The structural basis of hyperpromiscuity in a core combinatorial network of Type II toxin-antitoxin and related phage defence systems

Karin Ernits[1,*†], Chayan Kumar Saha[1†], Tetiana Brodiazhenko[2], Bhanu Chouhan[1,3], Aditi Shenoy[4,5], Julián J. Duque-Pedraza[1], Veda Bojar[1], Jose A. Nakamoto[1], Tatsuaki Kurata[1], Artyom Egorov[1], Lena Shyrokova[1], Marcus J. O. Johansson[1], Toomas Mets[2], Aytan Rustamova[2], Jelisaveta Džigurski[2], Tanel Tenson[2], Abel Garcia-Pino[5], Arne Elofsson[4], Vasili Hauryliuk[1,2,6,*], Gemma C. Atkinson[1,*]

† These authors contributed equally

* Corresponding authors

Karin Ernits: karin.ernits@med.lu.se

Vasili Hauryliuk: vasili.hauryliuk@med.lu.se

Gemma C. Atkinson: gemma.atkinson@med.lu.se

[1] Department of Experimental Medicine, University of Lund, 221 84 Lund, Sweden

[2] University of Tartu, Institute of Technology, 50411 Tartu, Estonia

[3] Department of Molecular Biology, Umeå University, 901 87 Umeå, Sweden

[4] Department of Biochemistry and Biophysics and Science for Life Laboratory, Stockholm University, 171 21 Solna, Sweden

[5] Cellular and Molecular Microbiology, Faculté des Sciences, Université libre de Bruxelles (ULB), Boulevard du Triomphe, Building BC, (1C4 203), 1050 Brussels, Belgium

[6] Science for Life Laboratory, Lund, Sweden

Abstract:

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 neutralising 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 neutralisation 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 neutralising mechanism of the hyperpromiscuous Panacea antitoxin domain. In antitoxins composed of standalone Panacea, the domain mediates direct toxin neutralisation, while in multidomain antitoxins the neutralisation is mediated by other domains, such as PAD1, Phd-C and ZFD. We hypothesise that Panacea acts as a sensor that regulates TA activation. We have experimentally validated 16 new NetFlax TA systems. We used functional domain annotations and with metabolic labelling assays to predict their potential mechanisms of toxicity (such as disruption of membrane integrity, inhibition of cell division and abrogation of protein synthesis) as well as biological functions (such as antiphage defence). The interactive version of the NetFlax TA network that includes structural predictions can be accessed at http://netflax.webflags.se/.

Significance statement:

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 defence 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 discover new toxic effectors. We validate the predicted toxin:antitoxin complex interfaces of four TA systems, uncovering the evolutionary malleable mechanism of toxin neutralisation by Panacea-containing PanA antitoxins. We find TAs are evolutionarily related to several other phage defence systems, cementing their role as important molecular components of the arsenal of microbial warfare.

Classification

BIOLOGICAL SCIENCES: Microbiology

Keywords

Toxin, Antitoxin, Panacea, protein evolution, phage defence, AlphaFold

Introduction

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 [2]. While TA antitoxins can be either RNA- or protein-based [2], the toxins are near-universally proteinaceous with the exception of rare RNA-based Type IV systems [3, 4]. 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 serendipitously in 1983 due to its stabilising effect on plasmid maintenance [5], and more TAs were identified *ad hoc* in the following decades [6-10]. 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 [11]. More recently, novel TAs have been identified through selection for phage immunity phenotypes [12-14]. As the number of sequenced genomes and known TAs have grown, sensitive sequence searching and "guilt by association" – i.e. conserved co-localisation of toxin and antitoxin as a bicistronic operon – have been used for *in silico* discovery of new TA systems [15-23]. Our bioinformatics-driven TA discovery relies on analysis of gene neighbourhood conservation using a sequence search approach that is sensitive enough to find remote similarity even in small, divergent proteins [19, 24, 25].

TA systems are ubiquitous in microbial life, being found encoded on mobile genetic elements and chromosomes of bacteria and archaea, as well as in genomes of temperate phages and prophages [19, 26, 27]. The wide distribution and extreme diversity of TAs has driven the search to discover the biological roles of these systems [28]. The functions that have been put forward over the years can be grouped into three main types. First, addiction or stability modules that support mobile genetic element or genomic region integrity [5, 29]. Second, "emergency brakes" that are activated to stop growth as part of stress responses [30, 31]. Finally, TAs have been discovered to mediate defence against phages via abortive infection [32-34]. The latter function has found substantial recent support. Numerous studies including large-scale exploratory and focused mechanistic approaches have rapidly advanced the field [12, 14, 35-39].

Being frequently horizontally transferred components of accessory genomes, toxin-antitoxin systems have patchy distributions across genomes [15, 40]. It has long been known that Type II toxins and antitoxins have a degree of modularity, in that they can swap partners through evolution [18, 21, 40-43]. The hyperpromiscuous antitoxin domain Panacea has provided a recent 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) [25]. This discovery suggested that the Panacea domain may have inherent properties that enable it to neutralise multiple unrelated toxins through an unknown mechanism [25]. However, a structural understanding of PanA-

mediated neutralisation has been lacking. It is not even clear if the Panacea domain mediates neutralisation directly or merely serves as a platform, with other structural elements providing the inhibitory function. 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 neighbourhood analysis tool FlaGs [24], followed by experimental validation and characterisation of novel TA systems. Searching 24,479 representative proteomes from the NCBI RefSeq database, we have identified 3,597 systems within which there are 278 distinct homologous clusters of proteins in 275 distinct combinations of two-gene modules. Validating the approach, we rediscover many classical Type II TA systems such as MqsR/A, VapB/C, RelE/B, Phd/Doc but also reveal novel lineages of unique domain combinations. We have structurally annotated our network of TA-like two-gene architectures through high-throughput prediction of TA complex structures using AlphaFold2 [44] implemented in the FoldDock pipeline [45]. Focusing on the Panacea node of the network, we have validated our structural predictions though extensive 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 neutralisation 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 toxin-antitoxin systems and antiphage systems, specifically those that include the AAA+ ATPase and OLD_TOPRIM endonuclease domains such as those seen in PARIS [35], AbiLi [46] the Septu system [47], and ImmA protease-containing systems as seen in RosmerTA systems [13]. We explore the network experimentally through validating 16 new systems in toxicity neutralisation assays and predict their potential mechanisms of toxicity through functional domain annotations and metabolic labelling assays.

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 toxin-antitoxin domain interaction network. The NetFlax principle is that if one partner gene of a TA system is found in a conserved two-gene neighbourhood 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. After testing different settings for how conserved a putative TA pair must be to be used as a query in the next hopping step (see **SI Text** and **Figure 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 though 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 discovered 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 [48], 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 particularly 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**, http://netflax.webflags.se/) represents the most conserved systems of the 24,474 representative predicted proteomes considered. 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 toxin-antitoxin 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**) [49], 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 considerable 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).

Phage defence 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 defence systems [43]. The latter search revealed that the AAA ATPase and ImmA zones are particularly enriched in phage defence systems. The AAA ATPase domain of node D29 (**Fig. 1** coordinates c6) is found in a number of defence systems, such as AriA of the PARIS system [35], GajA of the Gabija system [50] and PtuA of the Septu system [47]. GajA is a sequence-specific ATP-dependent DNA endonuclease that is inhibited by dNTP and NTP nucleotides [50]. 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 Lysogenisation Defect) families [50]. 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 [35]. the 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 [51], and the AbiLii protein, which is part of a plasmid-encoded phage abortive infection mechanism [46]. HHPred [52] 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 constitutes a similar two-domain PtuAB Septu system architecture previously identified in *B. thuringiensis* [47]. 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 defence RosmerTA systems, where different toxins (RmrTs) are paired with the protease domain-containing antitoxin (RmrA) [12, 13, 53] (**Fig. 1**). The ImmA domain is named after the protein encoded on a conjugative transposon of *Bacillus subtilis* [54]. ImmA is an anti-repressor that cleaves an HTH (Helix-Turn-Helix) domain-containing repressor (ImmR) to allow the expression of an integrase [54]. Indeed, proteins within the D25 node often possesses 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 [55, 56].

Phage defence domains appear in other parts of the NetFlax network. In addition to the AAA and ImmA/Rosmer zones that are clearly phage defence related, we have recently discovered that toxSASs can protect against phages [39]. Additionally, the identification of DefenceFinder domains in the Classical Type II zone, along with evidence for classical TA domains in phage defence [57, 58] shows how intricately TAs in general are associated with phage defence.

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 (**Fig. S2**, **Dataset S1**). Most NetFlax TAs were found in Pseudomonadota (formally known as Proteobacteria) – particularly Gammaproteobacteria, reflecting the bias of RefSeq towards 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 discovery of 13 NetFlax TAs in phages (**Dataset S1)** is likely a significant underestimate as many bacteria-encoded systems are likely resident on prophages integrated into the bacterial chromosome.

