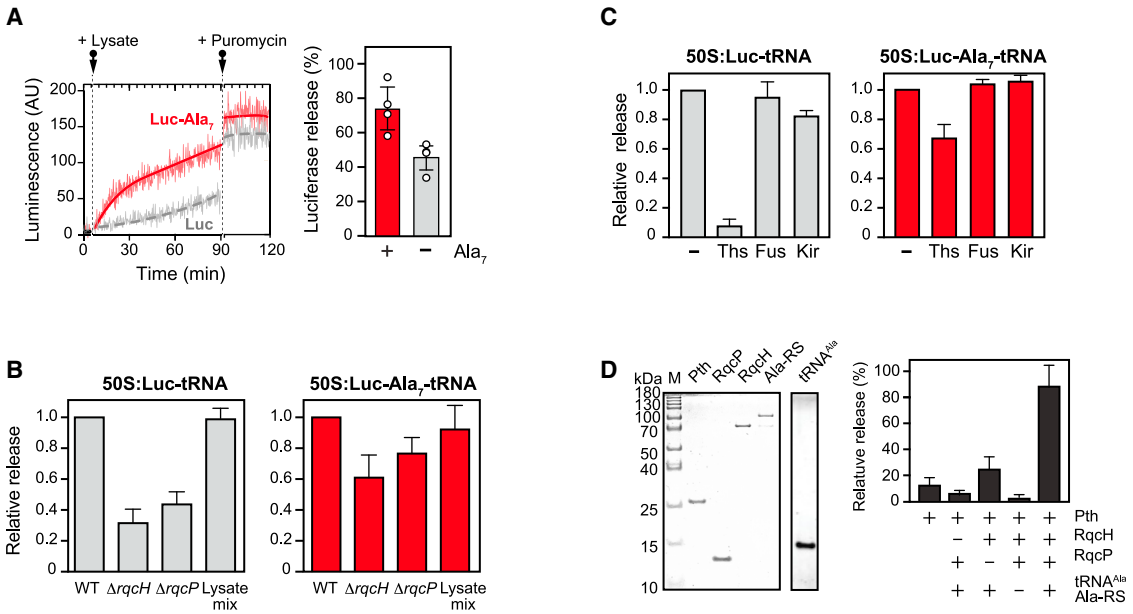


Figure 3. RQC-mediated oligoalanine tailing facilitates the Pth-mediated hydrolysis of pept-tRNA in the 50S-NCC

(A) Release of untailed (Luc) and tailed (Luc-Ala_7) luciferase from the 50S-NCC in *B. subtilis* lysates. Following incubation with lysates, complexes were treated with puromycin. Luminescence was achieved after puromycin treatment was set as 100%. Bar graph shows Pth-mediated luciferase release relative to that following puromycin treatment. See also Figures S2A and S3A.

(B) Efficiency of the release of untailed or Ala_7-tailed luciferase in lysates of wild-type, Δ*rqcH*, or Δ*rqcP B. subtilis* cells. "Lysate mix" sample had a 1:1 mixture of lysates of Δ*rqcH* and Δ*rqcP* cells.

(C) Effects of antibiotics that interfere with Ala tailing (thiostrepton [Ths]) or not (fusidic acid [Fus] or kirromycin [Kir]) on the Pth-promoted release of untailed (gray bars) or Ala_7-tailed (red bars) luciferase from the 50S-NCC in WT *B. subtilis* lysate. See also Figure S2B.

(D) Release of untailed luciferase from the 50S-NCC in an *in-vitro*-reconstituted Ala-tailing/release system composed of its purified protein components (RqcH, RqcP, Ala-RS, and Pth) and $tRNA^{Ala}$. Release efficiency was estimated relative to that afforded by puromycin treatment. The Coomassie- or ethidium bromide-stained gels of the purified components are shown on the left.

(A–C) Error bars show SD from four (A and B) or three (C and D) independent experiments.

sucrose gradient centrifugation (Figure S1A). We then treated the *B. subtilis* 50S-Luc-Ala_7-tRNA complexes with lysates of *B. subtilis pth*-KD cells with a normal or reduced level of Pth expression and followed protein release in real time by the appearance of luminescence. Luc-Ala_7 release was notably slower in lysates prepared from cells with diminished Pth expression than in those prepared from cells with normal Pth activity (Figure 2B). However, comparable levels of luminescence were reached in either of the lysates when 50S-NCCs were treated with puromycin, an antibiotic that facilitates nascent chain release by serving as a peptide acceptor in the PTC (reviewed in Aviner[45]). The results of these experiments implicate Pth in hydrolyzing oligo-Ala-tailed pept-tRNA trapped on 50S-NCCs, thereby supporting the notion that Pth may participate in the RQC pathway.

RQC-mediated Ala tailing of the nascent peptide facilitates its release from the 50S subunit

Because a significant fraction of nascent chains in the 50S-NCC undergo Ala tailing in RqcH-positive bacteria, we asked whether the ability of Pth to release Luc from the 50S-NCC depends upon the presence of the Ala tail at the C terminus of the nascent peptide. To address this question, we compared the kinetics of the release of untailed and Ala_7-tailed Luc. Remarkably, the presence of the Ala_7 tail markedly increased the rate of Luc release from *B. subtilis* 50S-NCC (Figure 3A), revealing the stimulatory role of the C-terminal oligo-Ala sequence in pept-tRNA hydrolysis. Consistently, even though an Ala-tailing machinery has not been described for *E. coli*, the Pth-catalyzed release of Luc from *E. coli* 50S-NCC was stimulated by the presence of the Ala_7 C-terminal tail (Figure S2A). The release of untailed Luc was even less efficient in lysates prepared from *B. subtilis pth*-KD or *E. coli* with reduced Pth activity (Figure S3), confirming that Pth is responsible for the nascent chain decoupling from tRNA.

We wondered whether the release of untailed Luc from the 50S-NCC was stimulated by the RqcH/RqcP-mediated Ala tailing. We found that, in comparison with wild-type (WT) *B. subtilis* lysate, the release of Luc from the 50S-NCC was significantly less efficient in lysates prepared from Δ*rqcH* or Δ*rqcP* strains (Figure 3B), which, although being deficient in Ala tailing, possessed WT levels of Pth activity. Restoring the Ala-tailing potential by mixing the lysates from Δ*rqcH* or Δ*rqcP* strains accelerated the release of the untailed Luc to levels comparable to those observed in WT cells lysate (Figure 3B). Not surprisingly, the release of the "pre-tailed" nascent protein from 50S:Luc-Ala_7-tRNA complexes was affected to a lesser extent by the absence of RqcH or RqcP in the cell lysates (Figure 3B).

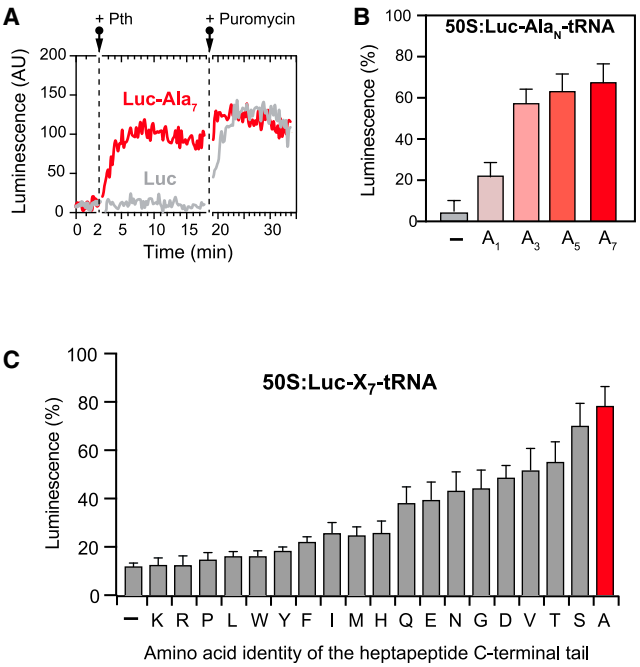


Figure 4. Properties of the C-terminal tail impact the efficiency of the Pth-mediated release of nascent proteins from the 50S-NCC

(A) Release of untailed (Luc) and tailed (Luc-A_7) luciferase from *B. subtilis* 50S-NCC by purified *B. subtilis* Pth and then by puromycin treatment. See also Figure S4.

(B and C) Release efficiencies of nascent luciferase in (B) *B. subtilis* 50S-NCC C-terminally extended with the indicated number of alanine residues or (C) nascent luciferase in *E. coli* 50S-NCC extended with different heptapeptide C-terminal tails. In (B) and (C), release carried out by purified respective Pth was estimated as a ratio of luminescence produced by Pth treatment relative to that following puromycin treatment. Error bars show SD from three independent experiments.

The antibiotic thiostrepton has been reported to inhibit Ala tailing by impeding the interaction of RqcH with 50S subunits.[22,46] Consistent with the stimulatory role of Ala tailing for the Pth-mediated pept-tRNA hydrolysis in the 50S-NCC, the addition of thiostrepton to cell lysates inhibited the release of untailed Luc from the 50S-NCC but did not importantly impact the release of the Ala-tailed protein (Figure 3C). Antibiotics that inhibit the elongation factors EF-Tu (kirromycin) or EF-G (fusidic acid), which are also thiostrepton targets, but are not involved in Ala tailing, had no effect on the release of untailed or tailed-nascent proteins (Figure 3C).

The presented data suggest that after ribosome splitting, 50S-NCC-associated Luc-tRNA remains largely inaccessible to Pth until the Ala tailing of the nascent protein catalyzed by the RQC machinery sensitizes pept-tRNA to Pth action. To directly test this hypothesis, we reconstituted *in vitro* the tailing/release system from purified *B. subtilis* proteins, i.e., each one of the individual components of the Ala-tailing apparatus ($tRNA^{Ala}$, alanyl-tRNA synthetase, and the RQC factors RqcH and RqcP) and Pth (Figure 3D). Although Pth alone, without Ala-tailing machinery, was unable to release untailed Luc from the 50S-NCC, efficient release was observed in the presence of RQC factors and Ala-tRNA. The stimulation of the Pth-mediated Luc release was likely due to the Ala-tailing reaction, rather than due to the mere association of RqcH and RqcP with the 50S-NCC, since the withdrawal of any single RQC component, including Ala-$tRNA^{Ala}$, from the tailing/release system resulted in background levels of pept-tRNA hydrolysis (Figure 3D). Combined, our data highlight the importance of the Ala-tailing machinery for the Pth-mediated release of the nascent protein from *B. subtilis* 50S-NCC.

Having established the importance of the nascent chain Ala tailing for the Pth-catalyzed pept-tRNA hydrolysis in 50S-NCCs, we next determined the length requirement of the C-terminal oligo-Ala sequence for efficient peptide release. Untailed Luc-tRNA in the 50S-NCC was largely impervious to hydrolysis by purified Pth (Figure 4A). The addition of a single alanine residue to the C terminus of Luc only slightly improved the Pth-mediated pept-tRNA hydrolysis, whereas the presence of Ala_3, Ala_5, or Ala_7 tails greatly stimulated the release of the nascent protein (Figure 4B). Because the stimulatory effect of the Ala tail plateaued at 5–7 Ala residues, most of the rest of the experiments were carried out with the Ala_7 tail. Importantly, Luc-tRNA and Luc-Ala_7-tRNA non-associated with ribosomes could be hydrolyzed by Pth with equally high efficiencies (Figure S4C), indicating that untailed Luc-tRNA is a poor Pth substrate only when within the large ribosomal subunit and that Ala tailing facilitates Pth-mediated hydrolysis specifically in the 50S-NCC.

Consistent with the conservation of the ribosome structure and properties across bacterial species, the presence of the Ala tail stimulated the Pth-catalyzed release of Luc also from *E. coli* 50S-NCC (Figure S4). Because of the similarity of the effects, we carried out some of the subsequent experiments using *E. coli* 50S-NCC, which can be more readily generated in the cell-free translation system.

The nature of the C-terminal tail is important for nascent peptide release from the 50S-NCC

We wondered whether it is the mere presence of extra C-terminal amino acids that is required for or whether the tail's amino acid identity is also important for stimulating the release of the nascent chain from the 50S-NCC. Strikingly, changing the Ala tail of Luc-Ala_7 to the other possible hepta homopeptides reduced the efficiency of the Pth-mediated release. Thus, although in the 50S-NCC context, Luc-Ser_7 was hydrolyzed nearly as efficiently as Luc-Ala_7, the release of Luc nascent chains appended with Tyr, Trp, Leu, Pro, Arg, or Lys tails was very inefficient and comparable to that of the untailed protein (Figure 4C). tRNAs esterified with Luc carrying Phe_7, Ile_7, Met_7, His_7, Gln_7, Glu_7, Asn_7, Gly_7, Asp_7, Val_7, or Thr_7 tails were hydrolyzed by Pth with intermediate efficiencies. Luc-Cys_7 lacked enzymatic activity, which prevented testing the efficiency of its release from 50S-NCCs. Importantly, among all the tested heptapeptide tails, Ala_7 was the most efficient in promoting pept-tRNA hydrolysis and thus 50S subunit recycling.

These results allowed us to conclude that the amino acid identity of the C-terminal tail introduced by the RQC system is important not only for the subsequent degradation of the truncated protein but also for the release of the nascent peptide from the 50S-NCC.

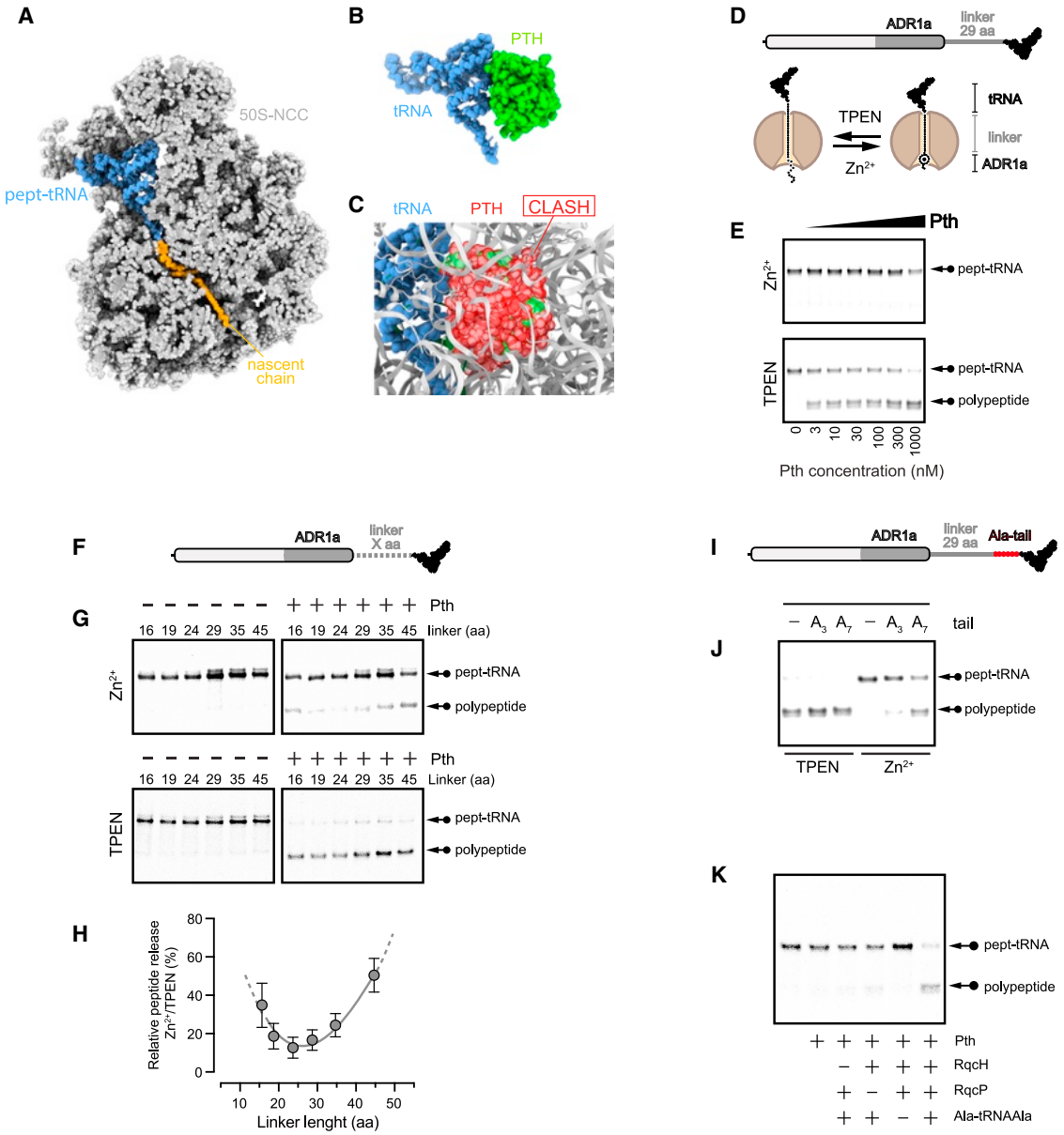


Figure 5. Nascent protein folding can interfere with Pth-catalyzed pept-tRNA hydrolysis in the 50S-NCC

(A) Model of a 50S-NCC representing 50S subunit associated with pept-tRNA (from the cryoelectron microscopy [cryo-EM] structure of a stalled translation complex, PDB: 3J9W[47]).

(B) Modeled structure of the Pth-tRNA complex (from the X-ray structure of the Pth complexed with the tRNA analog, PDB: 3VJR[48]).

(C) Superposition of the structures shown in (A) and (B) by aligning the tRNA bodies. Shown in red are Pth residues that would sterically clash with elements of the 50S subunit.

(D) Schematics of the experimental setup. With a 29-amino-acid-long linker, the Zn^{2+}-mediated folding of ADR1a in the vestibule of the ribosomal exit tunnel[49] should prevent nascent peptide retrosliding and pept-tRNA retraction.

(E, J, and K) Hydrolysis of the pept-tRNA in *B. subtilis* 50S-NCC by purified *B. subtilis* Pth. Reactions were performed in the absence (TPEN) or presence of Zn^{2+}, as indicated. [^{35}S]-labeled unreactive peptidyl-tRNA or Pth-hydrolyzed peptide product was gel separated and detected by autoradiography. Experiments shown in (E) and (J) were also carried out with *E. coli* 50S-NCC (see Figures S5A and S5D).

(E) Peptide release from the complex shown in (D).

(F) Constructs similar to those in (D), but with linkers of varying length.

(G) Peptide release from the complexes shown in (F) carried out using *E. coli* 50S-NCC and purified *E. coli* Pth.

(legend continued on next page)

Retraction of the pept-tRNA and exposure of its ester bond are required for its Pth-catalyzed hydrolysis in 50S-NCCs

We asked why Ala tailing facilitates pept-tRNA hydrolysis specifically in the context of 50S-NCCs. *In silico* structural analysis shows that the ester bond of the pept-tRNA residing in the ribosomal PTC is sterically inaccessible to Pth (Figures 5A–5C). Thus, for productive hydrolysis by Pth, the pept-tRNA ester bond needs to disengage from the PTC active site while retracting its nascent chain *backward* through the exit tunnel. Ala tailing may directly or indirectly facilitate the "retrosliding" of the nascent chain.

To test whether nascent chain retrosliding is a prerequisite for Pth-catalyzed pept-tRNA hydrolysis in 50S-NCCs, we exploited the Zn^{2+}-dependent folding of the Zn-finger ADR1a domain. A previous study has shown that the folding of ADR1a in the vestibule near the opening of the ribosomal exit tunnel exerts a pulling force on the peptide segment that tethers ADR1a to tRNA[49] and thus should prevent nascent protein retrosliding and pept-tRNA retraction (Figure 5D). We prepared a 50S-NCC, in which the tRNA moiety and the ADR1a domain of the pept-tRNA were separated by a 29-amino-acid-long linker—a length sufficient for Zn^{2}-mediated ADR1a folding in the vestibule section of the exit tunnel near its opening on the solvent side of the 50S subunit.[49] We then analyzed the efficiency of pept-tRNA hydrolysis by Pth in the presence of $ZnCl_2$ or of Zn^{2+} chelator *N*,*N*,*N*′,*N*′-tetrakis(2-pyridinylmethyl)-1,2-ethanediamine (TPEN). When $ZnCl_2$ was present and ADR1a could fold into a compact structure, pept-tRNA was resistant to the Pth-catalyzed hydrolysis. In contrast, it was readily hydrolyzed under Zn^{2+}-depletion conditions when ADR1a was unfolded, and the nascent chain could likely retroslide through the exit tunnel (Figures 5E and S6A). Importantly, the presence of $ZnCl_2$ had little effect upon the ability of Pth to release Luc-Ala_7 (whose folding is Zn^{2+} independent) from the 50S-NCC (Figure S5B) or to hydrolyze ribosome-free pept-tRNA (Figure S5C). Increasing the length of the linker separating tRNA from ADR1a facilitated pept-tRNA hydrolysis (Figures 5F–5H), likely because additional slack in the nascent peptide segment between the tRNA and ADR1a domain allows the retraction of the pept-tRNA from the PTC. Consistently, with a more abridged linker (16 amino acid long), which is too short to allow for the ADR1a domain to fold in the vestibule, the dependence of pept-tRNA hydrolysis by Pth in the presence of Zn^{2+} became less pronounced (Figures 5F–5H). These data are consistent with the model that the retraction of the pept-tRNA from the PTC is required for its hydrolysis by Pth and that protein folding within or outside of the exit tunnel can sterically block the retrosliding of the nascent chain, thereby preventing pept-tRNA release from the 50S-NCC and interfering with the recycling of the large ribosomal subunit.

We then reasoned that Ala tailing may promote the hydrolysis of the pept-tRNA in the Pth-resistant 50S-NCC by introducing extra slack in the nascent peptide and facilitating retrosliding that is otherwise blocked by protein folding. Indeed, the addition of Ala tails of increasing length to the C terminus of the ADR1a-containing constructs facilitated the Pth-catalyzed hydrolysis of the pept-tRNA in the 50S-NCC, even when the folding of the nascent chain would prevent its retrosliding through the exit tunnel (Figures 5I, 5J, and S6D). Finally, consistent with the idea that the hydrolysis of the 50S subunit-associated pept-tRNA is promoted by tailing, the preincubation of the Pth-resistant 50S-NCC in our *in vitro* reconstituted tailing/release system facilitated Pth-mediated peptide release when all the components of the Ala-tailing machinery were present (Figure 5K).

Thus, the Ala tailing of a non-retractable nascent protein in the 50S-NCC facilitates its release by making possible the interaction of the Pth active site with the pept-tRNA ester bond. The results obtained with the ADR1a-based reporter likely explain the sluggish release of untailed Luc from the 50S-NCC (Figures 3A and S2A), which is likely prevented by the cotranslational folding of the nascent Luc protein upon exit from the tunnel.

Antibiotics that stabilize pept-tRNA in the PTC and prevent its retraction interfere with the Pth-catalyzed release of the nascent chain from the 50S-NCC

Several antibiotics that interact with the large ribosomal subunit make direct contacts with either the P-site tRNA (sparsomycin and blasticidin S)[50,51] or the nascent chain that protrudes from the PTC into the exit tunnel (linezolid and chloramphenicol).[52,53] The preincubation of the 50S-NCC carrying Luc-Ala_7-tRNA with such antibiotics significantly inhibited pept-tRNA hydrolysis (Figure S2B). In contrast, antibiotics that target the initiation complex (retapamulin) (Figure S2B) or the elongation factors (kirromycin and fusidic acid) (Figure 3C) had no effect. Thus, small molecules that impede the disengagement of the pept-tRNA from the PTC or interfere with the retrosliding of the nascent peptide in the tunnel prevent the Pth-catalyzed hydrolysis of the 50S-NCC-associated pept-tRNA.

DISCUSSION

We have presented evidence here that the release of pept-tRNA from the 50S-NCC is catalyzed by Pth, revealing this enzyme as a critical component of the RQC pathway. We show that to become susceptible to Pth-catalyzed hydrolysis, the pept-tRNA needs to be dislodged from the PTC of the 50S-NCC and retracted so that its ester bond becomes accessible to the Pth active site.

Based on the presented data, we propose a model for the role of Pth in the mechanism of the ribosome and tRNA recycling as well as its interplay with the Ala-tailing machinery in RqcH-positive bacteria, such as *B. subtilis* (Figure 6). When translation arrest results in ribosome collision, RqcU promotes ribosome splitting.[19] Conceivably, ribosome splitting could also be collision

(H) Relative efficiency of peptidyl-tRNA hydrolysis as a function of the length of the linker estimated from the gels shown in (G). Error bars represent SD from three independent experiments.

(I) The construct is similar to that in (D) but C-terminally extended with the number of alanine residues indicated in (J).

(J) Peptide release from the 50S-NCC carrying peptidyl-tRNAs derived from the constructs shown in (I).

(K) Release of the peptide derived from the construct shown in (D) in the presence of the indicated purified components of the Ala-tailing/release system.

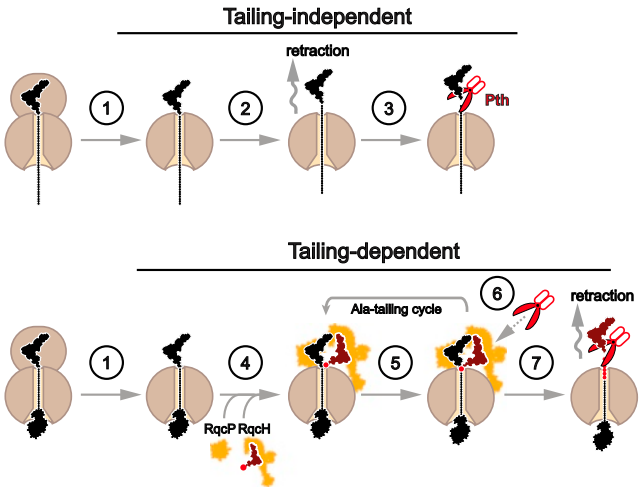


Figure 6. Role of Pth in tailing-independent and tailing-dependent ribosome rescue pathways

Tailing independent: 1, ribosome splitting; 2, pept-tRNA retraction; and 3, Pth-catalyzed pept-tRNA hydrolysis.

Tailing dependent: 1, ribosome splitting; 4, tailing machinery recruitment; 5, Ala tailing; and 6 and 7, Pth-catalyzed pept-tRNA hydrolysis attempted either (6) at each round of Ala-addition or (7) after the synthesis of the Ala tail of specific length.

and RqcU independent and caused by other factors. Irrespective of the cause, the dissociation of the translating ribosome into subunits results in the formation of the 50S-NCC (stage 1). If the nascent polypeptide is free to retroslide through the exit tunnel, pept-tRNA can easily detach from the PTC. In the absence of the small ribosomal subunit, pept-tRNA is retained in the 50S-NCC primarily due to the threading of the nascent peptide through the exit tunnel, whereas direct interactions of the tRNA with the PTC are limited to weak contacts of its CCA end with the 23S rRNA.[54,55] Therefore, with the freely sliding nascent chain, the thermal movement of pept-tRNA in the 50S-NCC would likely be sufficient to expose its ester bond to Pth (stage 2), leading to pept-tRNA hydrolysis (stage 3). However, the retrosliding of the nascent protein would likely be often prevented by its folding, specific contacts with the tunnel walls, anchoring within the translocon, or cotranslational interactions with chaperones or by jamming the tunnel with small molecule ligands, such as antibiotics or metabolites (reviewed in Wilson and Beckmann[56]). In this case, the ester bond of the pept-tRNA cannot be properly exposed; the binding of Pth to the 50S-NCC becomes futile; and, instead, the recruitment of the Ala-tailing machinery prevails (stage 4). The addition of Ala residues to the C terminus of the nascent chain (stage 5) may facilitate retrosliding by either displacing the tightly folded protein domains or the nascent chain cellular interacting partners away from the exit tunnel or by breaking interactions of the stuck polypeptide with the tunnel walls. In the process of Ala tailing, RqcP stabilizes pept-tRNA in the PTC, preventing a productive contact of Pth with the 50S-NCC. Depending on the processivity of Ala tailing and the kinetics of the binding/dissociation of RqcH/RqcP or Pth, none of which are currently known, Pth may probe the 50S-NCC at each cycle of Ala incorporation or after the addition of several Ala residues (stage 6). In any case, once the Ala tail reaches a length sufficient for the proper exposure of the pept-tRNA ester bond by retrosliding-mediated retraction, Pth manages to hydrolyze the pept-tRNA, and the nascent protein can be finally released (stage 7). Accordingly, we propose that in the RQC pathway, the Ala tailing of stuck peptides serves two functions: (1) promoting the rapid degradation of the truncated proteins by appending an oligo-Ala degron, as demonstrated previously,[18] and (2) facilitating the Pth-promoted release of the nascent polypeptide from the 50S-NCC, expediting the recycling of the tRNA and the large ribosomal subunit.

We found that in comparison with all other amino acid homo-oligomers, C-terminal tails composed of Ala residues are particularly beneficial for pept-tRNA hydrolysis in the 50S-NCC. Tailing with some other small (Ser, Thr, Val, and Gly) or negatively charged (Asp and Glu) amino acids also stimulates the release of non-retractable nascent chains. In contrast, positively charged (Arg and Lys) or bulky (Phe, Trp, and Tyr) amino acids, as well as Pro residues, do not stimulate 50S-NCC recycling. Why are Ala tails specifically favorable for pept-tRNA hydrolysis in the 50S-NCC? It has been shown that free N-acetyl-Ala-tRNA^{Ala} is particularly efficiently hydrolyzed by Pth in comparison with other N-acetyl-aminoacyl-tRNAs.[57] Thus, tailing with Ala may be favorable for Pth activity upon the 50S-NCC. However, because the rates of hydrolysis of other model pept-tRNA substrates with varying amino acids[57] do not correlate closely with their ability to stimulate the release of the stuck peptide from the 50S-NCC, we favor the hypothesis that, regardless of their merit as Pth substrates, Ala tails facilitate the nascent peptide retrosliding better than other C-terminal oligomers. Interactions of the nascent chain with the PTC-proximal segment of the exit tunnel are particularly important for translation.[40,58–60] Notably, amino acid residues composing the "release non-stimulating" tails are known to establish strong interactions with the PTC-proximal segment of the exit tunnel,[24,61,62] possibly preventing nascent chain retrosliding. The "non-sticky" nature of the oligo-Ala tail may be particularly favorable for mobilizing the nascent protein stuck in the ribosomal exit tunnel. Hence, the evolutionary selection of Ala tailing as a preferred C-terminal modification for faulty nascent proteins was likely driven by the ability of the Ala tail to target truncated protein for degradation as well as its capacity to facilitate pept-tRNA retraction, making it available for hydrolysis by Pth.

Pth was originally characterized as a protein responsible for of hydrolyzing free pept-tRNAs that prematurely dissociate from translating ribosomes.[39,40,63] The regeneration of the cellular pool of deacylated tRNAs necessary for translation is believed to underly the essentiality of Pth in bacteria.[64,65] Our present work and observations of others[23] have shown that the cellular functions of Pth extend beyond the hydrolysis of free-floating pept-tRNAs, revealing the role of this essential enzyme in the ribosome rescue pathway where it facilitates the recycling of not only tRNAs but also the large ribosomal subunits. How functions of Pth in ribosome rescue contribute to its essentiality in bacterial cells remains to be determined.

Although in RqcH-positive bacteria, Ala tailing targets truncated proteins for degradation by appending a C-terminal degron, Ala or CAT tailing in eukaryotes serves a similar goal

primarily by helping the ubiquitination of the nascent chains.[11,12] The non-programmed addition of C-terminal residues by the RQC machinery forces the extrusion of the lysine residues of the nascent protein chain buried in the exit tunnel, allowing their ubiquitylation by the 60S-NCC-associated Ltn1/Listerin ubiquitin ligase.[11,12] Conceivably, the ubiquitination of the nascent chain may sterically prevent its retrosliding through the tunnel. Hence, maybe not surprisingly, the release of ubiquitinated nascent chains from the 60S-NCC is catalyzed not by the hydrolysis of the pept-tRNA ester bond but by the cleavage of the tRNA's CCA end by Vms1/ANKZF1 nuclease, whose endonucleolytic activity does not require pept-tRNA retraction from the PTC.[34–37] Remarkably, *in vitro* data suggest that the release of the non-ubiquitinated pept-tRNA from the 60S-NCC can be promoted by Pth1, the eukaryotic ortholog of the bacterial Pth. Similar to Pth in bacteria, Pth1 action likely requires pept-tRNA retrosliding that would be abolished by the nascent chain ubiquitination.[37]

In *B. subtilis* and likely in other bacteria that exploit the RqcH/RqcP-dependent Ala tailing of the nascent peptide in 50S-NCCs, Pth operates as a component of the RQC pathway. However, many bacteria that lack RqcH and are thus presumably incapable of nascent peptide tailing still possess the RqcP ortholog Hsp15, which interacts specifically with 50S-NCCs.[26,27,66] The omnipresence of RqcP homologs suggests that 50S-NCC accumulation is common even in bacterial species lacking the Ala-tailing machinery. The splitting of the translating ribosome into subunits can result from heat shock and other stresses,[26,27,66] the action of specific ribosome-splitting factors,[25,29,67–70] and likely other effectors. Our data show that, similar to Pth in *B. subtilis*, *E. coli* Pth can also release the nascent protein from 50S-NCCs. These results are consistent with the previous reports showing that Pth is involved in releasing polypeptides from *E. coli* 50S-NCCs generated during the translation of the so-called ribosome destabilizing sequences.[23,24] Extrapolating our finding, it is reasonable to expect that the hydrolysis of the 50S-NCC-associated pept-tRNA in RqcH-negative bacteria also relies on the retrosliding of the nascent peptide in the exit tunnel and dislodging of pept-tRNA from the PTC. Some cellular polypeptides, for example, intrinsically unfolded proteins, can be particularly prone to retrosliding. The retrosliding of others, however, may be cumbersome. In our experiments, retrosliding could be prevented by the assembly of the compact ADR1a domain inside the exit tunnel and, likely, by the folding of the nascent Luc protein outside the ribosome. In the absence of the C-terminal tailing mechanism, other yet-unidentified factors might facilitate the exposure of the pept-tRNA ester bond to the Pth action on 50S-NCCs.

Our findings may help develop antibacterial therapies targeting bacteria-specific ribosome rescue pathways.[71,72] Given the difference in strategies for pept-tRNA release from 50S- and 60S-NCCs in bacteria and eukaryotes, respectively, and the lack of structural conservation of bacterial and eukaryotic pept-tRNA hydrolases, bacterial Pth emerges as an attractive antibiotic target. Due to the essentiality of Pth in bacteria, inhibiting its activity should arrest cell growth and proliferation by preventing the recycling of drop-off pept-tRNAs or hydrolysis of the pept-tRNA in the 50S-NCC. Furthermore, our data show that antibiotics that prevent pept-tRNA retraction either by interfering with the retrosliding of the protein chain through the exit tunnel or by locking pept-tRNA in the PTC interfere with the Pth-catalyzed release of the pept-tRNA from the 50S-NCC. Hence, these antibiotics emerge as not only direct blockers of protein synthesis but also as inhibitors of ribosome rescue. Consistently, diminishing Pth activity sensitizes bacteria to the action of several ribosome-targeting inhibitors.[73–75] Therefore, successful Pth inhibitors are expected to act synergistically with some of these drugs, potentiating their action against pathogenic microorganisms.

Limitations of the study

Although the extent of the pulling force upon the nascent protein chain afforded by the ADR1a domain folding in the tunnel's vestibule has been investigated,[49,76] the details of Luc folding outside the ribosome are unknown. Similarly, it remains to be investigated how the sequence of the nascent protein spanning the exit tunnel affects its retrosliding ability.

Finally, the details of the interplay of Pth and RqcH/RqcP with the 50S-NCC rely on the kinetics of the binding of these proteins to and dissociation of these proteins from the 50S subunit with the tRNA molecule exposed at the interface side of the subunit. Additional ensemble or single-molecule experiments will be needed to deduce these kinetics parameters, which will help to refine the proposed model.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANTS DETAILS
  - Strains, plasmids, and growth medium and primers
- METHOD DETAILS
  - DNA manipulations
  - Constructing DNA templates for *in vitro* transcription-translation
  - Phylogenetic analysis
  - Gene distance and neighborhood analysis
  - Construction of *B. subtilis* strains
  - Preparation of *B. subtilis* lysates
  - Preparation of 50S-NCC
  - Protein expression and purification
  - Luciferase release assay
  - Pth-mediated pept-tRNAs hydrolysis in ADR1a-containing 50S-NCC
  - Hydrolysis of ErmDL-tRNA in cell lysate
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.molcel.2023.12.002.

ACKNOWLEDGMENTS

We thank Vasili Hauryliuk and Tatsuaki Kurata (Lund University) for providing the Δ*rqcH B. subtilis* strain and plasmids encoding *B. subtilis* RqcH and RqcP; Gunnar von Heijne (Stockholm University) for the plasmids with the ADR1a constructs; Egor Syroegin and Alena Aleksandrova (University of Illinois at Chicago) for providing ErmDL-tRNA, $tRNA^{Ala}$, and Ala-RS; and Yury Polikanov (University of Illinois at Chicago) for help with the analysis of structural data and helpful discussions. This work was supported by grants from the National Science Foundation (MCB-1951405 to N.V.-L. and A.S.M.), the Knut and Alice Wallenberg Foundation (2020.0037 to G.C.A.), and the Swedish Research Council (2019-01085 and 2022-01603 to G.C.A.).

AUTHOR CONTRIBUTIONS

Conceptualization, M.S.S., A.S.M., and N.V.-L.; methodology, M.S.S., A.S.M., and G.C.A.; investigation, M.S.S., C.F.D., J.A.N., and H.A.S.; writing – original draft, M.S.S. and N.V.-L.; writing – final manuscript, review, and editing, A.S.M., N.V.-L., and M.S.S.; funding acquisition, N.V.-L., A.S.M., G.C.A., and D.N.W.; resources, N.V.-L., A.S.M., G.C.A., and D.N.W.; supervision, M.S.S., G.C.A., and D.N.W.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: August 25, 2023
Revised: November 8, 2023
Accepted: December 1, 2023
Published: January 5, 2024

REFERENCES

1. Filbeck, S., Cerullo, F., Pfeffer, S., and Joazeiro, C.A.P. (2022). Ribosome-associated quality-control mechanisms from bacteria to humans. Mol. Cell *82*, 1451–1466.
2. Müller, C., Crowe-McAuliffe, C., and Wilson, D.N. (2021). Ribosome rescue pathways in bacteria. Front. Microbiol. *12*, 652980.
3. Howard, C.J., and Frost, A. (2021). Ribosome-associated quality control and CAT tailing. Crit. Rev. Biochem. Mol. Biol. *56*, 603–620.
4. Becker, T., Franckenberg, S., Wickles, S., Shoemaker, C.J., Anger, A.M., Armache, J.P., Sieber, H., Ungewickell, C., Berninghausen, O., Daberkow, I., et al. (2012). Structural basis of highly conserved ribosome recycling in eukaryotes and archaea. Nature *482*, 501–506.
5. Pisareva, V.P., Skabkin, M.A., Hellen, C.U., Pestova, T.V., and Pisarev, A.V. (2011). Dissociation by Pelota, Hbs1 and ABCE1 of mammalian vacant 80S ribosomes and stalled elongation complexes. EMBO J. *30*, 1804–1817.
6. Franckenberg, S., Becker, T., and Beckmann, R. (2012). Structural view on recycling of archaeal and eukaryotic ribosomes after canonical termination and ribosome rescue. Curr. Opin. Struct. Biol. *22*, 786–796.
7. Buskirk, A.R., and Green, R. (2017). Ribosome pausing, arrest and rescue in bacteria and eukaryotes. Philos. Trans. R. Soc. Lond. B Biol. Sci. *372*, 20160183.
8. Joazeiro, C.A.P. (2019). Mechanisms and functions of ribosome-associated protein quality control. Nat. Rev. Mol. Cell Biol. *20*, 368–383.
9. Lyumkis, D., Oliveira dos Passos, D., Tahara, E.B., Webb, K., Bennett, E.J., Vinterbo, S., Potter, C.S., Carragher, B., and Joazeiro, C.A. (2014). Structural basis for translational surveillance by the large ribosomal subunit-associated protein quality control complex. Proc. Natl. Acad. Sci. USA *111*, 15981–15986.
10. Shen, P.S., Park, J., Qin, Y., Li, X., Parsawar, K., Larson, M.H., Cox, J., Cheng, Y., Lambowitz, A.M., Weissman, J.S., et al. (2015). Protein synthesis. Rqc2p and 60S ribosomal subunits mediate mRNA-independent elongation of nascent chains. Science *347*, 75–78.
11. Kostova, K.K., Hickey, K.L., Osuna, B.A., Hussmann, J.A., Frost, A., Weinberg, D.E., and Weissman, J.S. (2017). CAT-tailing as a fail-safe mechanism for efficient degradation of stalled nascent polypeptides. Science *357*, 414–417.
12. Osuna, B.A., Howard, C.J., Kc, S., Frost, A., and Weinberg, D.E. (2017). In vitro analysis of RQC activities provides insights into the mechanism and function of CAT tailing. eLife *6*, 10.7554.
13. Bengtson, M.H., and Joazeiro, C.A. (2010). Role of a ribosome-associated E3 ubiquitin ligase in protein quality control. Nature *467*, 470–473.
14. Brandman, O., Stewart-Ornstein, J., Wong, D., Larson, A., Williams, C.C., Li, G.W., Zhou, S., King, D., Shen, P.S., Weibezahn, J., et al. (2012). A ribosome-bound quality control complex triggers degradation of nascent peptides and signals translation stress. Cell *151*, 1042–1054.
15. Defenouillère, Q., Yao, Y., Mouaikel, J., Namane, A., Galopier, A., Decourty, L., Doyen, A., Malabat, C., Saveanu, C., Jacquier, A., et al. (2013). Cdc48-associated complex bound to 60S particles is required for the clearance of aberrant translation products. Proc. Natl. Acad. Sci. USA *110*, 5046–5051.
16. Thrun, A., Garzia, A., Kigoshi-Tansho, Y., Patil, P.R., Umbaugh, C.S., Dallinger, T., Liu, J., Kreger, S., Patrizi, A., Cox, G.A., et al. (2021). Convergence of mammalian RQC and C-end rule proteolytic pathways via alanine tailing. Mol. Cell *81*, 2112–2122.e7.
17. Patil, P.R., Burroughs, A.M., Misra, M., Cerullo, F., Dikic, I., Aravind, L., and Joazeiro, C.A.P. (2023). Mechanism and evolutionary origins of alanine-tail C-degron recognition by E3 ligases Pirh2 and CRL2-KLHDC10. Cell Rep. *42*, 113100.
18. Lytvynenko, I., Paternoga, H., Thrun, A., Balke, A., Müller, T.A., Chiang, C.H., Nagler, K., Tsaprailis, G., Anders, S., Bischofs, I., et al. (2019). Alanine tails signal proteolysis in bacterial ribosome-associated quality control. Cell *178*, 76–90.e22.
19. Cerullo, F., Filbeck, S., Patil, P.R., Hung, H.C., Xu, H., Vornberger, J., Hofer, F.W., Schmitt, J., Kramer, G., Bukau, B., et al. (2022). Bacterial ribosome collision sensing by a MutS DNA repair ATPase paralogue. Nature *603*, 509–514.
20. Park, E., Mackens-Kiani, T., Berhane, R., Esser, H., Erdenebat, C., Burroughs, A.M., Berninghausen, O., Aravind, L., Beckmann, R., Green, R., et al. (2023). *B. subtilis* MutS2 splits stalled ribosomes into subunits without mRNA cleavage. Preprint at bioRxiv.
21. Filbeck, S., Cerullo, F., Paternoga, H., Tsaprailis, G., Joazeiro, C.A.P., and Pfeffer, S. (2021). Mimicry of canonical translation elongation underlies alanine tail synthesis in RQC. Mol. Cell *81*, 104–114.e6.
22. Crowe-McAuliffe, C., Takada, H., Murina, V., Polte, C., Kasvandik, S., Tenson, T., Ignatova, Z., Atkinson, G.C., Wilson, D.N., and Hauryliuk, V. (2021). Structural basis for bacterial ribosome-associated quality control by RqcH and RqcP. Mol. Cell *81*, 115–126.e7.
23. Chadani, Y., Niwa, T., Izumi, T., Sugata, N., Nagao, A., Suzuki, T., Chiba, S., Ito, K., and Taguchi, H. (2017). Intrinsic ribosome destabilization underlies translation and provides an organism with a strategy of environmental sensing. Mol. Cell *68*, 528–539.e5.
24. Chadani, Y., Sugata, N., Niwa, T., Ito, Y., Iwasaki, S., and Taguchi, H. (2021). Nascent polypeptide within the exit tunnel stabilizes the ribosome to counteract risky translation. EMBO J. *40*, e108299.
25. Rudra, P., Hurst-Hess, K.R., Cotten, K.L., Partida-Miranda, A., and Ghosh, P. (2020). Mycobacterial HflX is a ribosome splitting factor that mediates antibiotic resistance. Proc. Natl. Acad. Sci. USA *117*, 629–634.
26. Jiang, L., Schaffitzel, C., Bingel-Erlenmeyer, R., Ban, N., Korber, P., Koning, R.I., de Geus, D.C., Plaisier, J.R., and Abrahams, J.P. (2009). Recycling of aborted ribosomal 50S subunit-nascent chain-tRNA complexes by the heat shock protein Hsp15. J. Mol. Biol. *386*, 1357–1367.
27. Safdari, H.A., Kasvandik, S., Polte, C., Ignatova, Z., Tenson, T., and Wilson, D.N. (2022). Structure of Escherichia coli heat shock protein Hsp15 in complex with the ribosomal 50S subunit bearing peptidyl-tRNA. Nucleic Acids Res. *50*, 12515–12526.

28. SenGupta, S., Mondal, A., Dutta, D., and Parrack, P. (2018). HflX protein protects *Escherichia coli* from manganese stress. J. Biosci. *43*, 1001–1013.

29. Bhattacharjee, S., Feng, X., Maji, S., Dadhwal, P., Zhang, Z., Brown, Z.P., and Frank, J. (2023). Time resolution in cryo-EM using a novel PDMS-based microfluidic chip assembly and its application to the study of HflX-mediated ribosome recycling. Preprint at bioRxiv.

30. Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G., Azevedo, V., Bertero, M.G., Bessières, P., Bolotin, A., Borchert, S., et al. (1997). The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. Nature *390*, 249–256.

31. ten Broeke-Smits, N.J., Pronk, T.E., Jongerius, I., Bruning, O., Wittink, F.R., Breit, T.M., van Strijp, J.A., Fluit, A.C., and Boel, C.H. (2010). Operon structure of *Staphylococcus aureus*. Nucleic Acids Res. *38*, 3263–3274.

32. Warrier, I., Ram-Mohan, N., Zhu, Z., Hazery, A., Echlin, H., Rosch, J., Meyer, M.M., and van Opijnen, T. (2018). The transcriptional landscape of *Streptococcus pneumoniae* TIGR4 reveals a complex operon architecture and abundant riboregulation critical for growth and virulence. PLoS Pathog. *14*, e1007461.

33. Peters, J.M., Colavin, A., Shi, H., Czarny, T.L., Larson, M.H., Wong, S., Hawkins, J.S., Lu, C.H.S., Koo, B.M., Marta, E., et al. (2016). A comprehensive, CRISPR-based functional analysis of essential genes in bacteria. Cell *165*, 1493–1506.

34. Verma, R., Reichermeier, K.M., Burroughs, A.M., Oania, R.S., Reitsma, J.M., Aravind, L., and Deshaies, R.J. (2018). Vms1 and ANKZF1 peptidyl-tRNA hydrolases release nascent chains from stalled ribosomes. Nature *557*, 446–451.

35. Yip, M.C.J., Keszei, A.F.A., Feng, Q., Chu, V., McKenna, M.J., and Shao, S. (2019). Mechanism for recycling tRNAs on stalled ribosomes. Nat. Struct. Mol. Biol. *26*, 343–349.

36. Su, T., Izawa, T., Thoms, M., Yamashita, Y., Cheng, J., Berninghausen, O., Hartl, F.U., Inada, T., Neupert, W., and Beckmann, R. (2019). Structure and function of Vms1 and Arb1 in RQC and mitochondrial proteome homeostasis. Nature *570*, 538–542.

37. Kuroha, K., Zinoviev, A., Hellen, C.U.T., and Pestova, T.V. (2018). Release of ubiquitinated and non-ubiquitinated nascent chains from stalled mammalian ribosomal complexes by ANKZF1 and Ptrh1. Mol. Cell *72*, 286–302.e8.

38. Svetlov, M.S. (2021). Ribosome-associated quality control in bacteria. Biochemistry (Mosc) *86*, 942–951.

39. Menninger, J.R. (1976). Peptidyl transfer RNA dissociates during protein synthesis from ribosomes of *Escherichia coli*. J. Biol. Chem. *251*, 3392–3398.

40. Nagao, A., Nakanishi, Y., Yamaguchi, Y., Mishina, Y., Karoji, M., Toya, T., Fujita, T., Iwasaki, S., Miyauchi, K., Sakaguchi, Y., et al. (2023). Quality control of protein synthesis in the early elongation stage. Nat. Commun. *14*, 2704.

41. Overbeek, R., Fonstein, M., D'Souza, M., Pusch, G.D., and Maltsev, N. (1999). The use of gene clusters to infer functional coupling. Proc. Natl. Acad. Sci. USA *96*, 2896–2901.

42. Saito, K., Kratzat, H., Campbell, A., Buschauer, R., Burroughs, A.M., Berninghausen, O., Aravind, L., Green, R., Beckmann, R., and Buskirk, A.R. (2022). Ribosome collisions induce mRNA cleavage and ribosome rescue in bacteria. Nature *603*, 503–508.

43. Kolb, V.A., Makeyev, E.V., and Spirin, A.S. (1994). Folding of firefly luciferase during translation in a cell-free system. EMBO J. *13*, 3631–3637.

44. Svetlov, M.S., Kommer, A., Kolb, V.A., and Spirin, A.S. (2006). Effective co-translational folding of firefly luciferase without chaperones of the Hsp70 family. Protein Sci. *15*, 242–247.

45. Aviner, R. (2020). The science of puromycin: from studies of ribosome function to applications in biotechnology. Comput. Struct. Biotechnol. J. *18*, 1074–1083.

46. Takada, H., Crowe-McAuliffe, C., Polte, C., Sidorova, Z.Y., Murina, V., Atkinson, G.C., Konevega, A.L., Ignatova, Z., Wilson, D.N., and Hauryliuk, V. (2021). RqcH and RqcP catalyze processive poly-alanine synthesis in a reconstituted ribosome-associated quality control system. Nucleic Acids Res. *49*, 8355–8369.

47. Sohmen, D., Chiba, S., Shimokawa-Chiba, N., Innis, C.A., Berninghausen, O., Beckmann, R., Ito, K., and Wilson, D.N. (2015). Structure of the Bacillus subtilis 70S ribosome reveals the basis for species-specific stalling. Nat. Commun. *6*, 6941.

48. Ito, K., Murakami, R., Mochizuki, M., Qi, H., Shimizu, Y., Miura, K., Ueda, T., and Uchiumi, T. (2012). Structural basis for the substrate recognition and catalysis of peptidyl-tRNA hydrolase. Nucleic Acids Res. *40*, 10521–10531.

49. Nilsson, O.B., Hedman, R., Marino, J., Wickles, S., Bischoff, L., Johansson, M., Müller-Lucks, A., Trovato, F., Puglisi, J.D., O'Brien, E.P., et al. (2015). Cotranslational protein folding inside the ribosome exit tunnel. Cell Rep. *12*, 1533–1540.

50. Schmeing, T.M., Huang, K.S., Kitchen, D.E., Strobel, S.A., and Steitz, T.A. (2005). Structural insights into the roles of water and the 2′ hydroxyl of the P site tRNA in the peptidyl transferase reaction. Mol. Cell *20*, 437–448.

51. Svidritskiy, E., Ling, C., Ermolenko, D.N., and Korostelev, A.A. (2013). Blasticidin S inhibits translation by trapping deformed tRNA on the ribosome. Proc. Natl. Acad. Sci. USA *110*, 12283–12288.

52. Tsai, K., Stojković, V., Lee, D.J., Young, I.D., Szal, T., Klepacki, D., Vázquez-Laslop, N., Mankin, A.S., Fraser, J.S., and Fujimori, D.G. (2022). Structural basis for context-specific inhibition of translation by oxazolidinone antibiotics. Nat. Struct. Mol. Biol. *29*, 162–171.

53. Syroegin, E.A., Flemmich, L., Klepacki, D., Vazquez-Laslop, N., Micura, R., and Polikanov, Y.S. (2022). Structural basis for the context-specific action of the classic peptidyl transferase inhibitor chloramphenicol. Nat. Struct. Mol. Biol. *29*, 152–161.

54. Moazed, D., and Noller, H.F. (1991). Sites of interaction of the CCA end of peptidyl-tRNA with 23S rRNA. Proc. Natl. Acad. Sci. USA *88*, 3725–3728.

55. Samaha, R.R., Green, R., and Noller, H.F. (1995). A base pair between tRNA and 23S rRNA in the peptidyl transferase centre of the ribosome. Nature *377*, 309–314.

56. Wilson, D.N., and Beckmann, R. (2011). The ribosomal tunnel as a functional environment for nascent polypeptide folding and translational stalling. Curr. Opin. Struct. Biol. *21*, 274–282.

57. Kössel, H. (1970). Purification and properties of peptidyl-tRNA hydrolase from Escherichia coli. Biochim. Biophys. Acta *204*, 191–202.

58. van der Stel, A.X., Gordon, E.R., SenGupta, A., Martínez, A.K., Klepacki, D., Perry, T.N., Herrero Del Valle, A., Vázquez-Laslop, N., Sachs, M.S., Cruz-Vera, L.R., et al. (2021). Structural basis for the tryptophan sensitivity of TnaC-mediated ribosome stalling. Nat. Commun. *12*, 5340.

59. Beckert, B., Leroy, E.C., Sothiselvam, S., Bock, L.V., Svetlov, M.S., Graf, M., Arenz, S., Abdelshahid, M., Seip, B., Grubmüller, H., et al. (2021). Structural and mechanistic basis for translation inhibition by macrolide and ketolide antibiotics. Nat. Commun. *12*, 4466.

60. Li, W., Chang, S.T., Ward, F.R., and Cate, J.H.D. (2020). Selective inhibition of human translation termination by a drug-like compound. Nat. Commun. *11*, 4941.

61. Huter, P., Arenz, S., Bock, L.V., Graf, M., Frister, J.O., Heuer, A., Peil, L., Starosta, A.L., Wohlgemuth, I., Peske, F., et al. (2017). Structural basis for polyproline-mediated ribosome stalling and rescue by the translation elongation factor EF-P. Mol. Cell *68*, 515–527.e6.

62. Chandrasekaran, V., Juszkiewicz, S., Choi, J., Puglisi, J.D., Brown, A., Shao, S., Ramakrishnan, V., and Hegde, R.S. (2019). Mechanism of ribosome stalling during translation of a poly(A) tail. Nat. Struct. Mol. Biol. *26*, 1132–1140.

63. Menninger, J.R., Mulholland, M.C., and Stirewalt, W.S. (1970). Peptidyl-tRNA hydrolase and protein chain termination. Biochim. Biophys. Acta *217*, 496–511.

