Computational and Structural Biotechnology Journal 20 (2022) 4517–4531

COMPUTATIONAL ANDSTRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

Genomic discovery and structural dissection of a novel type of polymorphic toxin system in gram-positive bacteria

Huan Li [a], Yongjun Tan [a], Dapeng Zhang [a,b,*]

[a] Department of Biology, College of Arts & Sciences, Saint Louis University, Saint Louis, MO 63103, USA
[b] Program of Bioinformatics and Computational Biology, College of Arts & Sciences, Saint Louis University, MO 63103, USA

ARTICLE INFO

Article history:
Received 12 May 2022
Received in revised form 15 August 2022
Accepted 15 August 2022
Available online 18 August 2022

Keywords:
Polymorphic toxins
S8 peptidases
Fold recognition
Function prediction
Gram-positive bacteria
Kin selection

ABSTRACT

Bacteria have developed several molecular conflict systems to facilitate kin recognition and non-kin competition to gain advantages in the acquisition of growth niches and of limited resources. One such example is a large class of so-called polymorphic toxin systems (PTSs), which comprise a variety of the toxin proteins secreted via T2SS, T5SS, T6SS, T7SS and many others. These systems are highly divergent in terms of sequence/structure, domain architecture, toxin-immunity association, and organization of the toxin loci, which makes it difficult to identify and characterize novel systems using traditional experimental and bioinformatic strategies. In recent years, we have been developing and utilizing unique genome-mining strategies and pipelines, based on the organizational principles of both domain architectures and genomic loci of PTSs, for an effective and comprehensive discovery of novel PTSs, dissection of their components, and prediction of their structures and functions. In this study, we present our systematic discovery of a new type of PTS (S8-PTS) in several gram-positive bacteria. We show that the S8-PTS contains three components: a peptidase of the S8 family (subtilases), a polymorphic toxin, and an immunity protein. We delineated the typical organization of these polymorphic toxins, in which a N-terminal signal peptide is followed by a potential receptor binding domain, BetaH, and one of 16 toxin domains. We classified each toxin domain by the distinct superfamily to which it belongs, identifying nine BECR ribonucleases, one Restriction Endonuclease, one HNH nuclease, two novel toxin domains homologous to the VOC enzymes, one toxin domain with the Frataxin-like fold, and several other unique toxin families such as Ntox33 and HicA. Accordingly, we identified 20 immunity families and classified them into different classes of folds. Further, we show that the S8-PTS-associated peptidases are analogous to many other processing peptidases found in T5SS, T7SS, T9SS, and many proprotein-processing peptidases, indicating that they function to release the toxin domains during secretion. The S8-PTSs are mostly found in animal and plant-associated bacteria, including many pathogens. We propose S8-PTSs will facilitate the competition of these bacteria with other microbes or contribute to the pathogen-host interactions.

1. Introduction

In the world of microbes, competition between these tiny organisms for resources is extremely fierce [1]. As a result, bacteria have developed a large repertoire of weaponries, collectively termed biological conflict systems (such as antibiotics/resistance systems [2], toxin-antitoxin systems [3], CRISPR-Cas systems [4], and polymorphic toxin systems [5]) against their rivals during the course of evolution. The discovery of polymorphic toxin systems (PTSs) a decade ago situates them as one of the most recently discovered biological conflict systems, and their characterization revealed that they are perhaps the most complicated and predominant conflict system in bacteria, making them an important subject of continued biological research [5–9]. The offensive abilities of PTSs rely on their various proteinaceous toxins. Remarkably, these toxin proteins display modular, but polymorphic domain architectures, including N-terminal regions, central repeats/linkers, and C-terminal domains [5]. The N-terminal regions of PTs often serve as recognition modules for specific secretion pathways to export these toxins, including the signal peptides for Type II Secretion System (T2SS) [10], ESPR and TpsA-SD module for T5SS [7,11], PAAR, HCP1 and VgrG modules for T6SS [5,12], and WXG, LXG and LDXD modules for T7SS [7,13,14]. There are also several

* Corresponding author at: Department of Biology, College of Arts & Sciences, Saint Louis University, Saint Louis, MO 63103, USA.

E-mail address: dapeng.zhang@slu.edu (D. Zhang).

https://doi.org/10.1016/j.csbj.2022.08.036
2001-0370/© 2022 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology.

specific N-terminal domains that are linked to other atypical secretion modes, such as MuF [15,16], PrsW [17], MafB [5,18] and TANDOR [19]. The C-terminal regions of PTs usually contain the toxin/effector modules that target the rival bacteria or other organisms. Comprehensive bioinformatics analysis of these toxin/effector domains has revealed an extraordinary catalytic spectrum, including nucleases, deaminases, peptidases, lipid and carbohydrate modifying domains, ADP-ribosyltransferases (ARTs), nucleotidyltransferases, glycosyltransferase, and many others [5–7,20].