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 co-evolutionary signal in the interface region [59]. AlphaFold2 is a particularly powerful structural prediction method that takes advantage of the co-evolution of residues for prediction [44]. We have run AlphaFold2 on a high throughput basis with the FoldDock pipeline [45] 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, rationalising 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 (**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 50-60% of the complexes are well modelled (**Fig. S4**).

Recurrent structural folds appear across the NetFlax network

The NetFlax algorithm includes a cross-checking step for the identified toxin and antitoxin clusters, to determine whether the potential cluster is unique or similar to any cluster discovered during the previous hopping rounds (see **Fig. S1B**). Despite this, we found multiple clusters in the classical Type II zone of the network with similar domain annotations, suggesting they may be homologous domains that are not clustered 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, **Fig. S5**), but that they are similar enough that they have the same fold. To systematically address this, we clustered all our predicted structures and annotated our network to show nodes that have the same fold (**Fig. S6**). We found 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 in the DefenseFinder database [43]. Structural alignments of the most common folds are shown in **Fig. S7**. This supports previous observations that toxins and antitoxins that are diverse at the sequence level can have the same structural fold [60]. A similar conservation of domains is also seen in phage defence systems [61].

PanA-containing TAs: the roles of individual antitoxin domains in toxin neutralisation

We focused on four previously experimentally validated PanTA TAs: *Bartonella choladocola* (previously *Bartonella apis*) PanT$_{D11}$:PanA, *Corynebacterium doosanense* Doc$_{D6}$:PanA, *Bacillus subtilis* Ia1a toxSAS PhRel2$_{D8}$:PanA and *Bacillus fungorum* MqsR$_{D2}$:PanA. The subscript D number refers to the node in **Fig. 1.** All NCBI protein accession numbers of TAs characterised in this paper are shown in **Table S1**. In all of the systems, the Panacea domain has the same compact architecture comprised of α-helices α1-α7 and β-strands β1 and β2 (**Fig. 3AB**). Despite these PanTAs having dramatically

different toxins, these four TA structures are predicted with confidence (pDockQ sores from 0.68 to 0.71). As selected PanTA systems differ in their antitoxin architecture (**Fig. 3C-F**, and see below), mutational analysis of the set allows us to interrogate the function of Panacea antitoxin in PanAs: does it mediate the toxin neutralisation directly or is it merely a ubiquitous accessory domain that has other functions, such as transcriptional regulation of the TA locus or sensing toxin-activating stimuli?

The structure of *B. choladocola* $PanT_{D11}$:PanA suggests that Panacea can, indeed, directly neutralise the toxin (**Fig. 3C**). 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 [25]. 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 neutralise the toxin, thus supporting the structural model. Both of these substituted residues are located in the conserved structural core of the Panacea domain.

In the case of *C. doosanense* PanTA, the two additional C-terminal helices decorating the Panacea core of the PanA antitoxin are predicted to make extensive contact with Doc_{D6} toxin (**Fig. 3D**). The globular Panacea domain itself is not predicted to be involved in neutralisation. 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 [62]. 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 the cellular protein synthesis [63]; *C. doosanense* Doc_{D6} similarly targets translation [25], and the active site residues are conserved amongst the two proteins (**Fig. 3D**). 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. 3D**). Expression of the isolated Phd-C domain does not neutralise the toxin, which could be due to the intrinsic instability of the element. To test this hypothesis, we fused Phd-C with a stabilising N-terminal SUMO tag, and as predicted, the resulting construct can readily neutralise Doc_{D6}, despite lacking the Panacea domain. No neutralisation was observed in the control experiment with SUMO alone. Collectively, these results suggest that Panacea can serve as an accessory domain, with neutralisation being mediated by a dedicated separate domain.

Next, we characterised the *B. subtilis* Ia1a $PhRel2_{D8}$:PanA system. In our previous analysis of the Panacea domain distribution, we discovered a novel domain that we named the PAD1 domain, standing for Panacea-associated domain 1 [25]. 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 *Bacillus subtilis* Ia1a, fused TA CapRel [39] showed that D27, PAD1 and pseudo-ZBD have the same fold, and share the same interface with the toxSAS toxin (**Fig. S8**). Importantly, it is PAD1 that forms most of the contacts with the $PhRel2_{D8}$ toxin (**Fig. 3E**). Strikingly, *B. subtilis* Ia1a PAD1 domain alone – with Panacea

removed – can neutralise the toxin, thus directly supporting the structural prediction. Furthermore, H52P substitutions that are predicted to break the PAD1:PhRel2$_{D8}$ interface completely abrogate the neutralisation, 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 neutralisation. To further support the role of PAD1 as a dedicated toxin-neutralising domain, we have, via toxicity neutralisation assays, validated the *Clostridium hylemonae* DSM 15053 TA system comprised of PhRel2$_{D8}$ and the PAD1$_{D27}$ antitoxin (**Fig. S9**). As the *C. hylemonae* antitoxin naturally lacks the Panacea domain, this observation further supports PAD1 being a directly neutralising antitoxin element.

Finally, we have dissected the *B. fungorum* MqsR$_{D2}$:PanA system (**Fig. 3F**). MqsR is an RNase [60, 64] that is neutralised by antitoxin MqsA comprised of an N-terminal Zn^{2+}-binding domain (ZBD) and C-terminal HTH [60]. While the ZBD interacts with MqsR and inhibits it without directly interacting with the RNase active site, the HTH region dimerises and acts as a transcriptional auto-regulator 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* PanTA, the ZBD directly mediates neutralisation of MqsR$_{D2}$. While both isolated ZBD and ZBD-HTH segments efficiently neutralise MqsR$_{D2}$, neither Panacea alone nor HTH-Panacea are sufficient for neutralisation.

Collectively, our results demonstrate that while Panacea can act as a direct toxin neutraliser, it is unlikely to act as such in PanA antitoxins that contain additional dedicated neutralisation 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 auto-regulator of *panAT* operons either, as *B. fungorum* PanA contains a dedicated DNA-binding regulatory domain, HTH (as do many other PanAs, [25]). Therefore, we favour the hypothesis that the Panacea domain acts as a sensor responding to – as yet unknown – TA-activating cues.

Experimental exploration of the NetFlax network

Finally, we explored our NetFlax TA network experimentally. We focused on novel TA pairs, i.e. the ones containing either i) novel toxin or/and antitoxin domains or ii) novel combinations of previously validated toxins and antitoxins. We have validated 15 TA pairs in toxicity neutralisation assays (**Fig. 1**). For 13 of them, we performed metabolic labelling 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) (**Fig. 1**).

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. 4A**). Similarly to *B. choladocola* PanT$_{D11}$:PanA, *G. mesophilus* PanA is comprised of a stand-alone Panacea domain that directly neutralises 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 labelling assays (**Fig. 4B**). The toxin contains an α-helical FtsL domain which is predicted to dimerise and be localised to the cell membrane (**Fig. 4C,D**). FtsL is an essential component of bacterial dividosome which forms a trimeric complex with FtsB and FtsQ via leucine zipper-like (LZ) motifs [65]. 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 cell division. Experiments with liquid cultures of *E. coli* expressing *G. mesophilus* toxFtsL$_D$ 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. 4E**).

The TA system from *Gordonia* phage Kita is a new member of the RosmerTA family [12, 13, 53] (**Fig. 5A**). The RmrA T_{D25} protease antitoxin is paired with a D_5 toxin which has no detectable similarity to other protein families. Metabolic labelling assays show rapid and dramatic abrogation of translation, transcription and replication upon expression of the Kita phage D_5 toxin (**Fig. 5B**). The toxin is not fully neutralised by the antitoxin and the structure of the TA complex can not be reliably predicted by AlphaFold (pDockQ score of 0.05) (**Fig. 5C**). The C-terminal region of the toxin is predicted to be localised in cellular membrane (**Fig. 65C**). In liquid culture experiments expression of RmrT$_{D5}$ immediatly abrogates bacterial growth without causing a consequent collapse of OD_{600}; the toxin is likely to share the mechanism of toxicity with membrane-depolarising *B. choladocola* PanT$_{D11}$ [25] that we used as a control (**Fig. 4E**). Following the nomenclature for RosmerTA toxins [12, 13, 53], we renamed the Kita phage toxin RmrT$_{D5}$. As other RosmerTA systems have been shown to be phage defence systems [12, 13, 53], it is likely that the Kita phage RmrTA has a similar function.

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 (**Fig. S6**), paired with an AAA ATPase A_{D29} antitoxin (**Fig. 6A**). We refer to the toxin as AarT for AAA-associated RNAse-like toxin. Metabolic labelling 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 [25]. Similar neutralisation architecture was predicted for D_{29} AAA antitoxins from *Clostridium algidicarnis* (**Fig. 6B**) and *Streptococcus agalactiae* 2603V R (**Fig. 6C**). 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 [46]. We speculate that the three AAA-neutralised TA pairs are also Abi phage defence systems. AlphaFold2 modelling does not give a convincing interface for AAA antitoxins and their toxins (pDockQ score of 0.24-0.35), which may be because phage defence AAA-containing systems form large multimeric complexes, as is seen with AAA-containing RADAR [66, 67].