64. Menninger, J.R. (1979). Accumulation of peptidyl tRNA is lethal to *Escherichia coli*. J. Bacteriol. *137*, 694–696.

65. Vivanco-Domínguez, S., Bueno-Martínez, J., León-Avila, G., Iwakura, N., Kaji, A., Kaji, H., and Guarneros, G. (2012). Protein synthesis factors (RF1, RF2, RF3, RRF, and tmRNA) and peptidyl-tRNA hydrolase rescue stalled ribosomes at sense codons. J. Mol. Biol. *417*, 425–439.

66. Korber, P., Stahl, J.M., Nierhaus, K.H., and Bardwell, J.C. (2000). Hsp15: a ribosome-associated heat shock protein. EMBO J. *19*, 741–748.

67. Jain, N., Dhimole, N., Khan, A.R., De, D., Tomar, S.K., Sajish, M., Dutta, D., Parrack, P., and Prakash, B. (2009). *E. coli* HflX interacts with 50S ribosomal subunits in presence of nucleotides. Biochem. Biophys. Res. Commun. *379*, 201–205.

68. Zhang, Y., Mandava, C.S., Cao, W., Li, X., Zhang, D., Li, N., Zhang, Y., Zhang, X., Qin, Y., Mi, K., et al. (2015). HflX is a ribosome-splitting factor rescuing stalled ribosomes under stress conditions. Nat. Struct. Mol. Biol. *22*, 906–913.

69. Coatham, M.L., Brandon, H.E., Fischer, J.J., Schümmer, T., and Wieden, H.J. (2016). The conserved GTPase HflX is a ribosome splitting factor that binds to the E-site of the bacterial ribosome. Nucleic Acids Res. *44*, 1952–1961.

70. Duval, M., Dar, D., Carvalho, F., Rocha, E.P.C., Sorek, R., and Cossart, P. (2018). HflXr, a homolog of a ribosome-splitting factor, mediates antibiotic resistance. Proc. Natl. Acad. Sci. USA *115*, 13359–13364.

71. Ramadoss, N.S., Alumasa, J.N., Cheng, L., Wang, Y., Li, S., Chambers, B.S., Chang, H., Chatterjee, A.K., Brinker, A., Engels, I.H., et al. (2013). Small molecule inhibitors of trans-translation have broad-spectrum antibiotic activity. Proc. Natl. Acad. Sci. USA *110*, 10282–10287.

72. Aron, Z.D., Mehrani, A., Hoffer, E.D., Connolly, K.L., Srinivas, P., Torhan, M.C., Alumasa, J.N., Cabrera, M., Hosangadi, D., Barbor, J.S., et al. (2021). trans-Translation inhibitors bind to a novel site on the ribosome and clear *Neisseria gonorrhoeae* in vivo. Nat. Commun. *12*, 1799.

73. Menninger, J.R., and Otto, D.P. (1982). Erythromycin, carbomycin, and spiramycin inhibit protein synthesis by stimulating the dissociation of peptidyl-tRNA from ribosomes. Antimicrob. Agents Chemother. *21*, 811–818.

74. Menninger, J.R., and Coleman, R.A. (1993). Lincosamide antibiotics stimulate dissociation of peptidyl-tRNA from ribosomes. Antimicrob. Agents Chemother. *37*, 2027–2029.

75. Tomasi, F.G., Schweber, J.T.P., Kimura, S., Zhu, J., Cleghorn, L.A.T., Davis, S.H., Green, S.R., Waldor, M.K., and Rubin, E.J. (2023). Peptidyl tRNA hydrolase is required for robust prolyl-tRNA turnover in *Mycobacterium tuberculosis*. mBio *14*, e0346922.

76. Wruck, F., Tian, P., Kudva, R., Best, R.B., von Heijne, G., Tans, S.J., and Katranidis, A. (2021). The ribosome modulates folding inside the ribosomal exit tunnel. Commun. Biol. *4*, 523.

77. Cruz-Vera, L.R., Toledo, I., Hernández-Sánchez, J., and Guarneros, G. (2000). Molecular basis for the temperature sensitivity of Escherichia coli pth(Ts). J. Bacteriol. *182*, 1523–1528.

78. Orelle, C., Carlson, E.D., Szal, T., Florin, T., Jewett, M.C., and Mankin, A.S. (2015). Protein synthesis by ribosomes with tethered subunits. Nature *524*, 119–124.

79. Kitagawa, M., Ara, T., Arifuzzaman, M., Ioka-Nakamichi, T., Inamoto, E., Toyonaga, H., and Mori, H. (2005). Complete set of ORF clones of *Escherichia coli* ASKA library (a complete set of E. coli K-12 ORF archive): unique resources for biological research. DNA Res. *12*, 291–299.

80. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780.

81. Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics *25*, 1972–1973.

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

83. Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M., and Barton, G.J. (2009). Jalview, version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics *25*, 1189–1191.

84. Saha, C.K., Sanches Pires, R., Brolin, H., Delannoy, M., and Atkinson, G.C. (2021). FlaGs and webFlaGs: discovering novel biology through the analysis of gene neighbourhood conservation. Bioinformatics *37*, 1312–1314.

85. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera–a visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612.

86. Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods *9*, 671–675.

87. Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., 3rd, and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods *6*, 343–345.

88. Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., and Vinh, L.S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. Mol. Biol. Evol. *35*, 518–522.

89. Kunst, F., and Rapoport, G. (1995). Salt stress is an environmental signal affecting degradative enzyme synthesis in Bacillus subtilis. J. Bacteriol. *177*, 2403–2407.

90. Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K., and Ueda, T. (2001). Cell-free translation reconstituted with purified components. Nat. Biotechnol. *19*, 751–755.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
B. subtilis 168	BGSC	1A1
B. subtilis dCas9	BGSC	1A1278
B. subtilis dCas9/sgRNA^{pth}	BGSC	BEC00530
B. subtilis ΔrqcH::spcR	Crowe-McAuliffe et al.[22]	VHB256
B. subtilis dCas9/sgRNA^{pth} *ΔrqcH::spcR*	This paper	
E. coli C600	Cruz-Vera et al.[77]	N/A
E. coli C600 *pth*(Ts)	Cruz-Vera et al.[77]	N/A
E. coli SQ171 pAM55-2	Orelle et al.[78]	N/A
E. coli JM109	Promega	PR-L2001
E. coli BL21(DE3)	Sigma	CMC-0016
Chemicals, peptides, and recombinant proteins		
Spectinomycin	Fisher Scientific	Cat#215899305
Chloramphenicol	Fisher Scientific	BP904-100
Erythromycin	Sigma-Aldrich	E6376
Ampicillin	Sigma-Aldrich	A9518
Xylose	Fisher Scientific	AC141001000
RNase-free DNase I	Roche	Cat#04716728001
Lysozyme	Sigma-Aldrich	L4919-1G
EasyTag™ EXPRESS35S Protein Labeling Mix, [35S]-	Perkin Elmer	NEG772002MC
D-Luciferin	Fisher Scientific	Cat#88293
Puromycin	Sigma-Aldrich	P7255
TPEN	Sigma-Aldrich	P4413
ErmDL-tRNA	Syroegin et al.[53]	N/A
Complete Protease Inhibitor cocktail	Sigma-Aldrich	Cat#11836153001
Critical commercial assays		
PURExpress Δ (aa, tRNAs) Kit	NEB	E6840S
PURExpress Δ Ribosome Kit	NEB	E3313S
Custom PURExpress Δ (aa, tRNAs) Δ Ribosome Kit	NEB	N/A
Phusion High-Fidelity DNA polymerase	NEB	M0530S
Gibson Assembly Master Mix	NEB	E2611S
Deposited data		
Deposited data	This paper	https://data.mendeley.com/datasets/kng9rzcx3d/1
Oligonucleotides		
See Table S2	This paper	N/A
Recombinant DNA		
pCA24N-*pth*(Ec)	Kitagawa et al.[79]	N/A
pCA24N-*pth*(Bs)	This paper	N/A
pET24d-rqcH-HTF	Takada et al.[46]	N/A
pET24d-rqcP-TEV-His$_6$	Takada et al.[46]	N/A
pLAS	Nilsson et al.[49]	N/A
pBESTluc™	Promega	L492A-C

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
MAFFT L-INS-i version 7.453	Katoh and Standley[80]	https://mafft.cbrc.jp/alignment/software/
HMMER v3.3.2	HMMER	hmmer.org
TrimAl v1.2	Capella-Gutierrez et al.[81]	http://trimal.cgenomics.org/downloads
IQ-TREE v1.6.12	Nguyen et al.[82]	http://www.iqtree.cibiv.univie.ac.at/
JalView v.2.11.2.0	Waterhouse et al.[83]	https://www.jalview.org/download/
FigTree v.1.4.4	Andrew Rambaut Group	https://github.com/rambaut/figtree/
FlaGs	Saha et al.[84]	https://github.com/GCA-VH-lab/FlaGs2
UCSF Chimera	Pettersen et al.[85]	https://www.cgl.ucsf.edu/chimera/download.html
Prism 9	GraphPad	https://www.graphpad.com/
ImageJ	Schneider et al.[86]	https://imagej.net/ij/index.html
R version 4.2.2	The R Project for Statistical Computing	https://cran.r-project.org/mirrors.html
Other		
Zirconium Beads	BioSpec	Cat#11079101Z
HisTrap HP column	GE Healthcare	Cat#17-5248-01
HiTrap Butyl FF column	GE Healthcare	Cat#17-5197-01
Microplate reader Infinite M200Pro	Tecan	Cat#30050303
Prep homogenizer	MP Biomedicals	Cat#116004500

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Maxim Svetlov (msvet2@uic.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

This paper analyzed publicly available genomic data from the National Center for Biotechnology Information using software described in the papers cited in the References section.

This paper does not report original code.

The original data for the gels and agar plates are deposited to Mendeley (Mendeley Data: https://data.mendeley.com/datasets/kng9rzcx3d/1).

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANTS DETAILS

Strains, plasmids, and growth medium and primers

The *B. subtilis* wild-type strain 168[30] and its derivative expressing dCas9 (designated as 'wild type') and dCas9/sgRNA^{pth} (designated as '*pth*-KD')[33] were obtained from the Bacillus Genetic Stock Center (BGSC). The *rqcH* gene in these strains was deleted as described below. The *B. subtilis ΔrqcH* and *ΔrqcP* strains[22] were provided by Vasilii Hauryliuk (Lund University). The *E. coli* strain JM109, used for expression of plasmid-encoded *B. subtilis* or *E. coli* Pth, and strain BL21(DE3) utilized for expression of recombinant *B. subtilis* RqcH and RqcP were obtained from Promega.

The plasmid pCA24N-*pth*(Ec) encoding the N-terminally His_6-tagged Pth was from the ASKA library.[79] The plasmids pET24d-rqcH-HTF and pET24d-rqcP-TEV-His_6 encoding recombinant RqcH and RqcP proteins were provided by Vasilii Hauryliuk (Lund University). The plasmid pLAS encoding the LepB-ADR1a-SecM sequence[49] was provided by Gunnar von Heijne (Stockholm University). The plasmid pCA24N-*pth*(Bs) encoding N-terminally His_6-tagged *B. subtilis* Pth was prepared in this study.

The lysogeny broth (LB) medium [1% yeast extract, 0.5% tryptone, 1% NaCl (all w/v)] was used for growth of *B. subtilis* and *E. coli* cells.

The sequences of all oligonucleotides are specified in Table S2, the relevant sequences of the templates used for in vitro translation are shown in Table S3.

METHOD DETAILS

DNA manipulations

Construction of the pCA24N-pth(Bs) plasmid

All PCR reactions were performed using Phusion®High-Fidelity DNA polymerase (New England Biolabs) and PCR products were purified using PCR-purification kit (Zymo Research).

The *B. subtilis pth* gene was PCR-amplified from chromosomal DNA of the strain 168 using a mixture of four primers: Fwd-PTH_{Bs}-1 (0.05 μM), Fwd-PTH_{Bs}-2 (0.5 μM), Rev-PTH_{Bs}-1 (0.05 μM), and Rev-PTH_{Bs}-2 (0.5 μM). The amplified gene was used to replace the *E. coli pth* gene in the pCA24N-*pth(Ec)* plasmid by Gibson assembly.[87]

Constructing DNA templates for *in vitro* transcription-translation

DNA templates for in vitro transcription-translation specifying non-stop mRNAs encoding firefly luciferase (Luc) (untailed or C-terminally tailed) were prepared by PCR-amplifying the *luc* gene from the pBESTluc™ plasmid (Promega) using primers Fwd-T7-Luc (0.5 μM), Fwd-Luc (0.05 μM), and 0.5 μM of one of the specific reverse primers designed to obtain luciferase proteins with the desired C-terminal tails (Table S2).

DNA templates for non-stop mRNAs encoding the peptides with the ADR1a domain were prepared in two consecutive PCR reactions. First, the ADR1a-encoding sequence was amplified from plasmid LepB-ADR1a-SecM[49] using forward primer Fwd-ADR1a and one of the specific Rev-ADR1a reverse primers to obtain the desired linker sequence. The resultant DNA products were then used as templates for the second round of PCR using Fwd-T7-ADR1a, introducing the T7 RNA polymerase promoter and the ribosome binding site, and one of the specific SecM reverse primers (Rev-SecM, Rev-SecM-A_3 or Rev-SecM-A_7) designed to obtain ADR1a-containing peptide with the desired C-terminal tails (Table S2).

Phylogenetic analysis

Representative sequences of RqcP and Hsp15 were obtained from a previous phylogenetic study (the mmc3 dataset) and aligned with MAFFT L-INS-i version 7.453.[80] Hidden Markov Models (HMMs) profiles were built using the alignments with HMMER v3.3.2 hmmbuild.

The proteome from complete bacterial reference genomes (May 4, 2023) were downloaded from NCBP FTP. RqcP and Hsp15 homologues were identified using the HMMs profiles with HMMER v3.3.2 hmmscan with an E value cut-off of $1e^{-10}$. Sequences were aligned with MAFFT L-INS-i version 7.453, with alignment positions with >50% gaps removed using TrimAl v1.2.[81]

Phylogenetic analysis was executed using IQ-TREE v1.6.12[82] (http://iqtree.cibiv.univie.ac.at/,) with default parameters and 1000 ultrafast bootstrap replicates.[88] Within the analysis, 'LG + I + G' was automatically selected and used as the best-fit substitution model. JalView v.2.11.2.0[83] and FigTree v.1.4.4 (https://github.com/rambaut/figtree/) were used respectively for alignment and phylogenetic tree visualization.

Gene distance and neighborhood analysis

Homologues from Pth and RqcH were identified using HMMER v3.3.2 Jackhmmer with a maximum number of 100 iterations and an E value cut-off of $1e^{-10}$. Gene distance, defined as the number of coding genes between *rqcP* and *pth*, was determined from the gtf genome files. Gene neighborhood analysis was performed using FlaGs[84] (https://github.com/GCA-VH-lab/FlaGs2) using Pth as query with 11 flanking genes. R version 4.2.2 was used to create and perform the gene distance distribution figures and statistical analyses.

Construction of *B. subtilis* strains

To delete the *rqcH* gene, the dCas9 (wild type) and dCas9/sgRNA^{pth} (*pth*-KD) *B. subtilis* strains were transformed according to the protocol described in Kunst and Rapoport[89] with a linear DNA amplified from the genome of the *B. subtilis rqcH::spc* strain[22] using primers Fwd-RqcH and Rev-RqcH. Transformants were selected on LB agar plates containing spectinomycin (100 μg/mL) and the deletion of the *rqcH* gene in the recipient cells was confirmed by PCR.

Preparation of *B. subtilis* lysates

Overnight cultures of *B. subtilis* cells were inoculated 1:1000 into 25 mL of fresh LB medium (containing or not 1% xylose) and grown at 37°C. Upon reaching exponential phase ($A_{600} \sim 0.7$), cultures were rapidly cooled down by adding equal volume of ice and cells were harvested by centrifugation at 4°C for 5 min at 5,500 rpm in JA-25.50 rotor (Beckman). Cell pellets were resuspended in 0.25 mL of lysis buffer (50 mM HEPES-KOH, pH 7.0, 100 mM NaCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$, 0.4% Triton X-100, 0.1% NP-40) containing 1 mg/mL lysozyme, 40 units of DNase I and 0.8% Complete Protease Inhibitor cocktail. Cell suspension was mixed with 290 mg of 0.1 mm zirconium beads (BioSpec Products) and lysed using Fast-Prep homogenizer for 1 min at 6.5 m/s (BaneBio). The resultant lysates were clarified by centrifugation at 4°C for 10 min at 21,000 rcf in the Eppendorf microcentrifuge, the supernatants were

collected, flash frozen in liquid nitrogen and stored at -80°C. Prior to use, the lysates were diluted with the cold lysis buffer to A_{260} of 1.0.

Preparation of 50S-NCC

The complexes of 70S ribosomes carrying different stalled peptidyl-tRNAs were generated by in vitro transcription-translation of the corresponding DNA templates in the PURExpress Δ ribosome cell-free system (New England Biolabs). The ribosomes were supplemented to a final concentration of 3.6 μM. Ribosomes from *B. subtilis* strain 168 or *E. coli* strain SQ171 pAM55-2[78] were isolated as described.[90] To generate stalled ribosomes carrying radiolabeled nascent peptides, the reactions contained 50 μM of unlabeled methionine and 1.1 μCi/μL of EasyTag Express ^{35}S Protein Labeling Mix (PerkinElmer) with a specific activity of 1175 Ci/mmol. After assembling the reaction components (final volume 50 μL), reactions were initiated by adding 0.5 μg of templates. To prevent tmRNA-mediated decay of 70S:peptidyl-tRNA complexes accumulated during transcription-translation, all reactions were supplemented with 5 μM of anti-tmRNA DNA oligonucleotide (Table S2). After 1h incubation at 37°C (for the reactions with *B. subtilis* ribosomes) or at 30°C (for the reactions with *E. coli* ribosomes), samples were diluted with an equal volume (50 μL) of cold 10/100 buffer containing 20 mM HEPES-KOH, pH 7.5, 10 mM $Mg(OAc)_2$, 100 mM KOAc, 2 mM DTT and loaded onto 5-20% sucrose gradient in ribosome dissociation buffer (20 mM HEPES-KOH pH 7.5, 1 mM $Mg(OAc)_2$, 100 mM KOAc, 2 mM DTT). Gradients were centrifuged at 40,000 rpm for 2.5 hours at 4°C in a SW41Ti rotor and fractionated using Gradient Fractionator (BioComp Instruments). Fractions containing the Luc-based 50S-NCC were collected, flash frozen in liquid nitrogen, and stored at -80°C. The similarly isolated 50S-NCCs from the reactions generating [^{35}S]-labeled nascent ADR1a peptides, were pelleted by centrifugation at 100,000 rpm for 5 h in a S110AT rotor (Hitachi). The pellets were resuspended in 25 μL of 10/100 buffer, flash frozen in liquid nitrogen, and stored at -80°C.

Protein expression and purification

B. subtilis RqcH and RqcP were expressed in the *E. coli* BL21(DE3) strain and purified generally following the described procedure[46] with minor modifications. Specifically, for RqcP purification, fractions containing RqcP eluted from a 5 mL HisTrap HP column were combined and dialyzed against buffer 25 mM HEPES-KOH, pH 7.5, 5 mM $MgCl_2$, 500 mM KCl, 10% glycerol, and 1 mM β-mercaptoethanol. The proteins was aliquoted, flash frozen in liquid nitrogen, and stored at -80°C.

B. subtilis and *E. coli* Pth proteins were expressed in *E. coli* JM109 cells. Colonies of the freshly transformed cells were inoculated into 1 L LB medium containing 20 μg/mL chloramphenicol and grown at 37°C with intense agitation. When the cultures reached A_{600} of 0.5, IPTG was added to the final concentration of 1 mM and growth continued for an additional 2 h at 37°C. Cells were harvested by centrifugation, resuspended in 30 mL binding buffer (20 mM HEPES-KOH, pH 7.5, 10 mM $Mg(OAc)_2$, 100 mM KOAc, and 0.5 mM DTT) supplemented with 1% Complete Protease Inhibitor cocktail, and lysed by one passage through a French press (SLM Aminco, Inc). Cell debris was removed by centrifugation at 17,000 rpm for 30 min at 4°C in a JA-25.50 rotor (Beckman Coulter). Clarified lysates were filtered through a 0.2 μm syringe filter and loaded onto a 5 mL HisTrap HP column (Cytiva) pre-equilibrated in binding buffer. The column was washed with 5 column volumes of binding buffer and then with 5 volumes of the same buffer supplemented with 20 mM imidazole. The Pth was eluted with a linear gradient of imidazole (20-500 mM) in binding buffer. Pth-containing fractions were combined and dialyzed against either 10/100 buffer (for *B. subtilis* Pth) or 10 mM Tris-HCl, pH 7.5 (for *E. coli* Pth). After dialysis, proteins were aliquoted, flash frozen in liquid nitrogen and stored at -80°C.

Luciferase release assay

Aliquots (25 μL) of Luc-based 50S-NCC were thawed on ice and mixed with an equal volume of the Assay Buffer (20 mM HEPES-KOH pH 7.5, 19 mM $Mg(OAc)_2$, 100 mM KOAc, 2 mM DTT, 2 mM ATP and 0.2 mM D-luciferin). The mixtures (50 μL) were transferred into wells of 384-well black/clear assay plates (Falcon) and placed into a microplate reader (Infinite M200Pro, Tecan) pre-set at 30°C. After recording the background luminescence for 2 min, the release of nascent luciferase from 50S-NCC was initiated by addition of 1 μL of bacterial lysate or purified Pth (0.15 μM). The time course of light emission reflecting accumulation of enzymatically active released protein was recorded over time.

The tailing/release reactions on Luc-tRNA 50S-NCCs were carried out at 30°C in the Buffer 20 mM HEPES-KOH pH 7.5, 10 mM $Mg(OAc)_2$, 100 mM KOAc, 2 mM DTT, 2 mM ATP supplemented with purified *B. subtilis* or *E. coli* components added to the following final concentrations: 3 nM Pth (*B. subtilis*), 30 nM RqcH (*B. subtilis*), 30 nM RqcP (*B. subtilis*), 120 nM alanyl-tRNA synthetase (*E. coli*), 50 μg/mL $tRNA^{Ala}$ (*E. coli*), and 0.2 mM L-alanine. After the specified times, puromycin was added to the reactions to the final concentration of 0.5 mM to release residual nascent Luc from 50S-NCC.

Pth-mediated pept-tRNAs hydrolysis in ADR1a-containing 50S-NCC

For Pth-mediated pept-tRNA hydrolysis, 33 pmol of 50S-NCC carrying [^{35}S]-labeled ADR1a nascent peptides were placed in 5 μL of 10/100 buffer containing 0.1 mM $ZnCl_2$ or TPEN (*N*,*N*,*N*′,*N*′-tetrakis(2-pyridinylmethyl)-1,2-ethanediamine). After incubation for 5 min at 30°C, 0.5 μL of 3 μM Pth was added to the mixtures and incubation continued for 15 min. The reactions were then mixed with 5 μL of 2x Tricine loading buffer (Bio-Rad) and resolved using Bis-Tris SDS-PAGE (https://openwetware.org/wiki/Sauer:bis-Tris_SDS-PAGE,_the_very_best) to prevent spontaneous hydrolysis of pept-tRNAs ester bonds. Gels were stained, dried, and exposed overnight to a phosphorimager screen. Radioactivity was visualized in the Typhoon Trio phosphorimager (GE Healthcare). The intensity of the bands was quantified using ImageJ.[86]

Hydrolysis of ErmDL-tRNA in cell lysate

Four pmol (100 ng) of ErmDL-tRNA[53] were incubated in 10 μL of 10/100 buffer with 0.5 μL of the *B. subtilis* lysate prepared as described above. After incubation for 15 min at 30°C, the reactions were mixed with 10 μL of 2x Tricine loading buffer (Bio-Rad) and resolved in Bis-Tris SDS gel ((https://openwetware.org/wiki/Sauer:bis-Tris_SDS-PAGE,_the_very_best). Gels were stained with ethidium bromide and tRNA bands were visualized in ChemiDoc MP Imaging System (Bio-Rad).

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical significance was assigned using Prism Version 9 (GraphPad). Each dot represents an individual experiment. The specific tests, number of experiments and dispersion and precision measures are indicated in the Figure legends.

PNAS RESEARCH ARTICLE | MICROBIOLOGY OPEN ACCESS

The structural basis of hyperpromiscuity in a core combinatorial network of type II toxin–antitoxin and related phage defense systems

Karin Ernits[a,1,2], Chayan Kumar Saha[a,1], Tetiana Brodiazhenko[b], Bhanu Chouhan[a,c], Aditi Shenoy[d], Jessica A. Buttress[e], Julián J. Duque-Pedraza[a], Veda Bojar[a], Jose A. Nakamoto[a], Tatsuaki Kurata[a], Artyom A. Egorov[a], Lena Shyrokova[a], Marcus J. O. Johansson[a], Toomas Mets[a,b], Aytan Rustamova[b], Jelisaveta Džigurski[b], Tanel Tenson[b], Abel Garcia-Pino[f], Henrik Strahl[e], Arne Elofsson[d], Vasili Hauryliuk[a,b,g,h,2], and Gemma C. Atkinson[a,h,2]

Edited by Marlene Belfort, University at Albany, State University of New York, Albany, NY; received April 4, 2023; accepted July 11, 2023

Toxin-antitoxin (TA) systems are a large group of small genetic modules found in prokaryotes and their mobile genetic elements. Type II TAs are encoded as bicistronic (two-gene) operons that encode two proteins: a toxin and a neutralizing antitoxin. Using our tool NetFlax (standing for Network-FlaGs for toxins and antitoxins), we have performed a large-scale bioinformatic analysis of proteinaceous TAs, revealing interconnected clusters constituting a core network of TA-like gene pairs. To understand the structural basis of toxin neutralization by antitoxins, we have predicted the structures of 3,419 complexes with AlphaFold2. Together with mutagenesis and functional assays, our structural predictions provide insights into the neutralizing mechanism of the hyperpromiscuous Panacea antitoxin domain. In antitoxins composed of standalone Panacea, the domain mediates direct toxin neutralization, while in multidomain antitoxins the neutralization is mediated by other domains, such as PAD1, Phd-C, and ZFD. We hypothesize that Panacea acts as a sensor that regulates TA activation. We have experimentally validated 16 NetFlax TA systems and used domain annotations and metabolic labeling assays to predict their potential mechanisms of toxicity (such as membrane disruption, and inhibition of cell division or protein synthesis) as well as biological functions (such as antiphage defense). We have validated the antiphage activity of a RosmerTA system encoded by *Gordonia* phage Kita, and used fluorescence microscopy to confirm its predicted membrane-depolarizing activity. The interactive version of the NetFlax TA network that includes structural predictions can be accessed at http://netflax.webflags.se/.

toxin | antitoxin | AlphaFold | phage | Panacea

Significance

Toxin–antitoxin systems are enigmatic components of microbial genomes, with their biological functions being a conundrum of debate for decades. Increasingly, TAs are being found to have a role in defense against bacteriophages. By mapping and experimentally validating a core combinatorial network of TA systems and high-throughput prediction of structural interfaces, we uncover the evolutionary scale of TA partner swapping and identify toxic effectors. We validate the predicted toxin:antitoxin complex interfaces of four TA systems, uncovering the evolutionary malleable mechanism of toxin neutralization by Panacea-containing PanA antitoxins. We find TAs are evolutionarily related to several other phage defense systems, cementing their role as important molecular components of the arsenal of microbial warfare.

Toxin-antitoxin (TA) systems typically consist of two adjacent, often overlapping genes that encode a toxin whose expression causes growth arrest and a cognate antitoxin that negates the toxic effect (1). Based on the nature and mode of action of the antitoxin, TA systems are classified into eight types depending on the mechanism of toxin neutralization and whether the components are RNA- or protein-based (2). The most common group of proteinaceous TA pairs is type II, where the protein antitoxin directly binds to the protein toxin to sequester it into an inert complex (2).

The first TA operon to be discovered, *ccdAB*, was identified due to its stabilizing effect on plasmids (3). The rate of TA discovery has dramatically increased as high-throughput approaches for TA identification have been developed. Systematic experimental discovery of TAs was first achieved using shotgun cloning for identification of toxic ORFs (4). As the number of sequenced genomes and known TAs have grown, sensitive sequence searching and "guilt by association"—i.e., conserved colocalization of toxin and antitoxin as a bicistronic operon—have been used for discovery of new TA systems (5–12). Our bioinformatics-driven TA identification relies on analysis of gene neighborhood conservation using an approach that is sensitive enough to find remote similarity even in small, divergent proteins (9, 13, 14).

TA systems are ubiquitous in microbial life, and their wide distribution and extreme diversity of TAs has driven the search to discover the biological roles of these systems (2). Increasingly, TAs are being discovered to mediate defense against phages (15, 16), and large-scale exploratory and focused mechanistic approaches have rapidly advanced the field (16–22).

Being frequently horizontally transferred components of accessory genomes, TA systems have patchy distributions across genomes (5). It has long been known that type II toxins and antitoxins have a degree of modularity, in that they can swap partners through evolution

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]K.E. and C.K.S. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: karin.ernits@med.lu.se, vasili.hauryliuk@med.lu.se, or gemma.atkinson@med.lu.se.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2305393120/-/DCSupplemental.

Published August 9, 2023.

(5, 8, 11, 24, 25). The hyperpromiscuous antitoxin domain Panacea is a striking example of how extensive TA partner swapping can be, with Panacea-containing antitoxins (PanA) being paired with dozens of different evolutionary and structurally unrelated toxin domains (PanTs) (14). This suggested that the Panacea domain may have inherent properties that enable it to neutralize multiple unrelated toxins through an unknown mechanism (14). However, a structural understanding of PanA-mediated neutralization has been lacking. Furthermore, while Panacea's hyperpromiscuity is remarkable, it is unclear just how much this is paralleled in other antitoxins.

In this study, we have systematically explored the TA partner swapping network using NetFlax (standing for Network-FlaGs for toxins and antitoxins), an iterative implementation of our gene neighborhood analysis tool FlaGs (13), followed by experimental validation and characterization of TA systems. We have identified 3,597 systems within which there are 278 distinct homologous clusters of proteins in 275 distinct combinations of two-gene modules. We have structurally annotated our network of TA-like two-gene architectures through high-throughput prediction of TA complex structures using AlphaFold2 (26) implemented in the FoldDock pipeline (27). Focusing on the Panacea node of the network, we have validated our structural predictions through mutagenesis. We establish that Panacea is an evolutionally malleable domain that can both inhibit toxins through direct interaction and as serve as a platform for toxin neutralization by Panacea-associated ZBD (Zn^{2+}-binding domain) and PAD1 (Panacea-Associated Domain 1) domains. The combinatorial network reveals close evolutionary relationships between classical type II TA systems and antiphage systems, specifically those that include the AAA ATPase and OLD_TOPRIM endonuclease domains such as those seen in PARIS (19), AbiLi (28) the Septu system (29), and ImmA protease-containing systems as seen in RosmerTA systems (30). We explore the network experimentally through validating 16 systems in toxicity neutralization assays and predict their potential mechanisms of toxicity through functional domain annotations and metabolic labeling assays. Finally, we validate the antiphage activity of a RosmerTA system encoded by *Gordonia* phage Kita, and use fluorescence microscopy to confirm its predicted membrane-depolarizing activity.

Results

The NetFlax Algorithm Reveals a Core Proteinaceous TA Network. To uncover a core framework of the network of TA pairs, we developed the computational tool NetFlax that identifies TA-like gene architectures in an unsupervised manner and generates a TA domain interaction network. The NetFlax principle is that if one partner gene of a TA system is found in a conserved two-gene neighborhood with an alternate partner, this is predicted as a new pair, and after "hopping" to this new partner, more partners can be found in the same way (*SI Appendix*, *SI Text* and Fig. S1). We set a requirement that each new pair must be conserved in at least eight representative genomes to be allowed to hop to a new node. As this stringency leads to missing some less well conserved systems, we improved sensitivity through adding a final guilt by association hop for each node, which only required a system to be conserved in two representative genomes.

NetFlax finished hopping after eight hopping steps, converging on a final network, having reached dead-ends for all the network lineages (Fig. 1). We initially identified 79 clusters conserved in a minimum of eight genomes. These we call D nodes (standing for central Domain nodes). After the subsequent less strict node analysis allowing conservation in two genomes with no onward hopping, we identified 234 additional nodes, which we refer to as M nodes, for Mininodes. In total, we identified 314 nodes. Toxin/antitoxin assignments are made by virtue of their lineages from the original Panacea antitoxin domain, assuming that the hopping goes from antitoxin to toxin to antitoxin etc. This assumption seems to work well on the whole; in the classical type II part of the network, our annotation of whether the cluster is a toxin or antitoxin domain matches that in the TADB database (31), and domain annotations (Dataset S1). However, we cannot be sure that our annotations hold true for the termini of the network. One extended lineage of three D nodes leading from the Rosmer/ImmA zone (D41, D95, D127, and D132 associated with 32 combined M nodes) became particularly complicated with node domain fusions, making our ability to predict toxins and antitoxins troublesome. Therefore, we decided on balance to "prune" this lineage from the core TA network of Fig. 1 (however, these lineages and their data are still available in the unpruned interactive network http://netflaxunpruned.webflags.se/ and Dataset S1).

Our final core TA network (Fig. 1 and http://netflax.webflags.se/) represents the most conserved systems of the 24,474 representative predicted proteomes. The network comprises 278 nodes, of which 107 are predicted to be toxins, and 171 are predicted to be antitoxins. These fall into 275 distinct TA node combinations. It is useful to roughly divide the network into five topological zones: i) the Panacea domain–containing systems at the core of the network, including six systems experimentally validated in our previous analyses (Dataset S1) (14), ii) the anti-toxSAS zone containing toxins related to RelA/SpoT alarmone synthetases that likely modify tRNA (nodes D8 and M14), plus their antitoxins, iii) a zone containing a hub AAA ATPase antitoxin domain (D29, Fig. 1 coordinates c6), iv) a zone containing a hub ImmA protease antitoxin domain (D25, Fig. 1 coordinates e6) and v) the largest zone of the network where many classical type II TA systems are found, with many interconnections among nodes indicating considerable partner swapping. Panacea-containing systems are the largest group of our TAs (Fig. 2), closely followed by D29 (AAA ATPase-like) and D31 (Phd-related antitoxins). Most node pairs favor a particular gene order of either toxin first or antitoxin first (the latter being the most common overall, Dataset S1). D25 (ImmA/Rosmer) and Panacea antitoxin nodes are unusual in their gene order variability (58% and 78% toxin first, respectively).

Phage Defense Systems Are Widespread in the NetFlax TA Network. For each node in the NetFlax network, we made functional predictions for a protein representative by searching with domain models including those from DefenceFinder, a database of phage defense systems (25). The latter search revealed that the AAA ATPase and ImmA zones are particularly enriched in phage defense systems. The AAA ATPase domain of node D29 (Fig. 1 coordinates c6) is found in a number of defense systems, such as AriA of the PARIS system (19), GajA of the Gabija system (32) and PtuA of the Septu system (29). GajA is a sequence-specific ATP-dependent DNA endonuclease that is inhibited by dNTP and NTP nucleotides (14). It is a two-domain protein, with an N-terminal AAA domain and a C-terminal TOPRIM (topoisomerase-primase) endonuclease domain, closely related to OLD (Overcome Lysogenization Defect) families (14). In our network, we see the latter domain can be associated with the AAA domain as a separate protein (node M70, coordinates c6). This is the same two-gene architecture as seen in AriAB of the PARIS system (19). Among the other nodes linked to D29 is D44, homologous to RloB. The RloB protein family has been observed in type I restriction-modification operons (33), and the AbiLii protein, which is part of a plasmid-encoded phage abortive infection mechanism (28). HHPred (34) indicates the RloB domain is also related to OLD_TOPRIM domains. The node D51 is homologous to the HNH nuclease domain, as seen in Septu protein PtuB (Dataset S1). Thus D29 and its cognate D51 together

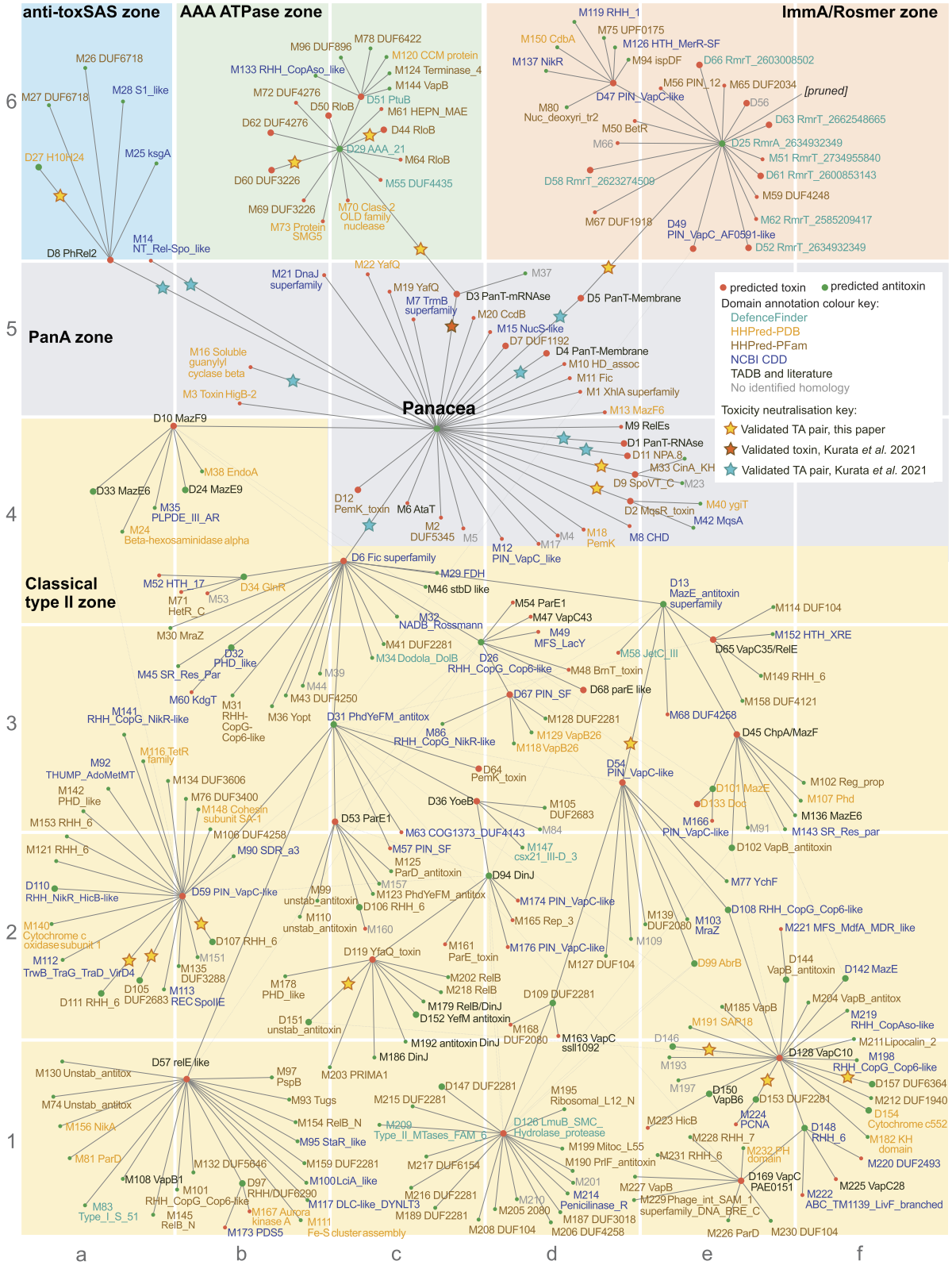


Fig. 1. The core proteinaceous TA network. The network shows connections of NetFlax-predicted toxin and antitoxin-like protein clusters across microbial life. The starting input was the Panacea domain. Green circles are predicted antitoxins and red circles are predicted toxins. Yellow stars show toxins and antitoxins validated here (Dataset S1). Dark orange and cyan stars show, respectively, toxins and TAs that have been previously validated in Kurata *et al*. (14). Predicted toxin and antitoxin domains are annotated based on sequence homology searches (see Materials and Methods for details and references).

constitutes a similar two-domain PtuAB Septu system architecture previously identified in *Bacillus thuringiensis* (29). The presence of the HEPN nuclease domain in node M61 indicates a general tendency for nuclease domains to be associated with AAA ATPase domains.

The ImmA protease domain which NetFlax predicts as an antitoxin (D25, Fig. 1 coordinates e6) was recently confirmed as such in the diverse phage defense RosmerTA systems, where different toxins (RmrTs) are paired with the protease domain–containing antitoxin (RmrA) (18, 30, 35) (Fig. 1). The ImmA domain is named

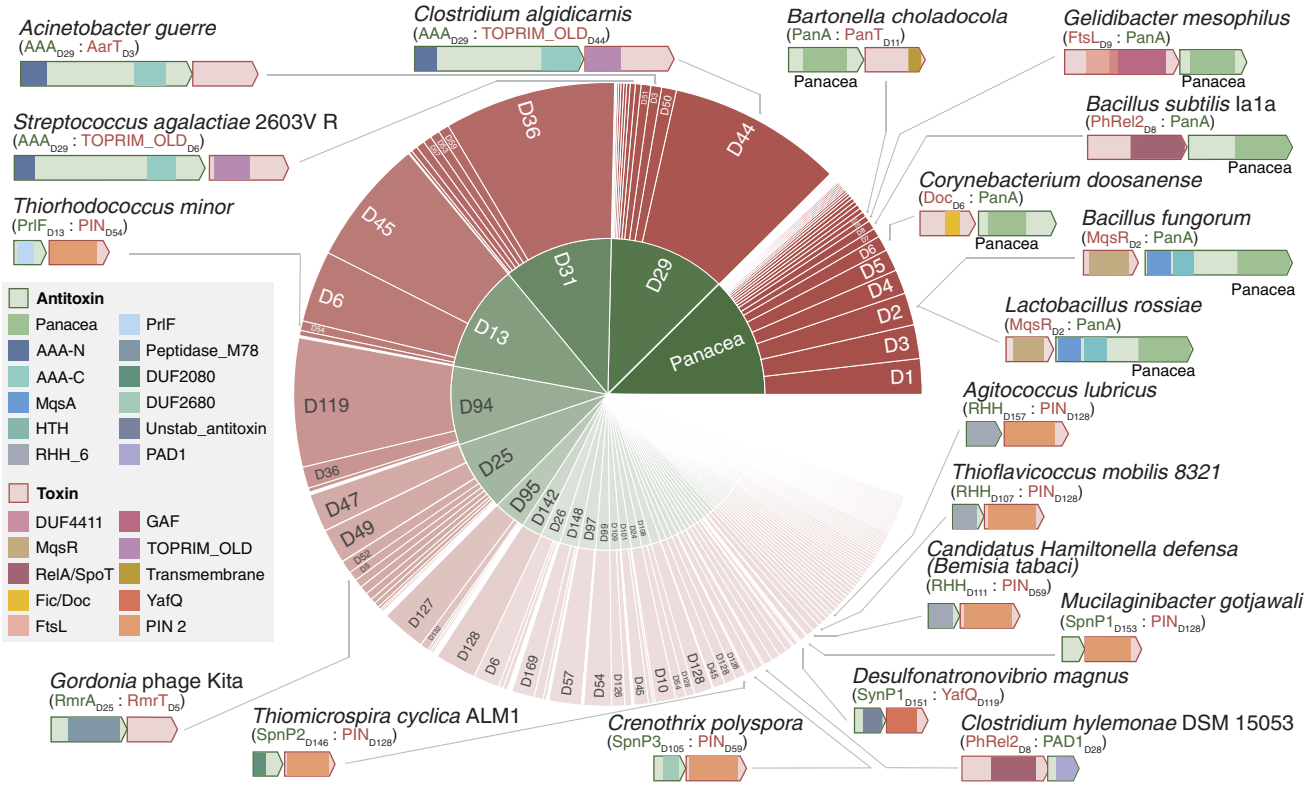


Fig. 2. The diversity of systems in the core NetFlax TA network. The size of the sections represents the number of proteins in each cluster (node) of Fig. 1. The inner ring shows antitoxin nodes, while the outer ring shows associated toxin nodes. The block arrows represent open reading frames, drawn to scale. Colored boxes within the arrows indicate domains, colored according to the legend on the left.

after the protein encoded on a conjugative transposon of *Bacillus subtilis* (36). ImmA is an antirepressor that cleaves an HTH (Helix-Turn-Helix) domain–containing repressor (ImmR) to allow the expression of an integrase (36). Indeed, proteins within the D25 node often possess a small N-terminal HTH domain in addition to the protease domain, suggesting a similar HTH cleavage mechanism of regulation in Rosmer-like TAs. Again, we see the involvement of nucleases in our predicted systems, this time in an association of D25 with PIN-like domains, the most diverse and ubiquitous nuclease superfamily, often seen as toxin components of TAs (37, 38).

Phage defense domains appear in other parts of the NetFlax network. In addition to the AAA and ImmA/Rosmer zones that are clearly phage defense related, we have recently found that toxSASs can protect against phages (22). Additionally, the identification of DefenceFinder domains in the Classical type II zone, along with evidence for classical TA domains in phage defense (39, 40) shows how intricately TAs in general are associated with phage defense.

NetFlax TAs Are Found across the Prokaryotic Tree of Life, and in Tailed Bacteriophages. NetFlax-predicted TAs are found in all major phyla of bacteria and archaea (*SI Appendix*, Fig. S2 and Dataset S1). Most NetFlax TAs were found in Pseudomonadota (formally known as Proteobacteria)—particularly Gammaproteobacteria, reflecting the bias of RefSeq toward these taxa. The pseudomonad *Thiobaca trueperi* has the most NetFlax-predicted TAs (eight). NetFlax TAs are also found across the archaeal tree of life, with representatives in the phyla Euryarchaeaota, Crenarchaeota, Thaumarchaeota, Candidatus Thermoplasmota, and Candidatus Koracheota. Within viruses, TAs were only predicted in Uroviricota (tailed bacteriophages). Our identification of 13 NetFlax TAs in phages (Dataset S1) is likely a significant underestimate as bacteria-encoded systems can be resident on prophages integrated into the bacterial chromosome, for example the toxSAS CapRel (22).

AlphaFold2 Confidently Predicts the Structure of Binary TA Complexes. TAs are excellent targets for modern deep-learning structural prediction methods—not just of single proteins—but of complexes. This is because type II systems necessarily form tight complexes to keep the toxin in check, with a coevolutionary signal in the interface region (41). We have run AlphaFold2 on a high throughput basis with the FoldDock pipeline (27) to predict the structure of all 3,597 protein pairs (3,277 after pruning). To keep predictions computationally feasible, we predict binary TA dimers, not higher order oligomers, rationalizing that even in larger complexes, there must be an interface between the toxin and antitoxin. The reliability of the structures of the complexes is assessed using the pDockQ score, which takes into account the number of interface contacts, and the plDDT reliability scores from AlphaFold2 for those regions. All structures and their scores are available on the interactive network (http://netflax.webflags.se/). Recapitulation of TA folds previously solved with X-ray crystallography indicates these predictions are reliable (*SI Appendix*, Fig. S3). We determined the distribution of model confidence (pDockQ scores) of TA pairs compared to random pairs. The TA predictions are much better than random predictions for the same set and roughly 50 to 60% of the complexes are well modeled (*SI Appendix*, Fig. S4).

Recurrent Structural Folds Appear across the NetFlax Network. The NetFlax algorithm includes a cross-checking step to determine whether each potential cluster is unique or similar to any cluster identified during the previous hopping rounds (*SI Appendix*, Fig. S1*B*). Despite this, we found multiple clusters in the classical type II zone of the network with similar annotations, suggesting they may be homologous despite not clustering together. For example, the Phd, ParE, and PIN domain appear multiple times in the network (Fig. 1). Sequence

alignments show that these unclustered but related nodes are clearly distinct in terms of sequence (including insertions and deletions, *SI Appendix*, Fig. S5), but that they are similar enough that they have the same three-dimensional fold. To systematically address this, we clustered all our predicted structures and annotated our network to show nodes that can be confidently classed as sharing a common fold (*SI Appendix*, Fig. S6). We find many of our predicted TAs can be clustered into 18 distinct folds, the most common being ATPase, MazF/PemK, PIN, RelE/ParE, Phd/YefM, and a common fold of Rosmer toxins (25). Structural alignments of the most common folds are shown in *SI Appendix*, Fig. S7. This supports previous observations that toxins and antitoxins that are diverse at the sequence level can have the same structural fold (42). A similar conservation of domains is also seen in phage defense systems (43).

PanA-Containing TAs: The Roles of Individual Antitoxin Domains in Toxin Neutralization. We focused on three previously experimentally validated PanAT TAs: *Bartonella choladocola* (previously *Bartonella apis*) $PanT_{D11}$:PanA, *Corynebacterium doosanense* Doc_{D6}:PanA, *B. subtilis* Ia1a toxSAS $PhRel2_{D8}$:PanA as well as one previously unexplored PanAT, *Bacillus fungorum* $MqsR_{D2}$:PanA. The subscript D number refers to the node in Fig. 1. All NCBI protein accession numbers of TAs characterized in this paper are shown in *SI Appendix*, Table S1. In all of the PanAT systems, the Panacea domain is predicted to have the same compact architecture comprised of α-helices α1-α7 and β-strands β1 and β2 (Fig. 3 *A* and *B*). Despite these PanATs having dramatically different toxins, these four TA structures are predicted with confidence (pDockQ scores from 0.68 to 0.71). As selected PanAT systems differ in their antitoxin architecture (Fig. 3 *C–E* and *G* and see below), mutational analysis of the set allows us to interrogate the function of the Panacea domain in PanAs: Does it mediate toxin neutralization directly or is this achieved by additional domains?

The structure of *B. choladocola* $PanT_{D11}$:PanA suggests that Panacea can, indeed, directly neutralize the toxin (Fig. 3*C*). In this case, Panacea is predicted to form a contact with the N-terminal unstructured region as well as a short α-helix that precedes the $PanT_{D11}$ predicted transmembrane region (14). Substitutions Y56A (N-terminally adjacent to β2) and H120E (C-terminal end of α5) that were designed to disrupt this interface do, indeed, render *B. choladocola* PanA unable to neutralize the toxin, thus supporting the structural model. Both of these substituted residues are located in the conserved structural core of the Panacea domain (Fig. 3*B*).

In the case of *C. doosanense* PanAT, the two additional C-terminal helices decorating the Panacea core of the PanA antitoxin are predicted to make extensive contacts with the Doc_{D6} toxin (Fig. 3*D*). The globular Panacea domain itself is not predicted to be involved in neutralization. These two C-terminal helices are structurally analogous to those found in the C-terminal extension of the *E. coli* Phd antitoxin that inhibits the Doc toxin (44). Therefore, we refer to this element of *C. doosanense* PanA as the Phd-C domain. *E. coli* Doc is a kinase that phosphorylates EF-Tu to abrogate cellular protein synthesis (45); *C. doosanense* Doc_{D6} similarly targets translation (14), and the active site residues are conserved among the two proteins (Fig. 3*D*). The Phd-C domain of PanA directly interacts with the active site of the toxin. Truncation of the Phd-C domain renders *C. doosanense* PanA unable to neutralize the toxin (Fig. 3*D*). Expression of the isolated Phd-C domain does not neutralize the toxin, which could be due to the intrinsic instability of the element. To test this hypothesis, we fused Phd-C with a stabilizing N-terminal SUMO tag, and as predicted, the resulting construct can readily neutralize Doc_{D6}, despite lacking the Panacea domain. No neutralization was observed in the control experiment with SUMO alone. Collectively, these results suggest that Panacea can serve as an accessory domain, with neutralization being mediated by a dedicated separate domain.

Next, we characterized the *B. subtilis* Ia1a $PhRel2_{D8}$:PanA system. In our previous analysis of the Panacea domain distribution, we identified a domain that we named the PAD1 domain, standing for Panacea-associated domain 1 (14). Apart from two strains of Ruminococcaceae where the putative toxin is an ATPase, PAD1-Panacea multidomain PanA antitoxins are only found paired with toxSASs such as $PhRel2_{D8}$, where it is the most widespread antitoxin for this kind of toxin in the NetFlax network. The second most widespread is NetFlax domain D27 (Fig. 1). Remarkably, structural alignment of the toxSAS:D27 of *Clostridium hylemonae* DSM 15053 with toxSAS:PAD1-PanA of *B. subtilis* Ia1a, fused TA CapRel (22) showed that D27, PAD1, and pseudo-ZBD are predicted to have the same fold, and share the same interface with the toxSAS toxin (*SI Appendix*, Fig. S8). Importantly, it is PAD1 that forms most of the contacts with the $PhRel2_{D8}$ toxin (Fig. 3*E*). Strikingly, *B. subtilis* Ia1a PAD1 domain alone—with Panacea removed—can neutralize the toxin, thus directly supporting the structural prediction. Furthermore, H52P substitutions that are predicted to break the PAD1:$PhRel2_{D8}$ interface completely abrogate the neutralization, both in the context of full-length PanA and isolated PAD1. Finally, the I178K substitution located on the Panacea: $PhRel2_{D8}$ interface did not affect the efficiency of neutralization. To further support the role of PAD1 as a dedicated toxin-neutralizing domain, we have, via toxicity neutralization assays, validated the *C. hylemonae* DSM 15053 TA system comprised of $PhRel2_{D8}$ and the $PAD1_{D27}$ antitoxin (Fig. 3*F*). As the *C. hylemonae* antitoxin naturally lacks the Panacea domain, this observation further supports PAD1 being a directly neutralizing antitoxin element. Substitutions predicted to compromise the PAD1:$PhRel2_{D8}$ interface either weakened (S47D) or completely abrogated the neutralization (I46P and I46E).

Finally, we have dissected the *B. fungorum* $MqsR_{D2}$:PanA system (Fig. 3*G*). MqsR is an RNase (42, 46) that is neutralized by antitoxin MqsA comprised of an N-terminal Zn^{2+}-binding domain (ZBD) and C-terminal HTH (42). Substitution of the predicted catalytic Y97 residue of *B. fungorum* $MqsR_{D2}$ renders it nontoxic, supporting the functional annotation. While the ZBD interacts with MqsR and inhibits it without directly interacting with the RNase active site, the HTH region dimerizes and acts as a transcriptional autoregulator of the *mqsRA* operon. *B. fungorum* PanA also contains the ZBD-HTH domain composition characteristic of the MqsA antitoxin, with the Panacea domain added to the C terminus. Our truncation analysis shows that, indeed, also in the case of *B. fungorum* PanAT, the ZBD directly mediates neutralization of $MqsR_{D2}$. While both isolated ZBD and ZBD-HTH segments efficiently neutralize $MqsR_{D2}$, neither Panacea alone nor HTH-Panacea are sufficient for neutralization. We used AlphaFold2 to predict homodimerization of our validated antitoxins (*SI Appendix*, Fig. S9). We see good support for HTH-mediated homodimerization of *B. fungorum* PanA. Note that in the HTH-containing PanA from *Lactobacillus rossiae* the Panacea domain also appears to be involved in homodimerization (*SI Appendix*, Fig. S9*A*).