Besides the modular toxins, another hallmark of PTSs is the presence of immunity proteins whose encoding genes are typically next to the toxin genes on genomic loci [5–7]. The immunity proteins are usually intracellular, single-domain proteins which can neutralize the cognate toxins by directly binding with them. When the toxins enter the surrounding bacteria, the kin bacteria encoding the cognate immunity protein remain unaffected while the non-kin bacteria, which do not carry the same cognate immunity protein will be destroyed [5,7,21], leading to kin selection or self/non-self recognition. Furthermore, several studies have demonstrated that the PTSs typically undergo rampant diversifications through both domain shuffling of toxin proteins and rapid remodeling of genomic loci [5–7,22]. As a result, even between closely related strains, the toxin modules, as well as the cognate immunity proteins, may belong to distinct families and display unrelated activities [5–9], despite other modules of toxin proteins remaining conserved.

Recent years have witnessed a great expansion of research on many molecular aspects of PTSs. This collective research has led to the identification of numerous toxins and their cognate immunity proteins [23–25], determination of their structures [26–29], characterizations of enzymatic activities [18,30–37], dissection of details of secretion and regulation [38–40], discovery of distinct secretion modes [41,42], as well as the demonstration of the profound roles of PTSs in kin selection and interbacterial competition [41,43], pathogenicity [44,45] and biofilm formation [46]. To be noted, many of these studies typically focus on a unique toxin system and its specific components, between which little homology is observed. This diversity serves as a testimony to the complexity within the PTS universe and indicates the presence of many more novel systems in nature that have yet to be discovered, reinforcing the need for a systematic discovery of PTSs. However, the traditional discovery process using experimental strategies is usually time-consuming. While sequence-based computational approaches can be useful in genome mining, this strategy often fails to identify novel toxin systems based on the existing PTS proteins, largely due to the high diversity among the PT systems (in terms of sequence, structure, and organization of genomic loci). This is especially true for many bacterial pathogens that have developed their own specific mechanisms of pathogenicity. To address the limitations in traditional toxin research, in recent years, we have been developing novel computational strategies, based on the organizational principles in both domain architectures and genomic loci of PTSs, leading to an effective and comprehensive discovery of novel PTSs, dissection of their components, and prediction of their structures and functions. We are confident that effective computational mining of novel PTSs will continue to accelerate the research in toxin-related studies, as demonstrated by several recent works [5–7,16,19,42,47].

In this study, we have discovered a novel PT system, S8-PTS (S8-peptidase associated Polymorphic Toxin System) by using our bioinformatic pipelines. The S8-PTS is featured by the presence of a single-domain S8 peptidase gene on the toxin loci, followed by one pair of polymorphic toxin and immunity genes. All the toxins in this system have a typical N-terminal domain which we refer to as BetaH. We extensively identified the toxin modules, which were classified into 16 distinct families, and the associated immunity proteins, which are found to belong to 20 distinct families. We also unified the majority of toxin and immunity domains in this system into superfamilies based on their shared structures/folds, conserved motifs, and similar configuration of the active sites. We show that the S8-PTS-associated peptidases are analogous to many other processing peptidases found in T5SS, T7SS, T9SS, and many proprotein-processing peptidases, indicating that they function to release the toxin domains during secretion. Further, we found that S8-PTS is only present in a clade of firmicutes (Bacillales) and several actinobacteria species, most of which belong to animal and plant-associated bacteria, including many pathogens. We predict that S8-PTS might facilitate the competition of these bacteria with other microbes or contribute to the pathogen-host interactions.

2. Materials and methods

2.1. Improved protein domain-centric strategy

Protein sequences that can be unified using sequence similarity-based methods, such as BLAST [48], PSIBLAST [49] and HMMER [50], are usually classified as a protein family. Historically, at the earlier stage of protein classification, the concept of family was applied to full length proteins. However, this is no longer valid since the discovery of modularity of the multidomain (or mosaic) proteins [51]. Domains are basic structural and functional units of proteins and may have separate evolutionary histories even inside the same protein [52]. This is especially true for proteins in PTSs which have extensive domain shuffling features. Hence, all the analyses in this study use domains instead of the full-length proteins. However, defining the boundary of domains using traditional methods is not always accurate as it typically refers to the overall conservation of the full-length sequence. Therefore, in this project, we will combine both the sequence conservation and structural information provided by AlphaFold2 [53] to help us dissect the domains more accurately and effectively (see below). Another bottleneck of domain analysis is the identification and classification of domain families. Here we will utilize CLANS [54] network analysis and structure-based similarity comparison (see below) to solve this problem.