Finally, we have validated 10 TA pairs of nuclease toxins (MqsR$_{D2}$, PIN/VapC-like and YafQ$_{D119}$) paired with diverse antitoxins (PanA, PerlF$_{D13}$, DUF2680$_{D105}$, RHH_6$_{D107}$, RHH_6$_{D111}$, DUF2080$_{D146}$, SynP1$_{D151}$ and RHH_6$_{D157}$), for which AF2 structural models suggest multiple mechanisms of direct and indirect

toxin neutralisation (**Fig. 7** and **Fig. S10**,**S11**). While the structures were predicted as binary complexes, RHH (Ribbon-Helix-Helix) domain is a well-characterised dimeric DNA-binding transcriptional regulator employed by numerous antitoxins such as CcdA [68] and FitA [69]. Dimers of RHH-containing antitoxins can be readily predicted by AlphaFold (**Fig. S12**). Multiple groups of translation-targeting RNase TA toxins have been characterised experimentally, and display considerable diversity, even within closely related groups; for example, different VapC PIN TA toxins can cleave either tRNA [70] or rRNA [71]. As expected for nucleases, metabolic labelling assays indicate the majority of the validated nuclease toxins do, indeed, target protein synthesis (**Fig. 7AB** and **Fig. S10**,**S11**). 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. 7C**). Although metabolic labelling 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 (**Fig. S10**).

Discussion

Classical TA antitoxins are modular proteins typically consisting of a DNA-binding domain involved in transcription autoregulation and a functionally independent neutralisation 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 neutralisation domains that enables antitoxin promiscuity, i.e. allows for neutralisation of toxins belonging to multiple protein families by different antitoxins possessing the same DNA-binding domain [72]. This model postulates that the fusion of a “linear” recognition motif that performs the neutralisation, 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. 3D**). Given the existence of multiple TA operons with single-domain antitoxins that consist of the neutralisation domain alone [73, 74], such fusion or exchange events constitute a plausible evolutionary pathway that could generate the complex toxin-antitoxin permutations observed in TAs (**Fig. 1**).

Importantly, the NetFlax network reveals the existence of a type of hyperpromiscuous antitoxin that defies this commonly accepted neutralisation paradigm. Such antitoxins are epitomised by Panacea and HTH domains [25, 60, 75-77] that possess within their structural fold an intrinsic capacity to specifically recognise and neutralise diverse toxins via three-dimensional epitopes (**Figs. 3C** and **8**). Furthermore, these antitoxin domains can also acquire linear epitopes consisting of intrinsically disordered regions or well-folded domains, to neutralise toxins in the “classical” manner (**Figs. 3EF** and **8**). In addition to this hyperpromiscuity, antitoxins have a capacity for moonlighting in other ways: PAD1-related domains are involved in both toxin neutralisation and phage detection [39], while HigA antitoxins

from TAC operons engage both the toxin and the dedicated regulator chaperone [78]. 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 defence.

In the case of PanAs that function via linear neutralisation without the direct involvement of the globular Panacea domain, the question is what then is the role of the Panacea domain. We hypothesise that its function is primarily sensory, reacting to a trigger and activating the toxin through an allosteric mechanism involving the neutralisation region. This activation may not even require dissociation of the PanTA complex; the fused toxSAS toxin-antitoxin CapRel shows that antitoxins do not have to dissociate in order to activate the toxin [39]. 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 [60]. However, conservation of fold may not mean conservation of function: the 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 [63, 79]. Similarly, toxSAS TAs can either inhibit bacterial growth by producing toxic alarmone (pp)pApp or pyrophosphorylating the 3′ CCA end of tRNA [19, 80]. 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, 81], but increasingly a role for these systems in phage defence is being discovered [12, 14, 35-38]. Here we show that this is also reflected in their evolutionary relationships: the core network of TAs is connected to domains other phage defence systems such as PARIS, AbiLi, Septu, Gabija and Rosmer systems [13, 35, 46, 47, 50]. Indeed, since it is common for multigene defence 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 module” TAs on mobile elements may actually have a role in defence 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 [82], HEPN-MNT [83], HicBA [84], and HipBA [85] escaped our prediction. Second, NetFlax only addresses conserved two-gene proteinaceous systems and therefore can not in this incarnation predict multi-gene 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 new avenues including the 20 TA systems we have

characterised, 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. The overall strategy is based on the pipeline of **Fig. S1**. Briefly, NetFlax works round by round; and each round follows three basic steps: i) scanning of a local database of proteomes using an HMM profile of toxin or antitoxin to identify its protein homologues, ii) prediction of conserved TA-like arrangements by analysing the neighbourhood of the protein homologues and identification of homologous clusters of toxins or antitoxins and iii) cross-checking if the predicted clusters are novel (never been identified in any previous round) and, if so, make their HMM profiles, to be used in the next round of scanning mentioned in step 1.

NetFlax uses a local database of 24,479 predicted proteomes downloaded from the NCBI RefSeq FTP server [86]. The local database includes one representative proteome per species of bacteria and archaea, along with all 10,449 available virus genomes (not limited to representatives). NetFlax uses conservation of gene neighbourhood analyses by FlaGs and combines it with the guilt-by-association strategy to identify TA-like pairs. NetFlax is multi-core aware and proceeds round by round hopping from antitoxin to toxin and toxin to antitoxin and continuing similarly until it finds the entire network (see **SI Text, Fig. 1** and – in more detail – **Fig. S1** for illustrations of how the algorithm works).

Protein function prediction

Protein domains and other functional predictions were carried out with searching the toxin-antitoxin database (TADB) [48], DefenceFinder [43], NCBI conserved domain database (CDD) [87] and HHPred [88] (with NCBI-CDD, Pfam-A and PDB as target databases).

Protein structure prediction and comparisons

Protein-protein complex structures were predicted with FoldDock [45]. Structures were clustered with FoldSeek v. 1.3 [89] and with resulting networks visualised with Cytoscape v. 3.5.0 [90]. Structural alignment was carried out with mTM-Align v. 20220104 [91]. Transmembrane prediction was carried out with DeepTMHMM [92]. More detailed methods are described in the **SI Text** document.

Experimental methods

Detailed experimental procedures are provided the **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 ORF-s were cloned into arabinose-inducible pBAD33 vector [93] either with or without Shine-Dalgarno sequence as required, for toxicity assay. For neutralization assay, antitoxins were expressed

from an IPTG inducible pMG25 vector [94]. Mutations and truncations were introduced to the antitoxin genes as described earlier [23].

Toxicity neutralisation 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 (repression conditions). Serial 10-fold dilutions were spotted (5 µL per spot) onto solid 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 with all possible combinations.

Metabolic labelling: metabolic labelling experiments using *E. coli* BW25113 strains co-transformed with pBAD33 derivatives as well as the empty pMG25 vector were performed as described earlier [25].

Acknowledgments

We are grateful to the Protein Expertise Platform at Umeå University for plasmid construction. The AlphaFold2/FoldDock computations were enabled by the supercomputing resource Berzelius provided by the National Supercomputer Centre at Linköping University and the Knut and Alice Wallenberg foundation. This work was supported by Knut and Alice Wallenberg Foundation (2020.0037 to GCA), the Swedish Research council (2019-01085 and 2022-01603 to GCA, 2021-01146 to VH, 2021-03979 to AE); Crafoord foundation (project grant Nr 20220562 to VH); Cancerfonden (20 0872 Pj to VH); Carl Tryggers Stiftelse för Vetenskaplig Forskning (CTS19:24 to GCA); the European Regional Development Fund through the Centre of Excellence for Molecular Cell Technology (VH and TT); Ragnar Söderberg foundation (VH); Fonds National de Recherche Scientifique (FNRS CDR J.0068.19, FNRS-EQP UN.025.19 and FNRS-PDR T.0066.18 to AG-P); ERC (CoG DiStRes, n° 864311 to AG-P).

Declaration of interests

AG-P is co-founder and stockholder of Santero Therapeutics.