Collectively, our results demonstrate that while Panacea can act as a direct toxin neutralizer, it is unlikely to act as such in PanA antitoxins that contain additional dedicated neutralization domains such as PAD1 or ZBD. Furthermore, the example of *B. fungorum* $MqsR_{D2}$:PanA system suggests that Panacea probably does not act as a transcriptional autoregulator of *panAT* operons either, as *B. fungorum* PanA contains a dedicated DNA-binding regulatory domain, HTH (as do many other PanAs, ref. 14). Therefore, we favor the hypothesis that the Panacea domain acts as a sensor responding to—as yet unknown—TA-activating cues.

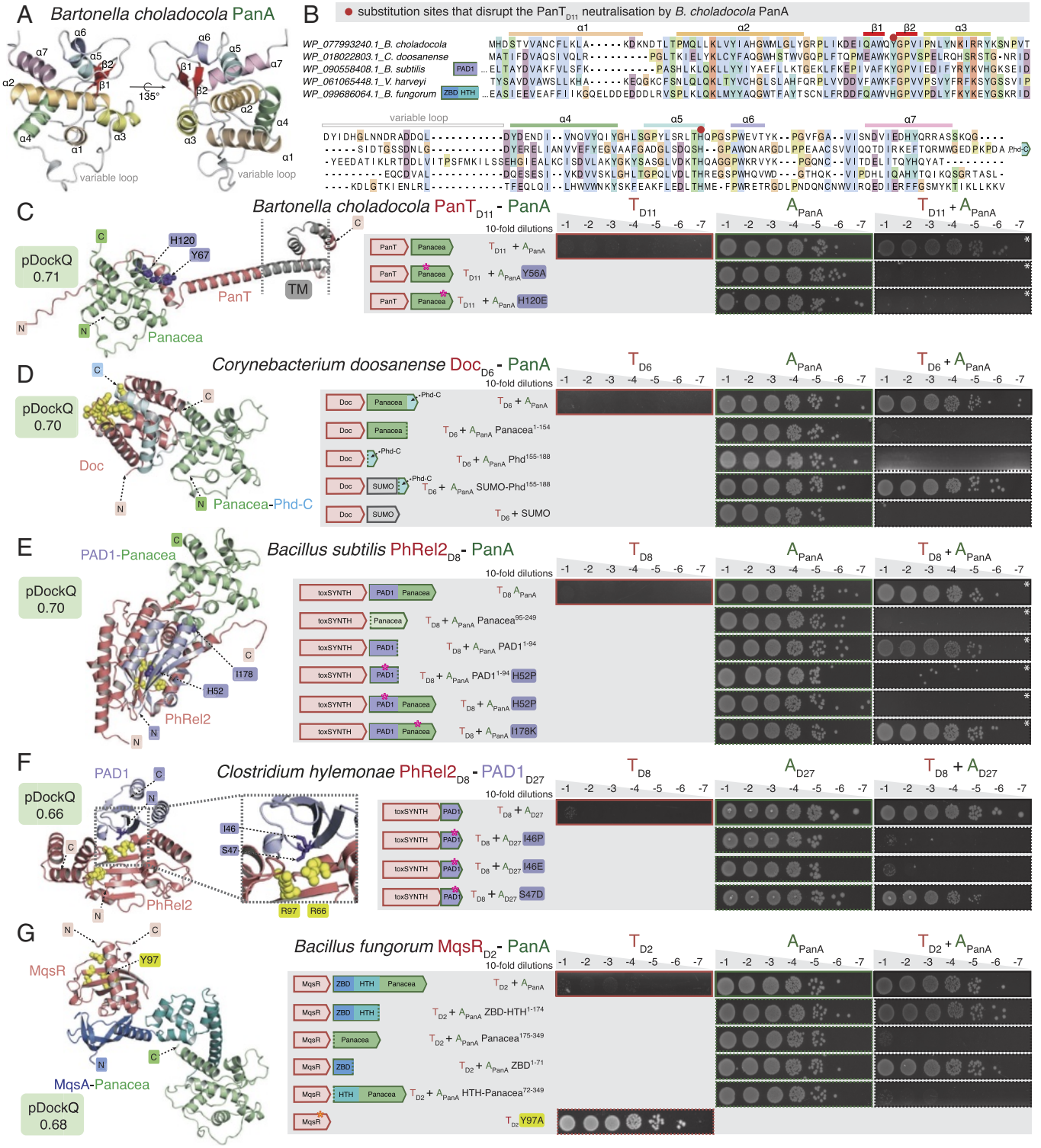


Fig. 3. PanA antitoxins neutralize PanT toxins either via Panacea domain directly or via additional N-terminal domains: Phd-C, PAD1, and MqsA. (*A*) AlphaFold-generated structural model of *B. choladocola* PanA. (*B*) Alignment of Panacea domains from representative PanA antitoxins. (*C*–*G*) Mutational probing of AlphaFold-generated structural models. In toxicity neutralization assays overnight cultures of *E. coli* strains transformed with pBAD33 and pMG25 vectors or derivatives expressing putative *panT* toxins and *panA* antitoxins, correspondingly, were adjusted to OD_{600} 1.0, serially diluted from 10^1- to 10^8-fold and spotted on LB medium supplemented with appropriate antibiotics and inducers [0.2% arabinose for toxin induction as well as either 50 or 500 μM (white asterisk) IPTG for antitoxin induction]. Predicted transmembrane domains (TM) are shown in gray, and the active site of the toxin is highlighted in yellow. Introduced substitutions in the antitoxin are shown in purple. Protein accessions are in *SI Appendix*, Table S1.

Experimental Exploration of the NetFlax Network. We have validated 16 TA pairs in toxicity neutralization assays (Fig. 1). For 13 of them, we performed metabolic labeling assays with ^{35}S methionine (a proxy for inhibition of translation), or ^{3}H uridine (a proxy for inhibition of transcription), or ^{3}H thymidine (a proxy for inhibition for replication). Ten additional systems did not show any toxicity in *E. coli* and were not pursued further (*SI Appendix*, Fig. S10).

The TA system from *Gelidibacter mesophilus* is comprised of a relatively large (281 aa) toxin T$_{D9}$ (toxFtsL$_{D9}$) paired with a PanA antitoxin (Fig. 4*A*). Similarly to *B. choladocola* PanT$_{D11}$:PanA,

G. mesophilus PanA is comprised of a stand-alone Panacea domain that directly neutralizes the toxin. While the toxin is clearly very efficient in abrogating the formation of bacterial colonies on solid LB plates, induction in liquid culture does not result in rapid growth inhibition nor do we see any dramatic effects in metabolic labeling assays (Fig. 4*B*). The toxin contains a GAF (cGMP-specific phosphodiesterases, adenylyl cyclases, and FhlA) domain and an α-helical FtsL-like domain, which is predicted to dimerize and be localized to the cell membrane (Fig. 4 *C* and *D*). FtsL is an essential component of bacterial divisome, which forms a trimeric complex with FtsB and FtsQ via leucine zipper-like motifs (47). Given the partial homology with FtsL, we propose naming the *G. mesophilus* T_{D9} toxin toxFtsL$_{D9}$. It is tempting to speculate that *G. mesophilus* T_{D9} could act by directly interfering with the cell division process. Experiments with liquid cultures of *E. coli* expressing *G. mesophilus* toxFtsL$_{D9}$ lend support to this hypothesis: After an hour of uninhibited growth, the OD_{600} increase stops and then the culture collapses, suggestive of cell lysis (Fig. 4*E*). Microscopy experiments show that expression of toxFtsL$_{D9}$ indeed results in cell filamentation and inhibition of divisome assembly (Fig. 4 *F* and *G*, *SI Appendix*, Fig. S11, and Movie S1). Crucially, this is not caused by partial membrane depolarization, which can interfere with the bacterial cell division process (*SI Appendix*, Fig. S12) (48). Inhibition of cell division is an established mode of action for TA toxins with *E. coli* toxin CbtA directly targeting FtsZ and MreB (49).

The TA system from *Gordonia* phage Kita is a new member of the RosmerTA family (18, 30, 35) (Fig. 5*A*). The RmrA$_{D25}$ protease antitoxin is paired with a D5 toxin, which has no detectable similarity to other protein families. Metabolic labeling assays show rapid and dramatic abrogation of translation, transcription, and replication upon expression of the Kita phage D5 toxin (Fig. 5*B*). The toxin is not fully neutralized by the antitoxin, and the structure of the TA complex cannot be reliably predicted by AlphaFold (pDockQ score of 0.05) (Fig. 5*C*). The C-terminal region of the toxin is predicted to be localized to the cellular membrane (Fig. 5*C*).

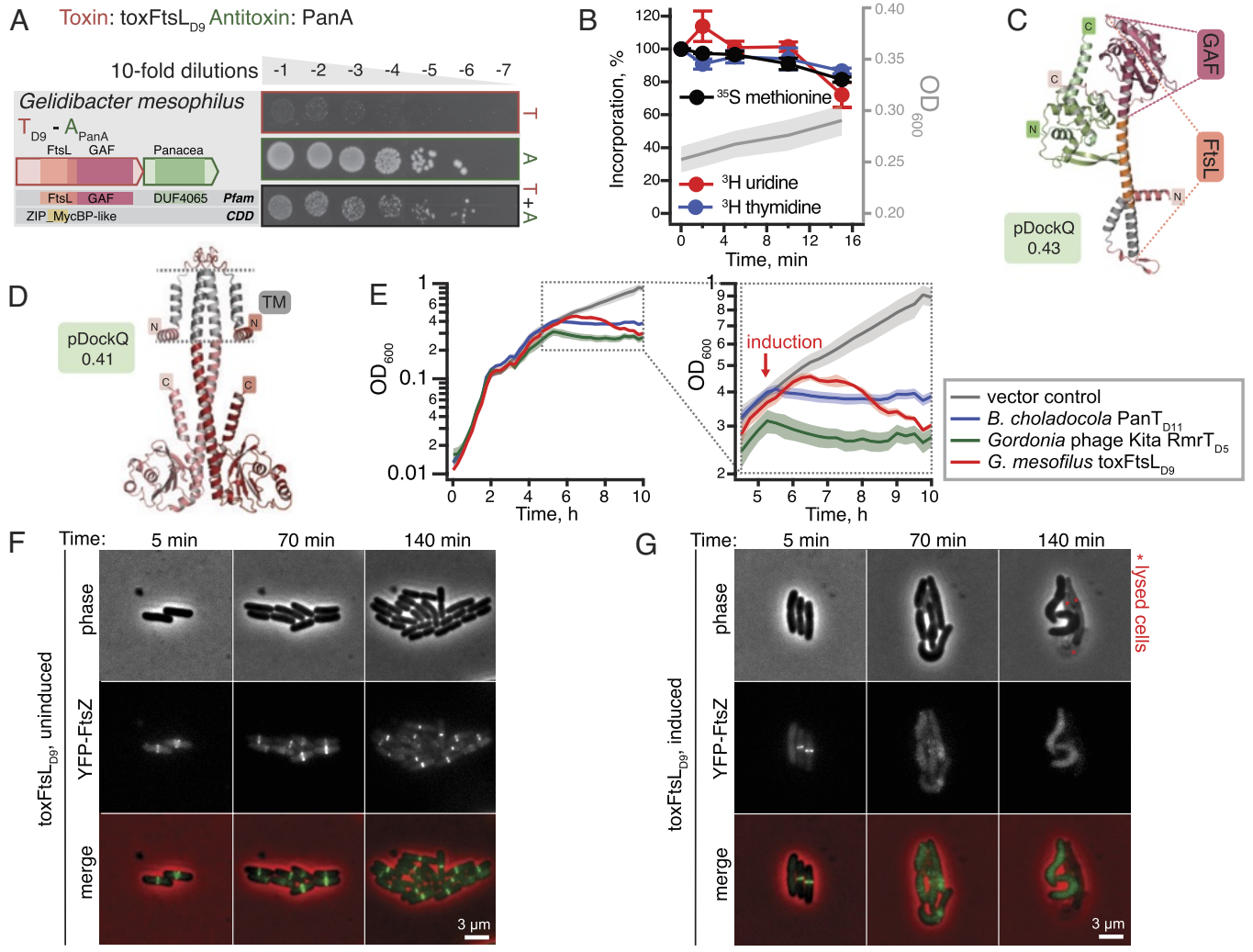


Fig. 4. *G. mesophilus* toxFtsL$_{D9}$ is a slow-acting PanT toxin with partial homology with the FtsL component of bacterial divisome. (*A*) Validation of the *G. mesophilus* toxFtsL$_{D9}$:PanA TA through toxicity neutralization assay. (*B*) Metabolic labeling assays with wild-type *E. coli* BW25113 expressing *G. mesophilus* toxFtsL$_{D9}$. (*C*) PanAT structural prediction of *G. mesophilus* toxFtsL$_{D9}$:PanA TA pair. toxFtsL$_{D9}$ is predicted to have two partially overlapping domains: The α-helical FtsL-like region is highlighted with orange dotted guide lines, and the GAF domain is indicated with dark pink dotted guide lines. (*D*) Structural prediction of the toxFtsL$_{D9}$ dimer. Predicted transmembrane (TM) helical regions are shown in gray on *C* and *D*. (*E*) Delayed growth inhibition and cell lysis by *G. mesophilus* toxFtsL$_{D9}$. Growth assays of *E. coli* BW25113 cells expressing *G. mesophilus* toxFtsL$_{D9}$, *B. choladocola* PanT$_{D11}$ or *Gordonia* phage Kita RmrT$_{D5}$ as well as a vector control strain harboring pBAD33 and pMG25 in MOPS liquid medium supplemented with 0.5% glycerol and 25 μg/mL each 20 amino acids. Expression of toxins was induced with 0.2% arabinose at OD_{600} of around 0.4. (*F* and *G*) Fluorescence and phase contrast time lapse microscopy of *E. coli* BW25113 cells expressing YFP-FtsZ in the absence (uninduced, *E*) and presence (induced with 0.2% arabinose, *F*) of *G. mesophilus* toxFtsL$_{D9}$. Note the toxFtsL$_{D9}$-induced delocalization of the main divisome scaffold protein FtsZ, and the associated cell elongation. For the complete time lapse data, see Movie S1. For quantification of cell elongation and additional controls, see *SI Appendix*, Figs. S11 and S12.

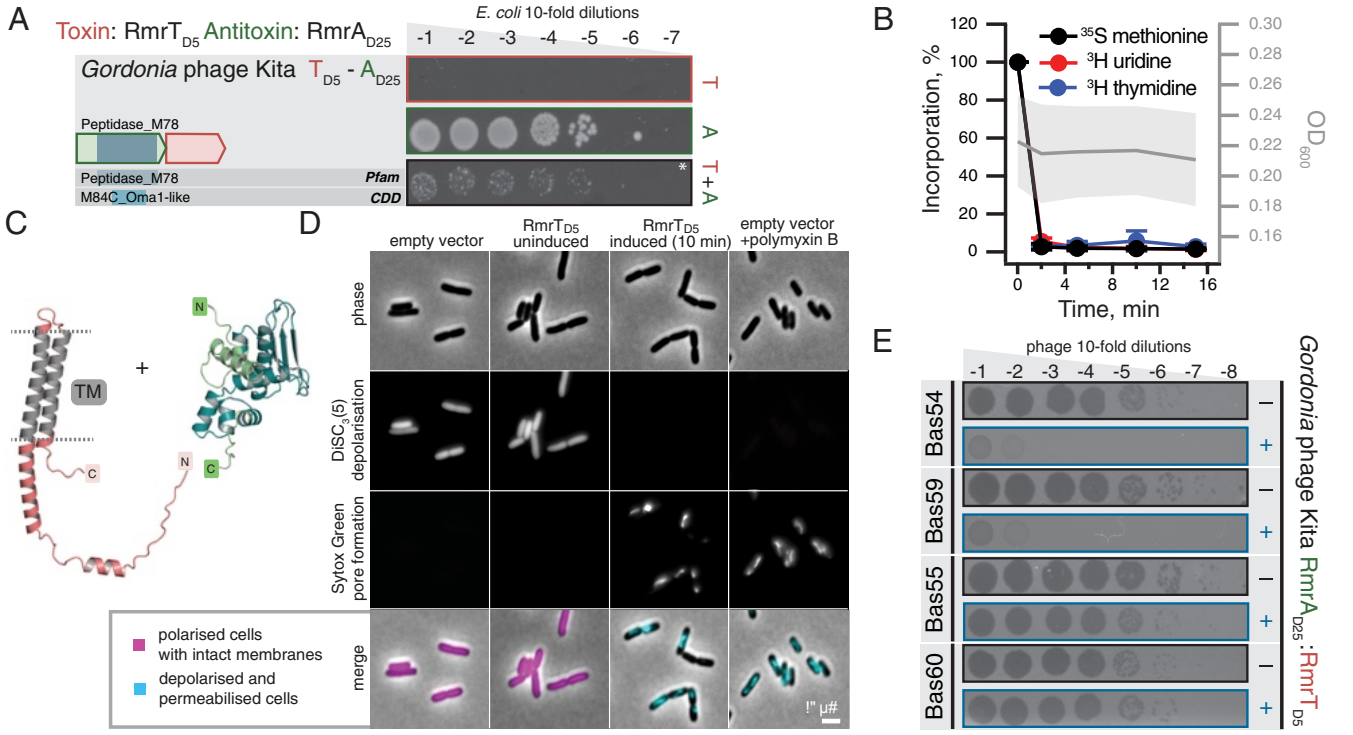


Fig. 5. The RosmerTA antiphage defense system from *Gordonia* phage Kita. Domain organization and TA validation through toxicity neutralization assays (*A*), metabolic labeling assays with toxins expressed in wild-type *E. coli* BW25113 (*B*) and AlphaFold-generated structural models (*C*) for RmrT$_{D5}$ and RmrA$_{D25}$ TA from *Gordonia* phage Kita. Predicted TM regions of the toxin are shown in grey. Liquid culture experiments with RmrT$_{D5}$ are shown on Fig. 4*E*. (*D*) Phase contrast and fluorescence images of *E. coli* BW25113 cells colabeled with the membrane potential-sensitive dye DISC$_3$(5) and the membrane permeability indicator Sytox Green. Note the depolarization and pore formation induced by RmrT$_{D5}$ that is comparable to the pore forming antibiotic Polymyxin B. For quantification of membrane depolarization and pore formation triggered by RmrT$_{D5}$, see *SI Appendix*, Fig. S13. (*E*) *E. coli* BW25113 transformed with either empty pBR322 derivative, pJD1423 (VHp1423), (−) or pJD1423 derivative expressing *Gordonia* phage Kita RosmerTA from P$_{tet}$ promoter (+) was challenged with ten-fold serial dilutions of BASEL coliphages (50). The full screen is shown on *SI Appendix*, Fig. S14, from which here we show results for Bas54 and Bas60 (defense) as well as Bas55 and Bas59 (lack of defense).

In liquid culture experiments, expression of RmrT$_{D5}$ immediately inhibits bacterial growth without causing a consequent collapse of OD$_{600}$; the toxin is likely to share the mechanism of toxicity with membrane-depolarizing *B. choladocola* PanT$_{D11}$ (14) that we used as a control (Fig. 4*E*). Indeed, the expression of RmrT$_{D5}$ results in rapid membrane depolarization and increased permeability indicating membrane pore formation (Fig. 5*D* and *SI Appendix*, Fig. S13). Following the nomenclature for RosmerTA toxins (18, 30, 35), we renamed the Kita phage toxin RmrT$_{D5}$. As other RosmerTA systems have been shown to be phage defense systems (18, 30, 35), it is likely that the Kita phage RmrTA has a similar function. To test this, we carried out phage infection assays using the BASEL coliphage collection (50) and find that RmrTA provides potent but narrow spectrum defense against *Myoviridae*: it counters Bas54 and Bas59 but not to closely related Bas55 and Bas60 (Fig. 5*E*)—or any other BASEL coliphages (*SI Appendix*, Fig. S14). We further confirm protection against Bas59 in liquid culture infection assays (*SI Appendix*, Fig. S15).

The TA system from *Acinetobacter guerrae* is composed of a toxin T$_{D3}$ that has no detectable hits with HHPred. However, it has the same fold as mRNA interferases (*SI Appendix*, Fig. S6), paired with an AAA ATPase A$_{D29}$ antitoxin (Fig. 6*A*). We refer to the toxin as AarT for AAA-associated RNAse-like toxin. Metabolic labeling experiments suggest that the toxin targets protein synthesis as its expression inhibits ^{35}S methionine incorporation with concurrent increase in ^{3}H uridine, a pattern that is characteristic for translation-targeting toxins and antibiotics, and supporting an identity as an mRNAse or tRNase (14). Similar neutralization architecture was predicted for D$_{29}$ AAA antitoxins from *Clostridium algidicarnis* (Fig. 6*B*) and *Streptococcus agalactiae* 2603V R (Fig. 6*C*). These two AAA antitoxins are paired with a TOPRIM_OLD domain. The *Lactococcus lactis* AbiL is a bicistronic plasmid-encoded phage system that acts thorough abortive infection elicited by the TOPRIM_OLD toxic effector AbiLii (28). We speculate that the three AAA-neutralized TA pairs are also Abi phage defense systems. AlphaFold2 modeling does not give a convincing interface for AAA antitoxins and their toxins (pDockQ score of 0.24 to 0.35), which may be because phage defense AAA-containing systems form large multimeric complexes, as is seen with AAA-containing RADAR (51, 52). Higher order complex formation of NetFlax-predicted AAA antitoxins is supported by our AlphaFold predictions of homodimerization (*SI Appendix*, Fig. S9*C*). A phage immunity screen of *E. coli* expressing *A. guerrae* AarT$_{D3}$:AAA$_{D29}$ versus the BASEL coliphage collection did not yield any hits. This does not rule out defense in the natural host; the mechanism of phage sensing and defense may be specific for *Acinetobacter* phages, or rely on host factors not present in *E. coli* (*SI Appendix*, Fig. S16).

Finally, we have validated 10 TA pairs of nuclease toxins (MqsR$_{D2}$, PIN/VapC-like and YafQ$_{D119}$) paired with diverse antitoxins (PanA, PerlF$_{D13}$, SpnP3/DUF2680$_{D105}$, RHH_6$_{D107}$, RHH_6$_{D111}$, SpnP2/DUF2080$_{D146}$, SynP1$_{D151}$ and RHH_6$_{D157}$), for which AF2 structural models suggest multiple mechanisms of direct and indirect toxin neutralization (Fig. 7 and *SI Appendix*, Figs. S17 and S18). While the structures were predicted as binary complexes, RHH (Ribbon-Helix-Helix) domain is a well-characterized dimeric DNA-binding transcriptional regulator employed by numerous antitoxins such as CcdA (53) and FitA (54). Dimers of RHH-containing antitoxins can be readily predicted by AlphaFold (*SI Appendix*, Fig. S9*E*). Multiple groups of translation-targeting RNase TA toxins have been characterized experimentally, and display considerable diversity, even

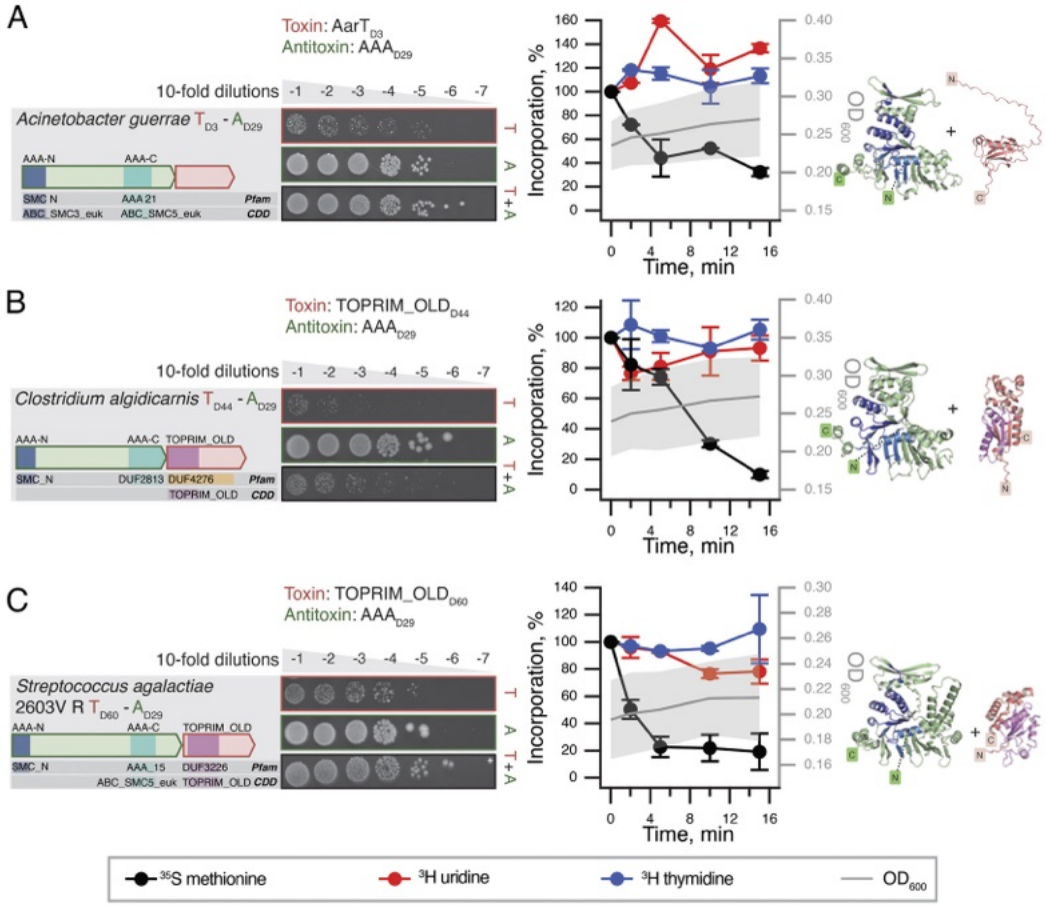


Fig. 6. AAA-neutralizd putative Abi phage defense systems. Domain organization and TA validation through toxicity neutralization assays (*Left*), metabolic labeling assays with toxins expressed in wild-type *E. coli* BW25113 (*Center*) and AlphaFold-generated structural models (*Right*) for AAA_{D29}-neutralized TAs: (*A*) *A. guerrae* $AarT_{D3}$:AAA_{D29} (*B*) *C. algidicarnis* $TOPRIM\_OLD_{D44}$:AAA_{D29} and (*C*) *S. agalactiae* 2603V R $TOPRIM\_OLD_{D60}$:AAA_{D29}.

within closely related groups; for example, different VapC PIN TA toxins can cleave either tRNA (55) or rRNA (56). As expected for nucleases, metabolic labeling assays indicate the majority of the validated nuclease toxins do, indeed, target protein synthesis (Fig. 7 *A* and *B* and *SI Appendix*, Figs. S17 and S18). However, unexpectedly, expression of *Thioflavicoccus mobilis* 8321 PIN_{D59} results in inhibition of incorporation of both ^{35}S methionine (abrogation of translation) and ^{3}H uridine (abrogation of transcription) (Fig. 7*C*). Despite this unexpected behavior, substitutions of predicted catalytic residues of the PIN_{D59} domain (D6 and E43) abrogate the toxicity. Although metabolic labeling assays with the other two PIN_{D59} toxins (from *Crenothrix polyspora* and *Candidatus Hamiltonella defensa* (formerly *Bemisia tabaci*)) do not yield clear-cut results, they are indicative of the PIN_{D59} toxin having additional toxic effects beyond specific inhibition of protein synthesis (*SI Appendix*, Fig. S18).

Discussion

Classical TA antitoxins are modular proteins typically consisting of a DNA-binding domain involved in transcription autoregulation and a functionally independent neutralization domain that folds upon binding to the toxin in most cases (1). It has been argued that it is the combination of, on one hand, the functional decoupling between these two structural modules and, on the other hand, the disordered nature of the neutralization domains that enables antitoxin promiscuity, i.e., allows for neutralization of toxins belonging to multiple protein families by different antitoxins possessing the same DNA-binding domain (57). This model postulates that the fusion of a "linear" recognition motif that performs the neutralization, along with a DNA-binding domain is sufficient to generate a functional TA operon. Indeed, as we show here for *C. doosanense* Doc_{D6}:PanA_Phd-C system, the fusion of the Phd-C region alone to a SUMO tag is sufficient to engineer a protein that efficiently counteract the Phd-C-cognate toxin in vivo (Fig. 3*D*). Given the existence of multiple TA operons with single-domain antitoxins that consist of the neutralization domain alone (58, 59), such fusion or exchange events constitute a plausible evolutionary pathway that could generate the complex TA permutations observed in TAs (Fig. 1).

Importantly, the NetFlax network reveals the existence of a type of hyperpromiscuous antitoxin that defies this commonly accepted neutralization paradigm. Such antitoxins are epitomized by Panacea and HTH domains (14, 42, 60–62) that possess within their structural fold an intrinsic capacity to specifically recognize and neutralize diverse toxins via three-dimensional epitopes (Figs. 3*C* and 8). Furthermore, these antitoxin domains can also acquire linear epitopes consisting of intrinsically disordered regions or well-folded domains, to neutralize toxins in the "classical" manner (Figs. 3 *D–G* and 8). In addition to this hyperpromiscuity, antitoxins have a capacity for moonlighting in other ways: PAD1-like domains are involved in both toxin neutralization and phage detection (22), while HigA antitoxins from toxin–antitoxin–chaperone (TAC) operons engage both the toxin and the dedicated regulator chaperone (63). The ability of antitoxins to readily remodel their domain combinations and evolve multifunctional moonlighting abilities of the constituent domains may be a core feature of TA roles in innate immunity involved in phage defense.

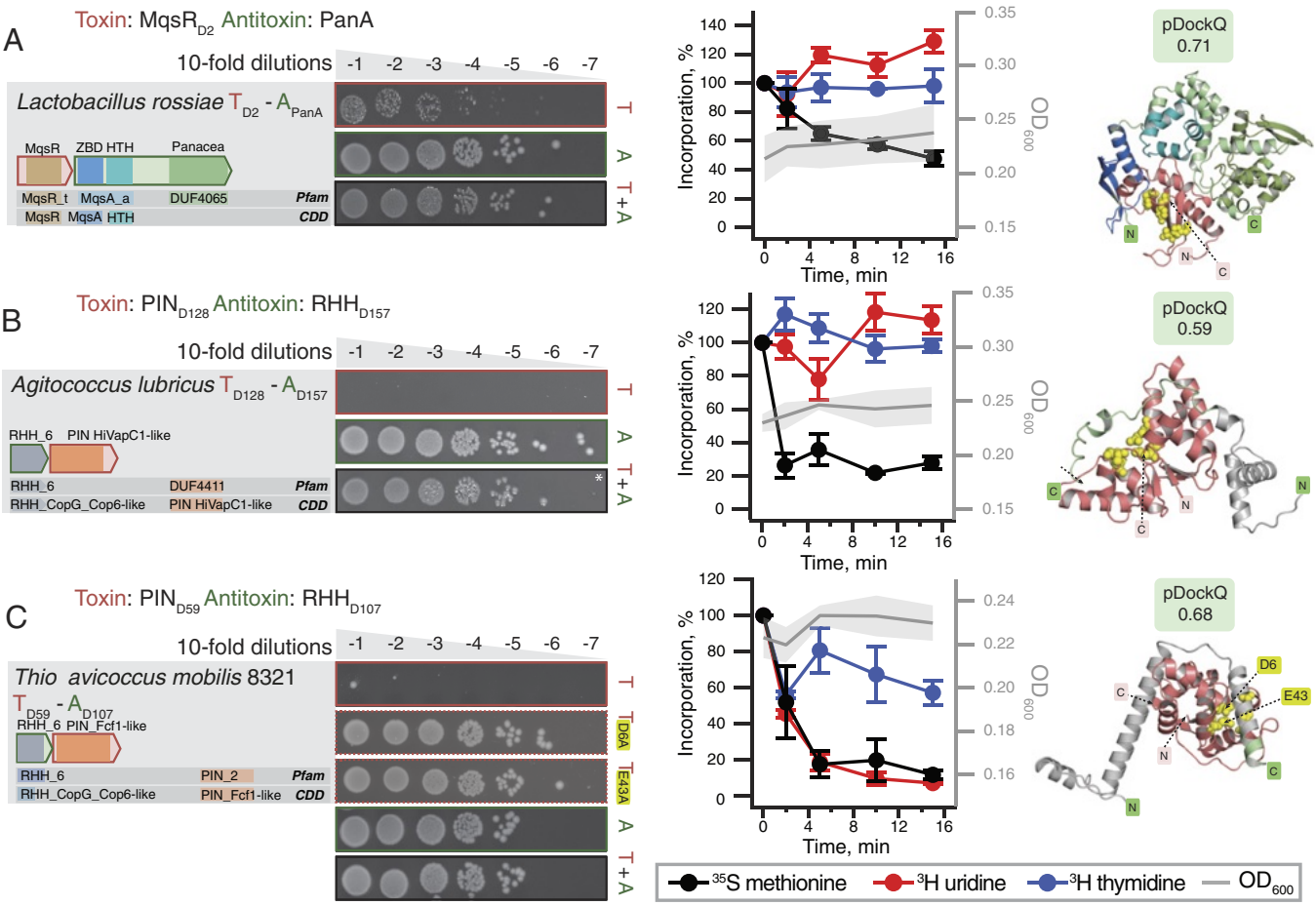


Fig. 7. Representative NetFlax TA systems with nuclease effectors: MqsR$_{D2}$, PIN$_{D128}$ and PIN$_{D59}$. Domain organization and TA validation through toxicity neutralization assays (*Left*), metabolic labeling assays with toxins expressed in wild-type *E. coli* BW25113 (*Center*) and AlphaFold-generated structural models (*Right*) for TAs with diverse nuclease effectors: (*A*) *L. rossiae* MqsR$_{D2}$:PanA, (*B*) *A. lubricus* PIN$_{D128}$:RHH$_{D157}$ and (*C*) *T. mobilis* 8321 PIM$_{D59}$: RHH$_{D107}$. Active sites are highlighted with yellow spheres on the structures.

In the case of PanAs that function via linear neutralization without the direct involvement of the globular Panacea domain, the question is what then is the role of the Panacea domain. We hypothesize that its function is primarily sensory, reacting to a trigger and activating the toxin through an allosteric mechanism involving the neutralization region. This activation may not even require dissociation of the PanAT complex; the fused toxSAS TA CapRel shows that antitoxins do not have to dissociate in order to activate the toxin (22). Further investigations are needed to determine what is the functional role of the Panacea domain.

Our structural prediction has allowed the clustering of predicted toxins and antitoxins into identifiable fold classes. This tendency of TA systems to reuse folds has previously been noted (42). However, conservation of fold may not mean conservation of function: Proteins can be divergent at the sequence level—even to the point of carrying out different biochemistry, despite being based on the same structural fold. For example, the Fic/Doc family of toxins contains members that can both NMPylate or phosphorylate their protein targets (45, 64). Similarly, RelE/ParE family members with the same fold can either cleave RNA or inhibit DNA gyrase (65). Recently, we found that homologous toxSAS TAs can inhibit bacterial growth by either producing the toxic alarmone (pp)pApp or pyrophosphorylating the 3′ CCA end of tRNA (9, 66). Our *T. mobilis* PIN$_{D59}$ domain toxin that seems to inhibit transcription as well as translation may indicate another example of a divergent function on a similar fold.

The biological function of TAs has remained a contentious subject for decades (1, 2, 67), but increasingly a role for these systems in phage defense is being discovered (15–21, 68). Here, we show that this is also reflected in their combinatorial evolutionary relationships: The core network of TAs is connected to domains in other phage defense systems. Indeed, since it is common for multigene defense systems to contain a toxic effector, it is unclear whether there is any meaningful distinction between the two kinds of system. This raises the question of whether classical "addiction

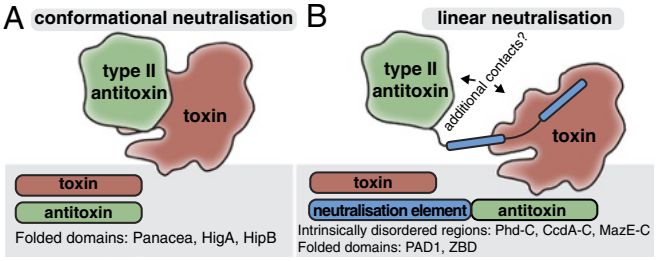


Fig. 8. Conformational and linear neutralization are two primary modes of toxin inactivation by hyperpromiscuous antitoxins in type II TA systems. (*A*) Conformational neutralization involves the direct recognition of toxins by a three-dimensional epitope of the antitoxin that forms part of the antitoxin fold and cannot be grafted to a different scaffold. Panacea and HTH domains represent members of this class. (*B*) Linear neutralization displayed by modular antitoxins characterized by the neutralization of toxins by exchangeable elements that can be intrinsically disordered or fully folded domains. Neutralizing elements are linearly attached to antitoxin domains at the N- or C-terminal ends. This is seen in the case of Panacea linked to Phd-C, ZBD or PAD1, which connects Panacea with the neutralization of Doc, MqsR and toxSAS toxins.

module" TAs on mobile elements may actually have a role in defense against phages or competition with other mobile elements, with addiction effects being a secondary consequence.

NetFlax is a broad stroke approach, which has its limitations and caveats that we acknowledge. First, it is limited to a representative set of proteomes, which means we are missing a substantial amount of diversity. Some known type II TAs such as DarTG (69), HEPN-MNT (70), HicBA (71), and HipBA (72) escaped our prediction. Second, NetFlax only addresses conserved two-gene proteinaceous systems and therefore cannot in this incarnation predict multigene toxin–containing systems. It is also at risk of predicting false positives due to spurious domain associations. Nevertheless, despite these caveats, the network is a starting point for exploring multiple avenues including the TA systems we have characterized, and more fine-grained prediction can be achieved through a subsequent focus on specific lineages.

Materials and Methods

NetFlax Strategy. TA pairs were predicted with the Python script NetFlax, which is a modification of our FlaGs program. Briefly, NetFlax works round by round; and each round follows three steps: i) scanning proteomes using an HMM profile of toxin or antitoxin to identify its protein homologues, ii) prediction of conserved TA-like arrangements and identification of homologous clusters of toxins or antitoxins and iii) cross-checking if the predicted clusters are novel (not encountered in any previous round) and, if so, make their HMM profiles, to be used in the next round of scanning.

NetFlax uses a local database of 24,474 predicted proteomes downloaded from the NCBI RefSeq FTP server (73). This includes one representative proteome per species of bacteria and archaea, along with all 10,449 available virus genomes (not limited to representatives). See *SI Appendix, SI Text* and *SI Appendix*, Fig. S1 for illustrations of how the algorithm works.

Protein sequence and structure analysis. Protein domains and other functional predictions were carried out with searching the toxin–antitoxin database (TADB) (31), DefenceFinder (25), NCBI conserved domain database (CDD) (74) and HHPred (75) (with NCBI-CDD, Pfam-A and PDB as target databases). Protein–protein complex structures were predicted with FoldDock (27). Structures were clustered with FoldSeek v. 5 (76) and with resulting networks visualized with Cytoscape v. 3.5.0 (77). Structural alignment was carried out with mTM-Align v. 20220104 (78). Transmembrane prediction was carried out with DeepTMHMM (79). More detailed methods are described in *SI Appendix, SI Text* document. All structures and scores are available at https://github.com/GCA-VH-lab/NetFlax_data (84).

Experimental Methods. Detailed experimental procedures are provided *SI Appendix, SI Text* document, with a summary below.

Plasmid construction. All bacterial strains, plasmids, and primers used in the study are listed in Dataset S2. Toxin ORFs were cloned into an arabinose-inducible pBAD33 vector (80) either with or without Shine–Dalgarno sequence as required, for toxicity assays. For neutralization assays, antitoxins were expressed from an IPTG-inducible pMG25 vector (81). Mutations and truncations were introduced as described earlier .

Toxicity neutralization assays. Toxicity-neutralization assays were performed on Lysogeny broth (LB) agar plates. First, the pBAD33 vector with the toxin ORF was transformed into competent cells of *E. coli* BW25113 strain with a pMG25 empty vector. A single colony with two plasmids was grown in liquid LB medium supplemented with 100 μg/mL ampicillin (Sigma-Aldrich) and 20 μg/mL chloramphenicol (AppliChem) as well as 0.2% glucose. Serial 10-fold dilutions were spotted (5 μL per spot) onto LB plates containing ampicillin and chloramphenicol under repressive (0.2% glucose) or induction conditions (0.2% arabinose combined with 0.05 or 0.5 mM IPTG). Plates were scored after an overnight incubation at 37 °C. After confirming toxin toxicity, another set of competent cells was produced with the cognate antitoxin in a pMG25 vector, and the spot test was repeated.

Metabolic labeling. Metabolic labeling experiments using *E. coli* BW25113 strains cotransformed with pBAD33 derivatives as well as the empty pMG25 vector were performed as described earlier (14).

Fluorescence microscopy. Fluorescence microscopy experiments with SYTOX Green (82) and $DiSC_3(5)$ (83) were performed as described previously (14).

Phage immunity assays. Phage immunity assays using BASEL coliphages and *E. coli* BW25113 strains transformed with pJD1423-based plasmids (pBR322 derivative) expressing TA systems under the control of P_{tet} promoter were performed as per Maffei *et al.* (50).

Data, Materials, and Software Availability. Code and structures data have been deposited in Github (https://github.com/GCA-VH-lab/NetFlax_data) (84). All study data are included in the article and/or supporting information.

ACKNOWLEDGMENTS. We are grateful to the Protein Expertise Platform at Umeå University for plasmid construction and to Alexander Harms for sharing the BASEL phage collection as well as protocols. The AlphaFold2/FoldDock computations were enabled by the supercomputing resources Berzelius provided by the National Supercomputer Centre (NSC) at Linköping University and the Knut and Alice Wallenberg foundation. Additional computational resources used were provided by the National Academic Infrastructure for Supercomputing in Sweden and the Swedish National Infrastructure for Computing at NSC and Chalmers University Centre for Computational Science and Engineering (C3SE), partially funded by the Swedish Research Council through grants 2018-05973 and 2022-06725. This work was supported by Knut and Alice Wallenberg Foundation (2020.0037 to G.C.A.), the Swedish Research council (2019-01085 and 2022-01603 to G.C.A., 2021-01146 to V.H., 2021-03979 to A.E.); Crafoord foundation (20220562 to V.H.); Cancerfonden (20 0872 Pj to V.H.); Carl Tryggers Stiftelse för Vetenskaplig Forskning (CTS19:24 to G.C.A.); Kempestiftelserna (SMK-2061.1 to G.C.A.); the Estonian Research Council (PRG335 to V.H. and T.T.); Fonds National de Recherche Scientifique (FNRS-CDR J.0068.19, FNRS-EQP UN.025.19 and FNRS-PDR T.0066.18 to A.G.-P.); ERC (CoG DiStRes, number 864311 to A.G.-P.); BBSRC (BB/X003035/1 and BB/T017570/1 to H.S.).

Author affiliations: [a]Department of Experimental Medicine, Lund University, Lund 221 84, Sweden; [b]Institute of Technology, University of Tartu, Tartu 50411, Estonia; [c]Department of Molecular Biology, Umeå University, Umeå 901 87, Sweden; [d]Department of Biochemistry and Biophysics and Science for Life Laboratory, Stockholm University, Solna 171 21, Sweden; [e]Centre for Bacterial Cell Biology, Biosciences Institute, Newcastle University, Newcastle upon Tyne NE2 4AX, United Kingdom; [f]Cellular and Molecular Microbiology, Faculté des Sciences, Université libre de Bruxelles, Brussels 1050, Belgium; [g]Science for Life Laboratory, Lund 221 84, Sweden; and [h]Lund University Virus Centre, Lund 221 84, Sweden

Author contributions: K.E., C.K.S., V.B., J.A.N., H.S., A.E., V.H., and G.C.A. designed research; K.E., C.K.S., T.B., B.C., A.S., J.A.B., J.J.D.-P., V.B., J.A.N., T.K., A.A.E., L.S., M.J.O.J., T.M., A.R., J.D., H.S., V.H., and G.C.A. performed research; C.K.S., B.C., A.S., J.A.N., V.B., A.A.E., A.E., and G.C.A. contributed new analytic tools; K.E., C.K.S., T.B., B.C., A.S., J.A.B., J.J.D.-P., V.B., J.A.N., T.K., A.A.E., L.S., M.J.O.J., T.M., A.R., J.D., T.T., A.G.-P., H.S., A.E., V.H., and G.C.A. analyzed data; and K.E., C.K.S., B.C., A.G.-P., V.H., and G.C.A. wrote the paper.

1. A. Harms, D. E. Brodersen, N. Mitarai, K. Gerdes, Toxins, targets, and triggers: An overview of toxin-antitoxin biology. *Mol. Cell* **70**, 768–784 (2018).
2. D. Jurėnas, N. Fraikin, F. Goormaghtigh, L. Van Melderen, Biology and evolution of bacterial toxin-antitoxin systems. *Nat. Rev. Microbiol.* **20**, 335–350 (2022), 10.1038/s41579-021-00661-1.
3. T. Ogura, S. Hiraga, Mini-F plasmid genes that couple host cell division to plasmid proliferation. *Proc. Natl. Acad. Sci. U.S.A.* **80**, 4784–4788 (1983).
4. H. Sberro *et al.*, Discovery of functional toxin/antitoxin systems in bacteria by shotgun cloning. *Mol. Cell* **50**, 136–148 (2013).
5. R. Leplae *et al.*, Diversity of bacterial type II toxin-antitoxin systems: A comprehensive search and functional analysis of novel families. *Nucleic Acids Res.* **39**, 5513–5525 (2011).
6. E. M. Fozo *et al.*, Abundance of type I toxin-antitoxin systems in bacteria: Searches for new candidates and discovery of novel families. *Nucleic Acids Res.* **38**, 3743–3759 (2010).
7. T. R. Blower *et al.*, Identification and classification of bacterial Type III toxin-antitoxin systems encoded in chromosomal and plasmid genomes. *Nucleic Acids Res.* **40**, 6158–6173 (2012).
8. H. Akarsu *et al.*, TASmania: A bacterial Toxin-Antitoxin Systems database. *PLoS Comput. Biol.* **15**, e1006946 (2019).
9. S. Jimmy *et al.*, A widespread toxin-antitoxin system exploiting growth control via alarmone signaling. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 10500–10510 (2020).
10. G. Horesh *et al.*, SLING: A tool to search for linked genes in bacterial datasets. *Nucleic Acids Res.* **46**, e128 (2018).
11. K. S. Makarova, Y. I. Wolf, S. Snir, E. V. Koonin, Defense islands in bacterial and archaeal genomes and prediction of novel defense systems. *J. Bacteriol.* **193**, 6039–6056 (2011).
12. V. Anantharaman, L. Aravind, New connections in the prokaryotic toxin-antitoxin network: Relationship with the eukaryotic nonsense-mediated RNA decay system. *Genome Biol.* **4**, R81 (2003).

13. C. K. Saha, R. Sanches Pires, H. Brolin, M. Delannoy, G. C. Atkinson, FlaGs and webFlaGs: Discovering novel biology through the analysis of gene neighbourhood conservation. *Bioinformatics* **37**, 1312–1314 (2021).
14. T. Kurata *et al.*, A hyperpromiscuous antitoxin protein domain for the neutralization of diverse toxin domains. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2102212119 (2022).
15. A. Kelly, T. J. Arrowsmith, S. C. Went, T. R. Blower, Toxin-antitoxin systems as mediators of phage defence and the implications for abortive infection. *Curr. Opin. Microbiol.* **73**, 102293 (2023).
16. M. LeRoux, M. T. Laub, Toxin-antitoxin systems as phage defense elements. *Annu. Rev. Microbiol.* **76**, 21–43 (2022), 10.1146/annurev-micro-020722-013730.
17. A. Fillol-Salom *et al.*, Bacteriophages benefit from mobilizing pathogenicity islands encoding immune systems against competitors. *Cell* **185**, 3248–3262.e20 (2022).
18. C. N. Vassallo, C. R. Doering, M. L. Littlehale, G. I. C. Teodoro, M. T. Laub, A functional selection reveals previously undetected anti-phage defence systems in the E. coli pangenome. *Nat. Microbiol.* **7**, 1568–1579 (2022).
19. F. Rousset *et al.*, Phages and their satellites encode hotspots of antiviral systems. *Cell Host Microbe.* **30**, 740–753.e5 (2022).
20. M. LeRoux *et al.*, The DarTG toxin-antitoxin system provides phage defence by ADP-ribosylating viral DNA. *Nat. Microbiol.* **7**, 1028–1040 (2022), 10.1038/s41564-022-01153-5.
21. C. K. Guegler, M. T. Laub, Shutoff of host transcription triggers a toxin-antitoxin system to cleave phage RNA and abort infection. *Mol. Cell* **81**, 2361–2373.e9 (2021).
22. T. Zhang *et al.*, Direct activation of a bacterial innate immune system by a viral capsid protein. *Nature* **612**, 132–140 (2022).
23. L. Zhu, J. D. Sharp, H. Kobayashi, N. A. Woychik, M. Inouye, Noncognate Mycobacterium tuberculosis toxin-antitoxins can physically and functionally interact. *J. Biol. Chem.* **285**, 39732–39738 (2010).
24. K. S. Makarova, Y. I. Wolf, E. V. Koonin, Comprehensive comparative-genomic analysis of type 2 toxin-antitoxin systems and related mobile stress response systems in prokaryotes. *Biol. Direct* **4**, 19 (2009).
25. F. Tesson *et al.*, Systematic and quantitative view of the antiviral arsenal of prokaryotes. *Nat. Commun.* **13**, 2561 (2022).
26. J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
27. P. Bryant, G. Pozzati, A. Elofsson, Improved prediction of protein-protein interactions using AlphaFold2. *Nat. Commun.* **13**, 1265 (2022).
28. Y. M. Deng, C. Q. Liu, N. W. Dunn, Genetic organization and functional analysis of a novel phage abortive infection system, AbiL, from Lactococcus lactis. *J. Biotechnol.* **67**, 135–149 (1999).
29. S. Doron *et al.*, Systematic discovery of antiphage defense systems in the microbial pangenome. *Science* **359**, eaar4120 (2018).
30. A. Millman *et al.*, An expanded arsenal of immune systems that protect bacteria from phages. *Cell. Host Microbe.* **30**, 1556–1569 e5 (2022).
31. Y. Xie *et al.*, TADB 2.0: An updated database of bacterial type II toxin-antitoxin loci. *Nucleic Acids Res.* **46**, D749–D753 (2018).
32. R. Cheng *et al.*, A nucleotide-sensing endonuclease from the Gabija bacterial defense system. *Nucleic Acids Res.* **49**, 5216–5229 (2021).
33. W. G. Miller *et al.*, Diversity within the Campylobacter jejuni type I restriction-modification loci. *Microbiology (Reading)* **151**, 337–351 (2005).
34. L. Zimmermann *et al.*, A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J. Mol. Biol.* **430**, 2237–2243 (2018).
35. R. K. Lau, E. Enustun, Y. Gu, J. V. Nguyen, K. D. Corbett, A conserved signaling pathway activates bacterial CBASS immune signaling in response to DNA damage. *EMBO J.* **22**, e111540 (2022).
36. B. Bose, A. D. Grossman, Regulation of horizontal gene transfer in Bacillus subtilis by activation of a conserved site-specific protease. *J. Bacteriol.* **193**, 22–29 (2011).
37. D. Matelska, K. Steczkiewicz, K. Ginalski, Comprehensive classification of the PIN domain-like superfamily. *Nucleic Acids Res.* **45**, 6995–7020 (2017).
38. V. L. Arcus, J. L. McKenzie, J. Robson, G. M. Cook, The PIN-domain ribonucleases and the prokaryotic VapBC toxin–antitoxin array. *Protein Eng. Des. Sel.* **24**, 33–40 (2010).
39. R. Hazan, H. Engelberg-Kulka, *Escherichia coli mazEF*-mediated cell death as a defense mechanism that inhibits the spread of phage P1. *Mol. Genet. Genomics* **272**, 227–234 (2004).
40. Y. Cui *et al.*, Bacterial MazF/MazE toxin-antitoxin suppresses lytic propagation of arbitrium-containing phages. *Cell Rep.* **41**, 111752 (2022).
41. C. D. Aakre *et al.*, Evolving new protein-protein interaction specificity through promiscuous intermediates. *Cell* **163**, 594–606 (2015).
42. B. L. Brown *et al.*, Three dimensional structure of the MqsR:MqsA complex: A novel TA pair comprised of a toxin homologous to RelE and an antitoxin with unique properties. *PLoS Pathog.* **5**, e1000706 (2009).
43. G. Mariano, T. R. Blower, Conserved domains can be found across distinct phage defence systems. *Mol. Microbiol.* **120**, 45–53 (2023), 10.1111/mmi.15047.
44. A. Garcia-Pino *et al.*, Doc of prophage P1 is inhibited by its antitoxin partner Phd through fold complementation. *J. Biol. Chem.* **283**, 30821–30827 (2008).
45. D. Castro-Roa *et al.*, The Fic protein Doc uses an inverted substrate to phosphorylate and inactivate EF-Tu. *Nat. Chem. Biol.* **9**, 811–817 (2013).
46. B. L. Brown, T. K. Wood, W. Peti, R. Page, Structure of the *Escherichia coli* antitoxin MqsA (YgiT/b3021) bound to its gene promoter reveals extensive domain rearrangements and the specificity of transcriptional regulation. *J. Biol. Chem.* **286**, 2285–2296 (2011).
47. N. Buddelmeijer, J. Beckwith, A complex of the *Escherichia coli* cell division proteins FtsL, FtsB and FtsQ forms independently of its localization to the septal region. *Mol. Microbiol.* **52**, 1315–1327 (2004).
48. H. Strahl, L. W. Hamoen, Membrane potential is important for bacterial cell division. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 12281–12286 (2010).
49. D. M. Heller, M. Tavag, A. Hochschild, CbtA toxin of Escherichia coli inhibits cell division and cell elongation via direct and independent interactions with FtsZ and MreB. *Plos Genet.* **13**, e1007007 (2017).
50. E. Maffei *et al.*, Systematic exploration of Escherichia coli phage-host interactions with the BASEL phage collection. *PLoS Biol.* **19**, e3001424 (2021).
51. Y. Gao *et al.*, Molecular basis of RADAR anti-phage supramolecular assemblies. *Cell* **186**, 999–1012. e20 (2023).
52. B. Duncan-Lowey *et al.*, Cryo-EM structure of the RADAR supramolecular anti-phage defense complex. *Cell* **186**, 987–998 e15 (2023).
53. T. Madl *et al.*, Structural basis for nucleic acid and toxin recognition of the bacterial antitoxin CcdA. *J. Mol. Biol.* **364**, 170–185 (2006).
54. K. Mattison, J. S. Wilbur, M. So, R. G. Brennan, Structure of FitAB from Neisseria gonorrhoeae bound to DNA reveals a tetramer of toxin-antitoxin heterodimers containing pin domains and ribbon-helix-helix motifs. *J. Biol. Chem.* **281**, 37942–37951 (2006).
55. K. S. Winther, K. Gerdes, Enteric virulence associated protein VapC inhibits translation by cleavage of initiator tRNA. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 7403–7407 (2011).
56. K. S. Winther, D. E. Brodersen, A. K. Brown, K. Gerdes, VapC20 of Mycobacterium tuberculosis cleaves the sarcin-ricin loop of 23S rRNA. *Nat. Commun.* **4**, 2796 (2013).
57. R. Loris, A. Garcia-Pino, Disorder- and dynamics-based regulatory mechanisms in toxin-antitoxin modules. *Chem. Rev.* **114**, 6933–6947 (2014).
58. Y. G. Sterckx *et al.*, The ParE2-PaaA2 toxin-antitoxin complex from Escherichia coli O157 forms a heterodocecamer in solution and in the crystal. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **68**, 724–729 (2012).
59. H. Takagi *et al.*, Crystal structure of archaeal toxin-antitoxin RelE-RelB complex with implications for toxin activity and antitoxin effects. *Nat. Struct. Mol. Biol.* **12**, 327–331 (2005).
60. S. Hadzi *et al.*, Ribosome-dependent Vibrio cholerae mRNAse HigB2 is regulated by a beta-strand sliding mechanism. *Nucleic Acids Res.* **45**, 4972–4983 (2017).
61. M. A. Schumacher *et al.*, Molecular mechanisms of HipA-mediated multidrug tolerance and its neutralization by HipB. *Science* **323**, 396–401 (2009).
62. A. Talavera *et al.*, A dual role in regulation and toxicity for the disordered N-terminus of the toxin GraT. *Nat. Commun.* **10**, 972 (2019).
63. P. Bordes *et al.*, SecB-like chaperone controls a toxin-antitoxin stress-responsive system in *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 8438–8443 (2011).
64. M. L. Yarbrough *et al.*, AMPylation of Rho GTPases by Vibrio VopS disrupts effector binding and downstream signaling. *Science* **323**, 269–272 (2009).
65. K. M. Dalton, S. Crosson, A conserved mode of protein recognition and binding in a ParD-ParE toxin-antitoxin complex. *Biochemistry* **49**, 2205–2215 (2010).
66. T. Kurata *et al.*, RelA-SpoT Homolog toxins pyrophosphorylate the CCA end of tRNA to inhibit protein synthesis. *Mol. Cell* **81**, 3160–3170.e9 (2021).
67. N. Fraikin, F. Goormaghtigh, L. Van Melderen, Type II toxin-antitoxin systems: Evolution and revolutions. *J. Bacteriol.* **202** (2020).
68. D. C. Pecota, T. K. Wood, Exclusion of T4 phage by the hok/sok killer locus from plasmid R1. *J. Bacteriol.* **178**, 2044–2050 (1996).
69. G. Jankevicius, A. Ariza, M. Ahel, I. Ahel, The toxin-antitoxin system DarTG catalyzes reversible ADP-ribosylation of DNA. *Mol. Cell* **64**, 1109–1116 (2016).
70. I. Songailiene *et al.*, HEPN-MNT toxin-antitoxin system: The HEPN ribonuclease is neutralized by OligoAMPylation. *Mol. Cell* **80**, 955–970.e7 (2020).
71. M. G. Jorgensen, D. P. Pandey, M. Jaskolska, K. Gerdes, HicA of Escherichia coli defines a novel family of translation-independent mRNA interferases in bacteria and archaea. *J. Bacteriol.* **191**, 1191–1199 (2009).
72. D. S. Black, A. J. Kelly, M. J. Mardis, H. S. Moyed, Structure and organization of hip, an operon that affects lethality due to inhibition of peptidoglycan or DNA synthesis. *J. Bacteriol.* **173**, 5732–5739 (1991).
73. N. A. O'Leary *et al.*, Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–D745 (2016).
74. A. Marchler-Bauer *et al.*, CDD: A Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res.* **39**, D225–229 (2011).
75. M. Steinegger *et al.*, HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinformatics* **20**, 473 (2019).
76. M. van Kempen *et al.*, Fast and accurate protein structure search with Foldseek. *Nat. Biotechnol.*, 10.1038/s41587-023-01773-0 (2023).
77. P. Shannon *et al.*, Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
78. R. Dong, Z. Peng, Y. Zhang, J. Yang, mTM-align: An algorithm for fast and accurate multiple protein structure alignment. *Bioinformatics* **34**, 1719–1725 (2018).
79. J. Hallgren, DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. bioRxiv [Preprint] (2022). https://doi.org/10.1101/2022.04.08.487609 (Accessed 2 January 2023).
80. L. M. Guzman, D. Belin, M. J. Carson, J. Beckwith, Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.* **177**, 4121–4130 (1995).
81. M. Jaskolska, K. Gerdes, CRP-dependent positive autoregulation and proteolytic degradation regulate competence activator Sxy of Escherichia coli. *Mol. Microbiol.* **95**, 833–845 (2015).
82. B. L. Roth, M. Poot, S. T. Yue, P. J. Millard, Bacterial viability and antibiotic susceptibility testing with SYTOX green nucleic acid stain. *Appl. Environ. Microbiol.* **63**, 2421–2431 (1997).
83. A. Waggoner, Optical probes of membrane potential. *J. Membr. Biol.* **27**, 317–334 (1976).
84. C. K. Saha *et al.*, Netflax_data. Github. https://github.com/GCA-VH-lab/NetFlax_data. Deposited 9 July 2023.