2.2. Homologous sequence search and analysis

In order to retrieve homologous sequences, iterative sequence-profile searches were run for each domain family until convergence via the Position-Specific Iterative Basic Local Alignment Search Tool (PSI-BLAST) [49] against the nonredundant protein (nr) database of the National Center for Biotechnology Information (NCBI), with a cut-off e-value of 0.001. For peptidase_S8, since it is a very large family, we ran PSI-BLAST searches against a custom prokaryotic database (11067 genomes with 40,046,196 protein sequences) built by selecting representative genomes of each species from the NCBI FTP site (ftp.ncbi.nih.gov/genomes/genbank/bacteria/and ftp.ncbi.nih.gov/genomes/genbank/archaea/) on May 2021. The set of homologous sequences for each domain family were used for downstream analyses, including taxonomic distribution and Multiple Sequence Alignment (MSA). For the former, the taxonomy information of the bacteria that contain the BetaH toxins was extracted from NCBI GenPept files and summarized based on phylum and class/order. We also classified the bacteria based on the information of their isolation source and host, which were extracted from the associated GenBank files (genome files) by using Entrez Direct [55] and a custom script. For the latter, the pairwise-based BLASTCLUST program was first used to select representative sequences for each family by removing relatively similar sequences. Then, representative sequences were used to generate MSAs by the KALIGN [56], MUSCLE [57] or PROMALS3D [58] alignment softwares. The MSAs of the families that were discovered in this study are available in Supplementary File S1. The consensus pattern and conserved residues of alignments were generated by a custom Perl script based on the classification of amino acids [59]. We used the CHROMA [60] tool for visualizing MSAs which were further edited via Microsoft Word.

2.3. Guilt by association—Domain architecture analysis

It is intuitive that domains inside the same protein cooperate with each other for a common functional aim. This helps us to infer functions of both the protein and its domains, especially given the conserved "secretion-toxin" domain architecture for toxin proteins from PTSs. In our study, the domain architectures for all proteins will be annotated via the HMMSCAN [50] program against domain profiles from both PFAM [61] and our custom PTS profile database. The regions that lack the annotations of existing domains will be further collected and studied to develop new domains and profiles. Other regions including signal peptide and transmembrane regions will be detected by the Phobius program [62].

2.4. Guilt by association—Gene neighborhood analysis

Polycistronic mRNA, a single copy of mRNA that can encode multiple proteins, is the hallmark of prokaryotic genomes, while the mRNAs of eukaryotic genomes are mostly monocistronic, meaning one mRNA encodes one protein [63]. The genes of polycistronic mRNA are organized under the same one promoter and the whole genomic locus is termed an operon [64]. The operon is the key concept of our gene neighborhood analysis since genes inside the same operon are generally functionally linked [65]. Specifically, we ran gene neighborhood analysis for BetaH-containing proteins by collecting their upstream and downstream genes (five genes for both sides of query) using the homologous sequences retrieved via PSI-BLAST as the queries. Then, both the BetaH proteins and proteins encoded by neighbor genes were clustered based on sequence similarities via the BLASTCLUST program (https://ftp.ncbi.nih.gov/blast/documents/blastclust.html). Each cluster and its respective proteins were then annotated by their domain architectures via the HMMSCAN program described above. The genes demonstrating a conserved association are determined by both the frequency and taxonomic distribution of their co-occurrence with BetaH genes. Specifically, for the taxonomic distribution, we look for associations conserved across at least two bacteria phyla or complete conservation in a single bacteria phylum.

2.5. Sequence network analysis

The representative sequences from multiple families were combined to run the CLAN program [54]. CLAN first conducts all-against-all BLASTP comparisons and generates a two-dimensional graph in which nodes represent sequences while edges represent detected pairwise similarities between sequences. Then, the layout of the graph is rearranged by the Fruchterman and Reingold force-directed graph drawing algorithm [66]. In short, the sequences with more intra-similarities will be clustered as a colony to represent a family. The final graph helped us to ensure that the families collected by PSIBLAST were not repeated collections for the same family, as unique families would separate into different colonies; this is especially important for identifying novel domains. However, we considered colonies close to each other as the same family if their sequences were collected via the same one PSIBLAST.