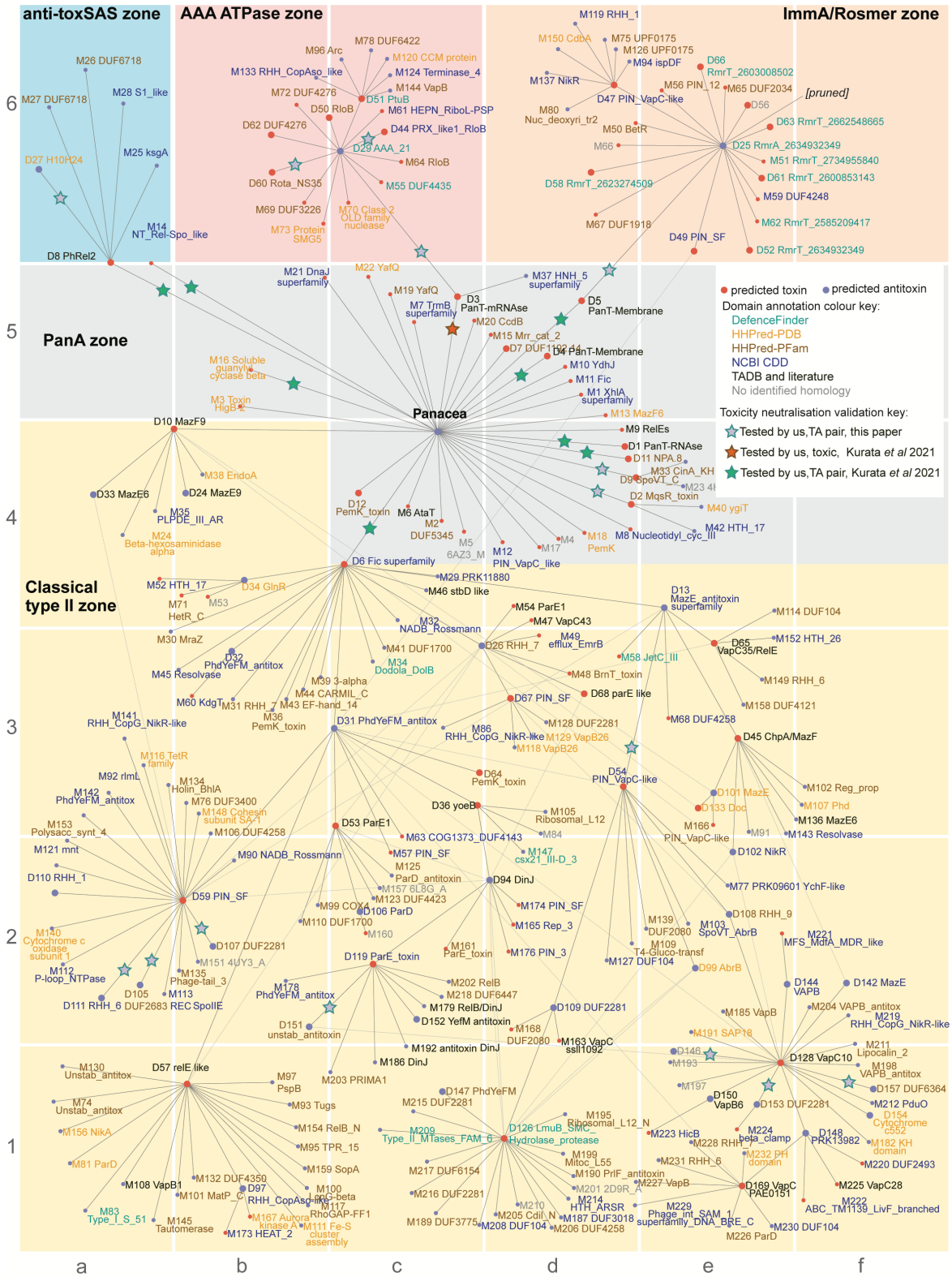

Figure 1. The core proteinaceous TA network.

The core network of NetFlax-predicted toxin and antitoxin-like domains across microbial life. The starting input was the Panacea domain. Blue circles are predicted antitoxins and red are predicted toxins. Yellow stars show groups that contain previously verified toxins and antitoxins (**Dataset S1**) [48]. Orange and green stars show, respectively toxins and TAs that have been previously validated,

including by us [19, 49]. Predicted toxin and antitoxin domains are annotated based on sequence homology searches (see text for references).

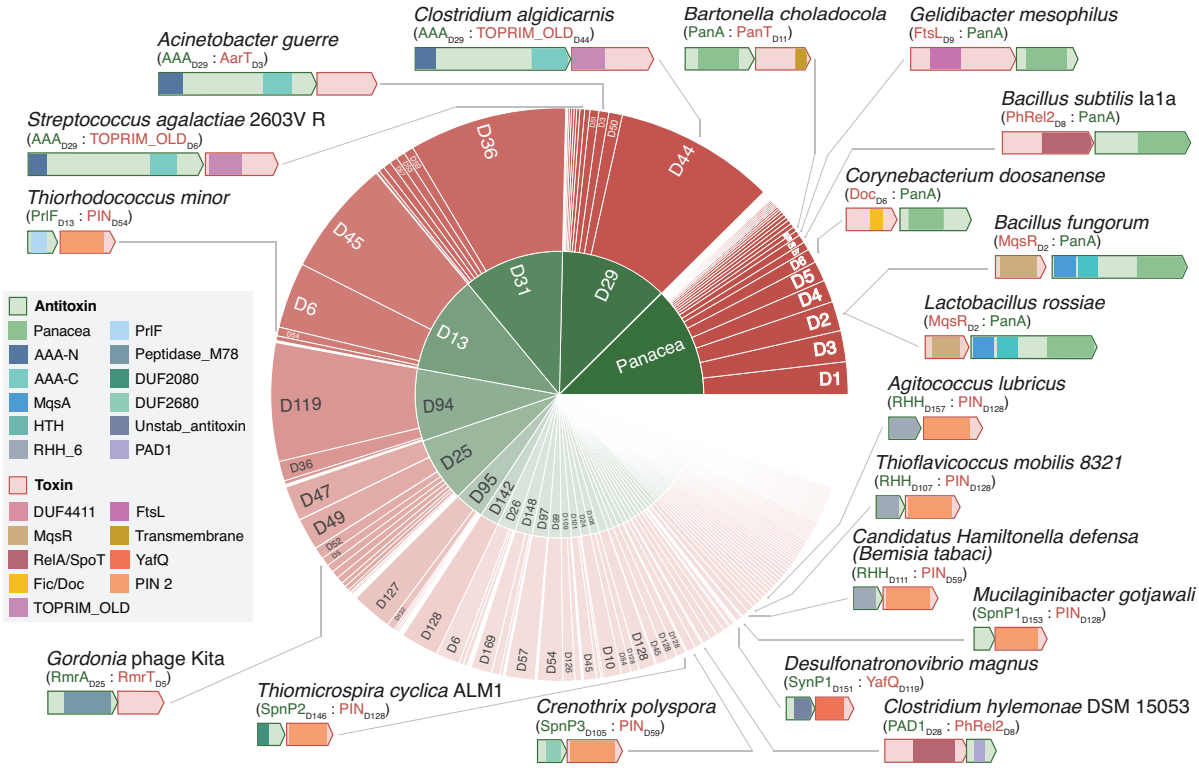

Figure 2. The diversity of systems in the core NetFlax TA network.

The size of the sections represents the number of proteins in each cluster (node) of **Fig. 1**. The block arrows represent open reading frames, drawn to scale. Coloured boxes within the arrows indicate domains, coloured according to the legend in the lower right.