Bioinformatics, 2023, **39(5)**, btad323
https://doi.org/10.1093/bioinformatics/btad323
Advance access publication 15 May 2023
Applications Note

OXFORD

Genome analysis

uORF4u: a tool for annotation of conserved upstream open reading frames

Artyom A. Egorov [1,*] **and Gemma C. Atkinson**[1,*]

[1]Department of Experimental Medical Science, Lund University, Sölvegatan 19, Lund, 223 62, Sweden

*Corresponding authors. Department of Experimental Medical Science, Lund University, Sölvegatan 19, Lund, 223 62, Sweden. E-mail: artem.egorov@med.lu.se (A.A.E.); gemma.atkinson@med.lu.se (G.C.A.)

Associate Editor: Tobias Marschall

Abstract

Summary: Upstream open reading frames (uORFs, often encoding so-called leader peptides) can regulate translation and transcription of downstream main ORFs (mORFs) in prokaryotes and eukaryotes. However, annotation of novel functional uORFs is challenging due to their short size of usually <100 codons. While transcription- and translation-level next-generation sequencing methods can be used for genome-wide functional uORF identification, this data are not available for the vast majority of species with sequenced genomes. At the same time, the exponentially increasing amount of genome assemblies gives us the opportunity to take advantage of evolutionary conservation in our predictions of functional ORFs.

Here, we present a tool for conserved uORF annotation in 5′ upstream sequences of a user-defined protein of interest or a set of protein homologs. It can also be used to find small conserved ORFs within a set of nucleotide sequences. The output includes publication-quality figures with multiple sequence alignments, sequence logos, and locus annotation of the predicted conserved uORFs in graphical vector format.

Availability and implementation: uORF4u is written in Python3 and runs on Linux and MacOS. The command-line interface covers most practical use cases, while the provided Python API allows usage within a Python program and additional customization. Source code is available from the GitHub page: github.com/GCA-VH-lab/uorf4u. Detailed documentation that includes an example-driven guide available at the software home page: gca-vh-lab.github.io/uorf4u. A web version of uORF4u is available at server.atkinson-lab.com/uorf4u.

1 Introduction

Functional upstream open reading frames (uORFs, encoding what are often referred to as leader peptides) regulate expression of downstream genes via translational and/or transcriptional attenuation in bacteria, archaea, eukaryotes, and viruses (Ito and Chiba 2013; Dever et al. 2020). Despite the fact that the structures of mRNA transcripts, and the mechanisms of translation initiation differ among the domains of life, the concept of uORFs as evolutionarily conserved cis-acting regulatory elements is universal (Dever et al. 2020). Regulatory uORFs typically act via condition-specific ribosome stalling on nascent peptides. However, some uORFs can be translated into functional proteins (Andreev et al. 2015; Brown et al. 2017; Chen et al. 2020; Jayaram et al. 2021). Prediction and annotation of potentially functional uORFs is essential for understanding complex regulation mechanisms, including inducible expression of antibiotic resistance genes upon an antibiotic challenge (Ramu et al. 2009; Ito and Chiba 2013). However, the prediction of functional uORFs is complicated by their properties, such as short length, variable distance to the main ORF (mORF), and unusual sequence composition. A breakthrough in genome-wide annotation of translated regions including uORFs came with the single-nucleotide resolution sequencing method ribosome profiling (Ribo-Seq) (Ingolia et al. 2009; Brar and Weissman 2015). Several bioinformatics tools were created for annotation of uORFs based on Ribo-Seq data, for example uORF-Tools (Scholz et al. 2019) and uORF-seqr (Spealman et al. 2021). However, Ribo-Seq data are available only for a limited number of usually model organisms, and annotation of short uORFs is often complicated by noise in phased signal tracks that appears due to the stochastic nature and specific cutting preferences of RNases (Gerashchenko and Gladyshev 2017). Another limitation of the Ribo-Seq approach to annotation is that translation of uORFs may only be induced in certain environmental or cell conditions; for example, translation of non-AUG uORFs that are only detectable under stress conditions (Andreev et al. 2015).

In the absence of ribosome profiling data, researchers often annotate potentially functional uORFs in a manual or semi-manual way. This involves retrieval of the 5′ upstream regions of protein-coding genes of interest from sequence databases, and visual inspection of the region, with or without the aid of a sequence alignment to indicate functional conservation. Tell-tale signatures of potentially functional uORFs are the presence of Shine-Dalgarno elements (in the case of prokaryotes), along with start and stop codons in the right context and frame (Sakiyama et al. 2021; Mangano et al. 2022; Takada et al. 2022). However, inspection of large sequence alignments by eye is a time-consuming and tedious task. Thus, various methods have been developed to automate

Received: November 17, 2022. **Revised:** April 3, 2023. **Editorial Decision:** May 8, 2023. **Accepted:** May 12, 2023

ORF annotation, even in the absence of expression data. The tool sORF finder (Hanada et al. 2010) takes advantage of nucleotide composition bias and can predict small eukaryotic ORFs of high coding potential within 10–100 amino acids length range. MiPepid (Zhu and Gribskov 2019) and csORFfinder (Zhang et al. 2022) are ML-based approaches trained on a limited set of eukaryotic organisms for prediction of micropeptides translated from small ORFs. We have found only one tool, uPEPperoni (Skarshewski et al. 2014), that when it was available, implemented conservation analysis for prediction. Other methods that take into account conservation analyses in functional uORF prediction are not distributed as tools (McGillivray et al. 2018; Spealman et al. 2018; Liu et al. 2023). Importantly, all these methods are designed for use with eukaryotic genomes.

Thus, there is currently a lack of a simple tool for functional uORF prediction in both prokaryotes and eukaryotes that leverages sequence conservation. To fill this gap, we set out to build a tool that also includes the following key properties:

1) Ease of installation and implementation with a command-line interface and Python API for higher customization.
2) Does not have a requirement to build or download large databases. The tool uses the NCBI API to access the RefSeq database (O'Leary et al. 2016) and is therefore always up-to-date.
3) Supports various input formats: a user-defined protein as the mORF, set of mORF homologs, or nucleotide sequences in FASTA format.
4) Thorough documentation with a home page that contains an example-driven guide and detailed API description.
5) Output that contains publication-ready and editable vector graphics.
6) Can be used for sequences across the tree of life (bacteria, archaea, eukaryotes, and viruses).

2 The uORF4u workflow

The architecture of the uORF4u workflow is defined on user input (Fig. 1A). If the input is a single RefSeq protein accession number, uORF4u performs a BlastP search (Camacho et al. 2009) against the online version of the RefSeq protein database (O'Leary et al. 2016). The retrieved list of homologs is saved to be used in the subsequent steps. Alternatively, a list of homologs previously curated by the user can be used as input. This is important for allowing the user to decide the breadth and depth of the search; uORFs may differ in their conservation levels across strains and species and therefore it might be necessary to test different input sets. Using the accession list, uORF4u retrieves the corresponding upstream sequences using the NCBI API as implemented in Biopython (Cock et al. 2009). For eukaryotes, the upstream region is the complete transcript's 5′ UTR sequence, and for non-eukaryotic microbes, the upstream region is a user-defined (default 500 nucleotides) length from the mORF start codon. The retrieved nucleotide sequences are saved as intermediate output in FASTA format. These sequences, as well as the list of homologs obtained in the previous step, can also be used as optional input for uORF4 in order to skip the previous steps. It is useful to note that when using nucleotide sequences as input uORF4u can be used as a general conserved ORF search tool. That is, to find ORFs that are not necessarily upstream of any particular mORF.

The next step after sequence retrieval is ORF annotation. An ORF is defined as a region between a start codon (alternative start codons can be included as well) and a downstream in-frame stop codon. The minimal length set by default is nine nucleotides (three codons). For prokaryotes, this step also includes Shine-Dalgarno (SD) sequence search within a 20-nucleotide window upstream of the start codon. SD sequence annotation is based on the calculation of the SD-antiSD interaction Gibbs free energy (Yang et al. 2016). For identified potential frames, the tool searches for conserved ORFs using a greedy algorithm: uORF4u iterates through sequences and tries to maximize the sum of pairwise alignment scores between uORFs. The detailed scheme of the algorithm is available at the Github and server home pages. The last step in our workflow is generation of multiple sequence alignments (MSAs) of the identified conserved uORFs, writing reports, and making results visualization files (annotation plots, sequence logos, and MSAs). To do this, we have made our own MSA visualization package, MSA4u, which is bundled with uORF4u (github.com/GCA-VH-lab/msa4u). Examples of output plots are shown on Fig. 1C–E, H, and G.

3 Implementation

uORF4u is written in Python3 and uses multiple python libraries: Biopython (Cock et al. 2009), configs, argparse, pandas, statistics, Logomaker (Tareen and Kinney 2020), matplotlib (Hunter 2007), reportlab, and msa4u.

The python uORF4u package is available in PyPI (*python3 -m pip install uorf4u*), and the source code is provided on the GitHub page (github.com/GCA-VH-lab/uorf4u). Detailed documentation with an installation guide, and an example-driven manual are available at the uORF4u home page (gca-vh-lab.github.io/uorf4u). Additionally, the web version of uORF4u is also available at server.atkinson-lab.com/uorf4u.

The command-line interface allows users to run the tool with various standard usage scenarios without any additional effort for user-side scripting. Furthermore, we provide a python API that allows additional customization.

4 Conclusion

The problem of novel functional uORF annotation requires specialized tools. Here, we present uORF4u, which performs database parsing, uORF searching, conservation analysis, and produces publication-quality images of the results. The utility of uORF4u has been demonstrated with the discovery of uORFs that have been validated to regulate expression of ABCF antibiotic resistance genes (Obana et al. 2023), and the rediscovery of known functional uORFs presented in Fig. 1B–G.

We believe that as well as identifying potentially functional uORFs of specific mORFs in targeted analyses that include experimental validation, our tool paves the way for systematic analysis of uORF genesis, conservation, and distribution on the scale of the whole proteomes.

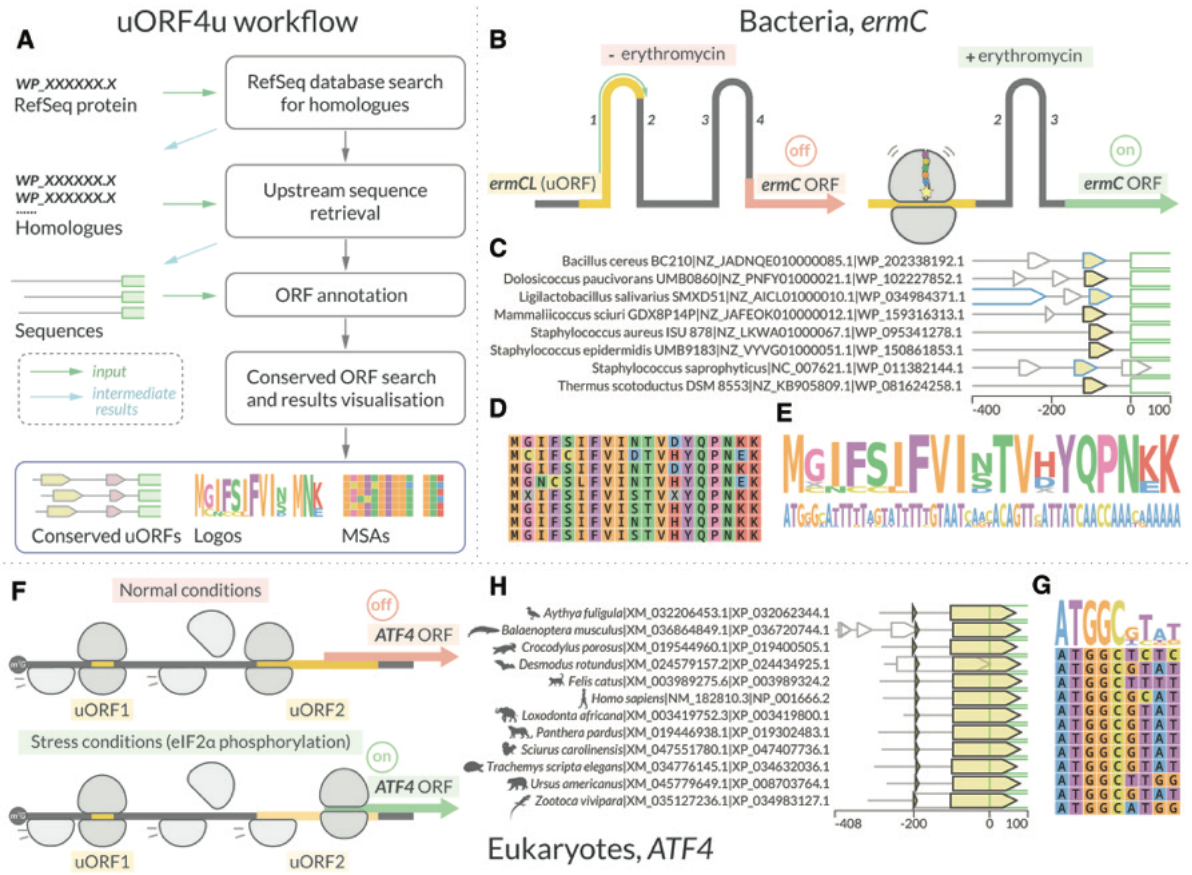


Figure 1. uORF4u workflow and usage examples. (A) The uORF4u workflow. The input data are indicated with green arrows. The simplest input is a user-defined NCBI RefSeq protein accession number but could also be a list of homologs (list of RefSeq accession numbers) or a FASTA format file of sequences. Depending on input, uORF4u consists of several steps: search for homologs against the RefSeq database with BlastP [writing the intermediate results (blue arrows) to output], upstream sequence retrieval with results written in an output FASTA file, ORF annotation, and, finally, conservation analysis and results visualization. (B–E) Results example using eight bacterial *ermC* proteins as a query (command: *uorf4u -hl WP_202338192.1 WP_102227852.1 WP_034984371.1 WP_159316313.1 WP_095341278.1 WP_150861853.1 WP_011382144.1 WP_081624258.1 -o ermC -c prokaryotes –annot*). The subset of query proteins was selected from the results obtained with an extended run, see homepage guide for details). (B) A bacterial uORF example: regulation of *ermC* expression by translational attenuation. Ribosome stalling on the *ermC* uORF (leader peptide *ermCL*; shown in yellow) is inducible by erythromycin. The arrest alters the regional mRNA structure, exposing the *ermC* SD sequence and allowing translation of the *ermC* mORF (Vazquez-Laslop et al. 2008). (C) Annotation plot of the upstream sequences with the conserved *ermCL* uORF. *ermCL* is shown in yellow, the 5′ end of the *ermC* mORF is shown with a green outline, grey outlines indicate other putative ORFs around this locus, and, finally, the blue outline shows ORFs annotated in the RefSeq database. In this case, three of the eight were already annotated, and five additional ORFs were annotated by uORF4u (black outline). uORF4u does not use RefSeq ORF annotations in its predictions, and its ability to rediscover both known and missing uORFs validates our strategy (D, E) Multiple sequence alignment and sequence logo visualization of the identified *ermCL* ORFs. (F–G) Results example using twelve eukaryotic *ATF4* proteins as a query (command: *uorf4u -hl NP_001666.2 XP_036720744.1 XP_024434925.1 XP_034632036.1 XP_008703764.1 XP_034983127.1 XP_019400505.1 XP_003989324.2 XP_003419800.1 XP_019302483.1 XP_047407736.1 XP_032062344.1 -c eukaryotes*). As with ermC, the list was built using the extended run results). (F) A eukaryotic uORFs example: the expression of *ATF4* (activating transcription factor) is regulated by two uORFs. After translation of the first uORF1, ribosomes are normally able to reinitiate translation at a downstream uORF2 after rebinding the initiating ternary complex (*eIF2-GTP-Met-tR- NA*). Reduced levels of the ternary complex during stress conditions leads to the ribosome scanning through the uORF2 start codon and instead reinitiating at the *ATF4* uORF (Vattem and Wek 2004). (H) Annotation plot of the 5′ UTRs with both conserved uORFs shown in yellow with black outline. (G) Multiple sequence alignment and sequence logo visualization of the identified uORF1.

Acknowledgements

The authors thank Jose Nakamoto and Veda Bojar for uORF4u testing and Vasili Hauryliuk for comments on the article.

Data availability

All code and example data is available from the uORF4u GitHub page (github.com/GCA-VH-lab/uorf4u).

Conflict of interest

The authors declare that there is no conflict of interest.

Funding

The work was supported by grants from the Swedish Research Council (Vetenskapsrådet) (2019-01085 and 2022-01603 to G.C.A.), the Knut and Alice Wallenberg Foundation (2020-0037 to G.C.A.), and Carl Tryggers Stiftelse för Vetenskaplig Forskning (CTS19:24 to G.C.A.).

References

Andreev DE, O'Connor PBF, Fahey C *et al.* Translation of 5' leaders is pervasive in genes resistant to eIF2 repression. Elife 2015;**4**:e03971.

Andreev DE, O'Connor PB, Zhdanov AV *et al.* Oxygen and glucose deprivation induces widespread alterations in mRNA translation within 20 minutes. *Genome Biol* 2015;**16**:1–14.

Brar GA, Weissman JS. Ribosome profiling reveals the what, when, where and how of protein synthesis. *Nat Rev Mol Cell Biol* 2015;**16**: 651–64.

Brown A, Rathore S, Kimanius D *et al.* Structures of the human mitochondrial ribosome in native states of assembly. *Nat Struct Mol Biol* 2017;**24**:866–9.

Camacho C, Coulouris G, Avagyan V *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* 2009;**10**:421.

Chen J, Brunner A-D, Cogan JZ *et al.* Pervasive functional translation of noncanonical human open reading frames. *Science* 2020;**367**: 1140–6.

Cock PJA, Antao T, Chang JT *et al.* Biopython: freely available python tools for computational molecular biology and bioinformatics. *Bioinformatics* 2009;**25**:1422–3.

Dever TE, Ivanov IP, Sachs MS. Conserved upstream open reading frame nascent peptides that control translation. *Annu Rev Genet* 2020;**54**:237–64.

Gerashchenko MV, Gladyshev VN. Ribonuclease selection for ribosome profiling. *Nucleic Acids Res* 2017;**45**:e6.

Hanada K, Akiyama K, Sakurai T *et al.* sORF finder: a program package to identify small open reading frames with high coding potential. *Bioinformatics* 2010;**26**:399–400.

Hunter JD. Matplotlib: a 2D graphics environment. *Comput Sci Eng* 2007;**9**:90–5.

Ingolia NT, Ghaemmaghami S, Newman JRS *et al.* Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* 2009;**324**:218–23.

Ito K, Chiba S. Arrest peptides: cis-acting modulators of translation. *Annu Rev Biochem* 2013;**82**:171–202.

Jayaram DR, Argov C., Chiba S. *et al.* Unraveling the hidden role of a uORF-encoded peptide as a kinase inhibitor of PKCs. *Proc Natl Acad Sci USA* 2021;**118**(40):e2018899118.

Liu Q, Peng X, Shen M *et al.* Ribo-uORF: a comprehensive data resource of upstream open reading frames (uORFs) based on ribosome profiling. *Nucleic Acids Res* 2023;**51**:D248–61.

Mangano K, Marks J, Klepacki D *et al.* Context-based sensing of orthosomycin antibiotics by the translating ribosome. *Nat Chem Biol* 2022;**18**:1277–86.

McGillivray P, Ault R, Pawashe M *et al.* A comprehensive catalog of predicted functional upstream open reading frames in humans. *Nucleic Acids Res* 2018;**46**:3326–38.

O'Leary NA, Wright MW, Brister JR *et al.* Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* 2016;**44**:D733–45.

Obana N, Takada H, Crowe-McAuliffe C *et al.* Genome-encoded ABCF factors implicated in intrinsic antibiotic resistance in gram-positive bacteria: vmlR2, Ard1 and CplR. *Nucleic Acids Res* 2023:gkad193.

Ramu H, Mankin A, Vazquez-Laslop N. Programmed drug-dependent ribosome stalling. *Mol Microbiol* 2009;**71**:811–24.

Sakiyama K, Shimokawa-Chiba N, Fujiwara K *et al.* Search for translation arrest peptides encoded upstream of genes for components of protein localization pathways. *Nucleic Acids Res* 2021;**49**:1550–66.

Scholz A, Eggenhofer F, Gelhausen R *et al.* uORF-Tools-Workflow for the determination of translation-regulatory upstream open reading frames. *PLoS One* 2019;**14**:e0222459.

Skarshewski A, Stanton-Cook M, Huber T *et al.* uPEPperoni: an online tool for upstream open reading frame location and analysis of transcript conservation. *BMC Bioinformatics* 2014;**15**:36.

Spealman P, Naik A, McManus J. uORF-seqr: a machine Learning-Based approach to the identification of upstream open reading frames in yeast. *Methods Mol Biol* 2021;**2252**:313–29.

Spealman P, Naik AW, May GE *et al.* Conserved non-AUG uORFs revealed by a novel regression analysis of ribosome profiling data. *Genome Res* 2018;**28**:214–22.

Takada H, Mandell ZF, Yakhnin H *et al.* Expression of *Bacillus subtilis* ABCF antibiotic resistance factor VmlR is regulated by RNA polymerase pausing, transcription attenuation, translation attenuation and (p)ppGpp. *Nucleic Acids Res* 2022;**50**:6174–89.

Tareen A, Kinney JB. Logomaker: beautiful sequence logos in python. *Bioinformatics* 2020;**36**:2272–4.

Vattem KM, Wek RC. Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells. *Proc Natl Acad Sci USA* 2004;**101**:11269–74.

Vazquez-Laslop N, Thum C, Mankin AS. Molecular mechanism of drug-dependent ribosome stalling. *Mol Cell* 2008;**30**:190–202.

Yang C, Hockenberry AJ, Jewett MC *et al.* Depletion of Shine-Dalgarno sequences within bacterial coding regions is expression dependent. *G3 (Bethesda)* 2016;**6**:3467–74.

Zhang M, Zhao J, Li C. *et al.* csORF-finder: an effective ensemble learning framework for accurate identification of multi-species coding short open reading frames. *Brief Bioinform* 2022;23(6):bbac392.

Zhu M, Gribskov M. MiPepid: microPeptide identification tool using machine learning. *BMC Bioinformatics* 2019;**20**:559.

Article

Direct activation of a bacterial innate immune system by a viral capsid protein

https://doi.org/10.1038/s41586-022-05444-z

Received: 26 May 2022

Accepted: 13 October 2022

Published online: 16 November 2022

Open access

Check for updates

Tong Zhang[1], Hedvig Tamman[2], Kyo Coppieters 't Wallant[3], Tatsuaki Kurata[4], Michele LeRoux[1], Sriram Srikant[1], Tetiana Brodiazhenko[5], Albinas Cepauskas[2], Ariel Talavera[2], Chloe Martens[3], Gemma C. Atkinson[4], Vasili Hauryliuk[4,5]✉, Abel Garcia-Pino[2,6]✉ & Michael T. Laub[1,7]✉

Bacteria have evolved diverse immunity mechanisms to protect themselves against the constant onslaught of bacteriophages[1–3]. Similar to how eukaryotic innate immune systems sense foreign invaders through pathogen-associated molecular patterns[4] (PAMPs), many bacterial immune systems that respond to bacteriophage infection require phage-specific triggers to be activated. However, the identities of such triggers and the sensing mechanisms remain largely unknown. Here we identify and investigate the anti-phage function of CapRel^{SJ46}, a fused toxin–antitoxin system that protects *Escherichia coli* against diverse phages. Using genetic, biochemical and structural analyses, we demonstrate that the C-terminal domain of CapRel^{SJ46} regulates the toxic N-terminal region, serving as both antitoxin and phage infection sensor. Following infection by certain phages, newly synthesized major capsid protein binds directly to the C-terminal domain of CapRel^{SJ46} to relieve autoinhibition, enabling the toxin domain to pyrophosphorylate tRNAs, which blocks translation to restrict viral infection. Collectively, our results reveal the molecular mechanism by which a bacterial immune system directly senses a conserved, essential component of phages, suggesting a PAMP-like sensing model for toxin–antitoxin-mediated innate immunity in bacteria. We provide evidence that CapRels and their phage-encoded triggers are engaged in a 'Red Queen conflict'[5], revealing a new front in the intense coevolutionary battle between phages and bacteria. Given that capsid proteins of some eukaryotic viruses are known to stimulate innate immune signalling in mammalian hosts[6–10], our results reveal a deeply conserved facet of immunity.

Innate immunity in eukaryotes relies on pattern recognition receptors that directly sense PAMPs, which are conserved molecules like bacterial lipopolysaccharide and flagellin, or viral RNA or DNA[4]. These innate immune signalling pathways must remain silent before infection, but be poised for rapid activation to defend against foreign invaders. Bacteria also encode innate immune systems to protect themselves against diverse invading bacteriophages, but how they sense infection is poorly understood. One exception is restriction-modification systems, but these distinguish self from non-self using DNA methylation patterns and thus do not need the nuclease effector to be specifically activated during phage infection. Similarly, for CRISPR–Cas systems, the adaptive immune system of some bacteria, guide RNAs enable a cell to specifically target foreign DNA without a need for infection-triggered activation. Dozens of bacterial defence systems have been identified in recent years[11–15], but unlike restriction modification and CRISPR–Cas, many of them must be specifically activated upon phage infection. This is particularly critical for abortive infection systems, in which a defence system uses a lethal effector to kill an infected cell and prevent propagation of the virus through a population[16]. The phage-encoded triggers for such bacterial immunity mechanisms are largely unknown.

Toxin–antitoxin systems are prevalent genetic elements in bacteria that are emerging as key components of anti-phage innate immunity[13,14,17,18], often serving as abortive infection modules that kill infected cells to prevent spread of phages through a population. How toxin–antitoxin systems sense and respond to phage infection remains poorly understood. For the *toxIN* system, toxin (ToxN) activation relies on efficient, phage-induced shutoff of host transcription coupled to the intrinsically fast turnover of the antitoxin *toxI*[19–21]. However, *toxI* is an RNA, whereas most toxin–antitoxin systems feature a protein antitoxin. For systems with a protein antitoxin, the mechanism of activation is often assumed to arise through antitoxin degradation. Although protein antitoxins are often more proteolytically unstable than their cognate toxins, their turnover may not be fast enough to enable toxin activation on the timescale of a phage infection[22], suggesting the existence of alternative mechanisms for toxin–antitoxin system activation. Bacterial retrons

[1]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA. [2]Cellular and Molecular Microbiology, Faculté des Sciences, Université Libre de Bruxelles, (ULB), Brussels, Belgium. [3]Centre for Structural Biology and Bioinformatics, Université Libre de Bruxelles (ULB), Bruxelles, Belgium. [4]Department of Experimental Medical Science, Lund University, Lund, Sweden. [5]Institute of Technology, University of Tartu, Tartu, Estonia. [6]WELBIO, Brussels, Belgium. [7]Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA, USA. ✉e-mail: vasili.hauryliuk@med.lu.se; abel.garcia.pino@ulb.be; laub@mit.edu

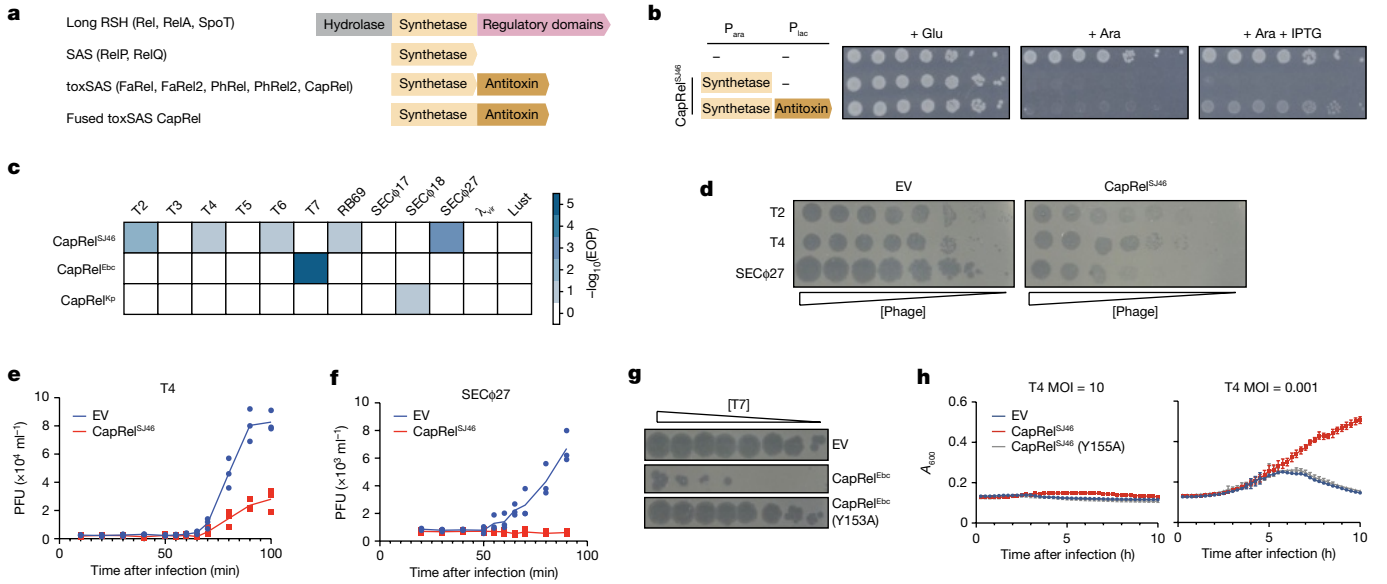


Fig. 1 | Fused CapRel homologues are toxin–antitoxin systems that can provide *E. coli* with robust defence against phages. a, Domain organization of long RelA-SpoT homologues (RSH), SAS, toxSAS and the fused subclass of toxSAS toxin–antitoxin systems including CapRel^{SJ46}. **b**, Cell viability assessed by serial dilutions for strains expressing the N-terminal toxin domain of CapRel^{SJ46} alone or with the C-terminal antitoxin domain. Ara, arabinose; glu, glucose; P_{ara}, arabinose-inducible, glucose-repressible promoter; P_{lac}, IPTG-inducible promoter. **c**, EOP data for the indicated phages when infecting cells producing CapRel^{SJ46}, CapRel^{Ebc} or CapRel^{Kp}. **d**, Serial, tenfold dilutions of the indicated phages spotted on lawns of cells harbouring plasmid expressing CapRel^{SJ46} or an empty vector (EV). Relative phage concentration is indicated by the height of the wedge. **e**,**f**, One-step growth curves measuring plaque-forming units (PFU) during the first round of infection by T4 (**e**) or SECΦ27 (**f**) in cells harbouring plasmid expressing CapRel^{SJ46} or an empty vector. **g**, Serial dilutions of T7 phage spotted on lawns of cells harbouring plasmids expressing CapRel^{Ebc}, CapRel^{Ebc}(Y153A) or an empty vector. **h**, Growth of cells producing CapRel^{SJ46} or CapRel^{SJ46}(Y155A) or harbouring an empty vector, following infection with T4 at a MOI of 10 or 0.001. Data are mean ± s.d. of eight plate replicates and representative of three independent experiments.

function as tripartite toxin–antitoxin systems and can be activated by overexpressing various prophage genes[23], but whether these activators function as such during phage infection is unknown.

CapRel^{SJ46} is an anti-phage fused toxin–antitoxin system

To investigate the molecular basis of phage-induced activation of bacterial immunity, we focused on toxSAS toxin–antitoxin systems, which feature toxins homologous to bacterial small alarmone synthetases (SAS) that pyrophosphorylate purine nucleotides[24]. Whereas most housekeeping alarmone synthetases produce the growth regulator (p)ppGpp[25,26], toxSAS toxins can synthesize (p)ppApp to deplete ATP[24,27] or pyrophosphorylate tRNAs to inhibit translation[24,28]. Their cognate antitoxins can either bind and neutralize the toxin or act as hydrolases to reverse toxin-catalysed pyrophosphorylation[24,28]. One subfamily of translation-inhibiting toxSAS is called CapRel, on the basis of their prevalence in Cyanobacteria, Actinobacteria, and Proteobacteria and sequence similarity to the (p)ppGpp synthetase/hydrolase Rel. The CapRel subfamily is the most broadly distributed subfamily of toxSAS, being found across multiple phyla of Gram-positive and Gram-negative bacteria. In addition to its prevalence in Cyanobacteria, Actinobacteria, and Proteobacteria, CapRel representatives are found in Spirochetes, Bacteroidetes, and Firmicutes[24]. This subfamily often features systems that are—in contrast to canonical bicistronic toxin–antitoxin systems—encoded by a single open reading frame, with an N-terminal domain homologous to toxSAS toxins and a C-terminal domain homologous to the corresponding antitoxins[29] (Fig. 1a and Extended Data Figs. 1 and 2a). This fused architecture is the predominant form of CapRel, except in Actinobacteria.

Prophages in bacterial genomes often encode anti-phage systems, helping both the temperate phage and its host to defend against other phages[11,30,31]. Here, we selected a fused CapRel encoded by the *Salmonella* temperate phage SJ46 and also encoded (with 100% amino acid sequence identity) in prophages of several *E. coli* strains (Extended Data Fig. 2b). The toxin and antitoxin-like regions of CapRel^{SJ46} are related to the PhRel toxSAS toxin and its antitoxin ATphRel, respectively, from the mycobacterial temperate phage Phrann[30] (Extended Data Fig. 1 and 2a). This Phrann-encoded system can inhibit superinfection by other temperate mycophages[30], although the molecular basis of PhRel activation is not known. To test whether CapRel^{SJ46} is a fused toxin–antitoxin system, we cloned the N-terminal region containing the conserved alarmone synthetase domain and the C-terminal region containing the putative antitoxin domain under the control of separate inducible promoters. Expression of the N-terminal fragment alone was toxic, and its toxicity was rescued in *trans* by co-expression with the C-terminal fragment (Fig. 1b and Extended Data Fig. 2c), suggesting that CapRel^{SJ46} is a fused toxin–antitoxin system.

To determine whether fused CapRels can defend against phages, we transformed *E. coli* MG1655 with three different systems expressed from their native promoters on low copy number plasmids, and then tested whether each conferred protection against a panel of 12 diverse coliphages. In addition to CapRel^{SJ46}, we also tested CapRel^{Ebc} from *Enterobacter chengduensis* and CapRel^{Kp} from *Klebsiella pneumoniae* (Fig. 1c and Extended Data Fig. 2b,d). CapRel^{SJ46} expressed from a low copy number plasmid or the *E. coli* genome decreased the efficiency of plaquing (EOP) for T2, T4, T6, RB69 and SECΦ27 by 10- to 1,000-fold (Fig. 1c,d and Extended Data Fig. 2e,f), indicating that this system provides strong protection against phages. T4 phage formed smaller plaques when plated onto CapRel^{SJ46}-containing cells, and one-step growth curves confirmed that CapRel^{SJ46} reduced the burst size of T4 by around 70% (Fig. 1e) and prevented bursting of SECΦ27 (Fig. 1f). CapRel^{Ebc} protected strongly against T7 and CapRel^{Kp} protected, albeit less efficiently, against SECΦ18 (Fig. 1g and Extended Data Fig. 2d,g).

Next, we tested whether CapRel^{SJ46} provides direct immunity or functions through abortive infection, in which an infected cell dies but

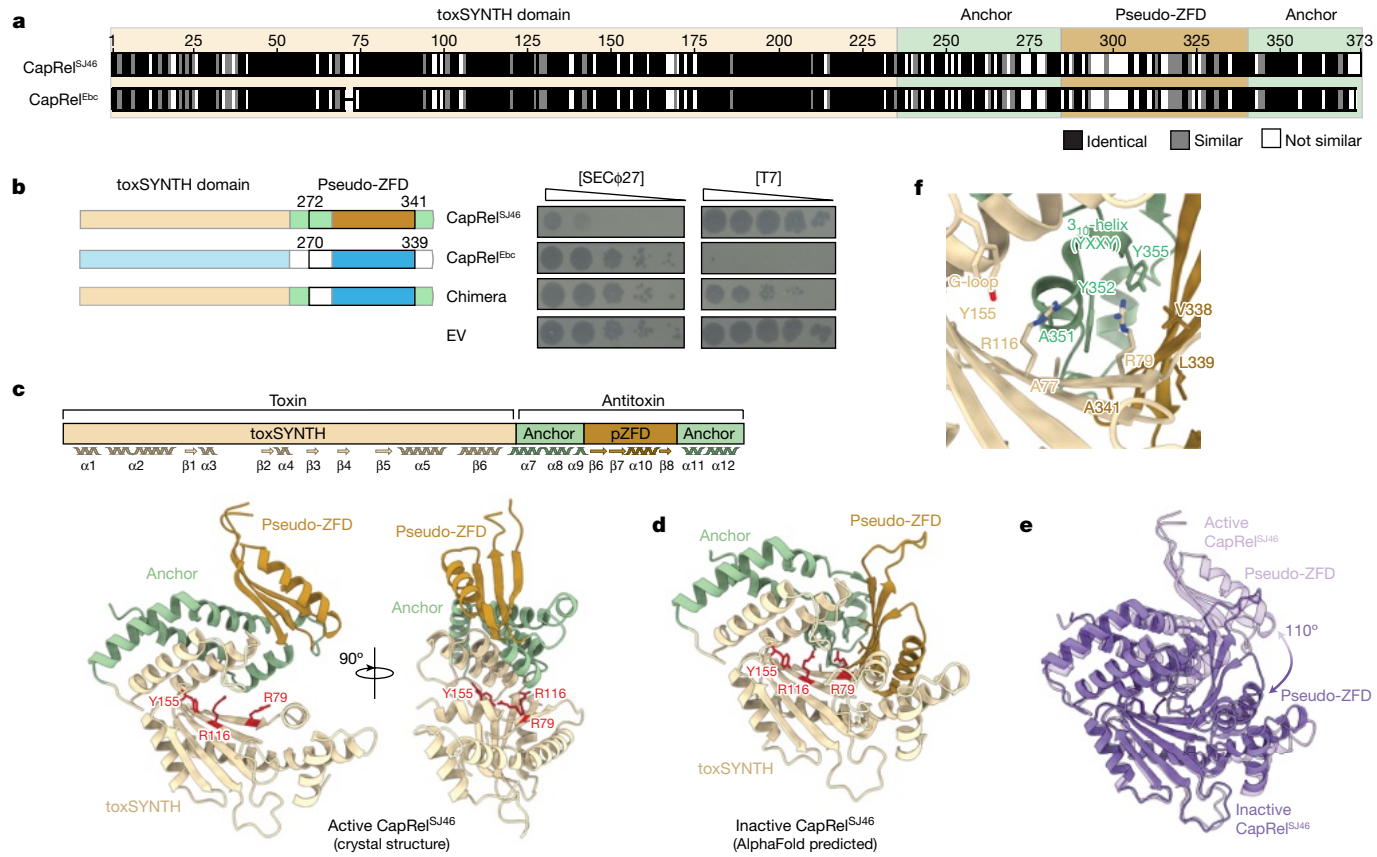


Fig. 2 | The pseudo-ZFD of CapRel confers phage specificity. a, Sequence alignment of CapRel^{SJ46} and CapRel^{Ebc}, with the more variable pseudo-ZFD labelled. **b**, Serial dilutions of the indicated phages spotted on lawns of cells harbouring the indicated CapRel constructs. Left, schematic of the CapRel constructs. **c**, Cartoon representation of the crystal structure of CapRel^{SJ46} with active site G-loop Y155 and the ATP-coordination residues R79 and R116 highlighted in red. Structural elements (toxSYNTH, pseudo-ZFD and the anchors) are coloured as in **a**. Wavy lines and arrows indicate alpha helices and beta strands, respectively. **d**, The closed conformation of CapRel^{SJ46} predicted by AlphaFold and coloured as in **c**. **e**, Superposition of the active (open, light purple) and inactive (closed, dark purple) states of CapRel^{SJ46} as observed in the crystal structure and predicted by AlphaFold. **f**, Details of the autoinhibited active site of CapRel^{SJ46} in the closed state. In this conformation, the YXXY neutralization motif blocks the adenine coordination site, preventing catalysis.

prevents the production of mature virions, thereby sparing uninfected cells in a population. To this end, we infected cells containing CapRel^{SJ46} with T4 at a multiplicity of infection (MOI) of either 10 or 0.001, and found that defence only manifested at the low MOI, indicating that CapRel^{SJ46} probably functions through abortive infection (Fig. 1h). Phage protection by CapRel^{SJ46} depended on the predicted enzymatic activity of the N-terminal synthetase domain, as substituting the conserved tyrosine (Y155A) in the G-loop that is critical for substrate binding abolished phage protection[32] (Fig. 1h). A similar catalysis-compromising substitution, Y153A in CapRel^{Ebc}, also abolished phage protection (Fig. 1g). Collectively, our results established that fused CapRels can provide anti-phage defence, with variable phage specificity.

To understand what determines the specificity of phage protection by fused CapRels, we compared CapRel^{SJ46} and CapRel^{Ebc}. These two proteins share 70% amino acid identity overall, but differ substantially in their C-terminal regions, which are only 47% identical (Fig. 2a). In addition, this region is the least conserved when we compared a more diverse set of fused CapRel homologues (Extended Data Fig. 3a). Because CapRel^{SJ46} and CapRel^{Ebc} protected against different phages, we made a chimera in which the C-terminal region of CapRel^{SJ46} was replaced by the corresponding region of CapRel^{Ebc}. This chimeric CapRel no longer protected against SECΦ27 and gained protection against T7, manifesting as a tenfold decrease in EOP and smaller plaques (Fig. 2b and Extended Data Fig. 3b). The chimera protected against T7 less efficiently than CapRel^{Ebc}, despite similar expression levels (Extended Data Fig. 3c), probably owing to a non-optimal interaction interface between the N and C termini. Nevertheless, this result indicates that the C-terminal region of CapRel is critical for phage specificity.

Structural analysis of CapRel^{SJ46}

To further understand the mechanistic basis of anti-phage defence by CapRel^{SJ46}, we solved a crystal structure to 2.3 Å resolution (Fig. 2c and Extended Data Table 1). CapRel^{SJ46} contains a conserved, N-terminal nucleotide pyrophosphokinase domain (toxSYNTH) that is also present in alarmone synthetases and tRNA-pyrophosphotransferase enzymes, which mediates toxicity. The smaller C-terminal antitoxin domain consists of a central antiparallel three-stranded β-sheet with an α-helix connecting β-strands β7 and β8 (Fig. 2c and Extended Data Fig. 3d,e). The antitoxin domain is topologically analogous to the classical zinc-finger domain (ZFD), but lacks the conserved cysteines (Extended Data Fig. 3f); we refer to this domain as a pseudo-ZFD. The pseudo-ZFD is connected to the toxSYNTH domain via α-helices α7, α8 and α9, and has a C-terminal α-helical extension that anchors the domain to α8 and α9 (Extended Data Fig. 3d). In this structure, the toxSYNTH and pseudo-ZFD are distant from each other, revealing the ATP donor nucleotide binding pocket and the conserved G-loop Y155 of toxSYNTH (Fig. 2c), indicating that this state may represent the active, toxic conformation of CapRel^{SJ46}. This open state captured by crystallography is probably stabilized by crystal lattice contacts that are incompatible with a closing of the active site owing to clashes with symmetry-related partners (Extended Data Fig. 3g).

To explore the conformational dynamics of the enzyme, we used AlphaFold[33] to predict possible alternative structures of CapRel^{SJ46}. In addition to predicting the open conformation observed in the crystal structure (Extended Data Fig. 3h), AlphaFold also predicted a closed conformational state in which the C-terminal domain folds back 110° onto the toxSYNTH central β-sheet and blocks the ATP-binding site (Fig. 2d). Comparison of the two states suggested that a conserved YXXY motif (Extended Data Fig. 3a) located in the hinge connecting the two C-terminal α-helices in the open state morphs into a short 3_{10}-helix in the closed state (Fig. 2e,f). This 3_{10}-helix projects into the toxSYNTH active site and intercalates between β1 R79 and β2 R116 to block the adenine coordination site (Fig. 2e,f).

We hypothesized that this closed-to-open switch underlies the activation of CapRel^{SJ46}, with the docking of the pseudo-ZFD onto toxSYNTH precluding substrate binding in the absence of phage infection (Fig. 2f). To test this hypothesis, we made single substitutions to the YXXY motif (Y352A and Y355A) and residues from the predicted interface that serves as a scaffold to orient and stabilize the 3_{10}-helix (A77K, R116A, V338A, L339A, A341K and A351K), which are highly conserved among diverse CapRel homologues (Extended Data Fig. 3a,i). Whereas wild-type CapRel^{SJ46} was not toxic when expressed in cells, each of the substitutions predicted to disrupt the intramolecular recognition interface, on either the N- or the C-terminal domain, rendered CapRel^{SJ46} toxic (Extended Data Fig. 3j,k). These substitutions probably lead to constitutive activation of CapRel^{SJ46} by disrupting an autoinhibited state. As a control, we showed that substitutions in different structural elements of the pseudo-ZFD but not pointing towards the interface did not lead to constitutive activation (Extended Data Fig. 3j,k). Collectively, our results indicate that the pseudo-ZFD docks onto the ATP-binding site of CapRel^{SJ46} to prevent switching to the open state captured in our crystal structure. Conservation of the YXXY motif and the interface residues suggests that this auto-inhibitory regulation is probably retained in other CapRels.

SECΦ27 capsid protein triggers CapRel^{SJ46}

Because full-length, wild-type CapRel^{SJ46} was not toxic when expressed in the absence of phage infection, we inferred that it must somehow be activated by phages. The toxins of some toxin–antitoxin systems are activated by the degradation of the more labile antitoxin[19,34,35]. To test whether the C-terminal antitoxin of CapRel^{SJ46} is proteolytically cleaved off and degraded upon phage infection, we N-terminally tagged CapRel^{SJ46} and first verified that the tagged protein still defends against phage (Extended Data Fig. 4a). We then tracked the size of CapRel^{SJ46} by immunoblotting following infection with SECΦ27. The overall protein levels of CapRel^{SJ46} remained constant and we observed only the full-length product, suggesting that CapRel^{SJ46} was not proteolytically processed (Fig. 3a and Extended Data Fig. 4b,c). Thus, we hypothesized that a specific phage product regulates the C-terminal domain of CapRel^{SJ46} to relieve autoinhibition. To identify such a factor, we sought to identify SECΦ27 mutants that escape CapRel^{SJ46} defence. As no spontaneous escape mutants could be isolated, we used an experimental evolution approach (Fig. 3b). In brief, we infected cells containing an empty vector or CapRel^{SJ46} with serial dilutions of phage in microtitre plates. After overnight incubation, we collected and pooled the phages from cleared wells, which indicated successful infection, and used these to seed the next round of infections. Initially, cells with the empty vector were infected more efficiently, but after 13 rounds, each phage population had evolved to infect cells containing empty vector or CapRel^{SJ46} to a similar degree (Fig. 3c and Extended Data Fig. 4d). We isolated ten mutant SECΦ27 clones from five independently evolved populations and sequenced their genomes. Remarkably, all ten clones contained a point mutation in the same gene that encodes a hypothetical protein, Gp57, with nine clones producing the same L114P substitution and one clone yielding an I115F substitution (Fig. 3d and Supplementary Table 1).

The structure of the hypothetical protein Gp57 predicted by AlphaFold[33] is highly similar (DALI *z*-score of approximately 17) to the HK97 fold commonly adopted by major capsid proteins of dsDNA viruses including bacteriophages and herpesviruses[36] (Fig. 3e). By performing mass spectrometry on wild-type and escape-mutant SECΦ27 phages, we identified this hypothetical protein as the most abundant protein in mature virions, consistent with it being the major capsid protein of SECΦ27 (Fig. 3f and Extended Data Table 2).

Our results suggested that wild-type Gp57 from SECΦ27 activates CapRel^{SJ46}, with the escape mutants preventing activation while retaining the ability to form a capsid. To test this hypothesis, we first examined whether Gp57 alone is sufficient to activate CapRel^{SJ46}. Indeed, co-producing wild-type Gp57 with wild-type CapRel^{SJ46} was highly toxic to cells in the absence of phage infection, whereas neither evolved variant (L114P or I115F) of Gp57 had a measurable effect on growth when co-produced with CapRel^{SJ46} (Fig. 3g and Extended Data Fig. 4e,f). As controls, we confirmed that expressing the wild-type or either Gp57 variant was not toxic on its own or if co-produced with a catalytically compromised CapRel^{SJ46} (Extended Data Fig. 4g).

To examine the basis of CapRel^{SJ46} toxicity, we first co-produced it with wild-type or the L114P variant of Gp57 and then measured the effects on bulk transcription or translation by pulse-labelling with ^{3}H-uridine or with ^{35}S-methionine and ^{35}S-cysteine, respectively. Active CapRel^{SJ46} produced with wild-type Gp57 robustly inhibited translation but not transcription (Fig. 3h and Extended Data Fig. 4h), whereas no effect was seen with Gp57(L114P). Similar effects were seen when overexpressing just the N-terminal domain of CapRel^{SJ46} (Extended Data Fig. 4i). We also measured bulk translation and transcription following SECΦ27 infection of CapRel^{SJ46}-containing cells and observed a decrease in translation but not transcription with wild-type SECΦ27 (Fig. 3i and Extended Data Fig. 4j), with a less pronounced decrease in translation compared to overproducing Gp57 likely because poor adsorption of SECΦ27 phage leads to substantial asynchrony of infection dynamics (Extended Data Fig. 4k). No effect on translation was seen with the evolved mutant phage producing Gp57(L114P) (Fig. 3i). To compare the timing of CapRel^{SJ46} activation during infection to the timing of Gp57 production, we expressed a haemagglutinin (HA)-tagged version of Gp57 from its native promoter in the bacterial genome and infected cells with SECΦ27. Immunoblotting indicated substantial accumulation of Gp57 by around 30 min after infection (Extended Data Fig. 4l), which coincides with the detected inhibition of translation (Fig. 3i).

Next, we measured the ability of full-length CapRel^{SJ46} to affect translation in vitro using the reconstituted in vitro transcription–translation system. Purified CapRel^{SJ46} inhibited synthesis of a control DHFR protein after pre-synthesizing SECΦ27 major capsid protein Gp57 (using a template encoding it), whereas no inhibition was seen for either the L114P or I115F variants of Gp57 (Fig. 3j). We also incubated wild-type or an evolved variant (L114P or I115F) of Gp57 with [γ-^{32}P]ATP and bulk *E. coli* tRNAs in the presence and absence of purified CapRel^{SJ46}. Wild-type Gp57 strongly stimulated the pyrophosphorylation of tRNAs by CapRel^{SJ46}, similar to the previously characterized toxSAS enzymes FaRel2 and PhRel2[28] (Fig. 3k). With the L114P or I115F variant of Gp57, tRNA pyrophosphorylation was reduced to the background levels seen with CapRel^{SJ46} alone. Together, our results demonstrate that Gp57, the major capsid protein of SECΦ27, is both necessary and sufficient to activate CapRel^{SJ46}, enabling it to pyrophosphorylate tRNAs and inhibit translation.