2.6. Protein structure modeling and analysis

The structural models for all selected proteins were predicted by AlphaFold2 [53,67]. AlphaFold2 is a recently published deep learning tool that can return accurate protein structures from just an amino acid sequence. All the predictions were based on the full length of the proteins, and boundaries of each domain were determined by both the structural models and PAE matrices provided by AlphaFold2. The pdb files of all the modeled structures are provided in Supplementary File S2. Next, to obtain structural homologs for understanding each domain, the domain structures were used as inputs for the DaliLite program [68] which searches against the PDB database [69]. The method of DaliLite we used in this study is a local structural comparison algorithm to search similar structures. This structural comparison strategy stems from the fact that structure is more conserved than sequence over the course of evolution [70]. Therefore, highly diverse homologous sequences that are undetectable by sequence- or profile-based methods can be retrieved by structural similarity, combined with the conserved motifs or active sites. The DaliLite search results can be found in Supplementary File S3. Further analyses and visualization of structures were conducted by PyMOL [71].

2.7. Phylogenetic analysis

For the phylogenetic analysis of peptidase_S8 family, we selected representative sequences filtered by the BLASTCLUST program and then added our sequences of interest. The added sequences were determined by their specific species, domain architectures or gene neighborhoods. Other sequences were added if they were necessary for the robustness of the tree. After constructing the MSA for the selected sequences, the phylogenetic calculations were conducted by the FastTree program with default parameters [72]. The final trees were visualized by MEGA7 program [73].

3. Results and discussion

3.1. Identification of a novel N-terminal domain of polymorphic toxins

Over the course of studying the domain architectures of previously identified toxin proteins [5], we found that several different C-terminal toxin domains, such as ColD, ColE3, Ntox21 and Ntox35, were coupled with the same, unknown module in their N-termini (Fig. 1A). Iterative sequence searches using PSIBLAST with this region (named the BetaH domain, explained later) as a query further retrieved ~700 homologs in the NCBI NR database. A global multiple sequence alignment of these homologs revealed that, despite the conservation of the N-terminal BetaH region, the C-terminal regions of these toxin domains displayed a strong variation, suggesting that they contain different domain families (Fig. 1A). Further, we found that almost all BetaH homologs contain a signal peptide at the N-terminus, suggesting that they are secreted out of the membrane through the Sec pathway. These features, including the secretion signal, C-terminal domain variation, and presence of several known toxin modules suggest that BetaH represents a characteristic domain of a new type of polymorphic toxins.

The BetaH domains are typically around 120 amino acids in length. Multiple sequence alignment (Fig. 1A and Supplementary File S1) revealed that no position and residue is universally conserved across the family. However, this domain is featured by con-