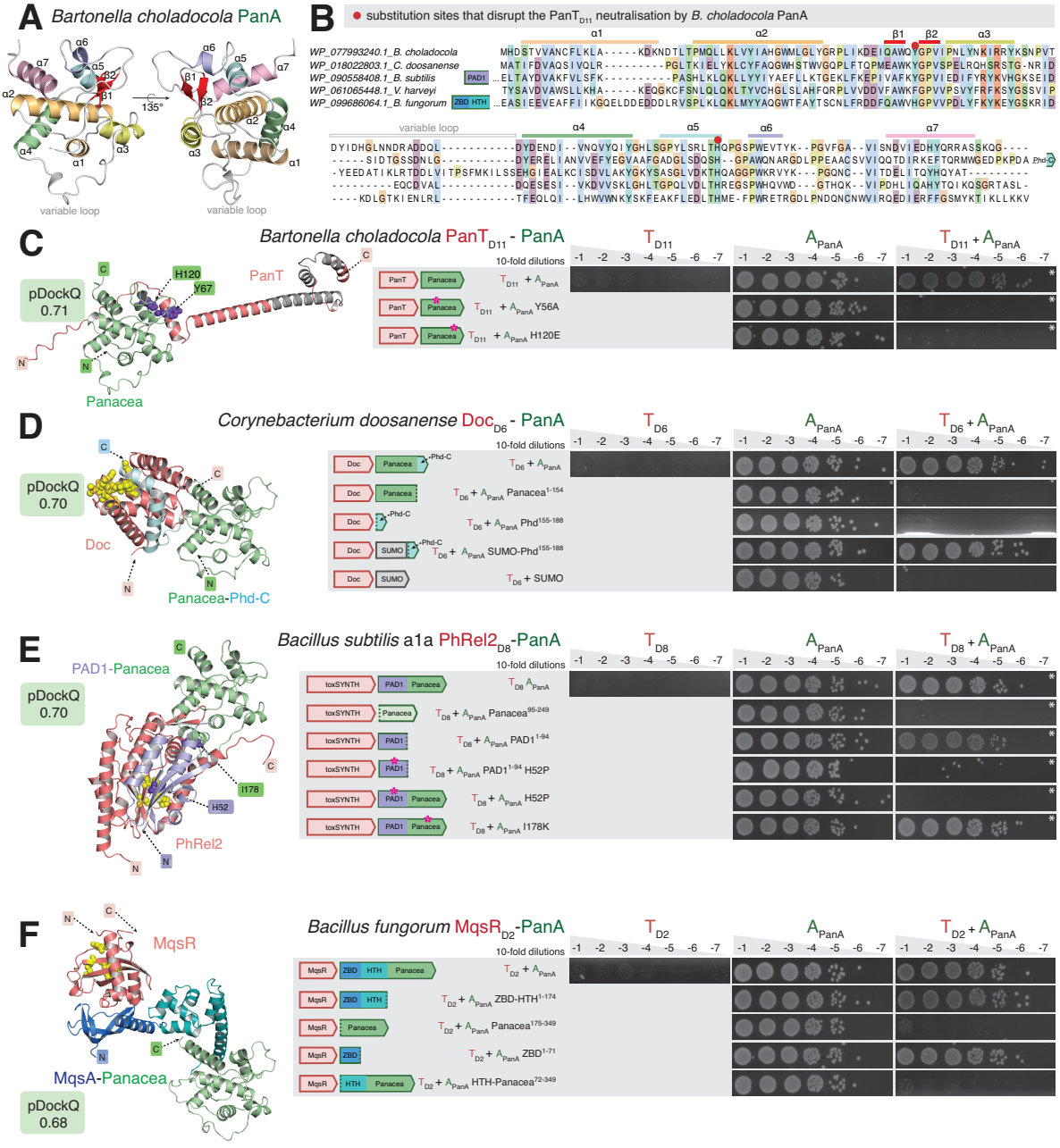

Figure 3. PanA antitoxins employ neutralise PanT toxins either via Panacea domain directly or via additional N-terminal domains: Phd-C, PAD1 and MqsA.

(**A**) AlphaFold-generated structural model of *Bartonella choladocola* PanA. Helixes and beta-sheets are labelled as per Kurata and colleagues [25]. (**B**) Alignment of Panacea domains from representative PanA antitoxins. (**C**-**D**) Mutational probing of AlphaFold-generated PanTA structural models. In toxicity neutralisation 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 *panT* induction and 0.05 or 0.5 (*) mM IPTG for *panA* induction). Predicted transmembrane domains are shown in grey and the active center of the toxin is highlighted in yellow. Introduced mutations on PanA antitoxin are shown in purple. Protein accessions are in **Table S1.**

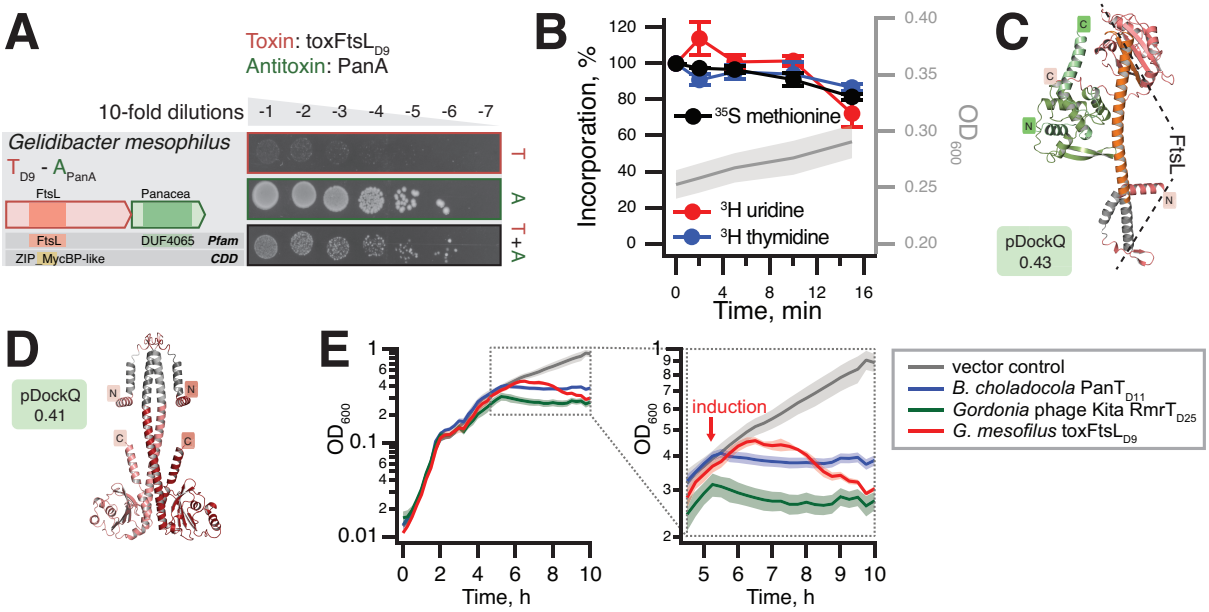

Figure 4. *G. mesophilus* toxFtsL is a slow-acting PanT toxin with partial homology with the FtsL component of bacterial dividosome.

(**A**) Validation of the *G. mesophilus* toxFtsL:PanA TA through toxicity neutralisation assay. (**B**) Metabolic labelling assays with wild-type *E. coli* BW25113 expressing *G. mesophilus* toxFtsL. (**C**) AlphaFold-generated PanTA structural model of *G. mesophilus* toxFtsL:PanA TA pair and (**D**) toxFtsL$_{D9}$ dimer model. Predicted transmembrane helical regions are shown in grey, and α-helical FtsL-like region is highlighted with a dotted line. (**E**) Delayed growth inhibition and cell lysis by *G. mesophilus* toxFtsL. Growth assays of *E. coli* cells expressing *G. mesophilus* toxFtsL, *B. cholaadocola* PanT$_{D11}$ or *Gordonia* phage Kita RmrT$_{D25}$ as well as a vector control strain harbouring 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.

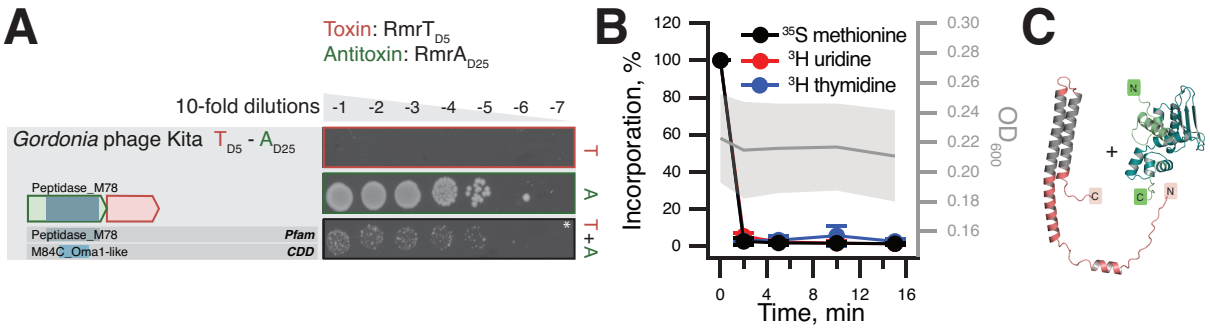

Figure 5. RosmerTA system from *Gordonia* phage Kita.

Domain organisation and TA validation though toxicity neutralisation assays (**A**), metabolic labelling assays with toxins expressed in wild-type *E. coli* BW25113 (**B**) and AlphaFold-generated structural models (**C**) for from $RmrT_{D5}$ and$RmrA_{D25}$ TA from Gordonia phage Kita. Predicted transmembrane regions of the toxin are shown in grey. Liquid culture experiments with $RmrT_{D5}$ are shown on **Fig. 4E**.