CapRel^{SJ46} binds SECΦ27 capsid protein

To test whether the SECΦ27 major capsid protein directly binds CapRel^{SJ46}, we first immunoprecipitated CapRel^{SJ46}–Flag from cells infected with wild-type phage or the mutant that produces Gp57(L114P) after verifying the tag does not affect CapRel^{SJ46} function (Extended Data Fig. 4a). We detected Gp57 that had co-precipitated with CapRel^{SJ46} by mass

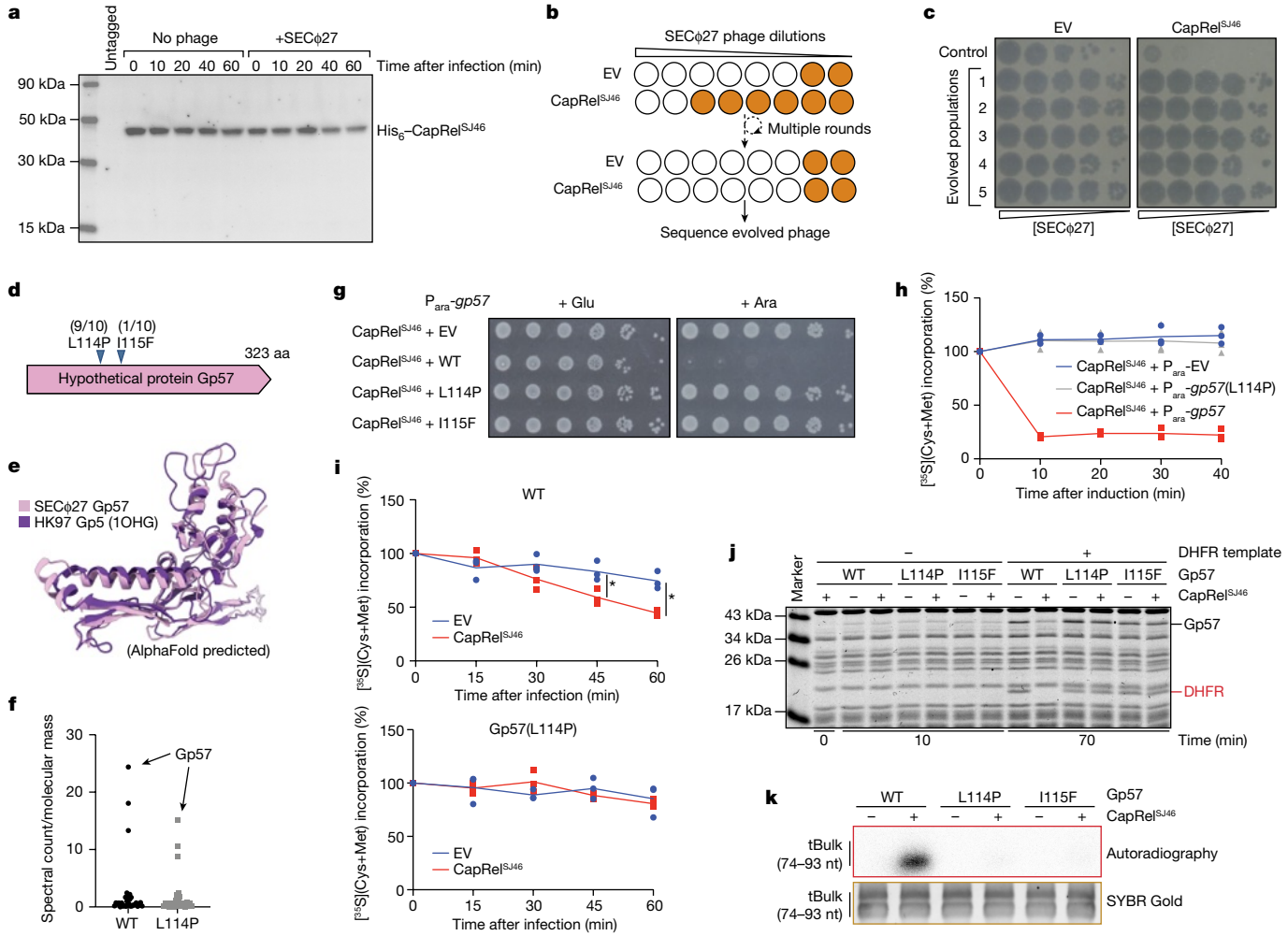


Fig. 3 | CapRel^{SJ46} is activated by the major capsid protein of SECΦ27 to pyrophosphorylate tRNAs and block translation. a, Immunoblot of His$_6$–CapRel^{SJ46} following infection with SECΦ27 compared with an uninfected control. Representative of two biological replicates. **b**, Schematic of the experimental evolution approach to identify SECΦ27 escape mutants that can infect CapRel^{SJ46}-containing cells. White wells indicate clearing by phages and brown wells indicate bacterial growth. **c**, Serial dilutions of five independently evolved populations of SECΦ27 and a control population spotted on cells harbouring an empty vector or a CapRel^{SJ46} expression vector. **d**, Summary of identified escape mutants, all of which map to a hypothetical protein encoded by gene 57 of SECΦ27. The fraction of each escape mutant is indicated. **e**, AlphaFold-predicted structure of Gp57 compared with the major capsid protein Gp5 from phage HK97. **f**, Mass spectrometry analysis of SECΦ27 phage lysates (wild type or a mutant producing Gp57(L114P)). Spectrum count normalized to molecular mass is shown for individual phage proteins. **g**, Serial dilutions of cells expressing CapRel^{SJ46} and the indicated variant of Gp57 from an arabinose-inducible promoter on media containing glucose or arabinose. **h**, Cells expressing CapRel^{SJ46} and Gp57 (wild type (WT) or L114P) from an arabinose-inducible promoter or harbouring empty vector were pulse-labelled with ^{35}S-cysteine and ^{35}S-methionine after addition of arabinose. **i**, As in **h**, but for cells harbouring a CapRel^{SJ46} expression vector or empty vector and after infection with SECΦ27 (top) or the escape mutant expressing Gp57(L114P) (bottom). Asterisks indicate $P = 0.022$ (45 min) or $P = 0.004$ (60 min) (unpaired two-tailed t-test). **j**, In vitro transcription–translation assays using DHFR production from a DNA template as the readout of expression activity. Purified CapRel^{SJ46} was added along with a template for producing Gp57. Representative of two biological replicates. **k**, Autoradiography of reactions in which purified CapRel^{SJ46} was incubated with [γ-^{32}P]ATP, bulk *E. coli* tRNAs and Gp57. SYBR Gold staining of bulk tRNAs serves as a loading control. Representative of two biological replicates.

spectrometry when cells were infected with wild-type phage, with a significant reduction in the mutant phage (Extended Data Fig. 5a,b). In addition, we co-produced CapRel^{SJ46}–Flag and Gp57–HA and found that wild-type, but not the L114P or I115F variant of the capsid protein, co-precipitated with CapRel^{SJ46}–Flag (Fig. 4a and Extended Data Fig. 5c). Finally, we purified both full-length CapRel^{SJ46} and Gp57, and used isothermal titration calorimetry (ITC) to show that they interact directly with an affinity of 350 nM in a 1:1 ratio (Fig. 4b). We confirmed that CapRel^{SJ46} and Gp57 form a complex with 1:1 stoichiometry using size-exclusion chromatography coupled to multi-angle light scattering (SEC–MALS) (Extended Data Fig. 5d). Introducing the escape substitutions L114P and I115F into Gp57 decreased affinity more than 60-fold (Extended Data Fig. 5e).

Consistent with this tight binding interaction, the ab initio AlphaFold prediction of the CapRel^{SJ46}–Gp57 complex has a large contact interface of around 1,800 Å^2 (Fig. 4c). Although further structural studies are needed to fully validate this complex structure, the AlphaFold prediction has CapRel^{SJ46} adopting the same open state seen in our crystal structure (Fig. 2c), with the pseudo-ZFD making extensive contacts with the β-sheet and spine α-helix of the peripheral (P)-domain of Gp57 (Fig. 4c and Extended Data Fig. 5f). Notably, this region of Gp57 contains the residues L114 and I115 identified in our escape mutants. The complex predicted further interactions of pseudo-ZFD β6–β7 loop with the β6–α5 and β8–β9 loops of the axial (A)-domain of Gp57. In this arrangement Gp57 prevents the recoil of pseudo-ZFD to block

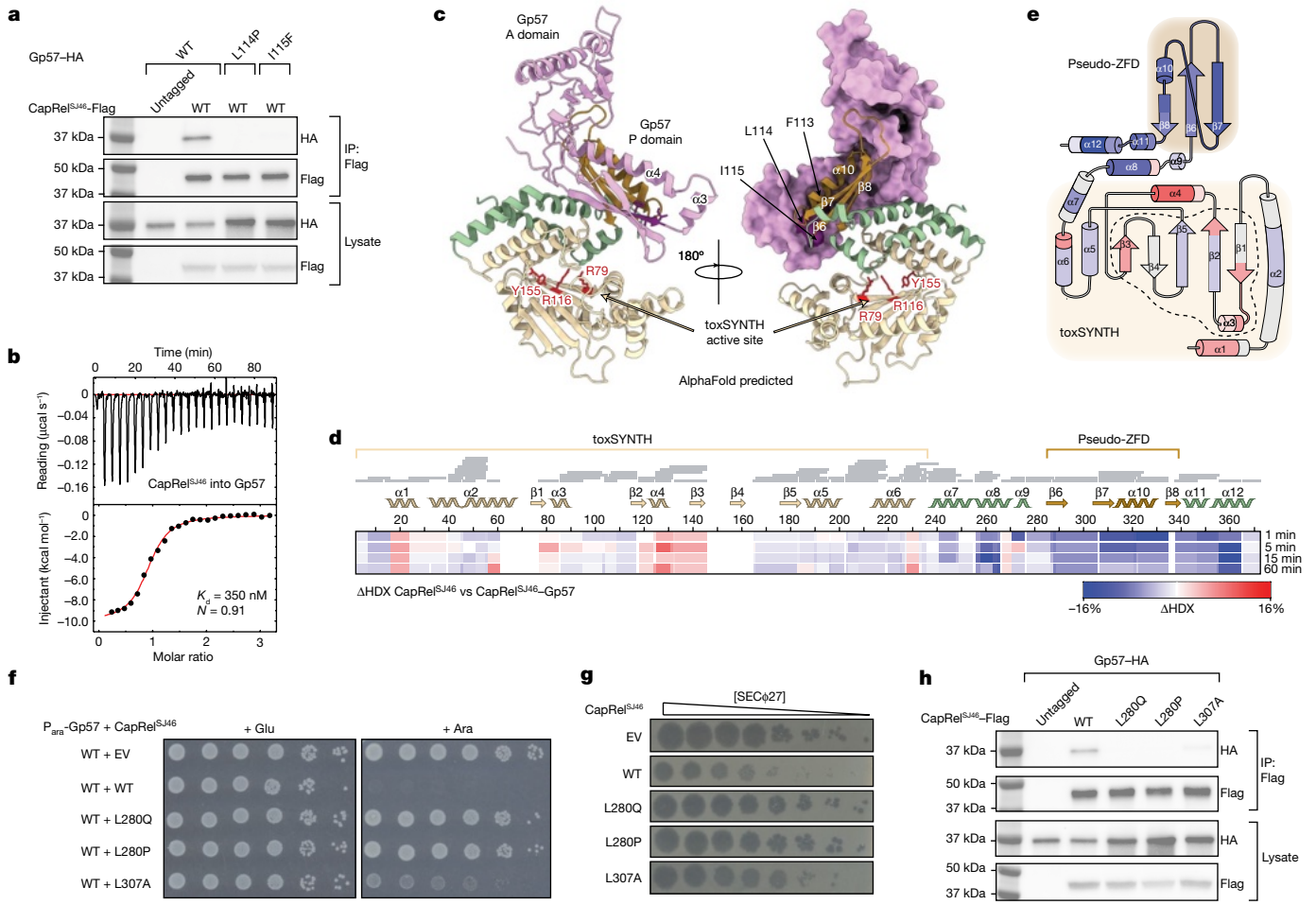


Fig. 4 | The SECΦ27 major capsid protein Gp57 binds directly to the pseudo-ZFD of CapRel^{SJ46}. **a**, CapRel^{SJ46}–Flag was immunoprecipitated from cells expressing CapRel^{SJ46}–Flag and Gp57–HA (wild-type or indicated mutant) and probed for the presence of the indicated Gp57 variant via the HA tag. Lysates used as input for the immunoprecipitation (IP) were probed as controls for expression levels. Images are representatives of three biological replicates. **b**, Binding of CapRel^{SJ46} to Gp57 monitored by ITC. K_d, binding affinity; N, stoichiometry. **c**, Structural model of the CapRel^{SJ46}–Gp57 complex predicted by AlphaFold. According to the model, the P-domain of Gp57 (pink) recognizes the pseudo-ZFD (orange) and anchor regions (green) of CapRel^{SJ46}. This interaction prevents the recoil of pseudo-ZFD to the active site and activates the enzyme. **d**, Differential HDX (ΔHDX) between CapRel^{SJ46} and CapRel^{SJ46}–Gp57 displayed as a difference heat map. Red indicates increased deuteration of CapRel^{SJ46} in the presence of Gp57; blue indicates lower deuteration. Grey bars indicate peptides identified in mass spectrometry analysis. **e**, Topological representation of CapRel^{SJ46}, coloured according to the ΔHDX. The active site of the enzyme is marked by a black dashed outline and the catalytic toxSYNTH domain and the phage-recognition pseudo-ZFD are shadowed in light yellow and light orange, respectively. **f**, Serial dilutions on media containing glucose or arabinose of cells expressing the indicated mutant of CapRel^{SJ46} from its native promoter and the wild-type Gp57 from an arabinose-inducible promoter. **g**, Serial dilutions of SECΦ27 phage spotted on cells expressing the indicated mutant of CapRel^{SJ46} or an empty vector. **h**, As in **a**, but with the indicated mutants of CapRel^{SJ46}–Flag. Images shown are representatives of three independent biological replicates.

the active site of the enzyme while stabilizing the YXXY motif in the non-neutralizing hinge conformation.

Results from hydrogen–deuterium exchange (HDX) monitored by mass spectrometry strongly supported the AlphaFold predictions. In the presence of Gp57, the pseudo-ZFD of CapRel^{SJ46} became more protected with the strongest protection mapping to α10, β8, and the C-terminal α-helical extension (Fig. 4d,e and Extended Data Fig. 5g,h). This overlaps the same region critical for phage specificity (Fig. 2b). The HDX data also confirmed the interface formed between Gp57 P-domain β5 and CapRel^{SJ46} pseudo-ZFD as well as the Gp57 A-domain β8–β9 loop and CapRel^{SJ46} β6–β7 loop. Finally, we observed increased deuterium uptake in CapRel^{SJ46} in residues 110–124 of β2 and 125–130 of α4, which are part of the adenine coordination pocket of toxSYNTH, thus confirming that interaction with Gp57 exposes the active site of the enzyme (Fig. 4d,e and Extended Data Fig. 5g,h).

To further validate the role of the pseudo-ZFD in binding and activating CapRel^{SJ46}, we performed error-prone PCR-based mutagenesis on this domain and screened for mutations that disrupted activation of CapRel^{SJ46} when it was co-produced with the capsid protein Gp57. The substitutions L280Q and L280P drastically reduced the toxicity of CapRel^{SJ46} in the presence of wild-type Gp57 (Fig. 4f and Extended Data Fig. 5i), and prevented CapRel^{SJ46} from protecting against SECΦ27 infection (Fig. 4g and Extended Data Fig. 5j). Importantly, these CapRel^{SJ46} variants still protected *E. coli* against phage T2 and T4, indicating that these variants retained structural integrity (Extended Data Fig. 5k). The substitution L307A had similar, but less pronounced, effects on CapRel^{SJ46} activity (Fig. 4f,g).

The crystal structure of CapRel^{SJ46} suggested that L280 and L307 in the wild-type protein promote the open, active state, with L280 stabilizing one of the hinge regions involving the pseudo-ZFD and L307 structuring the β6–β7 loop that interacts with Gp57 A-domain. The L280Q and L280P variants of CapRel^{SJ46} were unable to co-precipitate the major capsid protein of SECΦ27, and the L307A substitution significantly reduced binding in this assay (Fig. 4h). Using ITC, we also

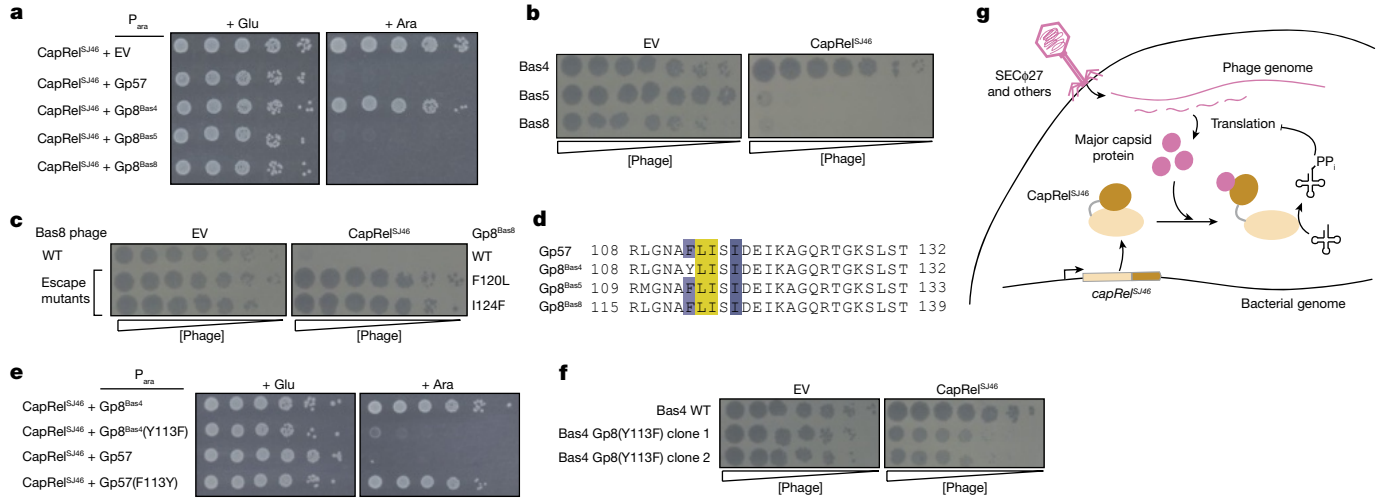


Fig. 5 | Evidence for the coevolution of CapRel^{SJ46} and the major capsid protein of SECΦ27 and related phages. a, Serial dilutions on media containing glucose or arabinose of cells expressing CapRel^{SJ46} from its native promoter and the major capsid protein homologue from the phage indicated from an arabinose-inducible promoter. **b**, Serial dilutions of the phages indicated spotted on lawns of cells expressing CapRel^{SJ46} or an empty vector. **c**, Serial dilutions of wild-type Bas8 phage or the escape mutants bearing the major capsid mutations indicated spotted on lawns of cells harbouring CapRel^{SJ46} or an empty vector. **d**, Alignment of the region of the major capsid protein in SECΦ27, Bas5 and Bas8 that triggers CapRel^{SJ46}, along with Bas4, which has a tyrosine at position 113 instead of phenylalanine. **e**, Serial dilutions on media containing glucose or arabinose of cells expressing CapRel^{SJ46} from its native promoter and the Bas4 or SECΦ27 major capsid protein variant indicated from an arabinose-inducible promoter. **f**, Serial dilutions of wild-type Bas4 or two mutant clones containing Y113F in the major capsid protein Gp8 spotted on lawns of cells expressing CapRel^{SJ46} or an empty vector. **g**, Model for the direct activation of CapRel^{SJ46} by the major capsid protein of SECΦ27 and related phages. After genome injection, the production of the major capsid protein triggers relief of autoinhibition by the C-terminal antitoxin of CapRel^{SJ46}, leading to pyrophosphorylation (PP_i) of tRNAs by activated CapRel^{SJ46}, which inhibits translation and restricts viral infection.

showed that the L280P variant of CapRel^{SJ46} could no longer bind to wild-type Gp57 (Extended Data Fig. 5e). In sum, our findings strongly support a model in which the C-terminal pseudo-ZFD of CapRel^{SJ46} directly recognizes the major capsid protein of SECΦ27, thereby triggering a relief of autoinhibition of the N-terminal toxSYNTH domain.

Other phage capsids activate CapRel^{SJ46}

The pseudo-ZFD of CapRel^{SJ46}, including residues L280 and L307, is the least well conserved region of the protein (Extended Data Fig. 3a,i). This variability may reflect a 'Red Queen dynamic', a hallmark of many host–pathogen interfaces that arises from cycles of selective pressure on pathogens to evade host immunity followed by selection on host immune factors to restore recognition of a pathogen[5]. As triggers of the CapRel defence system, phage capsid proteins are likely to be under pressure to diversify while retaining the ability to form a capsid, leading to a selective pressure on the pseudo-ZFD of CapRel to diversify and retain its interaction with the capsid proteins. To test this hypothesis, we examined three phages from the BASEL collection[37] (Bas4, Bas5 and Bas8) that are closely related to SECΦ27 and contain Gp8, a close homologue of Gp57 (Extended Data Fig. 6a). We first found that co-expressing the major capsid homologues from Bas5 and Bas8—but not that of Bas4—with CapRel^{SJ46} rendered CapRel^{SJ46} toxic, as with the SECΦ27 capsid protein (Fig. 5a and Extended Data Fig. 6b). We then tested whether CapRel^{SJ46} protects against these phages and found that it protected strongly against Bas5 and Bas8, but not Bas4 (Fig. 5b and Extended Data Fig. 6c).

To validate that defence against Bas8 requires activation of CapRel^{SJ46} by the capsid protein homologue of this phage, we isolated spontaneous mutants of Bas8 that escaped defence. Two mutant clones of Bas8 were no longer defended against by CapRel^{SJ46} and contained either an F120L or I124F substitution in the major capsid homologue Gp8^{Bas8} (Fig. 5c and Extended Data Fig. 6d). Each substitution significantly reduced the ability of the capsid protein to activate CapRel^{SJ46} when co-produced (Extended Data Fig. 6e,f). As these Gp8^{Bas8} variants were not toxic when expressed on their own (Extended Data Fig. 6g), the residual toxicity seen during co-production with CapRel^{SJ46} probably arises from their accumulation over extended periods of time, despite their reduced binding affinity for CapRel^{SJ46}. By contrast, each escape mutant almost completely evaded CapRel^{SJ46} defence (Fig. 5c), given the short timescale of phage infection. Notably, the two positions identified in Bas8 escape mutants were close to the positions of the escape mutants identified in SECΦ27 Gp57, further confirming that this region in the major capsid protein is important for activating CapRel^{SJ46} (Fig. 5d). In addition, we identified another CapRel homologue from *E. coli* strain HT2012018, called CapRel^{EcHT}, that defends against Bas8 (Extended Data Fig. 6h). Three different substitutions were identified in the same region of the major capsid protein (G111S, L116F and I124N) that enabled Bas8 to escape defence (Extended Data Fig. 6h), suggesting that CapRel^{EcHT}–like CapRel^{SJ46}—is activated by the major capsid protein.

Unlike Bas8, Bas4 was not defended against by CapRel^{SJ46}, and its capsid homologue did not activate CapRel^{SJ46} despite being 98% identical to SECΦ27 Gp57, with just 5 amino acid differences between the two. However, one difference is at position 113, near the region that is likely to bind to CapRel^{SJ46}. This residue is a phenylalanine in the capsid proteins of SECΦ27, Bas5 and Bas8, but a tyrosine in the Bas4 capsid homologue (Fig. 5d). We tested whether this residue is critical for activation by making a Y113F substitution in the Bas4 capsid homologue and found that it gained the ability to activate CapRel^{SJ46} when co-produced (Fig. 5e and Extended Data Fig. 6i). By contrast, a F113Y substitution in the SECΦ27 capsid protein abolished its ability to activate CapRel^{SJ46}. Additionally, we mutated Bas4 phage such that it produces major capsid protein with the Y113F substitution. This mutant phage could still produce mature virions, but was defended against by CapRel^{SJ46} (Fig. 5f and Extended Data Fig. 6j). These results support the notion of a Red Queen dynamic between the pseudo-ZFD of CapRel and the phage capsid proteins that directly bind and activate CapRel.

Conclusions

We propose the following model for CapRel^{SJ46} activation by SECΦ27 (Fig. 5g). Without phage infection, CapRel^{SJ46} adopts an inactive, closed conformation in cells with its C-terminal antitoxin domain autoinhibiting the N-terminal toxin domain. Upon infection, the phage's major capsid protein is produced and directly binds to CapRel^{SJ46} to stabilize the active, open state. This open state enables CapRel^{SJ46} to pyrophosphorylate tRNAs and inhibit translation, which is likely to prevent mature virion production, leading to an abortive infection that prevents propagation of phage through a population of cells. Notably, our results imply that type II toxin–antitoxin systems, which feature protein antitoxins, can be activated without proteolysis of the antitoxin, which is often asserted as their primary means of activation. Although our work focused on a fused type II toxin–antitoxin system, activation by the direct binding of phage proteins could also be a common mechanism for canonical, non-fused type II toxin–antitoxin systems. As noted above, although antitoxins are frequently less stable than their cognate toxins, their turnover may not occur quickly enough to respond to phage infection[22], whereas direct binding to phage-encoded triggers could allow rapid activation.

Major capsid proteins, such as Gp57 from SECΦ27, may be a common trigger for toxin–antitoxin systems and other anti-phage defence systems. Prior studies found that a short peptide called Gol within the major capsid protein Gp23 of T4 can activate the Lit protease in *E. coli*, which cleaves EF-Tu, if both Gol and Lit protease are overproduced[38,39]. For PifA, which allows the F plasmid to exclude T7, escape mutants mapped to the major capsid protein, but this interaction has not been studied biochemically[40]. Recent work reported that mutations in the major capsid protein of T5 allow it to overcome Pycsar-mediated defence, but the capsid protein alone is insufficient to activate Pycsar[41]. Finally, mutations in the gene encoding a major capsid protein of a *Pseudomonas aeruginosa* phage enables escape from CBASS-mediated defence, but whether the capsid protein is an activator of CBASS is not yet known[42]. Nevertheless, we anticipate that major capsid proteins may be common direct triggers for a diverse range of anti-phage defence systems. Other structural components of phages may also serve as triggers; for example, antiviral STAND defence systems were recently found to be directly triggered by terminase or portal proteins[43]. Although structural proteins may be effective triggers, the intense coevolution of phages and bacteria may also drive defence systems to rely on yet other triggers[44].

As with PAMPs in eukaryotes, relying on essential, abundant components of phages for activation may help ensure that an immune response is mounted only following an infection. Notably, the capsid proteins of some eukaryotic viruses stimulate mammalian innate immune pathways. For instance, HIV capsid protein is directly detected in the host cell cytoplasm and nucleus by TRIM5 and NONO, respectively, to trigger innate immune activation[7,9]. Thus, our results suggest that similar principles of pathogen detection underlie the function and molecular basis of innate immunity in both bacteria and eukaryotes.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-022-05444-z.

1. Bernheim, A. & Sorek, R. The pan-immune system of bacteria: antiviral defence as a community resource. *Nat. Rev. Microbiol.* **18**, 113–119 (2020).
2. Hampton, H. G., Watson, B. N. J. & Fineran, P. C. The arms race between bacteria and their phage foes. *Nature* **577**, 327–336 (2020).
3. Rostøl, J. T. & Marraffini, L. (Ph)ighting phages: how bacteria resist their parasites. *Cell Host Microbe* **25**, 184–194 (2019).
4. Fitzgerald, K. A. & Kagan, J. C. Toll-like receptors and the control of immunity. *Cell* **180**, 1044–1066 (2020).
5. Daugherty, M. D. & Malik, H. S. Rules of engagement: molecular insights from host–virus arms races. *Annu. Rev. Genet.* **46**, 677–700 (2012).
6. Fletcher, A. J. et al. Trivalent RING assembly on retroviral capsids activates TRIM5 ubiquitination and innate immune signaling. *Cell Host Microbe* **24**, 761–775.e6 (2018).
7. Lahaye, X. et al. NONO detects the nuclear HIV capsid to promote cGAS-mediated innate immune activation. *Cell* **175**, 488–501.e22 (2018).
8. Lin, Y.-T., Chen, Y.-P., Fang, C.-H., Huang, P.-Y. & Liang, S.-M. Capsid proteins of foot-and-mouth disease virus interact with TLR2 and CD14 to induce cytokine production. *Immunol. Lett.* **223**, 10–16 (2020).
9. Pertel, T. et al. TRIM5 is an innate immune sensor for the retrovirus capsid lattice. *Nature* **472**, 361–365 (2011).
10. Shepardson, K. M. et al. Induction of antiviral immune response through recognition of the repeating subunit pattern of viral capsids is toll-like receptor 2 Dependent. *mBio* **8**, e01356–17 (2017).
11. Rousset, F. et al. Phages and their satellites encode hotspots of antiviral systems. *Cell Host Microbe* **30**, 740–753.e5 (2022).
12. Doron, S. et al. Systematic discovery of antiphage defense systems in the microbial pangenome. *Science* **359**, eaar4120 (2018).
13. Millman, A. et al. An expanding arsenal of immune systems that protect bacteria from phages. Preprint at *bioRxiv* https://doi.org/10.1101/2022.05.11.491447 (2022).
14. 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).
15. Gao, L. et al. Diverse enzymatic activities mediate antiviral immunity in prokaryotes. *Science* **369**, 1077–1084 (2020).
16. Lopatina, A., Tal, N. & Sorek, R. Abortive infection: bacterial suicide as an antiviral immune strategy. *Annu. Rev. Virol.* **7**, 371–384 (2020).
17. LeRoux, M. & Laub, M. T. Toxin–antitoxin systems as phage defense elements. *Annu. Rev. Microbiol.* **76**, 21–43 (2022).
18. Song, S. & Wood, T. K. A primary physiological role of toxin/antitoxin systems is phage inhibition. *Front. Microbiol.* **11**, 1895 (2020).
19. Guegler, C. K. & Laub, M. T. Shutoff of host transcription triggers a toxin–antitoxin system to cleave phage RNA and abort infection. *Mol. Cell* **81**, 2361–2373.e9 (2021).
20. Fineran, P. C. et al. The phage abortive infection system, ToxIN, functions as a protein–RNA toxin–antitoxin pair. *Proc. Natl Acad. Sci. USA* **106**, 894–899 (2009).
21. Short, F. L., Akusobi, C., Broadhurst, W. R. & Salmond, G. P. C. The bacterial type III toxin–antitoxin system, ToxIN, is a dynamic protein–RNA complex with stability-dependent antiviral abortive infection activity. *Sci. Rep.* **8**, 1013 (2018).
22. LeRoux, M., Culviner, P. H., Liu, Y. J., Littlehale, M. L. & Laub, M. T. Stress can induce transcription of toxin–antitoxin systems without activating toxin. *Mol. Cell* **79**, 280–292.e8 (2020).
23. Bobonis, J. et al. Bacterial retrons encode phage-defending tripartite toxin–antitoxin systems. *Nature* **609**, 144–150 (2022).
24. Jimmy, S. et al. A widespread toxin–antitoxin system exploiting growth control via alarmone signaling. *Proc. Natl Acad. Sci. USA* **117**, 10500–10510 (2020).
25. Anderson, B. W., Fung, D. K. & Wang, J. D. Regulatory themes and variations by the stress-signaling nucleotide alarmones (p)ppGpp in bacteria. *Annu. Rev. Genet.* **55**, 115–133 (2021).
26. Bange, G., Brodersen, D. E., Liuzzi, A. & Steinchen, W. Two P or not two P: understanding regulation by the bacterial second messengers (p)ppGpp. *Annu. Rev. Microbiol.* **75**, 383–406 (2021).
27. Ahmad, S. et al. An interbacterial toxin inhibits target cell growth by synthesizing (p) ppApp. *Nature* **575**, 674–678 (2019).
28. Kurata, T. et al. RelA–SpoT homolog toxins pyrophosphorylate the CCA end of tRNA to inhibit protein synthesis. *Mol. Cell* **81**, 3160–3170 (2021).
29. Kurata, T. et al. A hyperpromiscuous antitoxin protein domain for the neutralization of diverse toxin domains. *Proc. Natl Acad. Sci. USA* **119**, e2102212119 (2022).
30. Dedrick, R. M. et al. Prophage-mediated defence against viral attack and viral counter-defence. *Nat. Microbiol.* **2**, 16251 (2017).
31. Owen, S. V. et al. Prophages encode phage-defense systems with cognate self-immunity. *Cell Host Microbe* **29**, 1620–1633.e8 (2021).
32. Steinchen, W. et al. Structural and mechanistic divergence of the small (p)ppGpp synthetases RelP and RelQ. *Sci. Rep.* **8**, 2195 (2018).
33. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
34. Van Melderen, L., Bernard, P. & Couturier, M. Lon-dependent proteolysis of CcdA is the key control for activation of CcdB in plasmid-free segregant bacteria. *Mol. Microbiol.* **11**, 1151–1157 (1994).
35. Koga, M., Otsuka, Y., Lemire, S. & Yonesaki, T. *Escherichia coli* rnlA and rnlB compose a novel toxin–antitoxin system. *Genetics* **187**, 123–130 (2011).
36. Duda, R. L. & Teschke, C. M. The amazing HK97 fold: versatile results of modest differences. *Curr. Opin. Virol.* **36**, 9–16 (2019).
37. Maffei, E. et al. Systematic exploration of *Escherichia coli* phage–host interactions with the BASEL phage collection. *PLoS Biol.* **19**, e3001424 (2021).
38. Kao, C., Gumbs, E. & Snyder, L. Cloning and characterization of the *Escherichia coli lit* gene, which blocks bacteriophage T4 late gene expression. *J. Bacteriol.* **169**, 1232–1238 (1987).
39. Bergsland, K. J., Kao, C., Yu, Y. T., Gulati, R. & Snyder, L. A site in the T4 bacteriophage major head protein gene that can promote the inhibition of all translation in *Escherichia coli*. *J. Mol. Biol.* **213**, 477–494 (1990).
40. Molineux, I. J., Schmitt, C. K. & Condreay, J. P. Mutants of bacteriophage T7 that escape F restriction. *J. Mol. Biol.* **207**, 563–574 (1989).

41. Tal, N. et al. Cyclic CMP and cyclic UMP mediate bacterial immunity against phages. *Cell* **184**, 5728–5739.e16 (2021).
42. Huiting, E. et al. Bacteriophages antagonize cGAS-like immunity in bacteria. Preprint at *bioRxiv* https://doi.org/10.1101/2022.03.30.486325 (2022).
43. Gao, L. A. et al. Prokaryotic innate immunity through pattern recognition of conserved viral proteins. *Science* **377**, eabm4096 (2022).
44. Stokar-Avihail, A. et al. Discovery of phage determinants that confer sensitivity to bacterial immune systems. Preprint at *bioRxiv* https://doi.org/10.1101/2022.08.27.505566 (2022).

Methods

Strains and growth conditions

All bacterial and phage strains used in this study are listed in Supplementary Table 2. *Escherichia coli* strains were routinely grown at 37 °C in Luria broth (LB) medium for cloning and maintenance. Phages were propagated by infecting a culture of *E. coli* MG1655 at an A_{600} of 0.1–0.2 with a MOI of 0.1. Cleared cultures were pelleted by centrifugation to remove residual bacteria and filtered through a 0.2-μm filter. Chloroform was then added to phage lysates to prevent bacterial growth. All phage infection experiments in liquid media and phage spotting experiments were performed in LB medium at 25 °C, except for spotting of T2 and T4 on strains producing CapRel^{SJ46} open reading frames were (6.4 g l^{-1} Na_2HPO_4-$7H_2O$, 1.5 g l^{-1} KH_2PO_4, 0.25 g l^{-1} NaCl, 0.5 g l^{-1} NH_4Cl medium supplemented with 0.1% casamino acids, 0.4% glycerol, 0.4% glucose, 2 mM $MgSO_4$, and 0.1 mM $CaCl_2$) at 30 °C. For liquid induction experiments from pBAD33 vectors, bacterial cells were grown in M9 medium. Antibiotics were used at the following concentrations (liquid; plates): carbenicillin (50 μg ml^{-1}; 100 μg ml^{-1}) and chloramphenicol (20 μg ml^{-1}; 30 μg ml^{-1}).

Plasmid construction

All plasmids are listed in Supplementary Table 3. All primers and synthesized gene sequences are listed in Supplementary Table 4.

pBR322-capRel constructs: DNA encoding CapRel^{SJ46}, CapRel^{Ebc}, CapRel^{Kp} or CapRel^{EcHT} open reading frames were expressed in *E. coli* and 100–200 bp of the upstream region from the source organism was added in each case for native expression (TZ-1 to TZ-5, TZ-92 and TZ-93). DNA was commercially synthesized by Integrated DNA Technology as gBlocks and assembled into a promoter-less backbone of pBR322 amplified with TZ-6 and TZ-7 by Gibson assembly. Mutations that produce the single amino acid substitutions CapRel^{SJ46}(Y155A), CapRel^{Ebc}(Y153A), CapRel^{SJ46}(L280Q), CapRel^{SJ46}(L280P) and CapRel^{SJ46}(L307A) were generated by site-directed mutagenesis using primers TZ-8 to TZ-11 and TZ-49 to TZ-54. To add an N-terminal His_6-tag or a C-terminal Flag tag to CapRel^{SJ46}, primers TZ-41 and TZ-42 or TZ-45 and TZ-46 were used to PCR-amplify pBR322-capRel^{SJ46} followed by Gibson assembly. pBR322-capRel-chimera was constructed by inserting *capRel^{Ebc}(270–339)* that had been PCR-amplified with TZ-22 and TZ-23 into pBR322-capRel^{SJ46} linearized with TZ-20 and TZ-21 using Gibson assembly. To add a C-terminal Flag tag to CapRel^{Ebc} or the CapRel chimera, primers TZ-90 and TZ-91, or TZ-45 and TZ-46 were used to PCR-amplify corresponding pBR322 vectors followed by Gibson assembly.

pBAD33-capRel^{SJ46} constructs: *capRel^{SJ46}(1–272)* or full-length *capRel^{SJ46}* was PCR-amplified with TZ-14 and TZ-15, or TZ-14 and TZ-24, respectively, and inserted into pBAD33 linearized with TZ-12 and TZ-13 using Gibson assembly. pBAD33-capRel^{SJ46} variants (A77K, R116A, V338A, L339A, A341K, A351K, Y352A or Y355A) were constructed by site-directed mutagenesis using primers TZ-25 to TZ-40. pBAD33-capRel^{SJ46} variants (R78A, K311A, R314A, E319A and K346A) were constructed by site-directed mutagenesis using primers TZ-80 to TZ-89.

pEXT20-capRel^{SJ46} construct: *capRel^{SJ46}(273–373)* was PCR-amplified with primers TZ-18 and TZ-19, and then inserted into linearized pEXT20 with TZ-16 and TZ-17 using Gibson assembly.

pBAD33-gp57 constructs: wild-type or mutant variant (L114P or I115F) *gp57* was PCR-amplified from the corresponding wild-type or escape-mutant SECΦ27 phage using primers TZ-43 and TZ-44, and inserted into linearized pBAD33 using Gibson assembly. A C-terminal HA tag was added to wild-type or mutant *gp57* using primers TZ-47 and TZ-48 to PCR-amplify the corresponding construct followed by Gibson assembly. The F113Y variant of *gp57* was generated by site-directed mutagenesis using primers TZ-63 and TZ-64.

pBAD33-gp8: the genes encoding the major capsid protein homologues Gp8^{Bas4}, Gp8^{Bas5} and Gp8^{Bas8} were PCR-amplified from the corresponding phage using primers TZ-55 to TZ-60 and inserted into linearized pBAD33 by Gibson assembly. The Y113F variant of *gp8^{Bas4}* was generated by site-directed mutagenesis using primers TZ-61 and TZ-62. The F120L and I124F variants of *gp8^{Bas8}* were cloned from the corresponding phage escape mutants using primers TZ-59 and TZ-60.

pET-gp57 constructs: a *gp57* fragment was PCR-amplified with primers TZ-65 and TZ-66 and the TZ-67 template using Gibson assembly, the resulting linear DNA fragment was inserted into linearized pET24d (without tag) using TZ-68 and TZ-69. Template TZ-67 was synthesized as a gBlock by Integrated DNA Technology. To construct L114P and I115F-substituted *gp57* variants, DNA fragments were PCR-amplified with primer pairs TZ-70 + TZ-71 and TZ-72 + TZ-73 (for L114P) or with pairs TZ-74 + TZ-71 and TZ75 + TZ-73 (for I115F) and pET-gp57 template. The resulting linear DNA fragments were assembled using Gibson assembly.

pET24d-His_{10}-SUMO-capRel^{SJ46} constructs: the *capRel^{SJ46}* open reading frame was PCR-amplified using primers TZ-76 and TZ-77 as well as pBAD-capRel^{SJ46} as template, and, using Gibson assembly, inserted into a linearized pET24d-His_{10}-SUMO plasmid using primers TZ-78 and TZ-79.

Strain construction

Plasmids described above were introduced into *E. coli* MG1655 or BW27783 by TSS transformation or electroporation.

A single copy of *capRel^{SJ46}*, *capRel^{SJ46}(Y155A)*, *His_6-capRel^{SJ46}*, *gp57* or *gp57-HA* with its native promoter was inserted onto the MG1655 chromosome at the HK022 attachment site using the CRIM system[45], using the pAH69 helper plasmid with the pAH144 vector containing the desired insert.

Bas4 mutant phage were generated using a CRISPR–Cas system for targeted mutagenesis as described[46]. In brief, sequences for RNA guides to target Cas9-mediated cleavage were designed using the toolbox in Geneious Prime 2021.2.2 and selected for targeting of *gp8^{Bas4}* but nowhere else in the Bas4 genome. The guides were inserted into the pCas9 plasmid and tested for their ability to restrict Bas4. An efficient guide was selected and the pCas9 guide plasmid was co-transformed into *E. coli* MG1655 with a high copy number repair plasmid containing *gp8^{Bas4}(Y113F)* with the guide mutated to prevent self-cutting. The wild-type Bas4 phage was plated onto a strain containing both the pCas9 guide and the repair plasmid, and single plaques were screened by Sanger Sequencing. Two clones that produce the Y113F substituted Gp8 were propagated twice on strains containing only pCas9 guide for further selection and genomes were sequence verified by Illumina sequencing as described below.

External data

Previously published structures are available from the Protein Data Bank (PDB) with IDs 5DEC, 6S2T and 2LVH. The UniRef90 database is publicly available. Reference phage genomes are publicly available: SECΦ27 (NC_047938.1), Bas04 (MZ501069.1), Bas08 (MZ501059.1).

Toxicity assays on solid media

For producing the CapRel^{SJ46} N- and C-terminal domains, single colonies of *E. coli* MG1655 containing pBAD33-capRel^{SJ46}(1–272) and pEXT20-capRel^{SJ46}(273–373) or the corresponding empty vectors were grown for 6 h at 37 °C in LB-glucose to saturation. 200 μl of each saturated culture was then pelleted by centrifugation at 4,000*g* for 10 min, washed once in 1× phosphate-buffered saline (PBS), and resuspended in 400 μl 1× PBS. Cultures were then serially diluted 10-fold in 1× PBS and spotted on M9L plates (M9 medium supplemented with 5% LB (v/v)) further supplemented with 0.4% glucose, 0.2% arabinose or 0.2% arabinose and 100 μM IPTG. Plates were then incubated at 37 °C overnight before imaging.

For producing full-length CapRel^{SJ46}, *E. coli* MG1655 containing pBAD33-capRel^{SJ46} or a mutant form of capRel^{SJ46} were grown to

saturation and processed as above. Cultures were plated onto 0.4% glucose and 0.2% arabinose and incubated at 37 °C overnight.

For co-producing CapRel^{SJ46} and the major capsid proteins from SECΦ27, Bas4, Bas5, or Bas8, *E. coli* MG1655 harbouring pBR322-capRel^{SJ46} or MG1655 containing genomic *capRel^{SJ46}* and pBAD33-capsid protein were grown to saturation and processed as above. Cultures were plated onto 0.4% glucose and 0.2% arabinose and incubated at 37 °C overnight.

For co-producing CapRel^{SJ46} and variants of the major capsid protein from Bas8, *E. coli* BW27783 harbouring pBR322-capRel^{SJ46} and pBAD33-gp8^{Bas8} (wild-type or a mutant variant) were grown to saturation and processed as above. Cultures were plated onto 0.4% glucose and 0.0002% arabinose and incubated at 37 °C overnight.

Phage spotting assays and EOP measurements

Phage stocks isolated from single plaques were propagated in *E. coli* MG1655 at 37 °C in LB. To titre phage, dilutions of stocks were mixed with *E. coli* MG1655 and melted LB + 0.5% agar and spread on LB + 1.2% agar plates and incubated at 37 °C overnight. For phage spotting assays, 40 µl of a bacterial strain of interest was mixed with 4 ml LB + 0.5% agar and spread on an LB + 1.2% agar + antibiotic plate. Phage stocks were then serially diluted in 1× FM buffer (20 mM Tris-HCl pH 7.4, 100 mM NaCl, 10 mM $MgSO_4$), and 2 µl of each dilution was spotted on the bacterial lawn. Plates were then incubated at 25 °C overnight before imaging. EOP was calculated by comparing the ability of the phage to form plaques on an experimental strain relative to the control strain. Experiments were replicated 3 times independently and representative images are shown.

For spotting phage T2 and T4 on strains producing CapRel^{SJ46} variants, 40 µl of a bacterial strain of interest was mixed with 4 ml M9 + 0.5% agar and spread on an M9 + 1.2% agar + antibiotic plate. Phage were serially diluted and spotted as described above. Plates were then incubated at 30 °C overnight before imaging.

Growth curves following phage infection in liquid culture

Single colonies of *E. coli* MG1655 pBR322 empty vector or pBR322-capRel^{SJ46} or pBR322-capRel^{SJ46}(Y155A) were grown in LB overnight. Cultures were then back-diluted to A_{600} = 0.1 in fresh LB and 100 µl of cells were added into each well of a 96-well plate. Ten microlitres of serial diluted T4 phage were added to each well at the indicated MOI and growth following phage infection was measured at 15 min intervals with orbital shaking at 25 °C on a plate reader (Biotek). Data reported are the mean and standard deviation of eight plate replicates and the growth curve experiment was replicated three times independently.

One-step growth curves

Single colonies of *E. coli* MG1655 pBR322 empty vector or pBR322-capRel^{SJ46} were grown overnight in LB. Overnight cultures were back-diluted to A_{600} = 0.05 in 25 ml fresh LB and grown to A_{600} ~0.3 at 25 °C. Ten millilitres of each culture was infected with T4 phage at an MOI of 0.05 or SECΦ27 phage at an MOI of 0.01 in LB at 25 °C and phages were allowed to adsorb for 10 min before serial dilution in LB three times (1:100, 1:10, 1:10 serial dilution) to three flasks. Then, at indicated time points, 100 µl of infected cells from the corresponding dilution flask were mixed with 100 µl of indicator cells MG1655 pBR322 empty vector (A_{600} ~0.3), and the mixtures were mixed with 4 ml of LB + 0.5% agar and spread on LB + 1.2% agar plates. Plates were incubated overnight at 25 °C and plaques were enumerated the following day. PFU was calculated based on the dilution flask samples were taken from. Data reported are the mean and individual data points from 3 biological replicates.

Adsorption assays

Single colonies of *E. coli* MG1655 pBR322-capRel^{SJ46} or pBR322 empty vector were grown overnight in LB. Overnight cultures were back-diluted to A_{600} = 0.05 in 25 ml of fresh LB and grown to A_{600} = 0.2 at 25 °C. Cells or a control containing only media were infected with phage T4 or SECΦ27 at MOI = 0.1 and incubated at 25 °C for 10 min to allow for adsorption. Cells were pelleted at 9,000*g* for 2 min and unadsorbed phages from the supernatant were serial diluted, mixed with indicator cells MG1655 pBR322 empty vector + top agar and plated. Plates were incubated overnight at 25 °C and plaques were enumerated the following day. Percent adsorption was calculated by normalizing unadsorbed phages from each sample to media control. Data reported are the mean and individual data points from 3 biological replicates.

Western blot of CapRel^{SJ46} and Gp57 after phage infection

For immunoblotting of CapRel^{SJ46}, single colonies of *E. coli* MG1655 pBR322-His$_6$-capRel^{SJ46} or *E. coli* MG1655 genomic *His$_6$-capRel^{SJ46}* under its native promoter were grown overnight in LB. Overnight cultures were back-diluted to A_{600} = 0.05 in 25 ml fresh LB and grown to A_{600} = 0.2 at 25 °C. Cells were infected with phage SECΦ27 at MOI = 100, and incubated at 25 °C during the experiment. At each indicated time point (0, 10, 20, 40, 60 min), A_{600} was measured and 1 ml of cells was pelleted at 21,000*g* for 2 min at 4 °C. Supernatant was removed and pellets were flash-frozen in liquid nitrogen. Pellets were thawed and resuspended in 1× Laemmli sample buffer (Bio-Rad) supplemented with 2-mercaptoethanol with A_{600} normalized. Samples were then boiled at 95 °C and analysed by 12% SDS–PAGE and transferred to a 0.45 µm PVDF membrane. Anti-His$_6$ antibody (Invitrogen) was used at a final concentration of 1:1000, and SuperSignal West Femto Maximum Sensitivity Substrate (Thermo Fisher) was used to develop the blots. Blots were imaged by a ChemiDoc Imaging system (Bio-Rad). Blots were stained with Coomassie stain and imaged as loading control. Image shown is a representative of two independent biological replicates.

For immunoblotting of Gp57, single colonies of *E. coli* MG1655 containing genomic *gp57-HA* under its native promoter were grown overnight in LB. Overnight cultures were back-diluted to A_{600} = 0.05 in 25 ml of fresh LB and grown to A_{600} = 0.2 at 25 °C. Cells were infected with phage SECΦ27 at MOI = 100, and incubated at 25 °C during the experiment. At each indicated time point (0, 10, 20, 30, 40, 50 min), A_{600} was measured and 1 ml of cells was pelleted at 21,000*g* for 2 min at 4 °C. Samples were processed as described above with A_{600} normalized. Samples were then boiled at 95 °C and analysed by 12% SDS–PAGE and transferred to a 0.45 µm PVDF membrane. Anti-HA antibody (Cell Signaling Technology) was used at a final concentration of 1:1000, and SuperSignal West Femto Maximum Sensitivity Substrate (Thermo Fisher) was used to develop the blots. Blots were imaged by a ChemiDoc Imaging system (Bio-Rad), then stained with Coomassie stain and imaged as loading control. The experiment was performed three times independently and band intensities were quantified using Fiji. Relative band intensities were calculated by normalizing the summed intensity of both Gp57 bands to the intensity of total proteins by Coomassie stain.

Western blot of CapRel^{SJ46}, CapRel^{EB} or chimera expression levels

Single colonies of *E. coli* MG1655 pBR322-capRel^{SJ46}, pBR322-capRel^{SJ46}-Flag, pBR322-capRel^{Ebc}-Flag or pBR322-chimera-Flag were grown overnight in LB. Overnight cultures were back-diluted to A_{600} = 0.05 in 5 ml fresh LB and grown to A_{600} = 0.2 at 37 °C. A_{600} was measured and 5 ml of cells was pelleted at 4,000*g* for 5 min. Supernatant was removed and pellets were resuspended in 1× Laemmli sample buffer (Bio-Rad) supplemented with 2-mercaptoethanol with A_{600} normalized. Samples were then boiled at 95 °C and analysed by 12% SDS–PAGE and transferred to a 0.45 µm PVDF membrane. Anti-Flag antibody (Cell Signaling Technology) and anti-GyrA antibody (Inspiralis) were used at a final concentration of 1:1,000, and SuperSignal West Femto Maximum Sensitivity Substrate (Thermo Fisher) was used to develop the blots. Blots were imaged by a ChemiDoc Imaging system (Bio-Rad). Image shown is a representative of 2 independent biological replicates.

Error-prone PCR mutagenesis of CapRel^{SJ46}

The C terminus of CapRel^{SJ46} was mutagenized using error-prone PCR-based mutagenesis as described previously[47]. In brief, primers TZ-94 and TZ-95 were used to amplify the C terminus of CapRel^{SJ46} using Taq polymerase (NEB) and 0.5 mM $MnCl_2$ was added to the reaction as the mutagenic agent. PCR products were treated with Dpn I, column purified, and inserted into pBR322-CapRel^{SJ46} backbone amplified with primer TZ-96 and TZ-97 using Gibson assembly. Gibson products were transformed into DH5α and grown overnight in LB at 37 °C. Overnight cultures were miniprepped to obtain the mutagenized library. Individual colonies were Sanger sequenced to assess the number of mutations. To perform the selection, mutagenized library was electroporated into *E. coli* MG1655 pBAD33-gp57, and plated onto LB plates containing 0.2% arabinose to select for survivors. Colonies were picked and sequenced to identify mutations in CapRel^{SJ46}, and further validated by constructing plasmid with only single mutants.

Isolation of phage escape mutants to infect CapRel^{SJ46}

The phage evolution experiment was conducted as described previously[48]. In brief, 5 independent populations were evolved in a 96-well plate containing a sensitive host *E. coli* MG1655 pBR322 empty vector and a resistant host *E. coli* MG1655 pBR322-capRel^{SJ46}. One control population was evolved with only the sensitive host. Overnight bacterial cultures were back-diluted to A_{600} = 0.1 in LB and 100 µl were seeded into each well. Cells were infected with tenfold serial dilutions of SECΦ27 phage with MOI from 10^2 to 10^{-4}, with one well uninfected to monitor for contamination. Plates were sealed with breathable plate seals and incubated at 25 °C for 6 h in a plate shaker at 1,000 rpm. Cleared wells from each population were pooled, pelleted at 4,000*g* for 20 min to remove bacteria, and the supernatant lysates were transferred to a 96 deep-well block with 40 µl chloroform added to prevent bacterial growth. Lysates were spotted onto both sensitive and resistant hosts to check the defence phenotype. Thirteen rounds of evolution were performed to allow all five populations to overcome CapRel^{SJ46} defence. Evolved clones from each evolved population were isolated by plating to single plaques on lawns of resistant host, and control clones from the control population were isolated on a lawn of the sensitive host. Two clones from each population were propagated using the corresponding host and sequenced as described below.