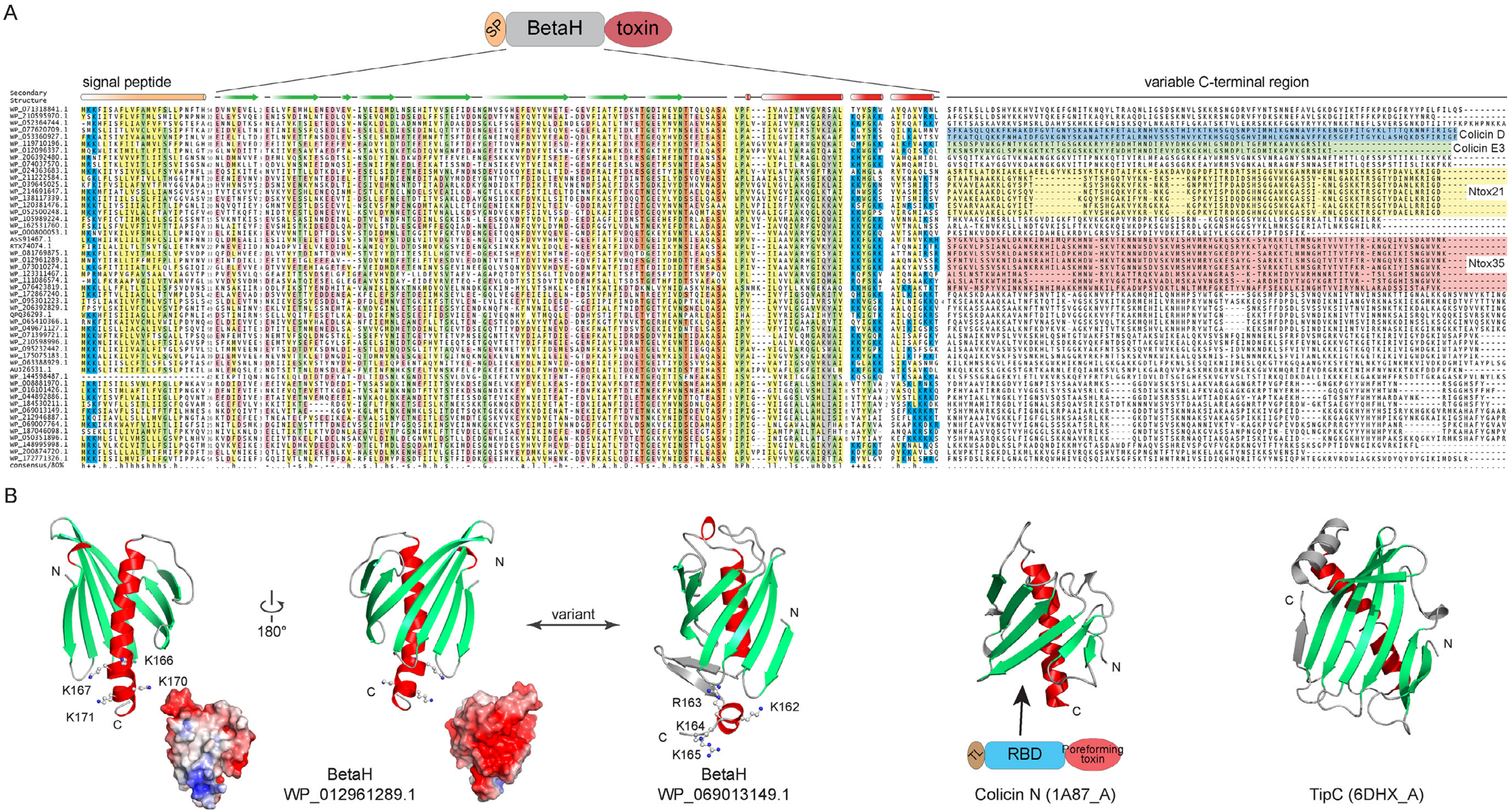


Fig. 1. (A) Multiple sequence alignment (MSA) of the BetaH-containing toxins. The general domain architecture of the BetaH toxins and a representative secondary structure excluding the C-terminal domains are shown above the MSA, while the consensus of amino acids at each column is shown below the MSA, where ‘h’ stands for hydrophobic residues highlighted in light yellow background, ‘l’ for aliphatic residues in light yellow, ‘a’ for aromatic residues in light yellow, ‘+’ for positive residues in blue, ‘-‘ for negative residues in light pink, ‘s’ for small residues in light green, ‘u’ for tiny residues in light green, ‘b’ for big residues in grey, and ‘o’ for alcohol residues in orange. In addition, the positively charged and negatively charged residues are highlighted in light purple and blue backgrounds, respectively. The C-terminal regions that contain known toxin domains are highlighted in different background. **(B)** The ribbon representation of the BetaH domains as compared to the receptor-binding domain (RBD) of the Colicin N (1A87_A) and the TipC immunity protein (6DHX_A). One representative of the BetaH domain is shown in front view and back view, as well as their electrostatic potential surface view, whereas the other variant only presents the back view to show the situation of presence of a β-hairpin insert in the ending α helix. α-helices are shown in red, β-sheets in green, and loops in grey. The basic residues located at the C-terminus of the BetaH domain are shown as sticks. Surface diagrams are colored by electrostatic potential as calculated by APBS in the PyMol program with positively charged regions (>5 mV) colored in blue and negatively charged regions (<–5 mV) in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

taining many acidic residues. Further, a peptidase cleavage motif, which is featured by dibasic residues, can be found at the C-terminus of the domains. A structural model predicted by AlphaFold2 showed that the BetaH domain adapts a globular structure comprised of a seven-stranded sheet that wraps around a C-terminal alpha helix (Fig. 1B). Interestingly, structural comparisons (Fig. 1B) revealed that BetaH shares a striking topological similarity with two structures, the N-terminal receptor-binding domain of the toxin Colicin N (ColN-RBD) [74], and an immunity family TipC found in T7SS systems [75]. Both ColN-RBD and TipC utilize the surface of the β sheet to bind protein components such as the target cell receptor [76] and the coupled toxin protein [75]. Similarly, BetaH is featured by a negatively charged surface on the β sheet, formed by the above-described acidic residues (Fig. 1B). Additionally, BetaH-containing proteins share a comparable domain organization as ColN; both are led by the secretion or translocation (TL) signal, followed by BetaH or RBD, and C-terminal toxins. Thus, both the structural and locational similarities suggest that BetaH might function as ColN-RBD in binding to the membrane receptor of the target cell. Given its structural feature, we named the domain BetaH (β sheet and α helix).

In order to comprehensively dissect the toxin systems that are associated with this unique BetaH domain, we next conducted a series of protein sequence, structure and genome analyses.