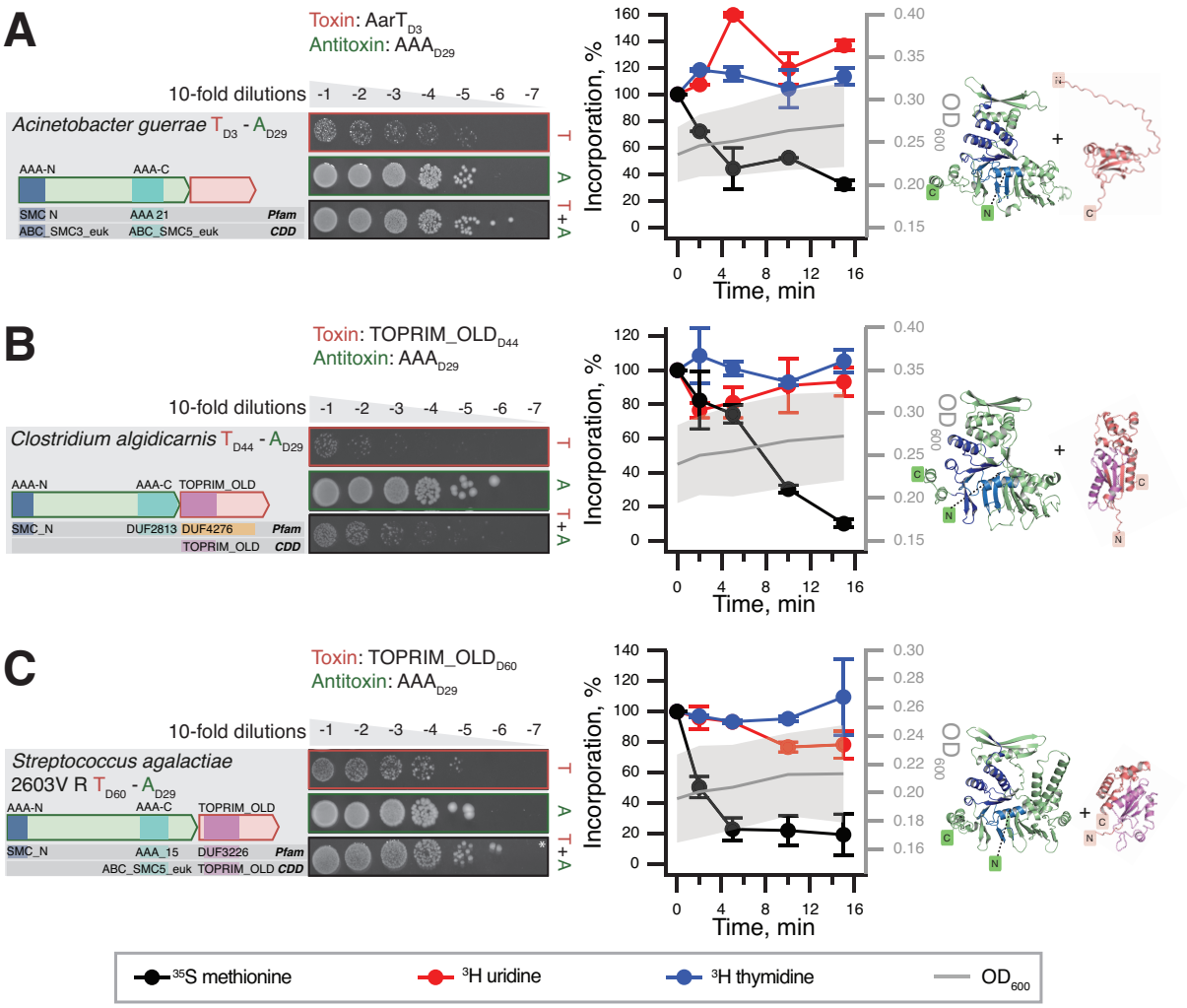

Figure 6. AAA-neutralised putative Abi phage defence systems.

Domain organisation and TA validation through toxicity neutralisation assays (*left*), metabolic labelling assays with toxins expressed in wild-type *E. coli* BW25113 (*center*) and AlphaFold-generated structural models (*right*) for AAA_{D29}-neutralised 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}.

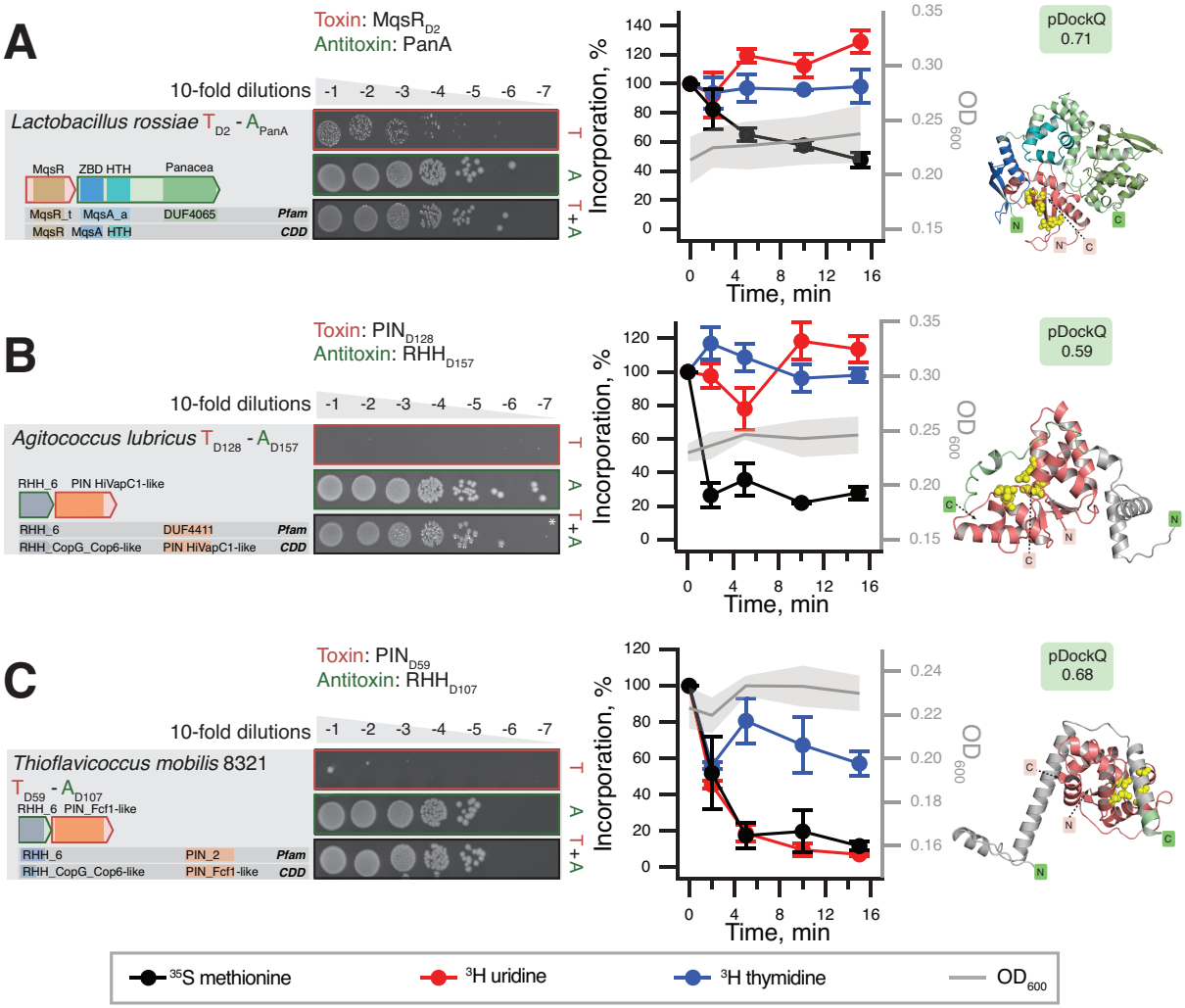

Figure 7. Representative NetFlax TA systems with nuclease effectors: MqsR$_{D2}$, PIN$_{D128}$ and PIN$_{D59}$.

Domain organisation and TA validation though toxicity neutralisation assays (*left*), metabolic labelling 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. mobilus* 8321 PIM$_{D59}$: RHH$_{D107}$. The active canter of various nucleases in highlighted with yellow spheres on AlphaFold structures.

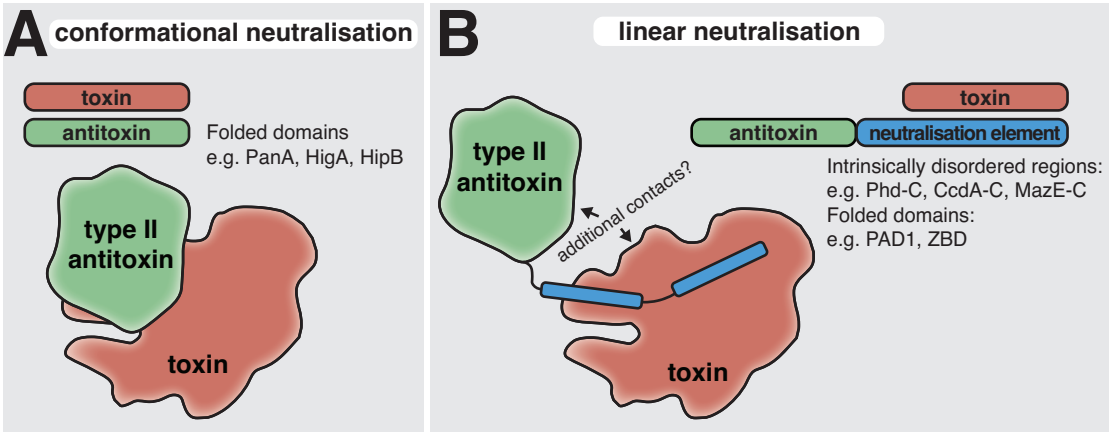

Figure 8. Conformational and linear neutralisation are two primary modes of toxin inactivation in Type II TA systems.