Bas8 escape mutants were isolated by plating a population of phage onto CapRel^{SJ46}- or CapRel^{EcHT}-containing cells. Twenty microlitres of 10^{11} PFU ml^{-1} Bas8 phage mixed with 40 µl overnight culture of *E. coli* MG1655 pBR322-capRel^{SJ46} or pBR322-capRel^{EcHT} were added to 4 ml LB + 0.5% agar and spread onto LB + 1.2% agar. Plates were incubated at 25 °C overnight. Single plaques were isolated and propagated using the same strain in LB at 25 °C. Amplified phage lysates were pelleted to remove bacteria, and then plated to single plaques and propagated similarly for a second round of isolation to improve purity and sequenced.

Phage DNA extraction and Illumina sequencing

To extract phage DNA, high-titre phage lysates (>10^6 PFU µl^{-1}) were treated with DNase I (0.001 U µl^{-1}) and RNase A (0.05 mg ml^{-1}) at 37 °C for 30 min. 10 mM EDTA was used to inactivate the nucleases. Lysates were then incubated with proteinase K at 50 °C for 30 min to disrupt capsids and release phage DNA. Phage DNA was isolated by ethanol precipitation. In brief, sodium acetate pH 5.2 was added to 300 mM followed by 100% ethanol to a final volume fraction of 70%. Samples were incubated at −80 °C overnight, pelleted at 21,000*g* for 20 min and supernatant removed. Pellets were washed with 100 µl isopropanol and 200 µl 70% (v/v) ethanol, and then aired dried at room temperature and resuspended in 25 µl 1× TE buffer (10 mM Tris-HCl, 0.1 mM EDTA, pH 8). Concentrations of extracted DNA were measured by NanoDrop (Thermo Fisher Scientific).

To prepare Illumina sequencing libraries, 100–200 ng of genomic DNA was sheared in a Diagenode Bioruptor 300 sonicator water bath for twenty 30 s cycles at maximum intensity. Sheared genomic DNA was purified using Ampure XP beads, followed by end repair, 3′ adenylation, and adapter ligation. Barcodes were added to both 5′ and 3′ ends by PCR with primers that anneal to the Illumina adapters. The libraries were cleaned by Ampure XP beads using a double cut to elute fragment sizes matching the read lengths of the sequencing run. Libraries were sequenced on an Illumina MiSeq at the MIT BioMicro Center. Illumina reads were assembled to the reference genomes using Geneious Prime 2021.2.2.

Mass spectrometry of phages

Wild-type or mutant (L114P in Gp57, evolved clone 1 from population 3) SECΦ27 phage were propagated in *E. coli* MG1655 for high-titre stocks. In brief, *E. coli* MG1655 (A_{600} = 0.2) in LB were infected with phages at MOI = 0.1 and incubated at 37 °C for 4 h. Cells were pelleted at 4,000*g* for 10 min and supernatant lysates were filtered through 0.2-µm filters. Five hundred microlitres of phage stocks (10^{10} PFU µl^{-1}) were further concentrated with Amicon Ultra filter (MW 100 kDa) and washed twice with 1× FM buffer (20 mM Tris-HCl pH 7.4, 100 mM NaCl, 10 mM $MgSO_4$). Concentrated phage lysates were boiled to denature virions and run on 4–20% SDS–PAGE. Each lane from the gel was excised. Proteins were reduced with 10 mM dithiothreitol (Sigma) for 1 h at 56 °C and then alkylated with 20 mM iodoacetamide (Sigma) for 1 h at 25 °C in the dark. Proteins were then digested with 12.5 ng µl^{-1} modified trypsin (Promega) in 50 µl 100 mM ammonium bicarbonate, pH 8.9 at 25 °C overnight. Peptides were extracted by incubating the gel pieces with 50% acetonitrile/5% formic acid then 100 mM ammonium bicarbonate, repeated twice followed by incubating the gel pieces with 100% acetonitrile then 100 mM ammonium bicarbonate, repeated twice. Each fraction was collected, combined, and reduced to near dryness in a vacuum centrifuge. Peptides were desalted using Pierce Peptide Desalting Spin Columns (Thermo) and then lyophilized. The tryptic peptides were separated by reverse phase HPLC (Thermo Ultimate 3000) using a Thermo PepMap RSLC C18 column over a 90 min gradient before nano-electrospray using an Exploris mass spectrometer (Thermo). Solvent A was 0.1% formic acid in water and solvent B was 0.1% formic acid in acetonitrile. Detected peptides were mapped to SECΦ27 protein sequences and the abundance of proteins were estimated by number of spectrum counts/molecular mass to normalize for protein sizes.

Co-immunoprecipitation analysis

For immunoprecipitation of CapRel^{SJ46} after phage infection, *E. coli* MG1655 containing pBR322-capRel^{SJ46}-Flag were grown overnight in LB. Overnight cultures were back-diluted to A_{600} = 0.05 in 175 ml of LB and grown to A_{600} ~0.3 at 25 °C. Cells were infected with wild-type or mutant (L114P in Gp57, evolved clone 1 from population 3) SECΦ27 at MOI = 100 and incubated at 25 °C. At the indicated time points (15 min or 40 min), A_{600} was measured and 50 ml of cells were pelleted at 6,000*g* for 5 min at 4 °C. Uninfected cells were collected at 0 min before phage infection. Supernatant was removed and cells were resuspended in 900 µl lysis buffer (25 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100 and 5% glycerol) supplemented with protease inhibitor (Roche), 1 µl per ml Ready-Lyse Lysozyme Solution (Lucigen) and 1 µl per ml benzonase nuclease (Sigma). Samples were lysed by two freeze-thaw cycles, and lysates were normalized by A_{600}. Lysates were pelleted at 21,000*g* for 10 min at 4 °C, and 850 µl of supernatant were incubated with pre-washed anti-Flag M2 magnetic beads (Sigma) for 1 h at 4 °C with end-over-end rotation. Beads were then washed 3 times with lysis buffer containing 350 mM NaCl but free of detergent. On-bead reduction, alkylation and digestion were performed. Proteins were reduced with 10 mM dithiothreitol (Sigma) for 1 h at 56 °C and then alkylated with 20 mM iodoacetamide (Sigma) for 1 h at 25 °C in the dark. Proteins were then digested with modified trypsin (Promega) at an enzyme/substrate ratio of 1:50 in 100 mM ammonium bicarbonate, pH 8 at

25 °C overnight. Trypsin activity was halted by addition of formic acid (99.9 %, Sigma) to a final concentration of 5 %. Peptides were desalted using Pierce Peptide Desalting Spin Columns (Thermo) then lyophilized. The tryptic peptides were subjected to LC–MS/MS as described above. Experiments were performed two times independently and spectral counts are reported. Ratio of spectral counts between Gp57 and CapRel^{SJ46} were calculated and plotted for normalization.

For co-producing CapRel^{SJ46} and Gp57, *E. coli* MG1655 containing pBR322-capRel^{SJ46} or pBR322-capRel^{SJ46}-Flag (wild type or mutants) and pBAD33-gp57-HA (wild type or mutants) were grown overnight in M9-glucose. Overnight cultures were back-diluted to A_{600} = 0.05 in 50 ml of M9 (no glucose) and grown to A_{600} ~0.3 at 37 °C. Cells were induced with 0.2% arabinose for 30 min at 37 °C, then A_{600} was measured and cells were pelleted at 4,000 *g* for 10 min at 4 °C. Supernatant was removed and cells were resuspended in 900 µl lysis buffer as described above. Samples were lysed by two freeze-thaw cycles, and lysates were normalized by A_{600}. Lysates were pelleted at 21,000*g* for 10 min at 4 °C, and 850 µl of supernatant were incubated with pre-washed anti-Flag M2 magnetic beads (Sigma) for 1 h at 4 °C with end-over-end rotation. Beads were then washed 3 times with lysis buffer containing 350 mM NaCl. Laemmli sample buffer (Bio-Rad) supplemented with 2-mercaptoethanol was added to beads directly to elute proteins. Samples were boiled at 95 °C and analysed by 12% SDS–PAGE and transferred to a 0.45 µm PVDF membrane. Anti-Flag and anti-HA antibodies (Cell Signaling Technology) were used at a final concentration of 1:1,000, and SuperSignal West Femto Maximum Sensitivity Substrate (Thermo Fisher) was used to develop the blots. Blots were imaged by a ChemiDoc Imaging system (Bio-Rad). Images shown are representatives of 3 independent biological replicates.

Incorporation assays

For co-producing CapRel^{SJ46} and Gp57, the SECΦ27 major capsid protein, single colonies of *E. coli* MG1655 containing pBR322-capRel^{SJ46} and pBAD33-gp57 (wild-type or L114P variant) or corresponding empty vectors were grown overnight in M9-glucose. Overnight cultures were back-diluted to A_{600} = 0.05 in 25 ml M9-glucose and grown to A_{600} ~0.3 at 37 °C. Cells were pelleted at 4,000*g* for 5 min at 4 °C and washed once with M9 (no glucose), and then back-diluted to A_{600} = 0.1 in 15 ml M9 (no glucose) and recovered for 45 min at 37 °C. At the beginning of the experiment, cells were induced with 0.2% arabinose. At the indicated time points (0, 10, 20, 30, 40 min), A_{600} was measured and an aliquot of 250 µl of cells was transferred to microcentrifuge tube containing [5,6-^{3}H]uridine (PerkinElmer) (4 µCi ml^{-1}) for transcription measurements or EasyTag EXPRESS-^{35}S Protein Labeling Mix, [^{35}S] (PerkinElmer) at 44 µCi ml^{-1} for translation measurements. Tubes were incubated at 37 °C for 2 min, then quenched by addition of nonradioactive uridine (1.5 mM) or cysteine and methionine (15 mM each) and incubated for an additional 2 min. Samples were then added to ice cold trichloroacetic acid (TCA) (10% w/v) and incubated at least 30 min on ice to allow for precipitation. Resulting samples were vacuum filtered onto a glass microfibre filter (Whatman, 1820-024) that had been pre-wetted with 5% w/v TCA. Filters were washed with 35× volume of 5% w/v TCA, then with 5× volume of 100% ethanol. Air dried filters were placed in tubes with scintillation fluid and measured in a scintillation counter (PerkinElmer). CPM (Counts Per Million) was normalized to A_{600} and percent incorporation at each time point was calculated by normalizing to t = 0. Data reported are the mean and individual data points from three independent biological replicates.

For producing the CapRel^{SJ46} N-terminal toxin domain, single colonies of *E. coli* MG1655 containing pBAD33-capRel^{SJ46}(1–272) or an empty vector were grown overnight in M9-glucose. Transcription and translation experiments were done as described above. Data reported are the mean and individual data points from three independent biological replicates.

For phage infection experiments, single colonies of *E. coli* MG1655 harbouring pBR322 empty vector or pBR322-capRel^{SJ46} were grown overnight in LB. Overnight cultures were back-diluted to A_{600} = 0.05 in 25 ml fresh LB and grown to A_{600} ~0.3 at 25 °C. Cells were then diluted to A_{600} = 0.1 in 10 ml LB and infected with wild-type or mutant (L114P in Gp57, evolved clone 1 from population 3) SECΦ27 at MOI = 100 and incubated at 25 °C. At the indicated time points (0, 15, 30, 45 and 60 min), A_{600} was measured and an aliquot of 250 µl of cells was transferred to a microcentrifuge tube containing [5,6-^{3}H]uridine (PerkinElmer) (32 µCi ml^{-1}) for transcription measurements or EasyTag EXPRESS-^{35}S Protein Labeling Mix, [^{35}S] (PerkinElmer) at 88 µCi ml^{-1} for translation measurements. Tubes were incubated at 25 °C for 4 min, then quenched by addition of nonradioactive uridine (1.5 mM) or cysteine and methionine (15 mM) and incubated for an additional 2 min. Samples were then processed same as above. Data reported are the mean and individual data points from three independent biological replicates. Statistical significance was determined by unpaired, two-tailed Student's t-test ($P < 0.05$).

Homology search, alignment, and conservation analysis

CapRel^{SJ46} was identified in the sequence database from our previous bioinformatic survey of RSH proteins[24] that included gene neighbourhood analysis to identify toxin–antitoxin systems[49]. Bacterial strains containing CapRel^{SJ46}, CapRel^{Ebc} or CapRel^{Kp} with 100% amino acid identity were found on NCBI database. Local genomic regions (±10 kb of CapRel) were extracted and annotated for all coding sequences. Prophage genes and intact prophage regions were identified by PHASTER[50]. Additional homologues of CapRel^{SJ46} were identified by ConSurf[51] using PSI-BLAST (default settings) to search UniRef90 database, yielding 44 homologues. For Extended Data Fig. 1, sequences were aligned with MAFFT L-INS-i v7.453 (ref. [52]) with manual curation of the C-terminal region guided by homology modelling of the stand-alone Phrann Gp30 antitoxin using Swiss-Model[53], and with our CapRel^{SJ46} predicted structure as a template. For ConSurf analysis, 52 homologues were used to generate the multiple sequence alignment by MAFFT and used as input. Conservation scores were calculated using the Bayesian method and default settings. For gene neighbourhood analysis, previously identified CapRel and PhRel sequences[24] were reduced to representatives with 65% amino acid identity using MMSeqs v13.45111 easy-cluster (default settings[54]), while retaining proteins of interest. The associated accession numbers were then used as input to FlaGs v1.2.6 (one flanking gene either side of the query and otherwise default settings)[49].

Homologues of the major capsid proteins in BASEL phages were identified by BLASTp[55] searches against each phage genome. Homologues of Gp57 (Gp8^{Bas4}, Gp8^{Bas5} and Gp8^{Bas8}) were aligned by MUSCLE[56].

CapRel^{SJ46} preparation for crystallization and HDX-MS

For the production of His$_{10}$–SUMO-tagged CapRel^{SJ46} and CapRel^{SJ46} variants, *E. coli* BL21 (DE3) cells were transformed with pET24d plasmids containing the gene of interest and grown in LB medium to A_{600} of 0.6. Expression of the protein of interest was induced by addition of 0.5 mM IPTG, and cells were grown for 3 h at 30 °C. The culture was then centrifuged, and pellet was resuspended in resuspension buffer (50 mM Tris-HCl pH 8.0, 1.5 M KCl, 2 mM $MgCl_2$, 1 mM TCEP, 0,002% mellitic acid and 1 pastil of protease inhibitors cocktail (Roche)).

Cells were disrupted using a high-pressure homogenizer (Emulsiflex) and the supernatant was separated from the pellet by centrifugation and filtered through 0.45 µm filters. Protein extracts were loaded onto a gravity-flow column (Cytiva) packed with HisPur Nickel resin (Thermo Fisher Scientific), washed with buffer A (50 mM Tris-HCl pH 8, 500 mM NaCl, 500 mM KCl, 1 mM TCEP, 0.002% Melitic acid) and stepwise eluted in buffer A supplemented with 500 mM imidazole. To remove remaining contaminants and imidazole, the elution fraction was immediately transferred to a size-exclusion chromatography (SEC) column Superdex 200 pgcolumn (Cytiva), previously equilibrated in the SEC buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 500 mM KCl, 2 mM $MgCl_2$, 1 mM TCEP, 0.002% mellitic acid). The fractions containing the protein were concentrated to around 1 mg ml^{-1} and the His tag was removed

by incubating with UlpI protease (1:50 molar ratio) at 4 °C for 30 min. The His_{10}–SUMO tag and the protease were then removed by passing the samples over a gravity-flow column (Cytiva) packed with HisPur Nickel resin (Thermo Fisher Scientific). Purity of the sample preparation was assessed spectrophotometrically and by SDS–PAGE. For all the purified protein samples, A_{260}/A_{280} ratio was below 0.6. Samples were stored at −20 °C or concentrated to 7 mg ml^{-1} and used directly in crystallization experiments.

For the purification of the complex containing His_{10}–SUMO–$CapRel^{SJ46}$ and His_{10}–SUMO–Gp57, *E. coli* BL21 (DE3) strain containing freshly transformed pET24d-His_{10}-SUMO-$capRel^{SJ46}$(Y155A) and pET21a-His_{10}-SUMO-gp57 were grown in LB medium to A_{600} of 0.2. This culture was then diluted in fresh LB media and grown until A_{600} of 0.6. Expression of the protein of interest was induced by addition of 0.5 mM IPTG, and cells were grown for overnight at 16 °C. The subsequent purification, Sumo tag cleavage and purity assessment steps were identical to the workflow described above for the all the $CapRel^{SJ46}$ protein variants.

Crystallization of $CapRel^{SJ46}$

The screening of crystallization conditions of $CapRel^{SJ46}$ was carried out using the sitting-drop vapour-diffusion method. The drops were set up in Swiss (MRC) 96-well two-drop UVP sitting-drop plates using the Mosquito HTS system (TTP Labtech). Drops of 0.1 µl protein and 0.1 µl precipitant solution were equilibrated to 80 µl precipitant solution in the reservoir. Commercially available screens LMB and SG1 (Molecular Dimensions) were used to test crystallization conditions. The condition resulting in protein crystals (LMB screen position C9 for $CapRel^{SJ46}$: 26% w/v PEG 2000 MME 0.1 M Bis-Tris 5.8) were repeated as 2 µl drops. The final crystallization drop was made by mixing in a 1:1 ratio the crystallization mother liquor at pH 5.8 and $CapRel^{SJ46}$ at pH 8.0. Crystals were collected using suitable cryo-protecting solutions (mother liquor supplemented with 20% glycerol) and vitrified in liquid N_2 for transport and storage before X-ray exposure. X-ray diffraction data was collected at the SOLEIL synchrotron (Gif-sur-Yvette, Paris, France) on the Proxima 1 (PX1) and Proxima 2A (PX2A) beamlines using an Eiger-X 16M detector. Because of the high anisotropic nature of the data from all the crystals we performed anisotropic cut-off and correction of the merged intensity data as implemented on the STARANISO server (http://staraniso.globalphasing.org/) using the DEBYE and STARANISO programs. The analysis of the data suggested a resolution of 2.31 Å (with 2.31 Å in a^*, 2.85 Å in b^* and 2.72 Å in c^*).

Structure determination

The data were processed with the XDS suite[57] and scaled with Aimless. In all cases, the unit cell content was estimated with the program MATTHEW COEF from the CCP4 program suite[58]. Molecular replacement was performed with Phaser[59]. The crystals of $CapRel^{SJ46}$ diffracted on average to ~2.3 Å. We used the coordinates of Rel_{Tt}^{NTD} (PDB ID 6S2T) as a search model for the toxSYNTH domain[60]. The MR solution from Phaser was used in combination with Rosetta as implemented in the MR-Rosetta[61] suit from the Phenix package[62]. After several iterations of manual building with Coot[63] and maximum likelihood refinement as implemented in Buster/TNT[64], the model was extended to cover all the residues (R/R_{free} of 21.5/26.0). Extended Data Table 1 details all the X-ray data collection and refinement statistics.

Isothermal titration calorimetry

For all ITC measurements $CapRel^{SJ46}$ samples were prepared from the pET24d-His_{10}-SUMO-$capRel^{SJ46}$ as detailed above. In the case of Gp57, *E. coli* BL21 (DE3) cells were transformed with pET21a-His_{10}-SUMO-gp57 and grown in LB medium to A_{600} of 0.2. This culture was then diluted in fresh LB media and grown until A_{600} of 0.6. Expression of His_{10}–SUMO–Gp57 was induced by addition of 0.1 mM IPTG, and cells were grown for overnight at 16 °C. The subsequent purification, SUMO-tag cleavage and purity assessment steps were identical to the workflow described above for the all the $CapRel^{SJ46}$ protein variants. After removing the SUMO tag, samples were concentrated to 10 µM and used directly for ITC immediately after purification.

All titrations were performed with an Affinity ITC (TA instruments) at 25 °C. For the titration, $CapRel^{SJ46}$ was loaded in the instrument syringe at 150 µM and Gp57 was used in the cell at 10 µM. The titration was performed in 50 mM HEPES pH 7.5; 500 mM KCl; 500 mM NaCl; 150 mM imidazole; 10 mM $MgCl_2$; 1 mM TCEP; 0.002% mellitic acid. Final concentrations were verified by the absorption using a Nanodrop One (ThermoScientific). All ITC measurements were performed by titrating 2 µl of $CapRel^{SJ46}$ (or a $CapRel^{SJ46}$ variant) into Gp57 (or a Gp57 variant) using a constant stirring rate of 75 rpm. All data were processed, buffer-corrected and analysed using the NanoAnalyse and Origin software packages.

Multiwavelength light scattering

The sample of the $CapRel^{SJ46}$-Gp57 complex used for the MALS measurements was prepared in the same way as the samples used for hydrogen–deuterium exchange mass spectrometry (HDX-MS): 50 mM HEPES pH 7.5; 500 mM KCl; 500 mM; NaCl; 10 mM $MgCl_2$; 1 mM TCEP; 0.002% mellitic acid at a concentration of 5 mg ml^{-1}. The measurement was performed in an HPLC Alliance system (Waters) connected to a 2998 PDA detector (Waters), a TREOS II MALS detector (Wyatt Technology) and a RI-501 refractive index detector (Shodex). The data were analysed with Astra 7 suite (Wyatt technology).

HDX-MS

HDX-MS experiments were performed on an HDX platform composed of a Synapt G2-Si mass spectrometer (Waters Corporation) connected to a nanoAcquity UPLC system. Samples of $CapRel^{SJ46}$ and $CapRel^{SJ46}$ complexed with Gp57 were prepared at a concentration of 20 to 50 µM. For each experiment 5 µl of sample ($CapRel^{SJ46}$ or $CapRel^{SJ46}$-Gp57) were incubated for 1 min, 5 min, 15 min or 60 min in 95 µl of labeling buffer L (50 mM HEPES, 500 mM KCl, 500 mM NaCl, 2 mM $MgCl_2$, 1 mM TCEP, 0.002% mellitic acid, pH 7.5) at 20 °C. The non-deuterated reference points were prepared by replacing buffer L by equilibration buffer E (50 mM HEPES, 500 mM KCl, 500 mM NaCl, 2 mM $MgCl_2$, 1 mM TCEP, 0.002% mellitic acid, pH 7.5). After labelling, the samples are quenched by mixing with 100 µl of pre-chilled quench buffer Q (1.2 % formic acid, pH 2.4). Seventy microlitres of the quenched samples are directly transferred to the Enzymate BEH Pepsin Column (Waters Corporation) at 200 µl min^{-1} and at 20 °C with a pressure 8.5 kPSI. Peptic peptides were trapped for 3 min on an Acquity UPLC BEH C18 VanGuard Pre-column (Waters Corporation) at a 200 µl min^{-1} flow rate in water (0.1% formic acid in HPLC water pH 2.5) before eluted to an Acquity UPLC BEH C18 Column for chromatographic separation. Separation was done with a linear gradient buffer (7–40% gradient of 0.1% formic acid in acetonitrile) at a flow rate of 40 µl min^{-1}. Peptides identification and deuteration uptake analysis was performed on the Synapt G2-Si in ESI ± $HDMS^E$ mode (Waters Corporation). Leucine enkephalin was applied for mass accuracy correction and sodium iodide was used as calibration for the mass spectrometer. $HDMS^E$ data were collected by a 20–30 V transfer collision energy ramp. The pepsin column was washed between injections using pepsin wash buffer (1.5 M guanidinium HCl, 4% (v/v) methanol, 0.8% (v/v) formic acid). A blank run was performed between each sample to prevent significant peptide carry-over. Optimized peptide identification and peptide coverage for all samples was performed from undeuterated controls (five replicates). All deuterium time points were performed in triplicate.

Data treatment and statistical analysis of HDX-MS

The non-deuterated references points were analysed by PLGS (ProteinLynx Global Server 2.5.1, Waters) to identify the peptic peptides belonging to

CapRel^{SJ46} or Gp57. Then, all the HDMS^{E} data including reference and deuterated samples were processed by DynamX 3.0 (Waters) for deuterium uptake determination. We chose the following filtering parameters: minimum intensity of 1,000, minimum and maximum peptide sequence length of 5 and 20, respectively, minimum MS/MS products of 3, minimum products per amino acid of 0.27, minimum score of 5, and a maximum MH+ error threshold of 15 p.p.m. Data were analysed at peptidic and overall level and manually curated by visual inspection of individual spectra. The overall level is based on the relative fractional uptake ($RFU_{a,t}$), which can be calculated by the following formula:

$$RFU_{a,t}(\%) = \frac{\gamma_{a,t}}{MaxUptake_a \times D}$$

where $\gamma_{a,t}$ is the deuterium uptake for peptide a at incubation time t, and $MaxUptake_a \times D$ is the theorical maximum uptake in deuterium value that peptide a can take. The ΔRFU compared RFU value between two different experimental conditions, in this case, this is the comparison between CapRel^{SJ46} and CapRel^{SJ46} + Gp57. Heat maps have been generated in DynamX. All the raw data can be accessed at: https://doi.org/10.6084/m9.figshare.19745089.

CapRel^{SJ46} expression and purification for biochemical assays

Full-length capRel^{SJ46} was overexpressed in freshly transformed *E. coli* BL21(DE3) pET24d-N-His_{10}-SUMO-capRel^{SJ46} pMG25-paSpo co-transformed with the plasmid encoding PaSpo small alarmone hydrolase (SAH) from *Salmonella* phage SSU5, which has been shown to neutralize the toxicity of other toxSAS toxins[24]. Fresh transformants were used to inoculate 800 ml of LB medium (final A_{600} of 0.03) supplemented with 50 μg ml^{-1} kanamycin, 20 μg ml^{-1} chloramphenicol and 0.2% arabinose. Bacterial cultures were grown at 37 °C until an A_{600} of 0.4–0.5 and protein expression was induced with 0.1 mM IPTG (final concentration). Cells were grown for additional 1 h at 30 °C and the biomass was collected by centrifugation (10,000 rpm, for 5 min, JLA-10.500 rotor (Beckman Coulter)).

Cell mass was resuspended in buffer A (750 mM KCl, 500 mM NaCl, 5 mM $MgCl_2$, 40 μM $MnCl_2$, 40 μM zinc acetate, 1 mM mellitic acid, 20 mM imidazole, 10% glycerol, 4 mM β-mercaptoethanol and 25 mM HEPES:KOH pH 8) supplemented with 0.1 mM PMSF and 1 U ml^{-1} of DNase I, and lysed by one passage through a high-pressure cell disrupter (Stansted Fluid Power, 150 MPa). Mellitic acid was added to buffers as it was earlier shown to stabilize *Thermus thermophilus* Rel stringent factor[65]. Cell debris was removed by centrifugation (25,000 rpm for 1 h at 4 °C, JA-25.50 rotor (Beckman Coulter)), the clarified lysate was filtered through a 0.22-μm syringe filter and loaded onto a HisTrap 5 ml HP column (Cytiva) pre-equilibrated in buffer A. The column was washed with 5 column volumes of buffer A, and the protein was eluted using a combination of stepwise and linear gradient (5 column volumes with 0–100% buffer B) of buffer B (750 mM KCl, 500 mM NaCl, 5 mM $MgCl_2$, 40 μM $MnCl_2$, 40 μM $Zn(OAc)_2$, 1 mM mellitic acid, 1 M imidazole, 10% glycerol, 4 mM β-mercaptoethanol, 25 mM HEPES:KOH pH 8). Fractions enriched in CapRel^{SJ46} (approximately 40% buffer B) were pooled, totalling approximately 5 ml. The sample was loaded on a HiLoad 16/600 Superdex 200 pg column pre-equilibrated with a high-salt buffer (buffer C; 2 M NaCl, 5 mM $MgCl_2$, 10% glycerol, 4 mM β-mercaptoethanol, 25 mM HEPES:KOH pH 8). The fractions containing CapRel^{SJ46} were pooled and applied on a HiPrep 10/26 desalting column (Cytiva) pre-equilibrated with storage buffer (buffer D; 720 mM KCl, 5 mM $MgCl_2$, 40 mM arginine, 40 mM glutamic acid, 10% glycerol, 4 mM β-mercaptoethanol, 25 mM HEPES:KOH pH 8). Fractions containing CapRel^{SJ46} were collected (about 14 ml in total) and the His_{10}–SUMO tag was cleaved off by addition of 10 μg of His_6–Ulp1 per 1 mg CapRel^{SJ46} followed by a 30-min incubation on ice. After the His_{10}–SUMO tag was cleaved off, the protein was passed through a 5 ml HisTrap HP pre-equilibrated with buffer D supplemented with 20 mM imidazole. Fractions containing CapRel^{SJ46} in the flow-through were collected and concentrated on an Amicon Ultra (Millipore) centrifugal filter device with a 10 kDa cut-off. The purity of protein preparations was assessed by SDS–PAGE. Protein preparations were aliquoted, frozen in liquid nitrogen and stored at −80 °C. Individual single-use aliquots were discarded after the experiment.

Cell-free translation

Experiments with PURExpress in vitro protein synthesis kit (NEB, E6800) were performed as per the manufacturer's instructions. All reactions were supplemented with 0.8 U μl^{-1} RNase Inhibitor Murine (NEB, M0314S). Purified CapRel^{SJ46} protein was used at a final concentration of 250 nM, with *gp57*, *gp57(L114P)* or *gp57(I115F)* as template plasmid at 10 ng μl^{-1}. As a mock control CapRel^{SJ46} was substituted for equal volume of HEPES:Polymix buffer[66], pH 7.5. After a 10-min incubation at 37 °C, a 1.34 μl aliquot of the reaction mixture was taken and quenched by addition of 13.66 μl of 2× sample buffer (100 mM Tris:HCl pH 6.8, 4% SDS, 0.02% bromophenol blue, 20% glycerol, 20 mM DTT and 4% β-mercaptoethanol), and DHFR template plasmid was added to the remaining reaction mixture at a final concentration of 20 ng μl^{-1}. After further incubation at 37 °C for 1 h, the reaction mixture was mixed with ninefold volume of 2× sample buffer, boiled at 98 °C for 5 min, and 5 μl of the mixture was resolved by 18% SDS–PAGE. The SDS–PAGE gel was fixed by incubating for 5 min at room temperature in 50% ethanol solution supplemented with 2% phosphoric acid, washed three times with water for 20 min at room temperature, and stained with 'blue silver' solution (0.12% Brilliant Blue G250 (Sigma-Aldrich, 27815), 10% ammonium sulfate, 10% phosphoric acid, and 20% methanol) overnight at room temperature. After washing with water for 3 h at room temperature, the gel was imaged on an Amersham ImageQuant 800 (Cytiva) imaging system. For tRNA pyrophosphorylation experiments, Gp57, Gp57(L114P), or Gp57(I115F) was produced in a similar reaction mixture without CapRel^{SJ46} and DHFR template at 37 °C for 2 h.

tRNA pyrophosphorylation by CapRel^{SJ46}

The reaction mixture containing 5 μM tRNA from *E. coli* MRE600 (Sigma-Aldrich, 10109541001), 500 μM [γ^{32}P]ATP, 250 nM CapRel^{SJ46} and 1/10 volume of either wild-type Gp57, Gp57(L114P), or Gp57(I115F) product from the PUREsystem in HEPES:Polymix buffer, pH 7.5 (5 mM Mg^{2+} final concentration) supplemented with 1 mM DTT was incubated at 37 °C for 10 min. To visualize phosphorylated tRNA, the reaction sample was mixed in 2 volumes of RNA dye (98% formamide, 10 mM EDTA, 0.3% bromophenol blue and 0.3% xylene cyanol), tRNA was denatured at 37 °C for 10 min and resolved on urea-PAGE in 1× TBE (8 M urea, 8% PAGE). The gel was stained with SYBR Gold (Life technologies, S11494) and exposed to an imaging plate overnight. The imaging plate was imaged by a FLA-3000 (Fujifilm).

Reporting summary

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

Data availability

Structural data from this study are available from the Protein Data Bank (PDB) under accession 7ZTB. HDX raw data can be accessed at https://doi.org/10.6084/m9.figshare.19745089. Sequencing data are available in the Sequence Read Archive (SRA) under BioProject PRJNA837951. Materials including strains and plasmids are available upon reasonable request. Source data are provided with this paper.

45. Haldimann, A. & Wanner, B. L. Conditional-replication, integration, excision, and retrieval plasmid-host systems for gene structure-function studies of bacteria. *J. Bacteriol.* **183**, 6384–6393 (2001).
46. Duong, M. M., Carmody, C. M., Ma, Q., Peters, J. E. & Nugen, S. R. Optimization of T4 phage engineering via CRISPR/Cas9. *Sci. Rep.* **10**, 18229 (2020).

47. Cadwell, R. C. & Joyce, G. F. Randomization of genes by PCR mutagenesis. *Genome Res.* **2**, 28–33 (1992).
48. Srikant, S., Guegler, C. K. & Laub, M. T. The evolution of a counter-defense mechanism in a virus constrains its host range. *eLife* **11**, e79549 (2022).
49. Saha, C. K., Sanches Pires, R., Brolin, H., Delannoy, M. & Atkinson, G. C. FlaGs and webFlaGs: discovering novel biology through the analysis of gene neighbourhood conservation. *Bioinformatics* **37**, 1312–1314 (2021).
50. Arndt, D. et al. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* **44**, W16–W21 (2016).
51. Ashkenazy, H. et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **44**, W344–W350 (2016).
52. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
53. Waterhouse, A. et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).
54. Hauser, M., Steinegger, M. & Söding, J. MMseqs software suite for fast and deep clustering and searching of large protein sequence sets. *Bioinformatics* **32**, 1323–1330 (2016).
55. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic Local Alignment Search Tool. *J. Mol. Biol.* **215**, 403–410 (1990).
56. Madeira, F. et al. Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res.* **50**, W276–W279 (2022).
57. Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).
58. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
59. McCoy, A. J. et al. *Phaser* crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
60. Tamman, H. et al. A nucleotide-switch mechanism mediates opposing catalytic activities of Rel enzymes. *Nat. Chem. Biol.* **16**, 834–840 (2020).
61. Terwilliger, T. C. et al. phenix.mr_rosetta: molecular replacement and model rebuilding with Phenix and Rosetta. *J. Struct. Funct. Genomics* **13**, 81–90 (2012).
62. Afonine, P. V. et al. Towards automated crystallographic structure refinement with *phenix.refine*. *Acta Crystallogr. D* **68**, 352–367 (2012).
63. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
64. Smart, O. S. et al. Exploiting structure similarity in refinement: automated NCS and target-structure restraints in *BUSTER*. *Acta Crystallogr. D* **68**, 368–380 (2012).
65. Van Nerom, K., Tamman, H., Takada, H., Hauryliuk, V. & Garcia-Pino, A. The Rel stringent factor from *Thermus thermophilus*: crystallization and X-ray analysis. *Acta Crystallogr. F* **75**, 561–569 (2019).
66. Takada, H. et al. The C-terminal RRM/ACT domain is crucial for fine-tuning the activation of 'long' RelA-SpoT homolog enzymes by ribosomal complexes. *Front. Microbiol.* **11**, 277 (2020).

Acknowledgements We thank A. Harms for sharing the BASEL phage collection; the MIT BioMicro Center and its staff for their support in sequencing; the MIT Biopolymers and Proteomics Core and its staff for their help in mass spectrometry experiments; W. Verseés for allowing the use of the biophysics facilities at the VUB; the Mass Spectrometry Laboratory at ULiège and Thomas Tilmant for support and assistance regarding MS data acquisition; K. Gozzi and B. Wang for comments on the manuscript; and all members of the Laub laboratory for helpful discussions. G.C.A. and V.H. were supported by the Swedish Research council (grant 2018-00956 within the RIBOTARGET consortium under the framework of JPIAMR, project grants 2017-03783 and 2021-01146 to V.H., project grant 2019-01085 to G.C.A.), the Knut and Alice Wallenberg Foundation (2020.0037 to G.C.A.), the Ragnar Söderberg Foundation (M23/14 to V.H.), the European Regional Development Fund through the Centre of Excellence for Molecular Cell Technology (V.H.), and the Estonian Science Foundation (project grant PRG335 to V.H.). A.G.-P. was supported by Fonds National de Recherche Scientifique (FRFS-WELBIO CR-2017S-03, FNRS CDR J.0068.19, FNRS-EQP UN.025.19 and FNRS-PDR T.0090.22), the European Research Council (CoG DiStRes, no. 864311), the Joint Programming Initiative on Antimicrobial Resistance (JPI-EC-AMR-R.8004.18), the Programme Actions de Recherche Concerté 2016-2021, Fonds Jean Brachet and the Fondation Van Buuren, Chargé de Recherches fellowship from the FNRS no. CR/DM-392 (H.T.). A.C. and K.C.W. are fellows of the FRIA, C.M. is supported as a Research Associate of the FRS–FNRS. C.M. was supported by grant F.4532.22 from the FRS–FNRS. The authors acknowledge the use of the PROXIMA 1 and 2A beamlines at the Soleil synchrotron (Gif-sur-Yvette, France). M.T.L. is an Investigator of the Howard Hughes Medical Institute.

Author contributions Experiments were conceived and designed by T.Z., T.K., G.C.A., V.H., A.G.-P. and M.T.L. Phage and bacterial experiments, as well as incorporation and co-immunoprecipitation assays, were done by T.Z. with assistance from M.L. and S.S. Metabolic labelling experiments were done by T.Z. and T.B. Cell-free translation and tRNA pyrophosphorylation assays were done by T.Z. and T.K. CapRel and Gp57 purification was done by T.Z., T.K., H.T., A.C. and A.T. ITC was performed by H.T. and A.C. MALS was done by A.C. and A.T. HDX-MS data acquisition and analysis was performed by C.M. and K.C.W. X-ray data collection and analyses was performed by H.T., A.T., A.C. and A.G.-P. Bioinformatic analyses were performed by T.Z. and G.C.A. Structural modelling was done by T.Z., A.T. and A.G.-P. Figure design, manuscript writing, and editing was done by T.Z., T.K., G.C.A., V.H., A.G.-P. and M.T.L. Project supervision and funding was provided by G.C.A., V.H., A.G.-P. and M.T.L.

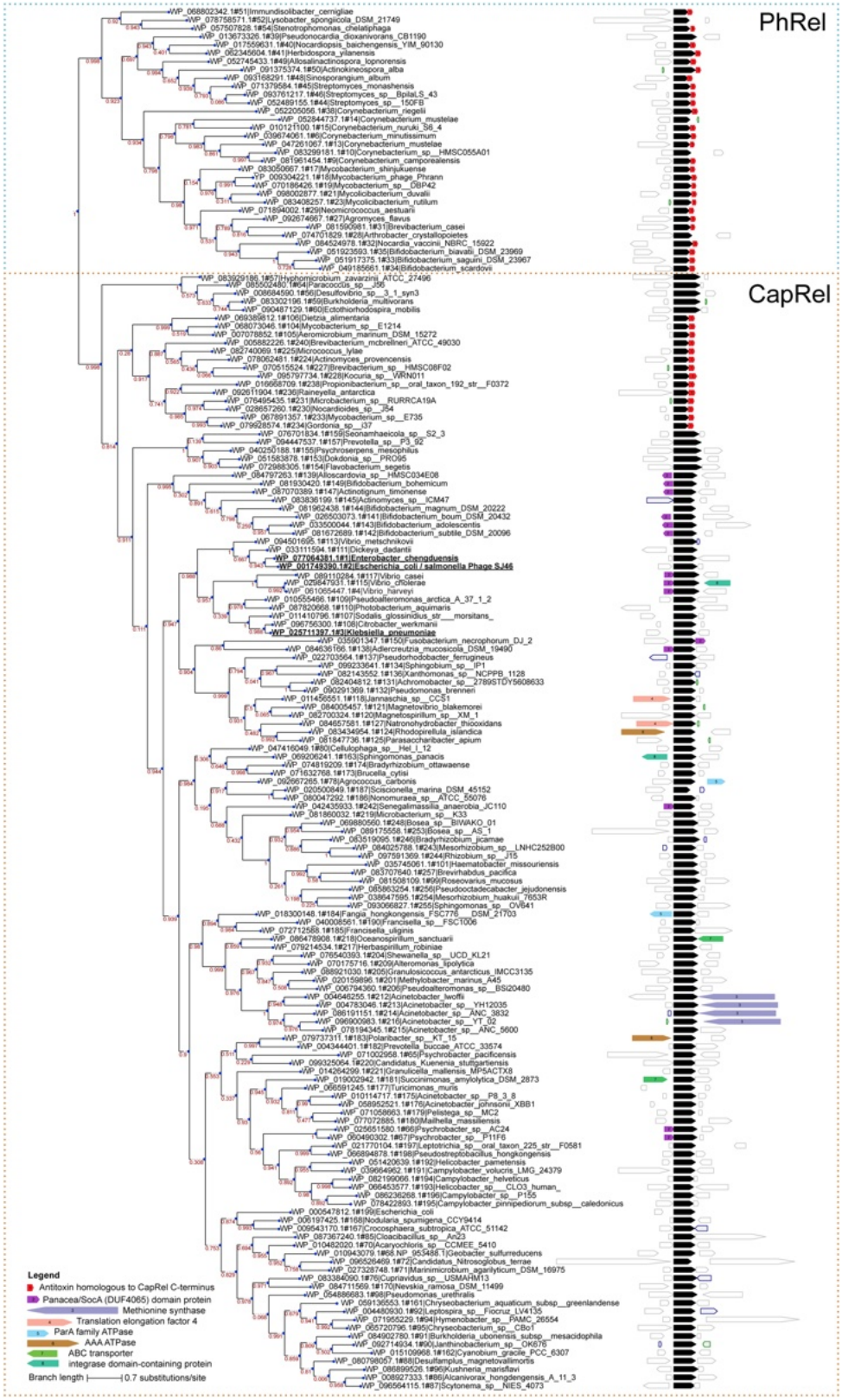
Competing interests A.G.-P. is co-founder and stockholder of Santero Therapeutics. the other authors declare no competing interests.

Additional information

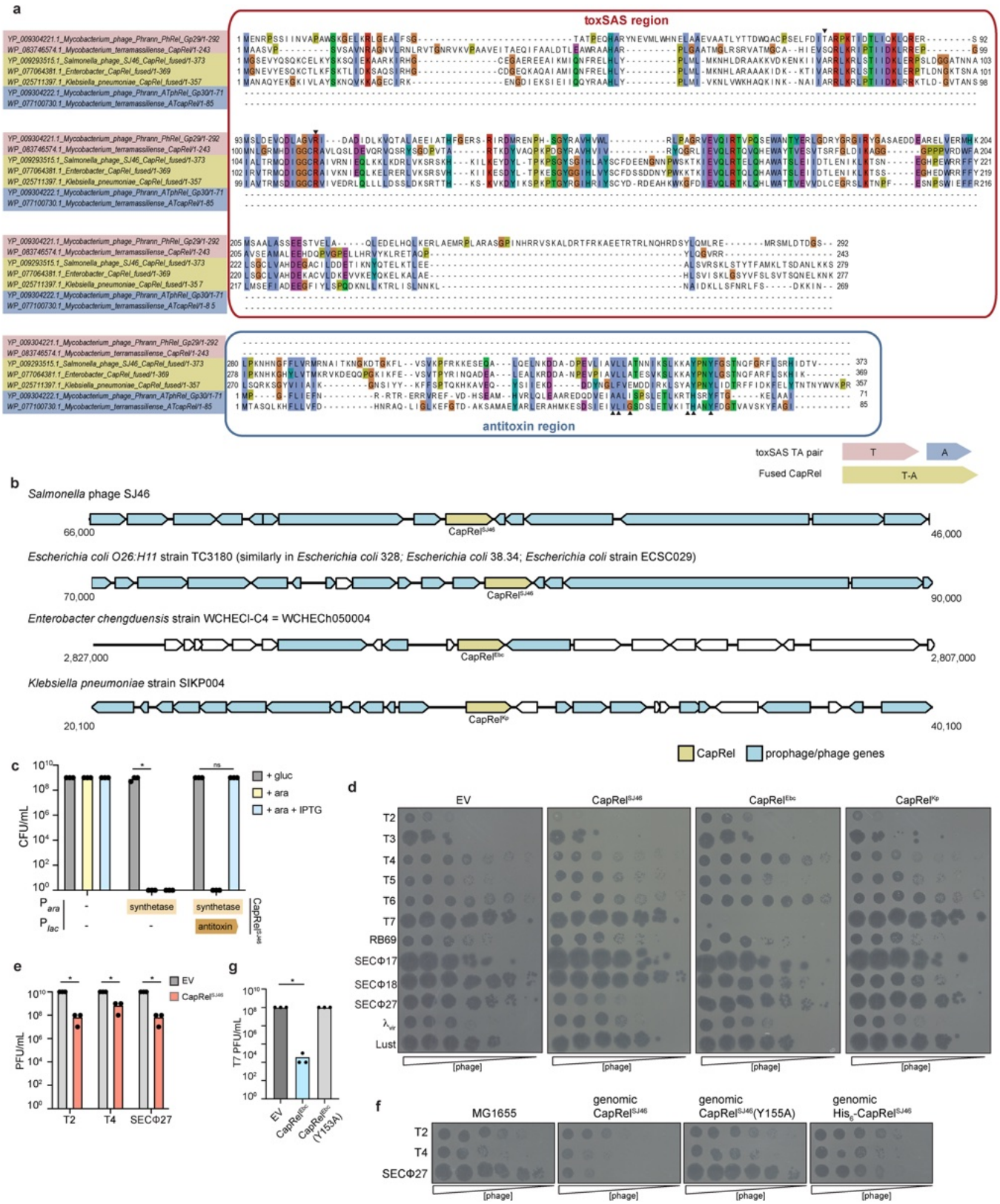
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-022-05444-z.

Correspondence and requests for materials should be addressed to Vasili Hauryliuk, Abel Garcia-Pino or Michael T. Laub.

Peer review information *Nature* thanks Philip Kranzusch and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer review reports are available.

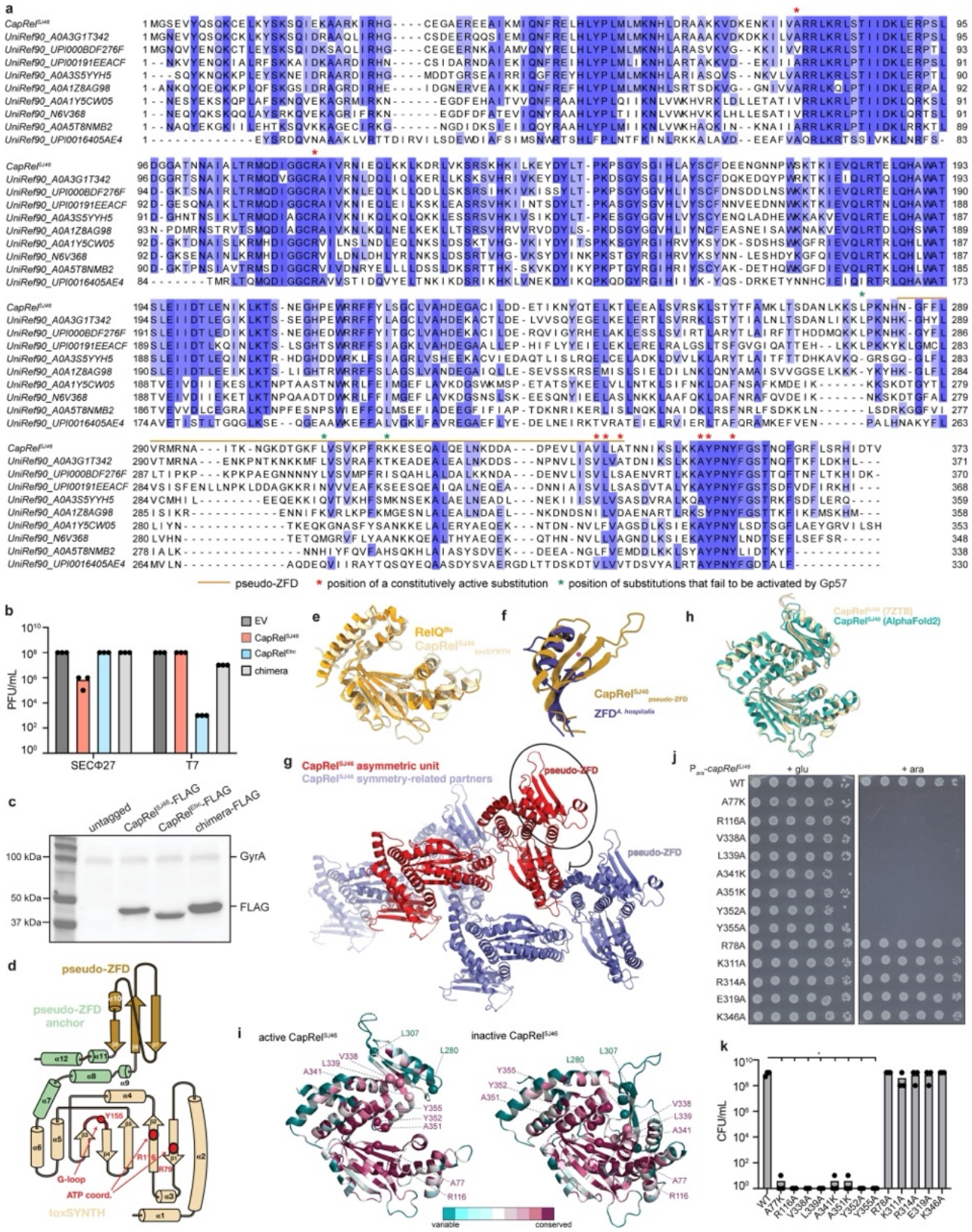


Extended Data Fig. 1 | CapRel is broadly distributed in different bacteria, and is usually a fused TA system. FlaGs output for CapRel and PhRel representatives (black arrows). Those proteins studied in this paper are in bold and underlined. Most CapRel systems are fused, with the exception of Actinobacteria where the TA pair is usually unfused. The unfused state is the most common form of the system in the closely related PhRel subfamily of toxSASs, and fusion and/or fission events appear to have occurred independently multiple times. Conserved flanking open reading frames are colored and numbered by homologous clusters: 1 (red): antitoxin homologous to the CapRel C-terminus, and 2 (purple): Panacea/SocA (DUF4065) domain-containing proteins. For other coloured open reading frames see the legend on the lower left. Arrows with no fill colour and with blue and green outlines are, respectively, pseudogenes and non-coding RNA genes. Red numbers on branches show Maximum Likelihood bootstrap support on a scale of 0-1, where 1 is 100% support.



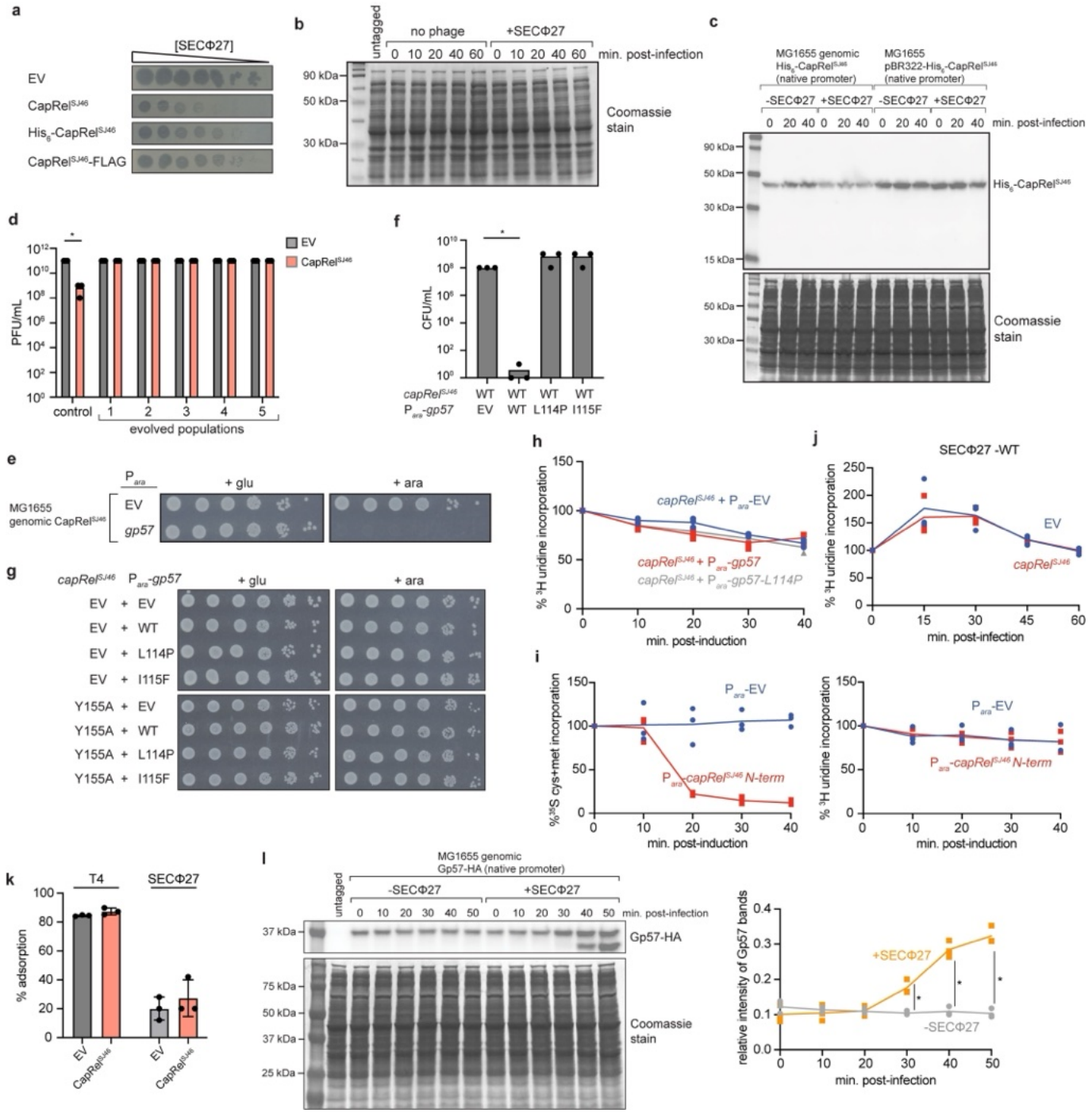
Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Analysis of CapRel homologs. (**a**) Sequence alignment comparing fused CapRel systems with related, unfused systems. Alignment of toxSAS PhRel and ATphRel from the *Mycobacterium* phage Phrann, non-fused CapRel and ATcapRel from *Mycobacterium terramassilience*, and the three fused systems CapRel^{SJ46}, CapRel^{Ebc} and CapRel^{Kp}. The N-terminal region of fused CapRel systems is a toxSAS toxin domain, while the C-terminal region is homologous to the antitoxins of the PhRel and unfused CapRel TA systems. Substituted sites that rendered CapRel^{SJ46} toxic (see Extended Data Fig. 3j) are indicated with black arrowheads. The inset diagram summarises the homologous regions of the bicistronic toxin-antitoxin and fused toxin-antitoxin systems considered here. (**b**) Genome maps of native locations of CapRel^{SJ46}, CapRel^{Ebc} and CapRel^{Kp} (+/− 10 kb) with predicted flanking prophage and phage genes. (**c**) Summary of 3 independent replicates of cell viability assay in Fig. 1b. Asterisks indicate p = 0.007 (unpaired two-tailed t-test). (**d**) Serial dilutions of the phages indicated spotted on lawns of cells producing CapRel^{SJ46}, CapRel^{Ebc}, or CapRel^{Kp} or harboring an empty vector (EV). (**e**) Summary of 3 independent replicates of phage spotting assay in Fig. 1d. Asterisks indicate p = 10^{-10} (T2, SECΦ27), 10^{-6} (T4) (unpaired two-tailed t-test). (**f**) Serial dilutions of the phages indicated spotted on lawns of cells containing genomic CapRel^{SJ46}, CapRel^{SJ46} (Y155A) or His_6-CapRel^{SJ46}. (**g**) Summary of 3 independent replicates of phage spotting assay in Fig. 1g. Asterisk indicates p = 10^{-22} (unpaired two-tailed t-test).