3.2. The BetaH domain is coupled with a variety of toxin domains

First, to identify the potential C-terminal domains of the BetaH-containing proteins, we extracted their C-terminal regions and conducted several PSIBLAST searches to collect their homologs. After removing the highly similar sequences, we used a CLAN network analysis to identify the major groups of the homologous sequences, each of which correspond to a domain family (Fig. 2A). To study the function of each domain family, we selected one representative and modelled its structure using AlphaFold2. The distant structure homologs were identified using the DaliLite program. We also identified the evolutionarily conserved residues for each domain and mapped them to the structure model to understand their potential catalytic activity (Supplementary File S1, Fig. 2B–F and Fig. 3). Altogether, we identified 16 major toxin domains among these proteins. Below we specifically describe the function and features of these toxin domains.

Nine ribonuclease toxins of the BECR (Barnase-EndoU-Colicin E5/D-RelE) fold

Among the BetaH-coupled toxin domains, nine belong to the BECR ribonuclease fold (Fig. 2B). BECR ribonucleases are metal-independent RNases that are usually found in PTSs and toxin-antitoxin systems [5]. They share a common structural core with one alpha-helix followed by a four stranded antiparallel β-sheet (αββββ). Based on the DaliLite searches, the newly identified BECR families display significant structural similarity with several known BECR families (Supplementary File S3). Specifically, Ntox21-like and ColD-like families both show significant sequence similarity and structural similarity with Ntox21 and Colicin D, respectively (Fig. 2A and B). In terms of catalytic mechanisms, known BECR RNases appear to utilize distinct catalytic configurations to fulfill the general acid-base catalysis in cutting RNA molecules [77]. For example, the typical BECR family, Barnase [78], utilizes one conserved negative residue (Glu) as a general base to activate the 2′-OH of RNA which then acts as a nucleophile to attack the adjacent 3′-phosphate, one conserved positive residue (His) as a general acid to stabilize the leaving 5′-OH, and several positive residues to stabilize the penta-covalent transition state [77]. Many other BECR families, e.g. the XendoU family, have active sites utilizing two conserved histidine residues since histidine can function as both a general acid and general base [79]. Aravind *et.al.* also observed that many BECR RNases, such as Colicin D, contain a new conserved alcoholic residue (S/T) in the fourth strand of the BECR core, in addition to two conserved basic residues [5]. Further, for the RelE toxin, a ribosome-dependent RNase that were mostly found in the TAs, three highly conserved residues, two arginine and a tyrosine, are used to configure the catalytic site [80]. For the 9 domains belonging to the BECR superfamily that were identified in the BetaH proteins, we predicted their active site configuration based on evolutionary conservation of residues and clustering of their structural location. The active sites appeared to display unique but comparable configurations, such as those found in Ntox35, S8-BECR1 and ColD-like commonly involving two highly conserved histidine residues (Fig. 2B). Hence, we proposed that these novel toxin families will function as RNases in S8-PTs to fight against rival bacteria.

HicA RNase

HicA is the other RNase domain found in this system (Fig. 2C). This domain adopts a αβββα topology of the double-stranded RNA-binding domain (dsRBD) fold and has a conserved histidine as the catalytic residue [81,82]. HicA is known for its primary presence in the TA systems, together with the HicB antitoxin [83]. However, we found the HicA domain is present at the C-termini of several BetaH-containing toxins and an RHS-type polymorphic toxin (i.e. RXZ81544.1). Sequence searches revealed that these HicA domains are further closely related to many HicA proteins belonging to the TA systems (i.e. approximately 52% identity to WP_173297211.1 and 50% identity to WP_139680861.1). Hence, it is possible that HicA toxin domains were recently acquired by the S8-PTS. This is also an example of component exchanges between PTSs and Toxin-Antitoxin systems.

DNases that belong to the HNH fold and the PD-(D/E)XK (restriction endonuclease) fold

We identified two toxin domains, belonging to the HNH and PD-(D/E)XK superfamilies, respectively. Both HNH and PD-(D/E)XK superfamilies are metal-dependent DNases. In our earlier studies, we identified several distinct versions of HNH nucleases in PT systems [5,7,84] which all adapt a typical treble-clef fold as a distinct group of zinc fingers [85], but display extensive variations in both structure and construction of catalytic residues. The XHH domain [24] is found in several BetaH toxins (Fig. 2D) and demonstrates the typical HNH fold (a β-hairpin followed by a C-terminal helix), embedded in several additional helices. Its catalytic site is configured by several residues (Fig. 2D), including a conserved HH dyad at the end of first strand of the fold, where the first histidine functions as one of the metal-chelating ligands and the other histidine is a general base to activate a water molecule as a nucleophile to hydrolyze a phosphoester.