(**A**) Conformational neutralisation 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, with the domains themselves sufficing to neutralise multiple toxins. (**B**) Linear neutralisation displayed by modular antitoxins characterised by the neutralisation of toxins by exchangeable elements that can be intrinsically disordered or fully folded domains. The hallmark of this neutralisation mode is that these neutralising elements are linearly attached to antitoxin domains at the N- or C-terminal ends. The simple permutation of these neutralising elements allows the same antitoxin to neutralise different types of toxins as seen in the case of Panacea linked to Phd_C, ZBD or PAD1, which connects Panacea with the neutralisation of Doc, MqsR and toxSAS toxins. Hyperpromiscuous antitoxins are involved in both types of neutralisation.

References

1. Harms, A., et al., *Toxins, Targets, and Triggers: An Overview of Toxin-Antitoxin Biology.* Mol Cell, 2018. **70**(5): p. 768-784.
2. Jurenas, D., et al., *Biology and evolution of bacterial toxin-antitoxin systems.* Nat Rev Microbiol, 2022.
3. Choi, J.S., et al., *The small RNA, SdsR, acts as a novel type of toxin in Escherichia coli.* RNA Biol, 2018. **15**(10): p. 1319-1335.
4. Li, M., et al., *Toxin-antitoxin RNA pairs safeguard CRISPR-Cas systems.* Science, 2021. **372**(6541).
5. Ogura, T. and S. Hiraga, *Mini-F plasmid genes that couple host cell division to plasmid proliferation.* Proc Natl Acad Sci U S A, 1983. **80**(15): p. 4784-8.
6. Gotfredsen, M. and K. Gerdes, *The Escherichia coli relBE genes belong to a new toxin-antitoxin gene family.* Mol Microbiol, 1998. **29**(4): p. 1065-76.
7. Aizenman, E., H. Engelberg-Kulka, and G. Glaser, *An Escherichia coli chromosomal "addiction module" regulated by guanosine 3',5'-bispyrophosphate: a model for programmed bacterial cell death.* Proc Natl Acad Sci U S A, 1996. **93**(12): p. 6059-63.
8. Magnuson, R., et al., *Autoregulation of the plasmid addiction operon of bacteriophage P1.* J Biol Chem, 1996. **271**(31): p. 18705-10.
9. Ruiz-Echevarria, M.J., M.A. de la Torre, and R. Diaz-Orejas, *A mutation that decreases the efficiency of plasmid R1 replication leads to the activation of parD, a killer stability system of the plasmid.* FEMS Microbiol Lett, 1995. **130**(2-3): p. 129-35.
10. Gerdes, K., J.E. Larsen, and S. Mølin, *Stable inheritance of plasmid R1 requires two different loci.* J Bacteriol, 1985. **161**(1): p. 292-8.
11. Sberro, H., et al., *Discovery of functional toxin/antitoxin systems in bacteria by shotgun cloning.* Mol Cell, 2013. **50**(1): p. 136-48.
12. Vassallo, C.N., et al., *A functional selection reveals previously undetected anti-phage defence systems in the E. coli pangenome.* Nat Microbiol, 2022.
13. Millman, A., et al., *An expanding arsenal of immune systems that protect bacteria from phages.* bioRxiv, 2022: p. 2022.05.11.491447.
14. Fillol-Salom, A., et al., *Bacteriophages benefit from mobilizing pathogenicity islands encoding immune systems against competitors.* Cell, 2022. **185**(17): p. 3248-3262 e20.
15. Pandey, D.P. and K. Gerdes, *Toxin-antitoxin loci are highly abundant in free-living but lost from host-associated prokaryotes.* Nucleic Acids Res, 2005. **33**(3): p. 966-76.
16. Fozo, E.M., et al., *Abundance of type I toxin-antitoxin systems in bacteria: searches for new candidates and discovery of novel families.* Nucleic Acids Res, 2010. **38**(11): p. 3743-59.
17. Blower, T.R., et al., *Identification and classification of bacterial Type III toxin-antitoxin systems encoded in chromosomal and plasmid genomes.* Nucleic Acids Res, 2012. **40**(13): p. 6158-73.
18. Akarsu, H., et al., *TASmania: A bacterial Toxin-Antitoxin Systems database.* PLoS Comput Biol, 2019. **15**(4): p. e1006946.
19. Jimmy, S., et al., *A widespread toxin-antitoxin system exploiting growth control via alarmone signaling.* Proc Natl Acad Sci U S A, 2020. **117**(19): p. 10500-10510.
20. Horesh, G., et al., *SLING: a tool to search for linked genes in bacterial datasets.* Nucleic Acids Res, 2018. **46**(21): p. e128.
21. Makarova, K.S., et al., *Defense islands in bacterial and archaeal genomes and prediction of novel defense systems.* J Bacteriol, 2011. **193**(21): p. 6039-56.
22. Anantharaman, V. and L. Aravind, *New connections in the prokaryotic toxin-antitoxin network: relationship with the eukaryotic nonsense-mediated RNA decay system.* Genome Biol, 2003. **4**(12): p. R81.
23. Tamman, H., et al., *Structure of SpoT reveals evolutionary tuning of catalysis via conformational constraint.* Nat Chem Biol, 2023. **19**(3): p. 334-345.
24. Saha, C.K., et al., *FlaGs and webFlaGs: discovering novel biology through the analysis of gene neighbourhood conservation.* Bioinformatics, 2021. **37**(9): p. 1312-1314.
25. Kurata, T., et al., *A hyperpromiscuous antitoxin protein domain for the neutralization of diverse toxin domains.* Proc Natl Acad Sci U S A, 2022. **119**(6).
26. Yamaguchi, Y., J.H. Park, and M. Inouye, *Toxin-antitoxin systems in bacteria and archaea.* Annu Rev Genet, 2011. **45**: p. 61-79.

27. Hallez, R., et al., *New toxins homologous to ParE belonging to three-component toxin-antitoxin systems in Escherichia coli O157:H7.* Mol Microbiol, 2010. **76**(3): p. 719-32.
28. Jurėnas, D., et al., *Biology and evolution of bacterial toxin–antitoxin systems.* Nature Reviews Microbiology, 2022.
29. Iqbal, N., et al., *Comprehensive Functional Analysis of the 18 Vibrio cholerae N16961 Toxin-Antitoxin Systems Substantiates Their Role in Stabilizing the Superintegron.* J Bacteriol, 2015. **197**(13): p. 2150-9.
30. Christensen, S.K., et al., *Toxin-antitoxin loci as stress-response-elements: ChpAK/MazF and ChpBK cleave translated RNAs and are counteracted by tmRNA.* J Mol Biol, 2003. **332**(4): p. 809-19.
31. Hazan, R., B. Sat, and H. Engelberg-Kulka, *Escherichia coli mazEF-mediated cell death is triggered by various stressful conditions.* J Bacteriol, 2004. **186**(11): p. 3663-9.
32. Otsuka, Y., H. Ueno, and T. Yonesaki, *Escherichia coli endoribonucleases involved in cleavage of bacteriophage T4 mRNAs.* J Bacteriol, 2003. **185**(3): p. 983-90.
33. Otsuka, Y. and T. Yonesaki, *Dmd of bacteriophage T4 functions as an antitoxin against Escherichia coli LsoA and RnlA toxins.* Mol Microbiol, 2012. **83**(4): p. 669-81.
34. Dy, R.L., et al., *A widespread bacteriophage abortive infection system functions through a Type IV toxin-antitoxin mechanism.* Nucleic Acids Res, 2014. **42**(7): p. 4590-605.
35. Rousset, F., et al., *Phages and their satellites encode hotspots of antiviral systems.* Cell Host Microbe, 2022. **30**(5): p. 740-753 e5.
36. LeRoux, M., et al., *The DarTG toxin-antitoxin system provides phage defence by ADP-ribosylating viral DNA.* Nat Microbiol, 2022.
37. LeRoux, M. and M.T. Laub, *Toxin-Antitoxin Systems as Phage Defense Elements.* Annu Rev Microbiol, 2022.
38. Guegler, C.K. and M.T. Laub, *Shutoff of host transcription triggers a toxin-antitoxin system to cleave phage RNA and abort infection.* Mol Cell, 2021. **81**(11): p. 2361-2373 e9.
39. Zhang, T., et al., *Direct activation of a bacterial innate immune system by a viral capsid protein.* Nature, 2022. **612**(7938): p. 132-140.
40. Leplae, R., et al., *Diversity of bacterial type II toxin-antitoxin systems: a comprehensive search and functional analysis of novel families.* Nucleic Acids Res, 2011. **39**(13): p. 5513-25.
41. Zhu, L., et al., *Noncognate Mycobacterium tuberculosis toxin-antitoxins can physically and functionally interact.* J Biol Chem, 2010. **285**(51): p. 39732-8.
42. Makarova, K.S., Y.I. Wolf, and E.V. Koonin, *Comprehensive comparative-genomic analysis of type 2 toxin-antitoxin systems and related mobile stress response systems in prokaryotes.* Biol Direct, 2009. **4**: p. 19.
43. Tesson, F., et al., *Systematic and quantitative view of the antiviral arsenal of prokaryotes.* Nat Commun, 2022. **13**(1): p. 2561.
44. Jumper, J., et al., *Highly accurate protein structure prediction with AlphaFold.* Nature, 2021. **596**(7873): p. 583-589.
45. Bryant, P., G. Pozzati, and A. Elofsson, *Improved prediction of protein-protein interactions using AlphaFold2.* bioRxiv, 2021: p. 2021.09.15.460468.
46. Deng, Y.M., C.Q. Liu, and N.W. Dunn, *Genetic organization and functional analysis of a novel phage abortive infection system, AbiL, from Lactococcus lactis.* J Biotechnol, 1999. **67**(2-3): p. 135-49.
47. Doron, S., et al., *Systematic discovery of antiphage defense systems in the microbial pangenome.* Science, 2018. **359**(6379).
48. Xie, Y., et al., *TADB 2.0: an updated database of bacterial type II toxin-antitoxin loci.* Nucleic Acids Res, 2018. **46**(D1): p. D749-D753.
49. Kurata, T., et al., *A hyperpromiscuous antitoxin protein domain for the neutralization of diverse toxin domains.* Proceedings of the National Academy of Sciences, 2022. **119**(6): p. e2102212119.
50. Cheng, R., et al., *A nucleotide-sensing endonuclease from the Gabija bacterial defense system.* Nucleic Acids Res, 2021. **49**(9): p. 5216-5229.
51. Miller, W.G., et al., *Diversity within the Campylobacter jejuni type I restriction-modification loci.* Microbiology (Reading), 2005. **151**(Pt 2): p. 337-351.
52. Zimmermann, L., et al., *A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core.* J Mol Biol, 2018. **430**(15): p. 2237-2243.
53. Lau, R.K., et al., *A conserved signaling pathway activates bacterial CBASS immune signaling in response to DNA damage.* EMBO J, 2022: p. e111540.