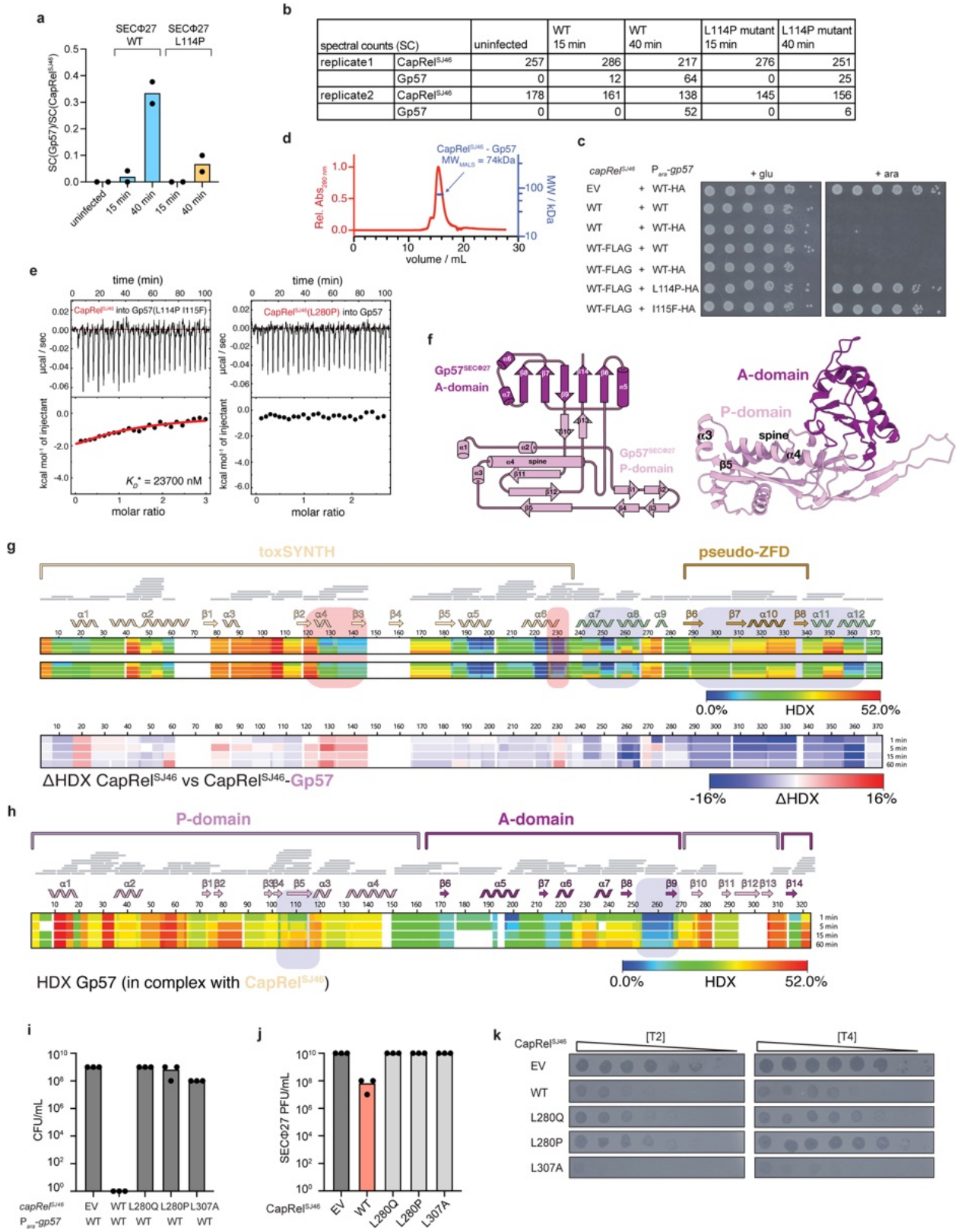
Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Structural analysis of CapRel^{SJ46}. (**a**) Alignment of CapRel^{SJ46} and diverse fused CapRel homologs, with labels indicating the pseudo-ZFD and location of substitutions that render CapRel^{SJ46} constitutively active or unable to be activated by Gp57, the SECΦ27 major capsid protein. (**b**) Summary of 3 independent replicates of phage spotting assay in Fig. 2b. (**c**) Immunoblot of untagged CapRel chimera or FLAG-tagged CapRel^{SJ46}, CapRel^{Ebc} or CapRel chimera. GyrA is included as a loading control. Image shown is a representative of 2 biological replicates. (**d**) Topology of CapRel^{SJ46}. The toxSYNTH domain is colored in light yellow, the pseudo-ZFD in dark gold and the regions that anchor pseudo-ZFD to toxSYNTH are in green. The adenine coordinating R79 and R116 are shown as red dots and the G-loop is colored in red. (**e**) Superposition of the toxSYNTH domain of CapRel^{SJ46} (colored in light yellow) onto RelQ (PDB ID: 5DEC, colored in light orange) from *Bacillus subtilis*. (**f**) Superposition of the pseudo-ZFD of CapRel^{SJ46} (colored in dark gold) onto the ZFD transcription factor of *Acidianus hospitalis* (2LVH, colored in purple). (**g**) Analysis of asymmetric unit and symmetry-related partners of CapRel^{SJ46} crystal packing. Black arrow indicates steric clash that would arise if CapRel^{SJ46} were in a closed conformation. (**h**) Superposition of the crystal structure of CapRel^{SJ46} (colored in light yellow) onto the structure of the open state predicted by AlphaFold (colored in green). (**i**) Structures of the open (*left*; from crystal structure) or closed (*right*; AlphaFold prediction) conformations of CapRel^{SJ46} color coded by the conservation score of each amino acid calculated by ConSurf. Substitutions that render CapRel^{SJ46} constitutively active or unable to be activated by Gp57 are labeled as spheres. (**j**) Serial dilutions of cells expressing the indicated variant of CapRel^{SJ46} from an arabinose-inducible promoter on media containing glucose (*left*) or arabinose (*right*). (**k**) Summary of 3 independent replicates of cell viability assay in Extended Data Fig. 3j under arabinose induction. Asterisks indicate p = 0.007 (unpaired two-tailed t-test).



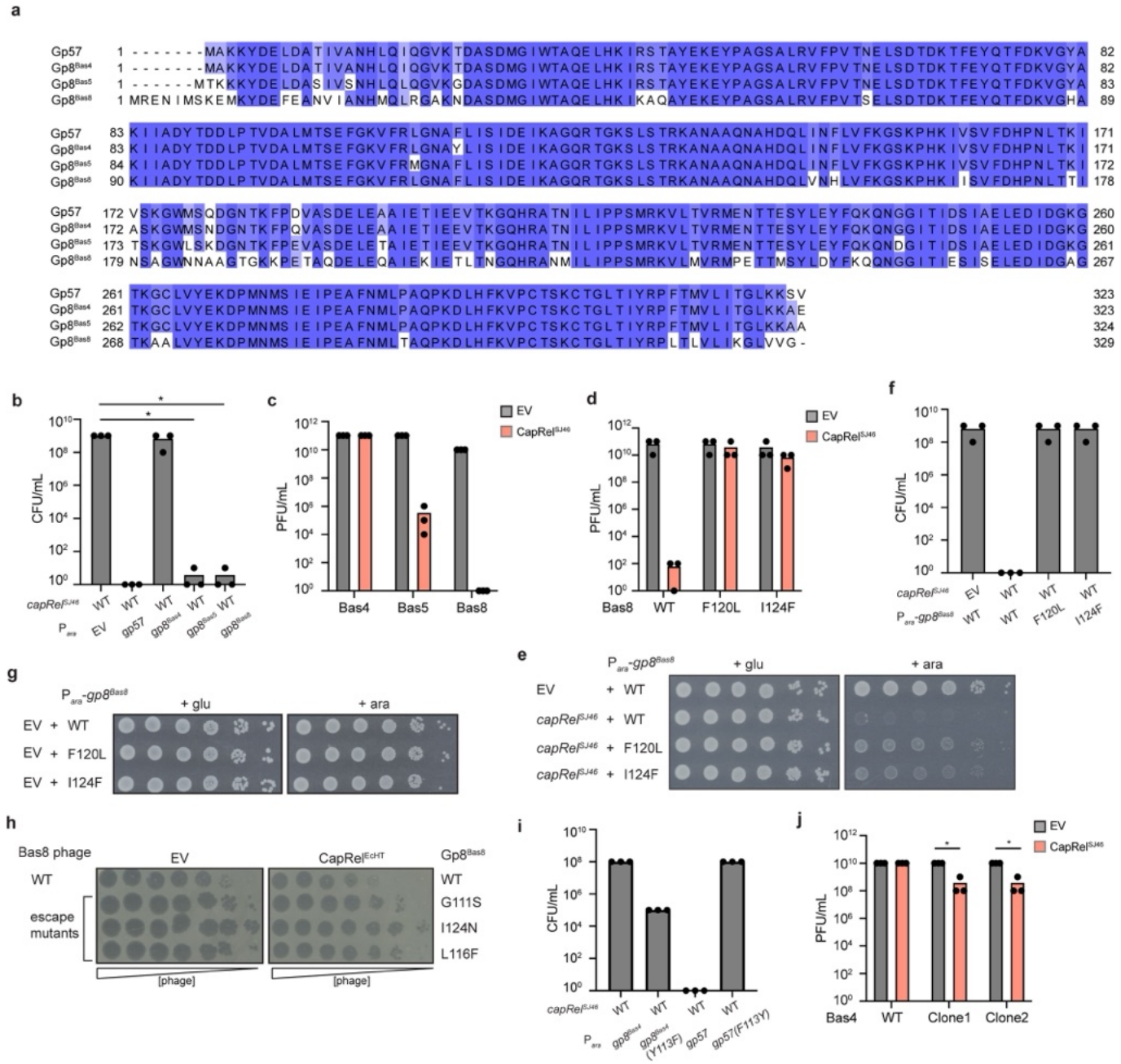
Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Gp57 from SECΦ27 triggers CapRel^{SJ46} to inhibit translation, not transcription. (**a**) Serial dilutions of phage SECΦ27 spotted on lawns of cells producing CapRel^{SJ46}, His$_6$-CapRel^{SJ46}, or CapRel^{SJ46}-FLAG, or harboring an empty vector (EV). (**b**) Loading control for Fig. 3a shown as total protein levels stained by Coomassie stain. Image shown is a representative of 2 biological replicates. (**c**) Immunoblot of His$_6$-CapRel^{SJ46} expressed from the bacterial genome or a low-copy number plasmid following infection with SECΦ27 (MOI = 100) compared to an uninfected control. Total protein levels stained by Coomassie stain is included as a loading control. Images shown are representatives of 2 biological replicates. (**d**) Summary of 3 independent replicates of phage spotting assay in Fig. 3c. Asterisk indicates $p = 10^{-10}$ (unpaired two-tailed t-test). (**e**) Serial dilutions on media containing glucose (*left*) or arabinose (*right*) of cells expressing CapRel^{SJ46} from its native promoter in the bacterial genome and expressing Gp57 from an arabinose-inducible promoter or an empty vector. (**f**) Summary of 3 independent replicates of cell viability assay in Fig. 3g under arabinose induction. Asterisk indicates $p = 10^{-30}$ (unpaired two-tailed t-test). (**g**) Serial dilutions on media containing glucose (*left*) or arabinose (*right*) of cells expressing CapRel^{SJ46}(Y155A) from its native promoter or an empty vector and expressing the indicated variant of Gp57 from an arabinose-inducible promoter. (**h**) Cells harboring CapRel^{SJ46} and producing the wild-type or L114P variant of Gp57 (expressed from an arabinose-inducible promoter) or harboring an empty vector were pulse-labeled with ^{3}H-uridine at the times indicated post-addition of arabinose. (**i**) Cells producing the CapRel^{SJ46} N-terminal toxin domain (expressed from an arabinose-inducible promoter) or harboring an empty vector were pulse-labeled with ^{35}S-Cys/Met (*left*) or ^{3}H-uridine (*right*) at the times indicated post-addition of arabinose. (**j**) Same as (**h**) but for cells carrying CapRel^{SJ46} or an empty vector and at times post-infection with SECΦ27 at MOI = 100. (**k**) Adsorption of T4 and SECΦ27 with cells containing CapRel^{SJ46} or an empty vector. Data reported are the mean +/− SD from 3 biological replicates. (**l**) Immunoblot of Gp57-HA expressed from its native promoter in the bacterial genome following infection with SECΦ27. Total protein level stained by Coomassie stain is included as a loading control (*left*). Quantification of the relative intensities of both Gp57-HA bands normalized to the loading control from 3 independent replicates (*right*). Asterisks indicate $p = 0.004$ (30 min), 0.0003 (40 min), 0.0002 (50 min) (unpaired two-tailed t-test).



Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Characterization of the CapRel^{SJ46}-Gp57 interaction. (**a**) Immunoprecipitation of CapRel^{SJ46}-FLAG from cells infected with wild-type SECΦ27 or mutant phage that produces Gp57(L114P), followed by mass spectrometry. Spectrum counts (SC) of Gp57 that had co-precipitated with CapRel^{SJ46} were normalized to the spectrum counts of CapRel^{SJ46}. Data reported are 2 biological replicates. (**b**) Same as in (**a**) but showing spectrum counts of CapRel^{SJ46} and Gp57 in two independent replicates. (**c**) Serial dilutions on media containing glucose (*left*) or arabinose (*right*) of cells producing CapRel^{SJ46} or CapRel^{SJ46}-FLAG, each expressed from its native promoter, and the indicated variant of untagged or HA-tagged version of Gp57, expressed from an arabinose-inducible promoter. (**d**) SEC-MALS analysis of CapRel^{SJ46}-Gp57 complex, revealing a molecular weight of 74 kDa. The monomers of CapRel^{SJ46} and Gp57 are predicted to be 42 and 36 kDa, respectively. (**e**) Binding of CapRel^{SJ46} to Gp57 (L114P I115F) *(left)* or CapRel^{SJ46}(L280P) to Gp57 *(right)* monitored by isothermal titration calorimetry (ITC). (**f**) Topology and cartoon representation of SECΦ27 Gp57. The P-domain is colored in pink and the A-domain in violet. (**g**) Heat maps representing the HDX of CapRel^{SJ46} (*top*) and CapRel^{SJ46}-Gp57 complex (*center*) and the ΔHDX (*bottom*). Regions involved in strong uptake such as residues 115-145 and 225-235 (which includes the active site β-strand β2 and the G-loop) are shaded in red and regions involved in strong protection 240-268 and 288-366 (which include both anchors and the pseudo-ZFD) are shaded in blue. (**h**) Heat map representing the HDX of Gp57 in the complex with CapRel^{SJ46}. Shaded regions highlight areas of variable HDX signal that indicate these regions are involved in the CapRel^{SJ46}-Gp57 interface. (**i**) Summary of 3 independent replicates of cell viability assay in Fig. 4f under arabinose induction. (**j**) Summary of 3 independent replicates of phage spotting assay in Fig. 4g. (**k**) Serial dilutions of T2 and T4 phage spotted on cells producing the indicated mutant of CapRel^{SJ46} or harboring an empty vector.



Extended Data Fig. 6 | The major capsid proteins from multiple, related phages activate CapRel^{SJ46}. (**a**) Multiple sequence alignment of the major capsid proteins from phages SECΦ27, Bas4, Bas5 and Bas8. (**b**) Summary of 3 independent replicates of cell viability assay in Fig. 5a under arabinose induction. Asterisks indicate $p = 10^{-34}$ (unpaired two-tailed t-test). (**c**) Summary of 3 independent replicates of phage spotting assay in Fig. 5b. (**d**) Summary of 3 independent replicates of phage spotting assay in Fig. 5c. (**e**) Serial dilutions on media containing glucose (*left*) or arabinose (*right*) of cells expressing CapRel^{SJ46} from its native promoter or harboring an empty vector and producing the indicated variant of the Bas8 major capsid protein (Gp8^{Bas8}) from an arabinose-inducible promoter. (**f**) Summary of 3 independent replicates of cell viability assay in Extended Data Fig. 6e under arabinose induction. (**g**) Serial dilutions on media containing glucose (*left*) or arabinose (*right*) of cells containing an empty vector and producing the indicated variant of the Bas8 major capsid protein from an arabinose-inducible promoter. (**h**) Serial dilutions of the phages indicated spotted on lawns of cells harboring CapRel^{EcHT} or an empty vector. (**i**) Summary of 3 independent replicates of cell viability assay in Fig. 5e under arabinose induction. (**j**) Summary of 3 independent replicates of phage spotting assay in Fig. 5f. Asterisks indicate $p = 10^{-6}$ (unpaired two-tailed t-test).

Extended Data Table 1 | X-ray data collection and processing

Sample	CapRel^{SJ46}
Diffraction source	Soleil PX1
Wavelength (Å)	0.9786
Temperature (K)	100.0
Detector	Eiger-X 16M
Crystal-detector distance (mm)	332.44
Rotation range per image (°)	0.01
Exposure time per image (s)	0.01
Space group	$P2_1$
a, *b*, *c* (Å)	49.6 136.2 57.8
α, β, γ (°)	90.0, 102.7, 90.0
Mosaicity (°)	0.20
Resolution range (Å)	68.10 – 2.31
Total N°. of reflections	147846 (7291)
N°. of unique reflections	20877 (1044)
Completeness (ellipsoidal %)	91.9 (63.1)
Redundancy	7.1 (7.0)
$\langle I/\sigma(I)\rangle$	11.2 (1.5)
$CC_{1/2}$	0.998 (0.712)
R_{pim}	0.051 (0.531)
Overall *B* factor / Wilson plot (Å^2)	54.7
R-factor (%)	21.2
R_{free}-factor (%)	26.5
Ramachandran profile (%)	
Core	97.9
Allowed	2.1
Outliers	0.0
R.m.s. deviations	
Bond lengths (Å)	0.013
Bond angles (°)	1.52
Number of atoms	5534
Macromolecules	5304
Solvent	228
Other	3
B-factors (Å^2)	
All atoms	68.7
Macromolecules	69.2
Solvent atoms	55.8
Other atoms	63.6
PDB ID	7ZTB

The $CC_{1/2}$ criterion was used to determine the resolution range. Values for the outer shell are given in parentheses.

Extended Data Table 2 | Mass spectrometry analysis of SECΦ27 phage lysates (wild type and mutant producing Gp57(L114P))

			WT Gp57			Gp57(L114P) mutant		
Gene product	annotation	MW (kDa)	Spectrum count (SC)	SC/MW	coverage (%)	Spectrum count (SC)	SC/MW	coverage (%)
Gp57	DUF2184 domain-containing protein	36	878	24.39	86	548	15.22	92
Gp55	HtjA	18	325	18.06	88	190	10.56	87
Gp63	tail protein/depolymerase	25	333	13.32	79	220	8.80	78
Gp66	tail tape-measure protein	98	197	2.01	58	168	1.71	56
Gp81	tail fiber protein	72	121	1.68	44	88	1.22	44
Gp52	portal protein	48	105	2.19	71	102	2.13	71
Gp71	tail fiber protein	132	99	0.75	39	73	0.55	35
Gp56	scaffolding protein	28	68	2.43	67	69	2.46	69
Gp77	recombinase	24	40	1.67	65	15	0.63	36
Gp20	hypothetical protein	16	11	0.69	43	13	0.81	43
Gp78	ssDNA-binding protein	17	24	1.41	77	16	0.94	66
Gp59	head-tail adaptor	16	11	0.69	26	10	0.63	17
Gp62	structural protein	15	20	1.33	44	16	1.07	44
Gp51	terminase large subunit	59	10	0.17	11	5	0.08	8.4
Gp54	major head subunit precursor	41	6	0.15	6.2	15	0.37	26
Gp68	minor tail protein L	28	15	0.54	33	9	0.32	21
Gp60	ribonucleoside-triphosphate reductase	14	10	0.71	29	9	0.64	29
Gp1	DNA adenine methyltransferase	28	4	0.14	12	6	0.21	26
Gp79	chaperone of endosialidase	26	14	0.54	41	9	0.35	33
Gp14	DNA-cytosine methylase	27	9	0.33	18	9	0.33	20
Gp67	minor tail protein	13	13	1.00	49	8	0.62	26
Gp75	exodeoxyribonuclease	41	5	0.12	12	7	0.17	18
Gp39	ATP-binding protein	23	5	0.22	20	5	0.22	20
Gp70	tail assembly protein	21	6	0.29	35	5	0.24	31
Gp72	hypothetical protein	22	4	0.18	16	0	0.00	0
Gp3	hypothetical protein	7	5	0.71	19	4	0.57	19
Gp53	head morphogenesis protein	29	5	0.17	11	2	0.07	8
Gp73	outer membrane protein	10	4	0.40	42	0	0.00	0
Gp80	hypothetical protein	11	3	0.27	21	2	0.18	21
Gp22	hypothetical protein	10	0	0.00	0	4	0.40	24
Gp69	C40 family peptidase	28	3	0.11	9.5	0	0.00	0

A hyperpromiscuous antitoxin protein domain for the neutralization of diverse toxin domains

Tatsuaki Kurata[a,1], Chayan Kumar Saha[a,b,1], Jessica A. Buttress[c], Toomas Mets[d], Tetiana Brodiazhenko[d], Kathryn J. Turnbull[e], Ololade F. Awoyomi[f], Sofia Raquel Alves Oliveira[d], Steffi Jimmy[g], Karin Ernits[h], Maxence Delannoy[i], Karina Persson[h], Tanel Tenson[d], Henrik Strahl[c], Vasili Hauryliuk[a,b,d,j,2], and Gemma C. Atkinson[a,b,2]

[a]Department of Experimental Medicine, University of Lund, 221 84 Lund, Sweden; [b]Department of Molecular Biology, Umeå University, 901 87 Umeå, Sweden; [c]Center for Bacterial Cell Biology, Biosciences Institute, Newcastle University, Newcastle upon Tyne NE2 4AX, United Kingdom; [d]University of Tartu, Institute of Technology, 50411 Tartu, Estonia; [e]Department of Clinical Microbiology, Rigshospitalet, 2200 Copenhagen, Denmark; [f]Department of Medical Biochemistry and Biophysics, Umeå University, 901 87 Umeå, Sweden; [g]Center for Structural Systems Biology, Deutsches Elektronen-Synchrotron, 22607 Hamburg, Germany; [h]Department of Chemistry, Umeå University, 901 87 Umeå, Sweden; [i]Département Génie Biologique, Campus SophiaTech, Université Nice Sophia Antipolis, 06000 Nice, France; and [j]Laboratory for Molecular Infection Medicine Sweden, Umeå University, 901 87 Umeå, Sweden

Edited by Gunnar von Heijne, Department of Biochemistry and Biophysics, Stockholms Universitet, Stockholm, Sweden; received May 7, 2021; accepted December 2, 2021, by Editorial Board Member Robert J. Collier

Toxin–antitoxin (TA) gene pairs are ubiquitous in microbial chromosomal genomes and plasmids as well as temperate bacteriophages. They act as regulatory switches, with the toxin limiting the growth of bacteria and archaea by compromising diverse essential cellular targets and the antitoxin counteracting the toxic effect. To uncover previously uncharted TA diversity across microbes and bacteriophages, we analyzed the conservation of genomic neighborhoods using our computational tool FlaGs (for flanking genes), which allows high-throughput detection of TA-like operons. Focusing on the widespread but poorly experimentally characterized antitoxin domain DUF4065, our in silico analyses indicated that DUF4065-containing proteins serve as broadly distributed antitoxin components in putative TA-like operons with dozens of different toxic domains with multiple different folds. Given the versatility of DUF4065, we have named the domain *Panacea* (and proteins containing the domain, PanA) after the Greek goddess of universal remedy. We have experimentally validated nine PanA-neutralized TA pairs. While the majority of validated PanA-neutralized toxins act as translation inhibitors or membrane disruptors, a putative nucleotide cyclase toxin from a *Burkholderia* prophage compromises transcription and translation as well as inducing RelA-dependent accumulation of the nucleotide alarmone (p)ppGpp. We find that Panacea-containing antitoxins form a complex with their diverse cognate toxins, characteristic of the direct neutralization mechanisms employed by Type II TA systems. Finally, through directed evolution, we have selected PanA variants that can neutralize noncognate TA toxins, thus experimentally demonstrating the evolutionary plasticity of this hyperpromiscuous antitoxin domain.

toxin | antitoxin | bacteriophage | evolution | panacea

Toxin–antitoxin systems (TAs) are diverse two-gene elements that are widespread in plasmids and chromosomes of bacteria and archaea (1, 2) as well as in genomes of temperate bacteriophages that prey on these microbes (3–6). The various protein toxins target different core processes of the encoding cell to dramatically inhibit growth while their cognate antitoxins efficiently neutralize the toxicity. Known TA toxins exert their toxicity in a variety of ways (1), often targeting translation through modification or cleavage of the ribosome, translation factors, transfer RNAs (tRNAs), or messenger RNAs (mRNAs). Similarly, antitoxins counteract the toxins through diverse mechanisms (1, 7). The means of neutralization are often classified into four or more subtypes, the main four being base pairing of the antitoxin RNA with the toxin mRNA (Type I TA systems), direct protein–protein binding and inhibition (Type II), inhibition of the protein toxin by the antitoxin RNA (Type III), or indirect nullification of toxicity (Type IV). Plasmid-encoded TA systems have long been known to function as addiction modules that promote plasmid maintenance and stability (8). The biological function of chromosomal TAs has been harder to pin down, with different TAs being implicated—with varying levels of confidence—in modulation of bacterial physiology in response to the environment, stabilization of genomic elements, and bacteriophage defense (1, 2, 9, 10).

We have recently discovered a class of TA systems that employs RelA/SpoT homologue (RSH) enzymes—so-called toxic Small Alarmone Synthetases (toxSASs)—as toxic enzymes to abrogate bacterial growth (4). Housekeeping RSH enzymes such as the *Escherichia coli* ribosome-associated amino acid starvation sensor RelA synthesize the nucleotide alarmone (p)ppGpp, a pyrophosphorylated derivative of GDP/GTP (11, 12). The toxicity of *Cellulomonas marina* toxSAS FaRel relies on the production of the related toxic alarmone (pp)pApp from housekeeping adenosine nucleotides AMP, ADP, and ATP (4). The accumulation of (pp)pApp results in dramatic depletion of ATP, which, in turn, leads to the cessation of

Significance

Toxin–antitoxin systems are enigmatic and diverse elements of bacterial and bacteriophage genomes. We have uncovered remarkable versatility in an antitoxin protein domain that has evolved to neutralize dozens of different toxin domains. We find that antitoxins carrying this domain—Panacea—form complexes with their cognate toxins, indicating a direct neutralization mechanism, and that Panacea can be evolved to neutralize a noncognate and nonhomologous toxin with just two amino acid substitutions. This raises the possibility that this domain could be an adaptable universal or semi-universal protein neutralizer with significant biotechnological and medical potential.

Author contributions: T.K., C.K.S., T.M., K.P., T.T., H.S., V.H., and G.C.A. designed research; T.K., C.K.S., J.A.B., T.M., T.B., K.J.T., O.F.A., S.R.A.O., S.J., K.E., M.D., and H.S. performed research; C.K.S. contributed new reagents/analytic tools; T.K., C.K.S., T.M., H.S., V.H., and G.C.A. analyzed data; and V.H. and G.C.A. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. G.v.H. is a guest editor invited by the Editorial Board.

[1]T.K. and C.K.S. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: vasili.hauryliuk@med.lu.se or gemma.atkinson@med.lu.se.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2102212119/-/DCSupplemental.

Published February 4, 2022.

transcription followed by the inhibition of translation and replication (4, 13). The synthesis of (pp)pApp is not the only mechanism of toxicity employed by toxSAS enzymes: we have found that the majority of experimentally explored toxSASs, such as PhRel2 from *Bacillus subtilis* strain Ia1a, act as specific protein synthesis inhibitors that pyrophosphorylate the 3′ CCA end of tRNA to abrogate aminoacylation (12). In the case of Mycobacterial phage Phrann protein Gp29, a translation-inhibiting toxSAS in the PhRel subfamily that likely pyrophosphorylates tRNA (12), the biological function appears to be defense against phage superinfection (5).

ToxSASs are neutralized by several different antitoxins that act via Type II and Type IV mechanisms. The cognate antitoxin of *B. subtilis* Ia1a PhRel2 (a tRNA-modifying toxSAS) belongs to a widespread domain family of unknown function designated by the Pfam database as DUF4065, in which DUF stands for domain of unknown function (14). Clues about the roles of DUF4065 are limited; however, it is found in so-called genetic element protein A, previously associated with TA loci (15, 16), and is also present in the proteolysis-promoting SocA antitoxin of the replication-inhibiting SocB toxin (17). This unusual mechanism of neutralization by an antitoxin is referred to as Type VI. We have earlier identified the DUF4065 domain in a putative alternative antitoxin to the ribonuclease (RNase) MqsR, but this was not tested experimentally (15).

We asked whether, given the broad distribution of DUF4065 across multiple phyla of bacteria and archaea, the analysis of the genomic neighborhood of DUF4065 can enable the prediction of novel TA systems. Using our tool FlaGs (for flanking genes) (18) to analyze diverse genomes across the tree of life, we find that DUF4065 is the predicted antitoxin counterpart of at least 1,268 different putative TA system families corresponding to at least 88 distinct putative toxin–DUF4065 domain combinations, found in diverse bacteria, archaea, and bacteriophages. While many of the toxins of these systems are related to classical TA toxins such as various mRNA interferases (19, 20), Fic/Doc-type protein modification enzymes (21), and toxSASs (4), others have little similarity to known domains or proteins with solved structures. We have experimentally verified nine DUF4065-containing antitoxins as neutralizers of their cognate toxin partners. These toxins include translation inhibitors, membrane disruptors, and a putative nucleotide cyclase that pleiotropically affects metabolism, compromising transcription and translation, as well as inducing RelA-dependent accumulation of the guanosine tetraphosphate alarmone nucleotide (p)ppGpp. Complex formation indicates DUF4065-containing antitoxins neutralize toxins via direct protein–protein interaction [that is, act as Type II TA systems (1, 2)], and we have identified substitutions that confer the ability of one antitoxin to neutralize a noncognate toxin. Given the versatility of the antitoxin function of DUF4065, we have named the domain *Panacea* after the Greek goddess of universal remedy.

Results

The Domain DUF4065 Is Found in Diverse TA-Like Loci across Bacteria, Archaea, and Bacteriophages. As DUF4065 has previously been associated with TA systems (15–17), we asked whether it may constitute a widespread antitoxin domain paired in operons with novel toxin domains. To answer this, we used sensitive sequence searching combined with an analysis of gene neighborhoods using our tool FlaGs (18) (see *SI Appendix*, Fig. S1 for a graphical overview of the procedure). Using the hidden Markov model (HMM) of the DUF4065 domain (14) to scan 20,209 genomes across cellular life and viruses, we identified 2,281 hits (Dataset S1) in prokaryotes and bacteriophages comprising 27 phyla of bacteria, 3 phyla of archaea, and 17 different bacteriophages (Dataset S1). Of those 2,281, 76 are present in complete prokaryotic genomes, allowing the determination of whether they are chromosome or plasmid encoded according to the genome annotations. All but two of our identified DUF4065 homologs in complete genomes are chromosome localized. The two exceptions annotated as plasmid encoded (but may be minichromosomes) are archaeal, found in Haloarchaea (protein accessions WP_050049451.1 and WP_049938427.1). Most DUF4065-carrying taxa only carry a single homolog; 217 taxa have two, 45 have three, 14 have four, 12 have five, and 5 have more than five. Of these five taxa, the taxon with the most DUF4065 homologs is the Mollicute bacterium "strawberry lethal yellows phytoplasma" strain NZSb11. This genome contains 25 DUF4065 homologs, of which three are predicted as being encoded in TA-like loci by our in silico analysis pipeline.

Adapting FlaGs for analyzing gene neighborhood conservation (*SI Appendix*, Fig. S1), we find that around half of the identified DUF4065-containing proteins can be detected as being encoded in two-gene loci that are conserved across multiple species, reminiscent of TA systems (Dataset S1, representatives in Fig. 1, Dataset S2, and *SI Appendix*, Fig. S2). In total, we predicted 1,313 preliminarily TA (pTA)-like loci using the criteria 1) that there should be a maximum distance of 100 nucleotides between the two genes, 2) that this architecture is conserved in two or more species, and 3) the conservation of the gene neighborhood does not suggest longer operons than three genes (*SI Appendix*, Fig. S1). We allowed three-gene architectures into our analysis, as TAs can sometimes be found with a conserved third gene, such as *mazG* in the case of *mazEF* (22), chaperones in the case of tripartite TA–chaperone modules (23), or transcriptional regulators in the case of the *paaR-paaA-parE* system (3). By allowing three-part clusters, we have identified 25 clusters that are conserved as a third gene in a subset of genomes that encode a particular predicted TA pair (Dataset S1). We call these accessory proteins, annotations of which include DNA/nucleotide and protein/amino acid modification enzymes, helicases, proteases, and nucleases. Each detected accessory third gene was only present in a small fraction of the genomes in which the main TA pair was identified, suggesting that whatever the role of these third genes, they probably do not play a general role in toxicity and neutralization.

It is conceivable that some homologous genes are found adjacent to DUF4065-encoding genes in multiple genomes purely by chance and are not part of genuine TA systems. Therefore, we used a BlastP-based (24) reciprocity test to filter out putative "toxins" that are at risk as being spurious hits (*SI Appendix*, Fig. S1*C*). From the 1,313 pTA-like loci, we determined that 67 proteins (of which 39 are predicted toxins and 28 are accessory proteins) are likely spurious hits (Dataset S1). Major classes of these spurious hits are transposases/integrases that are commonly found in TA-encoding neighborhoods and various ATPases that are captured into homologous clusters because of their well-conserved ATP-binding motifs (Dataset S1).

The remaining 1,268 putative TA loci that we predict to be relatively reliable correspond to 88 clusters of potential toxins. We number these clusters with a T prefix; for example, SocB is in cluster T10. The vast majority of these are annotated as "hypothetical protein," as they share only weak similarity to proteins of known function. Therefore, we searched the putative toxin protein sequences against the National Center for Biotechnology Information Conserved Domains Database (NCBI CDD) to detect the presence of known domains (Dataset S1). Of the 1,268 putative toxins, 938 sequences (belonging to 41 clusters) had no hit to a domain, and of the others, the most predominant domains were MqsR like (n = 90), Fic/Doc like (n = 32), and toxSAS like (domain names NT_Pol-beta-like, RelA_SpoT, and NT_Rel-Spo-like; n = 31). Other known toxin domains that were represented in the CDD results were PemK (mRNAse) and ParE (DNA gyrase inhibitor). For clusters that

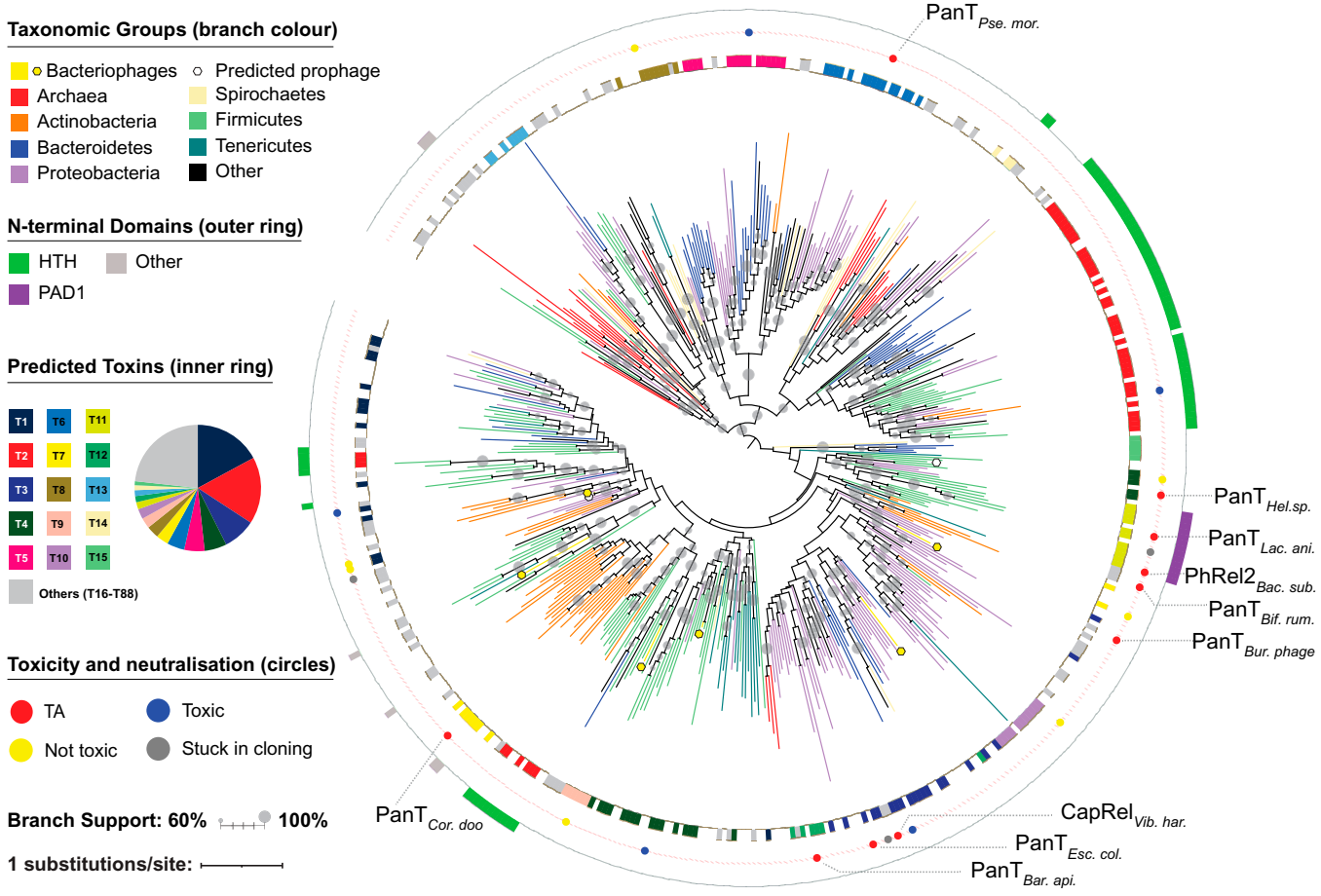


Fig. 1. The domain DUF4065/Panacea is found in a wide variety of TA-like loci across bacteria, archaea, and bacteriophages. Branches of the IQTree maximum likelihood phylogenetic tree of representative PanA sequences are colored by major taxonomic groupings as per the upper left key with an additional symbol to highlight bacteriophages. Rectangles in the outer and inner rings indicate the presence and absence of N-terminal domains in the PanA sequences and predicted associated toxin groups, respectively, according to the left-hand keys. Colored circles between the rings indicate putative TA pairs that have been tested in toxicity neutralization assays and the results of those assays. "TA" means the expression of the toxin compromises *E. coli* growth, and coexpression of the antitoxin either fully or partially counteracts the toxicity. "Toxic" means toxicity is confirmed, but the cognate PanA sequence does not rescue in this *E. coli* system. "Stuck in cloning" refers to cases in which the putative toxin genes could not be successfully chemically synthesized and plasmid subcloned, potentially because of the toxicity being too severe. Gray circles on the branches indicate branch support from IQTree ultrafast bootstrapping (53). Tree annotation was carried out with iTOL (54).

failed to find a hit in the CDD database, HHPred (25) was run with one to two representative sequences per cluster, revealing additional potential homology to proteins of known structures for 30 clusters (Dataset S1; see the following sections of *Results* for examples among our verified TAs).

The variety in the potential toxin domains suggests that the DUF4065 domain may be a universal or semi-universal antitoxin domain capable of neutralizing various different toxic proteins. In light of this, we suggest renaming DUF4065 to Panacea and abbreviate each Panacea-containing putative antitoxin and putative toxin protein as PanA and PanT, respectively. We refer to the two-gene system with the handle PanAT. In each PanAT system, the order of two genes can differ: either antitoxin first or toxin first. However, antitoxin first is the more common arrangement (943 versus 325) as is typical for Type II TA systems (1, 2).

Maximum likelihood phylogenetic analysis shows the PanA tree largely does not follow taxonomic relationships, reflecting a high degree of mobility (Fig. 1, *SI Appendix*, Fig. S2, and Dataset S2). While the deepest branches are poorly supported (not surprising for a small protein), there are a number of groups with medium to strong (over 60 to 100%) bootstrap support that include different bacterial—and sometimes archaeal—phyla. While Panacea is present broadly across prokaryotes, it does not appear to be present in eukaryotes. The only PanA we discovered in eukaryotes was in the Pharoah ant (*Monomorium pharaonic*; XP_028045404.1), and this appears to be a case of contamination, as an identical sequence is found in the bacterium *Stenotrophomonas maltophilia*. Surprisingly, a strongly supported clade of PanA sequences does not necessarily mean they all share the same PanT as shown by the inner ring in Fig. 1 and the toxin partner swapping in focus in Fig. 2*A* and *SI Appendix*, Fig. S2. Indeed, the exchange of toxin partners within a clade appears to be frequent. We refer to this kind of domain-level partner swapping as *hyperpromiscuity*, to distinguish from the promiscuity that can be seen when one single antitoxin sequence can nullify multiple con-cognate but homologous toxins (26–28).

Some—but not all—PanAs carry additional N-terminal domain regions (Fig. 1). Often, these match a known helix-turn-helix (HTH) domain, of which a number of variations exist in the NCBI CDD. We aligned all the identified regions with hits to HTH models to make our own updated HTH model. From this, we identified HTH domains in the N-terminal regions of 343 PanA sequences (Dataset S1). HTH domains are often DNA binding, are frequently found in transcription factors, and are common in Type II antitoxins (1, 2). This suggests that, like many other Type II systems, Panacea domain–containing

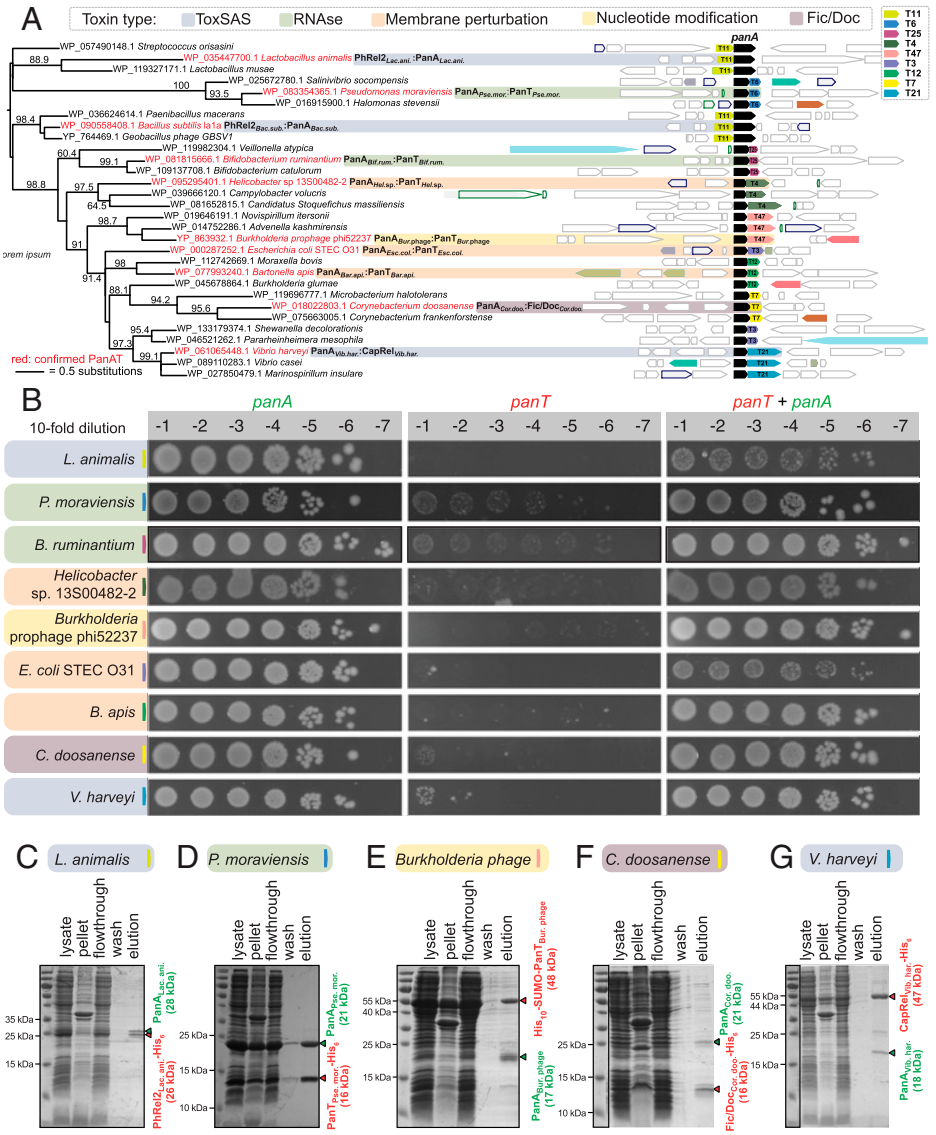


Fig. 2. PanA antitoxins form stable complexes with evolutionarily diverse TA toxins. (*A*) The maximum likelihood tree of PanA sequences, annotated with conserved gene neighborhoods generated with FlaGs (18). Numbers on branches show IQTree ultrafast bootstrap support (53). Genes belonging to homologous clusters are colored the same; the PanA antitoxin is universally shown in black. Numbers on genes preceded by a T indicate toxin clusters. (*B*) Validation of *panAT* TA pairs by toxicity neutralization assays. Overnight cultures of *E. coli* strains transformed with pBAD33 and pKK223-3 vectors or derivatives expressing putative *panT* toxins and *panA* antitoxins, correspondingly, were adjusted to OD_{600} 1.0, serially diluted, and spotted on LB medium supplemented with appropriate antibiotics and inducers (0.2% arabinose for *panT* induction and 1 mM IPTG for *panA* induction). (*C*–*G*) A pull-down assay demonstrates complex formation between PanA antitoxins and PanT toxins. Untagged PanA representatives were coexpressed in the *E. coli* BL21 DE3 strain together with N-terminally affinity-tagged (His_{10}-SUMO in the case of *Burkholderia* prophage phi52237 PanT; His_6 in all other cases) cognate PanT toxin. Filtered lysate was incubated with buffer-equilibrated Ni-beads, and PanAT complexes were eluted with 300 mM imidazole and resolved on 15% SDS-PAGE. The theoretical molecular weights for tagged toxins and antitoxins are indicated in red and green, respectively.

antitoxins can in some cases also regulate the TA function at the level of transcription. Apart from HTH domains, the only widely conserved N-terminal extension appears to correspond to a new domain, which we refer to as PanA-associated domain 1 (PAD1) (*SI Appendix*, Fig. S3). All but two of the TA-predicted PAD1 containing PanAs are paired with toxSAS-like toxins (the exception being putative ATPases from Clostridia [PanT group T62; Dataset S1]). The position at the N terminus and the presence of conserved histidines may indicate that PAD1 is a new DNA-binding domain, although it has no detectable homology with any known domain. PAD1 is also present in nine Panacea-containing proteins that do not meet the criteria for TA-like loci (Dataset S1). In all cases in which PanA contains the PAD1 domain and is in a TA-like locus, the toxin is encoded upstream of the antitoxin, the less common arrangement in the data set as a whole and in TA systems in general (1, 2).

PanA Is a Hyperpromiscuous Antitoxin Domain. Sampling broadly across PanA diversity, we selected 25 of the putative novel TAs for experimental validation in toxicity neutralization assays (Figs. 1 and 2*A*, Table 1, *SI Appendix*, Table S1, and Dataset S2). Putative toxins and antitoxins were expressed in *E. coli* strain BW25113 under the control of arabinose- and isopropyl β-D-1-thiogalactopyranoside (IPTG)-inducible promoters, respectively (4). For a gene pair to classify as a bona fide TA, two criteria need to be fulfilled: 1) the expression of the toxin should compromise *E. coli* growth, and 2) coexpression with the antitoxin should—either fully or partially—rescue from growth inhibition

Table 1. Summary of experimentally characterized PanAT pairs

Organism	Description	Toxin MOA/toxin family	Toxin accession	Antitoxin accession
Escherichia coli STEC O31	$PanA_{Esc.\ col.}$-$PanT_{Esc.\ col.}$	Membrane	WP_000019185.1	WP_000287252.1
Helicobacter sp. 13S00482-2	$PanA_{Hel.sp.}$-$PanT_{Hel.sp.}$	Membrane	WP_095295403.1	WP_095295401.1
Bartonella apis	$PanA_{Bar.\ api.}$-$PanT_{Bar.\ api.}$	Membrane	WP_077993242.1	WP_077993240.1
Burkholderia prophage phi52237	$PanA_{Bur.\ phage}$-$PanT_{Bur.\ phage}$	Nucleotide cyclase	YP_293707.1	YP_863932.1
Bifidobacterium ruminantium	$PanA_{Bif.\ rum.}$-$PanT_{Bif.\ rum.}$	RNase	WP_026646888.1	WP_081815666.1
Pseudomonas moraviensis	$PanA_{Pse.\ mor.}$-$PanT_{Pse.\ mor.}$	RNase	WP_083201923.1	WP_083354365.1[†]
Vibrio harveyi	$PanA_{Vib.\ har.}$-$CapRel_{Vib.\ har.}$	toxSAS	WP_061065447.1	WP_061065448.1
Bacillus subtilis Ia1a	$PanA_{Bac.\ sub.}$-$PhRel2_{Bac.\ sub.}$	toxSAS	WP_090558406.1	WP_090558408.1
Lactobacillus animalis	$PanA_{Lac.\ ani.}$-$PanT_{Lac.\ ani.}$	toxSAS	WP_052006344.1	WP_035447700.1
Corynebacterium doosanense	$PanA_{Cor.\ doo.}$-$PanT_{Cor.\ doo.}$	Fic/Doc	WP_018022804.1	WP_018022803.1

[†]The NBCI sequence of PanA from *P. moraviensis* appeared to be truncated at the N terminus relative to its homologues, and therefore we took an upstream start codon, equivalent to adding ten amino acids, MIFSEQKVAQ, to the N terminus.

by the toxin. In addition to the PanA-neutralized $PhRel2_{Bac.\ sub.}$ toxSAS from *B. subtilis* Ia1a that we have validated earlier (4), we have verified here nine PanAT pairs as being genuine TA loci (Table 1 and Fig. 2*B*).

PanA-neutralized toxins from *Lactobacillus animalis* ($PhRel2_{Lac.\ ani.}$) and *Vibrio harveyi* ($CapRel_{Vib.\ har.}$) belong to two different toxSAS subfamilies and, as we have shown recently, the majority of toxSAS target translation by inhibiting tRNA aminoacylation through the pyrophosphorylation of the 3′ CCA end of tRNA (12). Toxins from *Pseudomonas moraviensis* strain LMG 24280 ($PanT_{Pse.\ mor.}$) and *Bifidobacterium ruminantium* strain DSM 6489 ($PanT_{Bif.\ rum.}$) have no hits against the NCBI CDD but are predicted to be structurally similar to EndoA/PemK/MazF family RNases with HHPred (25) and thus may act as translational inhibitors similarly to the archetypal TA toxin MazF that cleaves mRNA at ACA nucleotide sequences (29). The *Corynebacterium doosanense* toxin ($PanT_{Cor.\ doo.}$) is predicted to be a member of the Fic/Doc protein family, which includes the Doc TA toxin that inhibits protein synthesis by phosphorylating the essential translation elongation factor EF-Tu (21). *Burkholderia* prophage phi52237 ($PanT_{Bur.\ phage}$) has no detectable homology to any protein domain in the NCBI CDD. However, HHpred predicts similarity to adenylate and guanylate cyclase with 97% probability, suggesting that its toxicity could be via the production of a toxic cyclic nucleotide species. Finally, many of the predicted toxin genes encode putative small peptides with predicted transmembrane helices (*SI Appendix*, Fig. S4). Of the verified TAs, the toxins with putative membrane spanning segments are those originating from *E. coli* strain STEC O31 ($PanT_{Esc.\ col.}$), *Helicobacter* sp. 13S00482-2 ($PanT_{Hel.\ sp.}$), and *Bartonella apis* strain BBC0122 ($PanT_{Bar.\ api.}$) (Table 1 and *SI Appendix*, Fig. S4). The clusters containing $PanT_{Esc.\ col.}$ (T3) and $PanT_{Bar.\ api.}$ (T12) have similar sequence compositions consisting of a charged N-terminal region followed by a hydrophilic C-terminal region where the transmembrane regions are predicted (*SI Appendix*, Figs. S2*A* and S4). It is possible that T3 and T12 are homologous, although they are dissimilar enough that they are not clustered together by FlaGs (*SI Appendix*, Fig. S2*A*). The transmembrane helices of $PanT_{Hel.\ sp.}$ are found at its C terminus, while its N-terminal region is similar to coiled-coil regions found in the synaptonemal complex protein 1 superfamily (30) and a *Salmonella* phage tail needle protein (25). For additional confirmation that PanTs vary substantially on the protein fold level, we de novo–predicted the structures of representative validated PanTs using trRosetta, a deep learning–based method (31) (*SI Appendix*, Fig. S5). In agreement with the HHpred results, the PanTs are predicted to adopt different, structurally unrelated folds.

Of the potential TA pairs that were selected and could not be verified, three of the putative toxin genes could not be successfully plasmid subcloned by the commercial provider (*SI Appendix*, Table S1). While we cannot be sure of the reason for this, it is likely that their toxicity was too severe to allow cloning in *E. coli*. Five PanTs were toxic but were not rescued by their cognate PanA, and in three of these cases, PanA itself was toxic (*SI Appendix*, Table S1 and Fig. S6 *A–D*). For example, while the PanA-associated mRNAse MqsR from *Herbaspirillum frisingense* GSF30 was—as we predicted earlier (15)—toxic, its toxicity was not countered by its cognate PanA when coexpressed in *E. coli* (*SI Appendix*, Fig. S6*C*). Finally, eight PanTs were not toxic when tested in *E. coli*—but this does not rule out the possibility of toxicity in the original host (*SI Appendix*, Table S1).

PanAT Pairs Are Type II TA Systems. The Panacea domain–containing SocA antitoxin of *Caulobacter crescentus* acts as a proteolytic adaptor, bringing the toxin SocB into contact with the protease ClpXP (17). To test whether other PanAs act as such adapters, we repeated our neutralization assays in *E. coli* strains lacking ClpXP and Lon proteases. These proteases are not necessary for neutralization by PanA (*SI Appendix*, Fig. S7). We therefore hypothesized that the general neutralization mechanism of PanA is through direct binding and inhibition typical of classical Type II systems. To test this, we carried out pull-down assays using coexpressed cognate native PanA antitoxins together with N-terminally affinity-tagged PanT toxins (with either His_6 or His_{10}-SUMO tags). We validated stable complex formation for five PanAT pairs: *L. animalis* $PhRel2_{Lac.\ ani.}$:$PanA_{Lac.\ ani.}$ (Fig. 2*C*), *P. moraviensis* $PanT_{Pse.\ mor.}$:$PanA_{Pse.\ mor.}$ (Fig. 2*D*), *Burkholderia* prophage phi52237 $PanT_{Bur.\ phage}$: $PanA_{Bur.\ phage}$ (Fig. 2*E*), *C. doosanense* $Fic/Doc_{Cor.\ doo.}$: $PanA_{Cor.\ doo.}$ (Fig. 2*F*), and *V. harveyi* $CapRel_{Vib.\ har.}$: $PanA_{Vib.\ har.}$ (Fig. 2*G*).

Protein Synthesis Is a Major Target of PanT Toxins. To address the molecular mechanisms of PanT toxicity, we assayed the effects of PanT expression on macromolecular synthesis by following the incorporation of ^{35}S methionine in proteins, ^{3}H uridine in RNA, and ^{3}H thymidine in DNA, comparing to the effects of *E. coli* MazF RNase as a positive control (*SI Appendix*, Fig. S8*A*). As predicted, five of the identified PanT—*L. animalis* $PhRel2_{Lac.\ ani.}$ and *V. harveyi* $CapRel_{Vib.\ har.}$ toxSAS, putative RNases $PanT_{Pse.\ mor.}$ and $PanT_{Bif.\ rum.}$ and *C. doosanense* Fic/Doc toxin, $Fic/Doc_{Cor.\ doo.}$—specifically inhibit protein synthesis (Fig. 3 *A–E*). The mechanism of action of all the protein synthesis–inhibiting toxins can be predicted by homology. ToxSASs *L. animalis* PhRel2 and *V. harveyi* CapRel are closely related to other representatives we have characterized earlier (12) and almost certainly pyrophosphorylate the CCA end of tRNA. The *C. doosanense* Fic/Doc toxin $Fic/Doc_{Cor.\ doo.}$ presumably modifies EF-Tu as observed for other Doc enzymes (32). Predicted RNases $PanT_{Bif.\ rum.}$ and $PanT_{Pse.\ mor.}$ likely inhibit translation by cleaving mRNA or tRNA as do their—albeit distant—relatives (33).

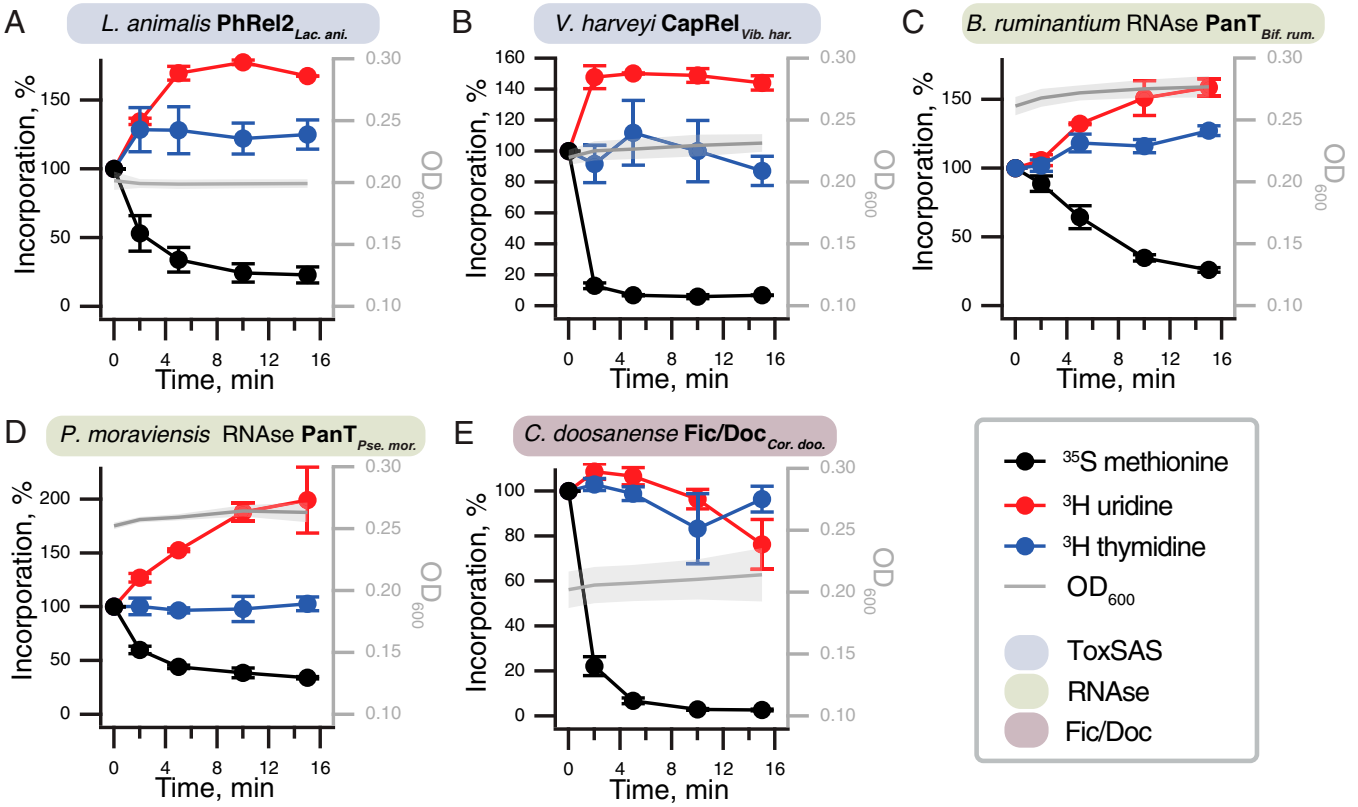


Fig. 3. Protein synthesis is a major target of PanT toxins. Metabolic-labeling assays following the incorporation of ^{35}S methionine (black traces), ^{3}H uridine (red), and ^{3}H thymidine (blue) upon expression of translation-inhibiting PanT representatives: (*A*) *L. animalis* PhRel2 and (*B*) *V. harveyi* CapRel toxSAS; putative RNases (*C*) PanT$_{Bif.\ rum.}$ and (*D*) PanT$_{Pse.\ mor.}$; and (*E*) *C. doosanense* Fic/Doc toxin, Fic/Doc$_{Cor.\ doo.}$ Expression of PanTs in *E. coli* BW25113 was induced with 0.2% L-arabinose.

***Burkholderia* Prophage phi52237 PanT Is a Pleiotropic Toxin that Induces the RelA-Mediated Stringent Response.** The *Burkholderia* prophage PanT$_{Bur.\ phage}$ toxin is unique among our verified toxins in that it predominantly inhibits transcription, with weaker effects on translation and even weaker on replication (Fig. 4*A*). The mode of inhibition is reminiscent of that of *C. marina* FaRel toxSAS (4) and *P. aeruginosa* Type VI secretion system RSH effector Tas1 (12, 13) that act though production of the toxic nucleotide alarmone (pp)pApp, leading to dramatic depletion of ATP and GTP. Therefore, we used our high performance liquid chromatography (HPLC)-based approach to study the effects of PanT$_{Bur.\ phage}$ toxin expression on *E. coli* nucleotide pools (34). In contrast to the drastic drop in GTP and ATP seen upon expression of *C. marina* FaRel toxSAS (4), expression of PanT$_{Bur.\ phage}$ results

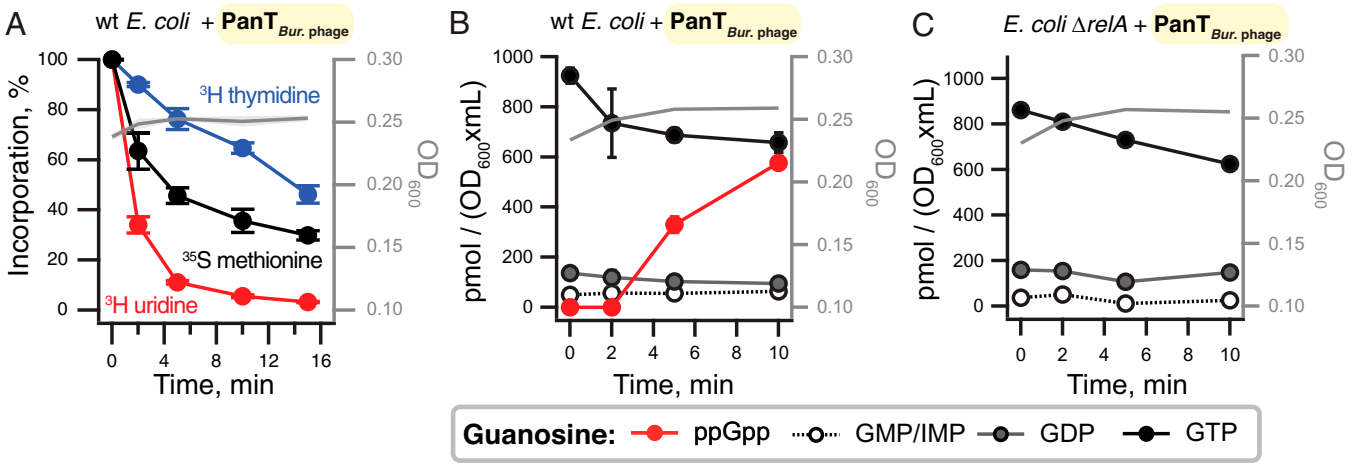


Fig. 4. PanT$_{Bur.\ phage}$ toxin from *Burkholderia* prophage phi52237 compromises transcription and translation as well as inducing the RelA-mediated stringent response. (*A*) Metabolic-labeling assay using wild-type *E. coli* BW25113 expressing PanT$_{Bur.\ phage}$ toxin. (*B* and *C*) Guanosine nucleotide pools in either wild-type (*B*) or Δ*relA* (*C*) *E. coli* BW25113 expressing PanT$_{Bur.\ phage}$ toxin. Cell cultures were grown in defined minimal MOPS medium supplemented with 0.5% glycerol at 37 °C with vigorous aeration. Expression of PanT$_{Bur.\ phage}$ toxin was induced with 0.2% L-arabinose at the OD$_{600}$ 0.2. Intracellular nucleotides are expressed in pmol per OD$_{600}$ × mL as per *inset*. Error bars indicate the SE of the arithmetic mean of three biological replicates.

in only a slight decrease in GTP (Fig. 4*B*) without affecting the ATP levels (*SI Appendix*, Fig. S9*A*). Surprisingly, despite having no detectable sequence or structural homology with RSH protein family members, PanT$_{Bur.\ phage}$ expression causes an accumulation of the alarmone nucleotide ppGpp (Fig. 4*B*). This suggests that either 1) the toxin activates cellular RSH enzymes—given the strength of the effect, likely the most potent *E. coli* (p)ppGpp synthetase RelA—or 2) the PanT$_{Bur.\ phage}$ toxin itself is capable of producing the alarmone. To distinguish between the two scenarios, we analyzed nucleotide levels upon toxin expression in an *E. coli* strain lacking *relA*. No accumulation of ppGpp is detected upon PanT$_{Bur.\ phage}$ expression in the *relA*-deficient strain (Fig. 4*C*), and just as in the case of wild type, there is no effect on ATP levels (*SI Appendix*, Fig. S9*B*). Therefore, we conclude that expression of this toxin directly or indirectly induces ppGpp production by RelA. To deconvolute the direct effects of *Burkholderia* prophage PanT$_{Bur.\ phage}$ toxin on ^{35}S methionine, ^{3}H uridine, and ^{3}H thymidine incorporation from the secondary effects caused by RelA-dependent ppGpp accumulation, we performed metabolic labeling in the Δ*relA E. coli* strain (*SI Appendix*, Fig. S9*C*). Just as in the wild-type strain, the main target is transcription, closely followed by translation. Thus, the growth inhibition and metabolic-labeling effects observed upon PanT$_{Bur.\ phage}$ expression are not related to ppGpp accumulation.

The Cell Membrane Is Another Major Target of PanT Toxins. Next, we performed ^{35}S methionine, ^{3}H uridine, and ^{3}H thymidine metabolic labeling experiments with the predicted transmembrane domain harboring toxins PanT$_{Esc.\ col.}$ (Fig. 5*A*), PanT$_{Bar.\ api.}$ (Fig. 5*B*), and PanT$_{Hel.\ sp.}$ (Fig. 5*C*). Unlike the toxins of the previous two sections of *Results* that predominantly target translation or transcription, expression of these toxins indiscriminately inhibited transcription, translation, and DNA replication, consistent with a more general shutdown of metabolic activities caused by membrane disruption. Indeed, a comparable response was observed with the induction of membrane-depolarizing *E. coli* HokB TA toxin (35) (*SI Appendix*, Fig. S8*B*) and treatment with the membrane-targeting inhibitor of oxidative phosphorylation carbonyl cyanide 3-chlorophenylhydrazone (*SI Appendix*, Fig. S8*C*).

To directly test this hypothesis, we analyzed the integrity of cell membranes upon toxin induction using a combination of the membrane potential–sensitive dye "DiSC$_3$(5)" (36) and inner membrane permeability indicator SYTOX Green (37). A strong membrane depolarization combined with an increased SYTOX Green permeability was observed for PanT$_{Bar.\ api.}$ and PanT$_{Esc.\ col.}$ (Fig. 5 *D*–*F*). Expression of PanT$_{Hel.\ sp.}$, in contrast, triggered strong depolarization without an increase in SYTOX Green permeability. Thus, we conclude PanT$_{Esc.\ col.}$, PanT$_{Hel.\ sp.}$, and PanT$_{Bar.\ api.}$ exert their toxic activity through membrane depolarization which, in the case of PanT$_{Esc.\ col.}$ and PanT$_{Bar.\ api.}$, is caused by large pore formation. Finally, weak membrane depolarization was also observed for PanT$_{Bif.\ rum.}$ and PanT$_{Pse.\ mor.}$, although these are not predicted to contain transmembrane helices and are instead predicted to be RNases. Therefore, the effect of these toxins on cell membranes is more likely to be indirect through disturbances in respiration or central carbon metabolism. A potential membrane-spanning region is predicted for PanAT$_{Bur.\ phage}$, albeit with relatively weak support (55%) (*SI Appendix*, Fig. S4*D*). As this protein does not appear to affect membrane integrity, its toxicity that is particularly striking in its effect on transcription as described above is more likely to result from its enzymatic activity, putatively cyclic nucleotide synthesis.

While PanAs Are Naturally Specific for their Cognate PanT Toxins, Their PanT Neutralization Spectrum Can Be Expanded through Directed Evolution. We have earlier shown that Type II antitoxins neutralizing toxSAS toxins—such as *B. subtilis* Ia1a PanA$_{Bac.\ sub.}$ neutralizing PhRel2$_{Bac.\ sub.}$—are specific for their cognate toxins (4). PanA is clearly a versatile domain that can evolve to neutralize—and become specific for—a range of different toxin domains. Therefore, we performed exhaustive cross-inhibition testing, resulting in a 10 × 10 cross-neutralization matrix (Fig. 6*A* and *SI Appendix*, Fig. S10). A clear diagonal signal is indicative of PanA antitoxins naturally efficiently protecting only from cognate toxins—even within groups of evolutionary related toxic effectors such as toxSAS CapRel$_{Vib.\ har.}$, PhRel2$_{Lac.\ ani.}$, and PhRel2$_{Bac.\ sub.}$ Conversely, on the evolutionary timescale, Panacea changes its toxin specificity and swaps partners, which raises the questions of what the structurally important regions for neutralization are and how a new specificity profile can be evolved.

The Panacea domain is not identifiably homologous to any protein with a known structure. Therefore, we have de novo predicted the structure of PanA$_{Vib.\ har.}$ using trRosetta, a deep learning–based method (31) (Fig. 6*B*). The model has a confidence categorized as "very high," with an estimated template modeling (TM) score of 0.704. Independent structural prediction with AlphaFold2 (10) indicates the same overall fold, with an RMSD of 1.05 Å as calculated by PyMol (*SI Appendix*, Fig. S11). The structure is comprised of a central helix (α2) surrounded by five further helices and a small three-strand β-sheet that contains a strongly conserved GPV amino acid sequence motif in the β2 strand proximal to the central helix α2 (Fig. 6*B* and *SI Appendix*, Fig. S12). The β3 and α2 elements are particularly well conserved in the sequence alignment (*SI Appendix*, Fig. S12). We probed the functional importance of the GPV motif in toxicity neutralization assays. While individual G62A and V64S substitutions did not affect the ability of *panA*$_{Vib.\ har.}$ to neutralize its cognate toxin *capRel*$_{Vib.\ har.}$, the G62A V64S double substitution resulted in the loss of neutralization activity (Fig. 6*C*), supporting that the GPV motif is, indeed, functionally important.

Next, we subjected a pair of toxSAS:PanA TA systems with effectors belonging to two distinct toxSAS subfamilies—PhRel2 and CapRel—to directed evolution experiments and screened for mutant variants of PanA$_{Vib.\ har.}$ that are able to neutralize *B. subtilis* PhRel2$_{Bac.\ sub.}$ Even though the amino acid identity between PanA$_{Vib.\ har.}$ and PanA$_{Bac.\ sub.}$ proteins is only 30 to 40%, just two substitutions—T36M and Q131L—were sufficient for cell viability as judged by colony-counting experiments (Fig. 6*D* and *SI Appendix*, Fig. S13*A*). Individual T36M and Q131L substitutions are not sufficient to elicit cross-reactivity (*SI Appendix*, Fig. S13*A*). T36 is part of the well-conserved central helix α2, while Q131 is located in a small, variable β3 strand. The β2 strand containing the conserved GPV motif is sandwiched between these structural elements (Fig. 6*B*). Notably, the T36M Q131L PanA$_{Vib.\ har.}$ variant is still capable of protecting from the cognate CapRel$_{Vib.\ har.}$ toxin. However, the protection from PhRel2$_{Bac.\ sub.}$ toxicity is less efficient than that conferred by the cognate PanA$_{Bac.\ sub.}$ antitoxin: the bacterial colonies are smaller, indicative of incomplete detoxification (*SI Appendix*, Fig. S13*A*). Therefore, we hypothesized that the T36M Q131L double substitution does not result in specificity switching in a strict sense but rather relaxes the specificity, thus allowing the neutralization of noncognate toxins. To probe this hypothesis, we tested if T36M Q131L PanA$_{Vib.\ har.}$ could protect from other noncognate PanTs (Fig. 6*E* and *SI Appendix*, Fig. S13*B*). We found that T36M Q131L PanA$_{Vib.\ har.}$ can protect from the noncognate cell membrane–targeting PanT$_{Esc.\ col.}$ (Fig. 6*E*), although incompletely, as evident from the smaller colony size (*SI Appendix*, Fig. S13*B*); no increased protection from other noncognate PanTs was detected (*SI Appendix*, Fig. S13*C*).

Discussion

Type II TAs are highly specific at the sequence level; however, small changes can result in promiscuous intermediates allowing the neutralization of additional homologous but noncognate toxins

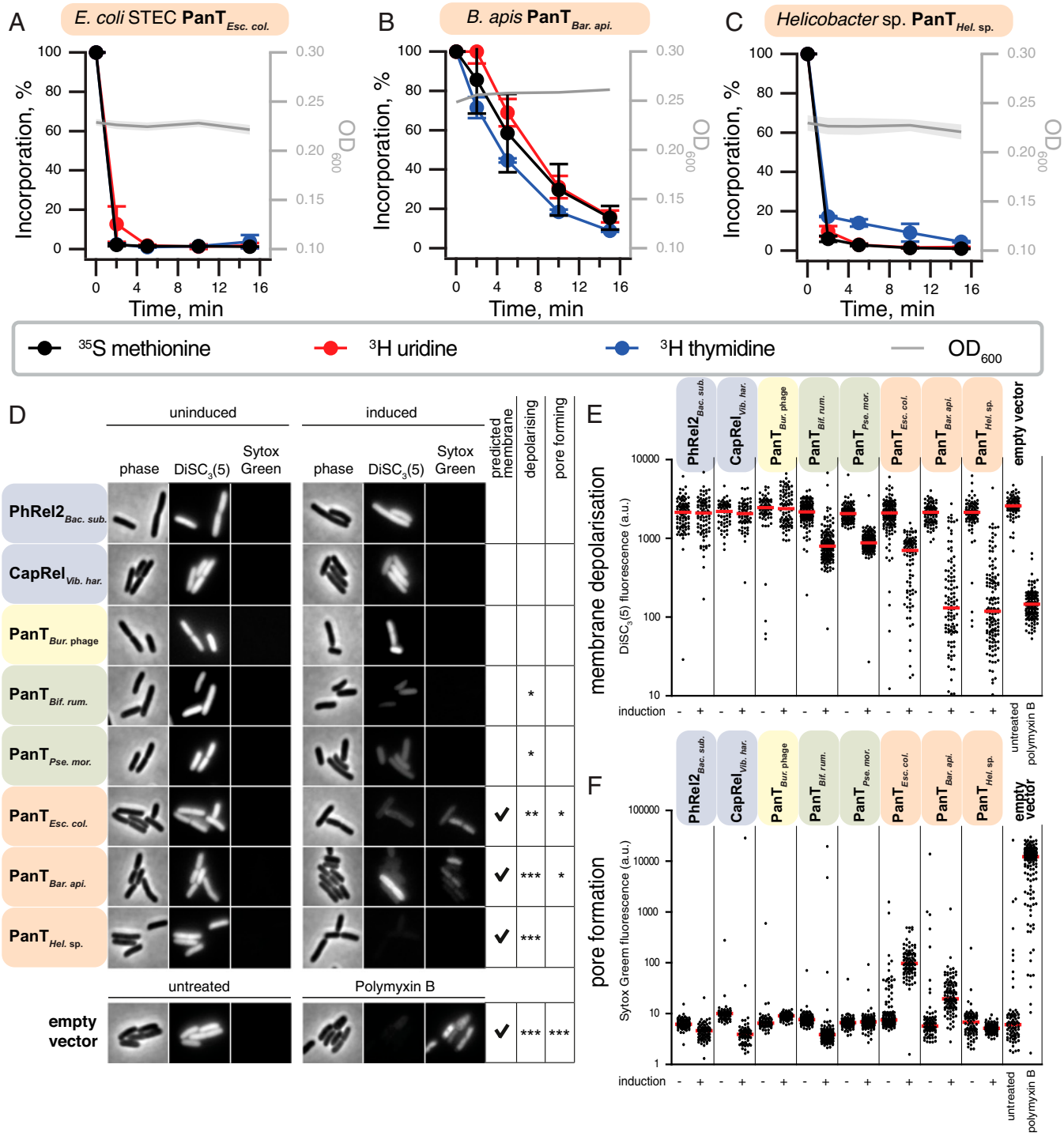


Fig. 5. Membrane integrity is a major target of PanT toxins. (*A*–*C*) Metabolic-labeling assays with wild-type *E. coli* BW25113 expressing (*A*) $PanT_{Esc.\ col.}$, (*B*) $PanT_{Bar.\ api.}$, or (*C*) $PanT_{Hel.\ sp.}$ toxins. (*D*) Phase-contrast (*Left*) and fluorescence images (*Middle* and *Right*) of *E. coli* cells costained with membrane potential–sensitive dye $DiSC_3(5)$ and membrane permeability indicator SYTOX Green. Depicted are representative cells carrying either an empty or PanT-expressing vector under uninducing (no L-arabinose) or inducing (30-min induction with 0.2% L-arabinose) conditions. As a positive control, cells containing empty vector (in MOPS-glucose medium) were incubated for 15 min with membrane-depolarizing and pore-forming antibiotic Polymyxin B. High $DiSC_3(5)$ fluorescence levels indicate high membrane potential levels found in well-energized, metabolically active cells. High SYTOX Green levels, in contrast, indicate formation of pores in the inner membrane. (*E* and *F*) Quantification of (*E*) $DiSC_3(5)$ and (*F*) SYTOX Green fluorescence for individual cells from the same imaging dataset ($n = 92$ to 165 cells). Median fluorescence intensity is indicated with a red line. Shading indicates toxin type as per Fig. 2A .

(28, 38, 39). Through selection experiments, we have demonstrated that via just two amino acid substitutions, Panacea-containing antitoxins can be made to neutralize not just noncognate but *nonhomologous* noncognate toxins that have different cellular targets and mechanisms of action. This reveals a remarkable versatility of the Panacea domain. We suggest describing the ability of an antitoxin domain to evolve to neutralize different toxin domains as *hyperpromiscuity*, distinguishing from *promiscuity*, in which one individual antitoxin can neutralize noncognate but homologous toxins sharing the same structural fold (Fig. 7). A naturally occurring

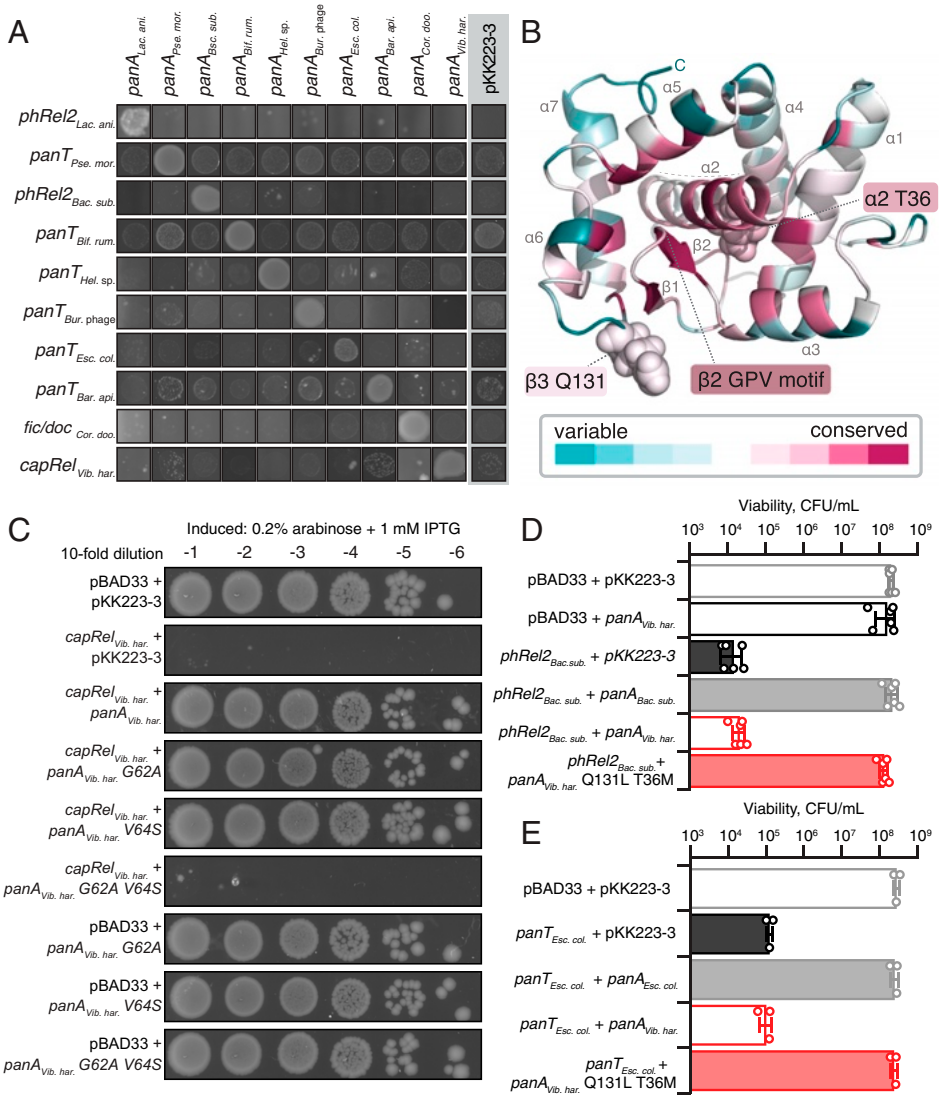


Fig. 6. PanA specificity can be readily evolved though directed evolution. (*A*) Exhaustive cross-neutralization testing establishes the strict specificity of PanA antitoxins toward their cognate toxins. The overnight cultures of *E. coli* strains transformed with pBAD33 and pKK223-3 vectors or derivatives thereof expressing toxin and PanA antitoxins was adjusted to 1.0, cultures serially diluted from 10^1- to 10^6-fold and spotted on LB agar medium supplemented with appropriate antibiotics as well as inducers (0.2% arabinose for toxin induction and 1 mM IPTG for induction PanA variants); 10^1-fold dilution is shown. (*B*) trRosetta-predicted structure (TM score 0.70, "very high confidence") of the $PanA_{Vib.\ har.}$ antitoxin colored by degree of conservation as per *SI Appendix*, Fig. S11. (*C*) Effect of amino acid substitutions in conserved GPV motif on of $PanA_{Vib.\ har.}$ antitoxin functionality. (*D*) Neutralization of $PhRel2_{Bac.\ sub.}$ toxin by evolved noncognate $PanA_{Vib.\ har.}$ T36M Q131L antitoxin. To quantify the effects of PanA/PanT coexpression on bacterial viability, the overnight cultures were subdiluted and spread on the LB agar, and individual colonies were counted. Analogous experiments with single-substituted T36M and Q131L $PanA_{Vib.\ har.}$ variants are shown in *SI Appendix*, Fig. S11. (*E*) Neutralization of $PanT_{Esc.\ col.}$ toxin by wild-type and T36M Q131L $PanA_{Vib.\ har.}$ variants.

example of the latter can be seen in the bacteriophage T4 antitoxin Dmd that neutralizes the homologous mRNase toxins RnlA and LsoA (26, 27).

Other versatile antitoxin domains have also previously been observed in computational analyses to be associated with multiple toxin-like domains (16, 40, 41), indicating potentially similar plasticity and hyperpromiscuity. One example is the Phd-related antitoxin domain found in proteins that can neutralize RelE-like mRNases, in addition to those that neutralize the EF-Tu phosphorylating toxin Doc (40). DUF4065/Panacea has previously avoided identification as a widespread antitoxin domain, despite its broad distribution in prokaryotic chromosomes. Further bioinformatic investigations of TA systems are required to understand just how unique Panacea is in its hyperpromiscuity and how many more hyperpromiscuous TAs are waiting for discovery in the vast wealth of microbial genomes.

A number of other outstanding questions about PanA remain. Firstly, how is one single domain able to neutralize so many different toxins while insulating itself against noncognate interactions? The answer to this will come from structural analyses of multiple PanAs—both alone and in complex with cognate toxins—combined with additional directed evolution experiments using different PanA pairs and an analysis of sequence coevolution. A structural analysis of complexes will also reveal the molecular function of the conserved GPV motif that is a signature of the Panacea domain and is critical for neutralization. The second question is just how much of a role proteases play in the function of PanA in some species—given the previously observed function of the Panacea domain–containing antitoxin SocA in the proteolytic degradation of toxin SocB in *Caulobacter* (17). While our results are most consistent with a Type II direct mechanism of inactivation rather than the indirect Type VI–like

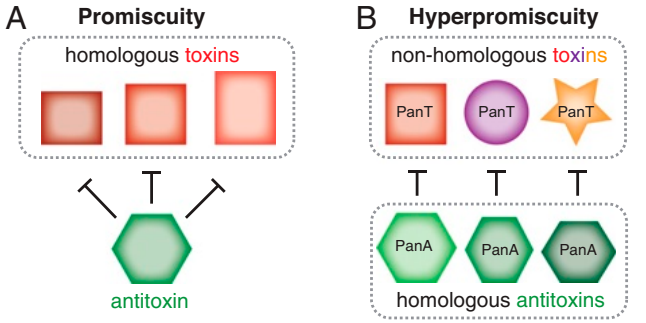


Fig. 7. Antitoxin promiscuity versus hyperpromiscuity. (*A*) A promiscuous antitoxin has relaxed neutralization specificity toward its target toxin and can neutralize a range of related toxins which all share the same structural fold. Examples include cross regulation of RelBE-like modules in *Mycobacterium tuberculosis* (55) and promiscuous ParD antitoxins generated through directed evolution that neutralize noncognate ParE toxins (28). (*B*) A hyperpromiscuous antitoxin domain, as exemplified by Panacea, can evolve to neutralize unrelated toxins that share neither structural fold nor mechanism of action.

mechanism observed for SocA, the two means of neutralization may not be mutually exclusive. Dual antitoxin functions have previously been observed; for example, DarG of the DarTG system both removes a toxic modification from DNA (Type IV) and directly binds to and inhibits the modifying enzyme DarT (Type II) (42). Finally, the evolutionary forces that drive and enable such ready partner swapping of PanAT pairs are unclear. One answer to this is hinted at in the kinds of proteins that are encoded near PanATs (Dataset S1) and in the analysis of TA (although not PanAT) gene locations near recombination sites of Tn3 transposases (43). We have found that many PanATs are encoded in close enough vicinity to transposase genes for the latter to be predicted as third component TA system genes or even false positive potential toxins that were filtered out by our pipeline (Dataset S1). It is not surprising, nor is it a new observation, that TAs can be associated with transposons; they can potentially act as addiction modules, similar to their role on plasmids (2). It is tempting to speculate that the presence of PanATs near hotspots of genomic rearrangements involving transposons and prophages—combined with an inbuilt versatility of the Panacea domain—could be driving the recombining of pairs that we observe.

Materials and Methods

Identification of PanA in Proteomes across the Tree of Life. From the NCBI genomes index (http://ftp.ncbi.nlm.nih.gov/genomes), we downloaded 20,209 predicted proteomes, selecting all viruses and one representative proteome per species for archaea, bacteria, and eukaryotes. The full taxonomy was also retrieved from NCBI. To detect the presence of PanA across the tree of life, we used the HMM of the DUF4065 domain from the Pfam database (14). We used HMMer v3.1b2 (44) to scan our database of proteomes with the DUF4065 HMM using thresholds set to the HMM profile's gathering cutoffs. We found that the DUF4065 domain was present in 2,281 identified sequences. We stored the sequences, taxonomy of the source organism, and domain composition in a MySQL database. We used this dataset and subsets of it for further phylogenetic analysis (*SI Appendix, Methods*: *Representative sequence dataset assembly* and *Phylogenetic analysis*).

Prediction of Sequence Features and Structure. Structural modeling was carried out with the trRosetta server (31). This prediction is based on de novo folding, guided by deep learning restraints. The model was colored by conservation using the Consurf server and an alignment of the sequences shown in Fig. 2 (45). Additional structural prediction was carried out for PanA$_{Vib.\ har.}$ with the AlphaFold2 (46) Colab notebook with default settings ("advanced" version; https://github.com/sokrypton/ColabFold). RSMD was calculated using structural alignment in PyMol v2.4.2 (pymol.org). Transmembrane regions were predicted with the TMHMM 2.0 sever (default settings). See *SI Appendix, Methods*: *Prediction of sequence features and structure* for details of sequence analyses for the prediction of protein domains and the identification of prophage-like genomic regions.

Prediction of TA Loci. Our Python tool FlaGs (18), which takes advantage of the sensitive sequence search method Jackhmmer (44), was adapted to identify conserved two- or three-gene conserved architectures that are typical of TA loci. Full details of the method are described in *SI Appendix, Methods*: *Prediction of TA loci*, with a schematic of the workflow shown in *SI Appendix*, Fig. S1. All scripts and datasets are available at https://github.com/GCA-VH-lab/Panacea.

Metabolic Labeling with ^{35}S Methionine, ^{3}H Uridine, or ^{3}H Thymidine. Metabolic-labeling assays were performed as described previously (4). For details, see *SI Appendix, Methods*: *Metabolic labeling*.

Construction of Plasmids. All bacterial strains and plasmids used in the study are listed in Dataset S3, and details can be found in *SI Appendix, Methods*: *Construction of plasmids*.

HPLC-Based Nucleotide Quantification. *E. coli* strain BW2511324 and *E. coli* BW25113 Δ*relA* were transformed with PanT$_{Bur.\ phage}$–expressing plasmid (pBAD33—*Burkholderia* prophage phi52237) as well as empty pKK223-3 vector. The starter cultures were pregrown overnight at 37 °C with vigorous shaking (200 rpm) in Neidhardt MOPS minimal media supplemented with 1 μg/mL thiamine, 1% glucose, 0.1% caa, 100 μg/mL carbenicillin, and 20 μg/mL chloramphenicol. The overnight cultures were diluted to optical density $(OD)_{600}$ 0.05 in 115 mL prewarmed medium MOPS supplemented with 0.5% glycerol as carbon source and grown until OD_{600} ~ 0.2 at 37 °C, 200 rpm. At this point, 0.2% arabinose was added to induce the expression of the toxin. A total of 26-mL samples were collected for HPLC analyses at 0, 2, 5, and 10 min after the addition of arabinose and 1 mM IPTG. Nucleotide extraction and HPLC analyses were performed as described previously (34). The OD_{600} measurements were performed in parallel with a collection of the samples for HPLC analyses.

Toxicity Neutralization Assays. Toxicity-neutralization assays were performed on Lysogeny broth (LB) medium (Lennox) plates (VWR). *E. coli* BW25113 strains transformed with pBAD33 derivative plasmids encoding toxins [medium copy number, p15A origin of replication, Cm^{R}, toxins are expressed under the control of a P_{BAD} promoter (47)] and pKK223-3 derivatives encoding antitoxins [medium copy number, ColE1 origin of replication, Amp^{R}, antitoxins are expressed under the control of a P_{Tac} promoter (48)] were grown in liquid LB medium (BD) supplemented with 100 μg/mL carbenicillin (AppliChem) and 20 μg/mL chloramphenicol (AppliChem) as well as 1% glucose (repression conditions). Serial 10-fold dilutions were spotted (5 μL per spot) on solid LB plates containing carbenicillin and chloramphenicol in addition to either 1% glucose (repressive conditions) or 0.2% arabinose combined with 1 mM IPTG (induction conditions). Plates were scored after an overnight incubation at 37 °C.

To quantify bacterial viability (colony forming units, CFU), overnight cultures were diluted to OD_{600} either in the range from 0.1 to 0.01 (for the strains expressing PhRel2$_{Bac.\ sub.}$, with and without coexpression of wild-type PanA$_{Vib.\ har.}$) or OD_{600} ranging from 1.0×10^{-4} to 1.0×10^{-5} (all other strains) and spread on the LB agar medium as described above for the spot-test toxicity neutralization assay. The final CFU/mL estimates were normalized to OD_{600} 1.0.

PanAT Complex Formation. The plasmids were transformed into the *E. coli* BL21 DE3 strain. Fresh transformants were washed from an LB (BD Difco-Fisher Scientific) agar plate and used to inoculate a 1-L culture LB supplemented with kanamycin (50 μg/mL). The cells were grown on 37 °C until OD_{600} reached 0.4 to 0.5 and induced with 0.5 mM IPTG. The cells were harvested after overnight cultivation on 18 °C, 220 rpm. The cells were opened with sonication in binding buffer (BB: 25 mM Hepes, pH 7.5; 300 mM NaCl; 10 mM imidazole; 2 mM $CaCl_2$; 2 mM β-ME). Filtered lysate was incubated with 1 mL previously buffer-equilibrated Ni-beads (His60 Ni Superflow Resin, TaKaRa) for 30 min. Bound protein was washed with BB on a gravity column and eluted with 300 mM imidazole. Fractions were resolved on a 15% sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS-PAGE) gel and stained with SimplyBlue SafeStain.

Fluorescence Microscopy. Overnight cultures were grown at 37 °C in MOPS minimal medium supplemented with 1% glucose. Next morning, the cells were washed, resuspended, and diluted in MOPS medium supplemented with 0.5% glycerol in order to remove glucose followed by incubation at 37 °C until

an OD_{600} of 0.3. For maintaining the plasmids, all cultures were grown in the presence of 34 μg/mL chloramphenicol. Toxin production was induced by the addition of arabinose (0.2%) for 30 min followed by staining with 200 nM of the membrane permeability indicator SYTOX Green (37) alongside the induction and 250 nM membrane potential–sensitive dye $DiSC_3(5)$ (49) for the last 5 min (36, 50). The samples were immobilized on microscope slides covered with a thin layer of H_2O/1.2% agarose and imaged immediately. As a positive control for pore formation, BW25113 *E. coli* cells transformed with the empty pBAD33 vector were incubated with 10 μg/mL polymyxin B for 15 min (51). Microscopy was performed using a Nikon Eclipse Ti equipped with a Nikon Plan Apo 100×/1.40 Oil Ph3 objective, CoolLED pE-4000 light source, Photometrics BSI sCMOS camera, and Chroma 49002 (excitation [EX] 470/40, dichroic mirror [DM] 495 lpxr, and emission [EM] 525/50) and Semrock Cy5-4040C (EX 628/40, DM 660 lp, EM 692/40) filter sets. The images were acquired with Metamorph 7.7 (MolecularDevices) and analyzed with Fiji (52).

Selection of Cross-Inhibiting PanA Mutants. An error-prone PCR mutant library of *Vibrio Harveyii* PanA antitoxins was created as described in *SI Appendix, Methods: Selection of cross-neutralizing PanAs: preparation of the antitoxin mutant library*. A total of 5 μL (around 1 μg) antitoxin mutant library was transformed into the BW25113 *E. coli* strain carrying a noncognate toxin expression plasmid $PhRel2_{Bac.\ sub.}$ toxSAS toxin from *B. subtilis* Ia1a (VHp303). The transformants were let to recover for 1 h in 1 mL SOC media at 37 °C and added to 20 mL LB media supplemented with ampicillin (100 μg/mL), chloramphenicol (25 μg/mL), 0.2% L-arabinose, and 1 mM IPTG. The bacteria were grown overnight at 37 °C while expressing both toxin and antitoxin. Next day, the plasmid was extracted from 3 mL culture using a Favorprep Plasmid Extraction Mini Kit (Favorgen Biotech Corp.). A total of 500 ng plasmid mix was again transformed into BW25113 carrying a toxin expression plasmid and let to recover as before. A total of 100 μL recovery culture was spread on LB agar plates containing corresponding antibiotics as well as 0.2% glucose (control of transformation efficiency), and the rest of the culture was collected by centrifugation and spread on an LB agar plate containing corresponding antibiotics as well as 0.2% L-arabinose and 1 mM IPTG.

Overnight cultures were started from selected colonies for further testing of cross-inhibition. The plasmids were extracted with Favorprep Plasmid Extraction Mini Kit and cleaved with FastDigest *Sac*I restriction enzyme (Thermo Scientific) to eliminate the toxin plasmids. To ensure the purity of the antitoxin mutant mix, it was transformed into the *E. coli* DH5α strain, and the plasmids were extracted from the offspring of a single colony. The *E. coli* BW25113 strain expressing the cognate or noncognate toxin was then transformed with 500 ng mutated plasmid. Again, 100 μL recovery culture was spread onto LB supplemented with corresponding antibiotics as well as 0.2% glucose agar plates, and the rest of the bacteria was collected and spread on LB agar plates supplemented with corresponding antibiotics, 0.2% arabinose and 1 mM IPTG. Phusion High-Fidelity DNA Polymerase (Thermo Scientific) was used to amplify the *panA* mutant (pK223_fwd_CPEC and pK223_rev_CPEC primers) and toxin genes (pBAD_fwd and pBAD_rev primers) with colony PCR and sequenced using pK223_rev_CPEC or pBAD_fwd primer correspondingly. The plasmid mixes and bacterial colonies were tested for possible contamination at various steps using FIREPol DNA Polymerase (Solis BioDyne): antitoxins were tested with the combination of pK223_rev_CPEC and STEC_panA_ctrl2, VH_panA_ctrl1, or Bsup_panA_ctrl1 primers and toxins with the combination of pBAD_fwd and STEC_TOX_ctrl1, VH_TOX_ctrl1, and Bsup_TOX_ctrl1 (Dataset S3).

Data Availability. Python code and text files of alignments, trees, and HMMs have been deposited in GitHub (https://github.com/GCA-VH-lab/Panacea) (56). All other study data are included in the article and/or supporting information.

ACKNOWLEDGMENTS. We are thankful to Tsigereda Ghebretnsae Ghebrelul and Constantine Stavropoulos for their help with microbiological assays, Niilo Kaldalu and Villu Kasari for sharing an *E. coli* strain lacking Lon and ClpXP, Michael Laub for sharing SocA/B-encoding plasmids, and to Bhanu Chouhan for structural prediction with the AlphaFold2 Colab notebook. We are also grateful to Mikael Lindberg and the Umeå University Protein Expression Platform. This research was supported by grants from the Knut and Alice Wallenberg Foundation (2020.0037 to G.C.A.), Umeå University Medical Faculty (Biotechnology Grant to G.C.A.), Umeå Center for Microbial Research (Gender Policy Grant to G.C.A.), Kempestiftelserna (SMK-1858.3 to G.C.A. and Scholarship Fall 2020 to C.K.S.), Carl Tryggers Stiftelse för Vetenskaplig Forskning (CTS19:24 to G.C.A.), the Swedish Research Council (Vetenskapsrådet) Grants (2017-03783 to V.H. and 2019-01085 to G.C.A.), Ragnar Söderbergs Stiftelse (to V.H.), the European Union from the European Regional Development Fund through the Center of Excellence in Molecular Cell Engineering (2014 to 2020.4.01.15-0013 to T.T. and V.H.), Molecular Infection Biology Estonia—Research Capacity Building Project (H2020-WIDESPREAD-2018-2020/GA: 857518 to T.T. and V.H.), and the Estonian Research Council (PRG335 to T.T. and V.H.). The V.H. group was also supported by the Swedish Research Council (2018-00956 to V.H.) within the RIBOTARGET consortium under the framework of The Joint Programming Initiative on Antimicrobial Resistance (JPIAMR). H.S. was supported by UKRI (UK Research and Innovation) Biotechnology and Biological Sciences Research Council Grant (BB/S00257X/1) and J.A.B. by UKRI Medical Research Council Grant MR/N013840/1.

1. A. Harms, D. E. Brodersen, N. Mitarai, K. Gerdes, Toxins, targets, and triggers: An overview of toxin-antitoxin biology. *Mol. Cell* **70**, 768–784 (2018).
2. N. Fraikin, F. Goormaghtigh, L. Van Melderen, Type II toxin-antitoxin systems: Evolution and revolutions. *J. Bacteriol.* **202**, e00763-19 (2020).
3. R. Hallez *et al.*, New toxins homologous to ParE belonging to three-component toxin-antitoxin systems in *Escherichia coli* O157:H7. *Mol. Microbiol.* **76**, 719–732 (2010).
4. S. Jimmy *et al.*, A widespread toxin-antitoxin system exploiting growth control via alarmone signaling. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 10500–10510 (2020).
5. R. M. Dedrick *et al.*, Prophage-mediated defence against viral attack and viral counter-defence. *Nat. Microbiol.* **2**, 16251 (2017).
6. Y. Guo *et al.*, RalR (a DNase) and RalA (a small RNA) form a type I toxin-antitoxin system in *Escherichia coli*. *Nucleic Acids Res.* **42**, 6448–6462 (2014).
7. P. De Bruyn, Y. Girardin, R. Loris, Prokaryote toxin-antitoxin modules: Complex regulation of an unclear function. *Protein Sci.* **30**, 1103–1113 (2021).
8. T. Ogura, S. Hiraga, Mini-F plasmid genes that couple host cell division to plasmid proliferation. *Proc. Natl. Acad. Sci. U.S.A.* **80**, 4784–4788 (1983).
9. A. Lopatina, N. Tal, R. Sorek, Abortive infection: Bacterial suicide as an antiviral immune strategy. *Annu. Rev. Virol.* **7**, 371–384 (2020).
10. S. Song, T. K. Wood, A primary physiological role of toxin/antitoxin systems is phage inhibition. *Front. Microbiol.* **11**, 1895 (2020).
11. W. A. Haseltine, R. Block, W. Gilbert, K. Weber, MSI and MSII made on ribosome in idling step of protein synthesis. *Nature* **238**, 381–384 (1972).
12. T. Kurata *et al.*, RelA-SpoT Homolog toxins pyrophosphorylate the CCA end of tRNA to inhibit protein synthesis. *Mol. Cell* **81**, 3160–3170.e9 (2021).
13. S. Ahmad *et al.*, An interbacterial toxin inhibits target cell growth by synthesizing (p)ppApp. *Nature* **575**, 674–678 (2019).
14. S. El-Gebali *et al.*, The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**, D427–D432 (2019).
15. N. Kaldalu, V. Kasari, G. Atkinson, T. Tenson, "Type II toxin-antitoxin loci: The unusual mqsRA locus" in *Prokaryotic Toxin-Antitoxins*, K. Gerdes, Ed. (Springer Berlin Heidelberg, Berlin, Heidelberg, 2013), pp. 93–105.
16. K. S. Makarova, Y. I. Wolf, E. V. Koonin, Comprehensive comparative-genomic analysis of type 2 toxin-antitoxin systems and related mobile stress response systems in prokaryotes. *Biol. Direct* **4**, 19 (2009).
17. C. D. Aakre, T. N. Phung, D. Huang, M. T. Laub, A bacterial toxin inhibits DNA replication elongation through a direct interaction with the β sliding clamp. *Mol. Cell* **52**, 617–628 (2013).
18. C. K. Saha, R. Sanches Pires, H. Brolin, M. Delannoy, G. C. Atkinson, FlaGs and webFlaGs: Discovering novel biology through the analysis of gene neighbourhood conservation. *Bioinformatics* **37**, 1312–1314 (2021).
19. Y. Yamaguchi, J. H. Park, M. Inouye, MqsR, a crucial regulator for quorum sensing and biofilm formation, is a GCU-specific mRNA interferase in *Escherichia coli*. *J. Biol. Chem.* **284**, 28746–28753 (2009).
20. S. K. Christensen, K. Pedersen, F. G. Hansen, K. Gerdes, Toxin-antitoxin loci as stress-response-elements: ChpAK/MazF and ChpBK cleave translated RNAs and are counteracted by tmRNA. *J. Mol. Biol.* **332**, 809–819 (2003).
21. D. Castro-Roa *et al.*, The Fic protein Doc uses an inverted substrate to phosphorylate and inactivate EF-Tu. *Nat. Chem. Biol.* **9**, 811–817 (2013).
22. M. Gross, I. Marianovsky, G. Glaser, MazG – A regulator of programmed cell death in *Escherichia coli*. *Mol. Microbiol.* **59**, 590–601 (2006).
23. P. Bordes *et al.*, SecB-like chaperone controls a toxin-antitoxin stress-responsive system in *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 8438–8443 (2011).
24. C. Camacho *et al.*, BLAST+: Architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
25. L. Zimmermann *et al.*, A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J. Mol. Biol.* **430**, 2237–2243 (2018).
26. F. L. Short, T. R. Blower, G. P. Salmond, A promiscuous antitoxin of bacteriophage T4 ensures successful viral replication. *Mol. Microbiol.* **83**, 665–668 (2012).
27. Y. Otsuka, T. Yonesaki, Dmd of bacteriophage T4 functions as an antitoxin against *Escherichia coli* LsoA and RnlA toxins. *Mol. Microbiol.* **83**, 669–681 (2012).
28. C. D. Aakre *et al.*, Evolving new protein-protein interaction specificity through promiscuous intermediates. *Cell* **163**, 594–606 (2015).
29. Y. Zhang *et al.*, MazF cleaves cellular mRNAs specifically at ACA to block protein synthesis in *Escherichia coli*. *Mol. Cell* **12**, 913–923 (2003).
30. A. Marchler-Bauer *et al.*, CDD/SPARCLE: Functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.* **45**, D200–D203 (2017).
31. J. Yang *et al.*, Improved protein structure prediction using predicted interresidue orientations. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 1496–1503 (2020).

32. S. Veyron, G. Peyroche, J. Cherfils, FIC proteins: From bacteria to humans and back again. *Pathog. Dis.* **76**, fty012 (2018).
33. H. Masuda, M. Inouye, Toxins of prokaryotic toxin-antitoxin systems with sequence-specific endoribonuclease activity. *Toxins (Basel)* **9**, 140 (2017).
34. V. Varik, S. R. A. Oliveira, V. Hauryliuk, T. Tenson, HPLC-based quantification of bacterial housekeeping nucleotides and alarmone messengers ppGpp and pppGpp. *Sci. Rep.* **7**, 11022 (2017).
35. K. Gerdes *et al.*, The hok killer gene family in gram-negative bacteria. *New Biol.* **2**, 946–956 (1990).
36. J. D. Te Winkel, D. A. Gray, K. H. Seistrup, L. W. Hamoen, H. Strahl, Analysis of antimicrobial-triggered membrane depolarization using voltage sensitive dyes. *Front. Cell Dev. Biol.* **4**, 29 (2016).
37. B. L. Roth, M. Poot, S. T. Yue, P. J. Millard, Bacterial viability and antibiotic susceptibility testing with SYTOX green nucleic acid stain. *Appl. Environ. Microbiol.* **63**, 2421–2431 (1997).
38. L. R. Walling, J. S. Butler, Structural determinants for antitoxin identity and insulation of cross talk between homologous toxin-antitoxin systems. *J. Bacteriol.* **198**, 3287–3295 (2016).
39. T. V. Lite *et al.*, Uncovering the basis of protein-protein interaction specificity with a combinatorially complete library. *eLife* **9**, e60924 (2020).
40. R. Leplae *et al.*, Diversity of bacterial type II toxin-antitoxin systems: A comprehensive search and functional analysis of novel families. *Nucleic Acids Res.* **39**, 5513–5525 (2011).
41. H. Akarsu *et al.*, TASmania: A bacterial toxin-antitoxin systems database. *PLoS Comput. Biol.* **15**, e1006946 (2019).
42. G. Jankevicius, A. Ariza, M. Ahel, I. Ahel, The toxin-antitoxin system DarTG catalyzes reversible ADP-ribosylation of DNA. *Mol. Cell* **64**, 1109–1116 (2016).
43. G. Lima-Mendez *et al.*, Toxin-antitoxin gene pairs found in Tn*3* family transposons appear to be an integral part of the transposition module. *MBio* **11**, e00452-20 (2020).
44. S. R. Eddy, Accelerated profile HMM searches. *PLoS Comput. Biol.* **7**, e1002195 (2011).
45. H. Ashkenazy *et al.*, ConSurf 2016: An improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **44**, W344–W350 (2016).
46. J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
47. L. M. Guzman, D. Belin, M. J. Carson, J. Beckwith, Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.* **177**, 4121–4130 (1995).
48. J. Brosius, A. Holy, Regulation of ribosomal RNA promoters with a synthetic lac operator. *Proc. Natl. Acad. Sci. U.S.A.* **81**, 6929–6933 (1984).
49. A. Waggoner, Optical probes of membrane potential. *J. Membr. Biol.* **27**, 317–334 (1976).
50. B. Kepplinger *et al.*, Mode of action and heterologous expression of the natural product antibiotic vancoresmycin. *ACS Chem. Biol.* **13**, 207–214 (2018).
51. M. Vaara, Agents that increase the permeability of the outer membrane. *Microbiol. Rev.* **56**, 395–411 (1992).
52. J. Schindelin *et al.*, Fiji: An open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
53. L. T. Nguyen, H. A. Schmidt, A. von Haeseler, B. Q. Minh, IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
54. I. Letunic, P. Bork, Interactive Tree Of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res.* **47**, W256–W259 (2019).
55. M. Yang, C. Gao, Y. Wang, H. Zhang, Z. G. He, Characterization of the interaction and cross-regulation of three *Mycobacterium tuberculosis* RelBE modules. *PLoS One* **5**, e10672 (2010).
56. C. K. Saha, G. C. Atkinson, GCA-VH-lab/Panacea. Panacea data and scripts. https://github.com/GCA-VH-lab/Panacea. Deposited 15 November 2021.