54. Bose, B. and A.D. Grossman, *Regulation of horizontal gene transfer in Bacillus subtilis by activation of a conserved site-specific protease.* J Bacteriol, 2011. **193**(1): p. 22-9.
55. Matelska, D., K. Steczkiewicz, and K. Ginalski, *Comprehensive classification of the PIN domain-like superfamily.* Nucleic Acids Res, 2017. **45**(12): p. 6995-7020.
56. Arcus, V.L., et al., *The PIN-domain ribonucleases and the prokaryotic VapBC toxin–antitoxin array.* Protein Engineering, Design and Selection, 2010. **24**(1-2): p. 33-40.
57. Hazan, R. and H. Engelberg-Kulka, *Escherichia coli mazEF-mediated cell death as a defense mechanism that inhibits the spread of phage P1.* Mol Genet Genomics, 2004. **272**(2): p. 227-34.
58. Cui, Y., et al., *Bacterial MazF/MazE toxin-antitoxin suppresses lytic propagation of arbitrium-containing phages.* Cell Rep, 2022. **41**(10): p. 111752.
59. Aakre, C.D., et al., *Evolving new protein-protein interaction specificity through promiscuous intermediates.* Cell, 2015. **163**(3): p. 594-606.
60. Brown, B.L., 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, 2009. **5**(12): p. e1000706.
61. Mariano, G. and T.R. Blower, *Conserved domains can be found across distinct phage defence systems.* Mol Microbiol, 2023.
62. Garcia-Pino, A., et al., *Doc of prophage P1 is inhibited by its antitoxin partner Phd through fold complementation.* J Biol Chem, 2008. **283**(45): p. 30821-7.
63. Castro-Roa, D., et al., *The Fic protein Doc uses an inverted substrate to phosphorylate and inactivate EF-Tu.* Nat Chem Biol, 2013. **9**(12): p. 811-7.
64. Brown, B.L., et al., *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, 2011. **286**(3): p. 2285-96.
65. Buddelmeijer, N. and 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, 2004. **52**(5): p. 1315-27.
66. Gao, Y., et al., *Molecular basis of RADAR anti-phage supramolecular assemblies.* Cell, 2023. **186**(5): p. 999-1012 e20.
67. Duncan-Lowey, B., et al., *Cryo-EM structure of the RADAR supramolecular anti-phage defense complex.* Cell, 2023. **186**(5): p. 987-998 e15.
68. Madl, T., et al., *Structural basis for nucleic acid and toxin recognition of the bacterial antitoxin CcdA.* J Mol Biol, 2006. **364**(2): p. 170-85.
69. Mattison, K., et al., *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, 2006. **281**(49): p. 37942-51.
70. Winther, K.S. and K. Gerdes, *Enteric virulence associated protein VapC inhibits translation by cleavage of initiator tRNA.* Proc Natl Acad Sci U S A, 2011. **108**(18): p. 7403-7.
71. Winther, K.S., et al., *VapC20 of Mycobacterium tuberculosis cleaves the sarcin-ricin loop of 23S rRNA.* Nat Commun, 2013. **4**: p. 2796.
72. Loris, R. and A. Garcia-Pino, *Disorder- and dynamics-based regulatory mechanisms in toxin-antitoxin modules.* Chem Rev, 2014. **114**(13): p. 6933-47.
73. Sterckx, Y.G., 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, 2012. **68**(Pt 6): p. 724-9.
74. Takagi, H., et al., *Crystal structure of archaeal toxin-antitoxin RelE-RelB complex with implications for toxin activity and antitoxin effects.* Nat Struct Mol Biol, 2005. **12**(4): p. 327-31.
75. Hadzi, S., et al., *Ribosome-dependent Vibrio cholerae mRNAse HigB2 is regulated by a beta-strand sliding mechanism.* Nucleic Acids Res, 2017. **45**(8): p. 4972-4983.
76. Schumacher, M.A., et al., *Molecular mechanisms of HipA-mediated multidrug tolerance and its neutralization by HipB.* Science, 2009. **323**(5912): p. 396-401.
77. Talavera, A., et al., *A dual role in regulation and toxicity for the disordered N-terminus of the toxin GraT.* Nat Commun, 2019. **10**(1): p. 972.
78. Bordes, P., et al., *SecB-like chaperone controls a toxin-antitoxin stress-responsive system in Mycobacterium tuberculosis.* Proc Natl Acad Sci U S A, 2011. **108**(20): p. 8438-43.
79. Yarbrough, M.L., et al., *AMPylation of Rho GTPases by Vibrio VopS disrupts effector binding and downstream signaling.* Science, 2009. **323**(5911): p. 269-72.
80. Kurata, T., et al., *RelA-SpoT Homolog toxins pyrophosphorylate the CCA end of tRNA to inhibit protein synthesis.* Mol Cell, 2021. **81**(15): p. 3160-3170 e9.

81. Fraikin, N., F. Goormaghtigh, and L. Van Melderen, *Type II Toxin-Antitoxin Systems: Evolution and Revolutions.* J Bacteriol, 2020. **202**(7).
82. Jankevicius, G., et al., *The Toxin-Antitoxin System DarTG Catalyzes Reversible ADP-Ribosylation of DNA.* Mol Cell, 2016. **64**(6): p. 1109-1116.
83. Songailiene, I., et al., *HEPN-MNT Toxin-Antitoxin System: The HEPN Ribonuclease Is Neutralized by OligoAMPylation.* Mol Cell, 2020. **80**(6): p. 955-970 e7.
84. Jorgensen, M.G., et al., *HicA of Escherichia coli defines a novel family of translation-independent mRNA interferases in bacteria and archaea.* J Bacteriol, 2009. **191**(4): p. 1191-9.
85. Black, D.S., et al., *Structure and organization of hip, an operon that affects lethality due to inhibition of peptidoglycan or DNA synthesis.* J Bacteriol, 1991. **173**(18): p. 5732-9.
86. O'Leary, N.A., et al., *Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation.* Nucleic Acids Res, 2016. **44**(D1): p. D733-45.
87. Marchler-Bauer, A., et al., *CDD: a Conserved Domain Database for the functional annotation of proteins.* Nucleic Acids Res, 2011. **39**(Database issue): p. D225-9.
88. Steinegger, M., et al., *HH-suite3 for fast remote homology detection and deep protein annotation.* BMC Bioinformatics, 2019. **20**(1): p. 473.
89. van Kempen, M., et al., *Foldseek: fast and accurate protein structure search.* bioRxiv, 2022: p. 2022.02.07.479398.
90. Shannon, P., et al., *Cytoscape: a software environment for integrated models of biomolecular interaction networks.* Genome Res, 2003. **13**(11): p. 2498-504.
91. Dong, R., et al., *mTM-align: an algorithm for fast and accurate multiple protein structure alignment.* Bioinformatics, 2018. **34**(10): p. 1719-1725.
92. Hallgren, J., et al., *DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks.* bioRxiv, 2022: p. 2022.04.08.487609.
93. Guzman, L.M., et al., *Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter.* J Bacteriol, 1995. **177**(14): p. 4121-30.
94. Jaskolska, M. and K. Gerdes, *CRP-dependent positive autoregulation and proteolytic degradation regulate competence activator Sxy of Escherichia coli.* Mol Microbiol, 2015. **95**(5): p. 833-45.