The PD-(D/E)XK superfamily is also known as a Restriction Endonuclease (REase) superfamily since it is the primary type of nuclease in the restriction-modification systems [86]. This fold consists of a sandwich-like αβββαβ core with the first three strands forming a antiparallel sheet while the fourth strand is parallel with the third strand to form a mixed four-stranded sheet in total [87]. The second and third strands are bent heavily towards each other to form a Y-shaped crevice where the active site is located [88]. The typical active site of the PD-(D/E)XK superfamily is configured by five residues involving D/E-D--D/ExK-Q [88], in which the first D/E is located at the N-terminal α-helix, followed by D on the second strand, a (D/E)xK motif on the third strand, and a Q on a following α-helix. Our previous analysis has found a great diversity in the configuration of catalytic residues between the PD-(D/E)XK nucleases [5,84]. The identified toxin family, namely S8-REase, represents a new example, as the typical D/E residue on the first helix has been replaced by a conserved residue D232 on the second helix of the core (Fig. 2E).

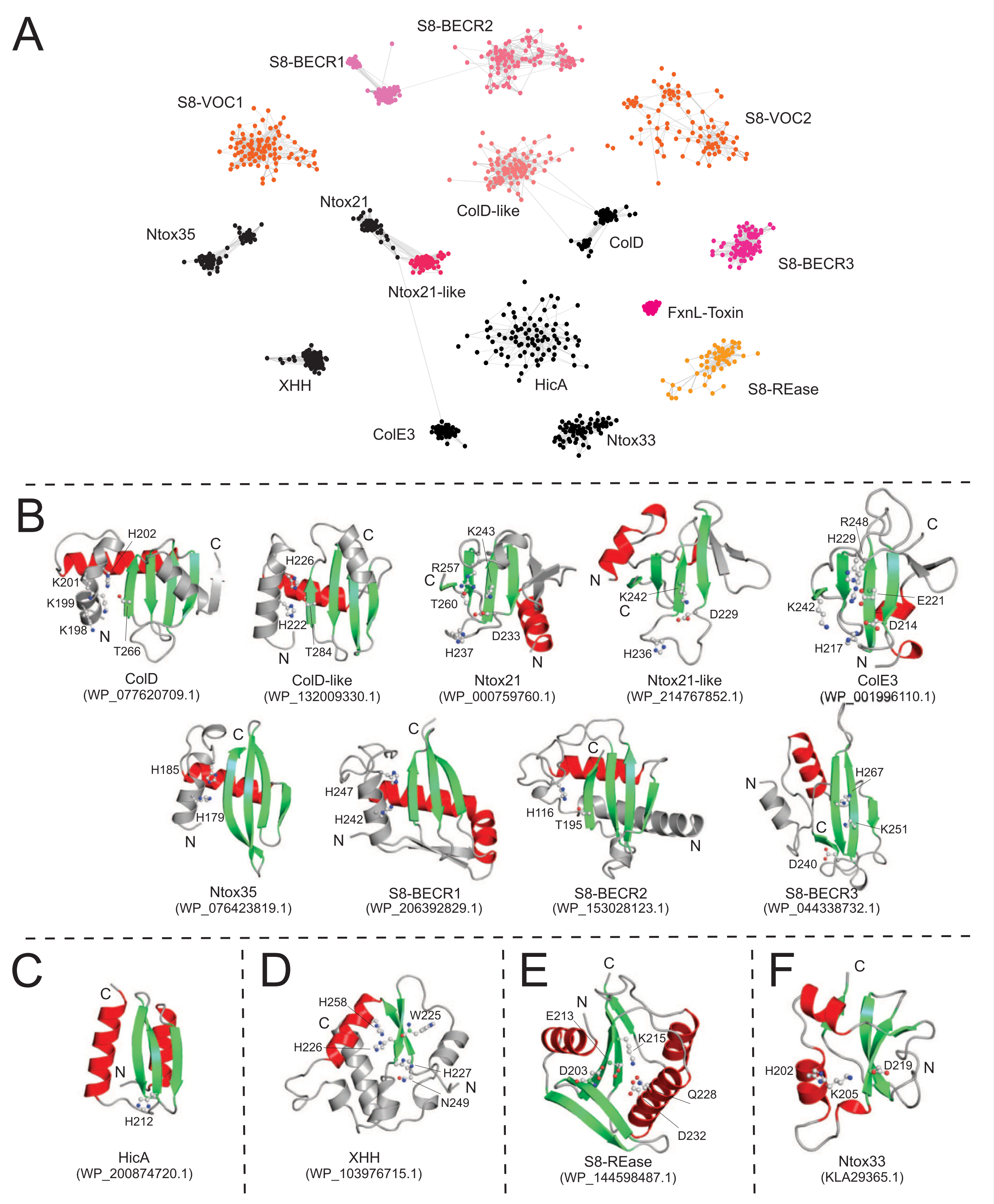
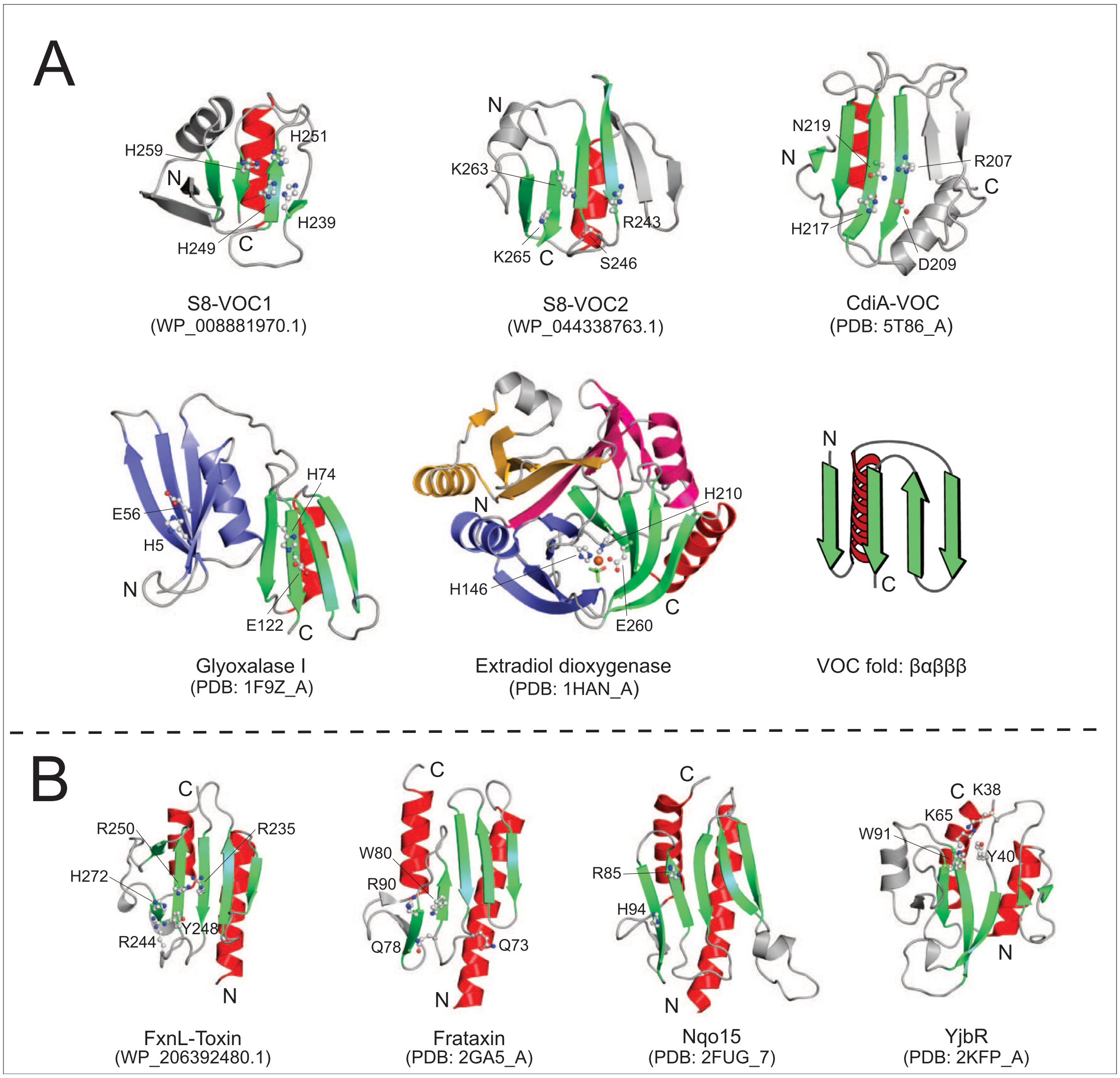


Fig. 2. **(A)** A network clustering analysis of the C-terminal domains of the BetaH toxins. Each node corresponds to a sequence. Straight lines indicate significant high-scoring segment pairs (HSPs) detected by all-against-all BLASTP searches with scoring matrix BLOSUM62 and an e-value cutoff of 0.0001. The graph was generated by the CLANS program, which uses the Fruchterman and Reingold graph drawing algorithm. The known toxin families are labeled in black while the novel toxin families are labeled in various colors. **(B-F)** Representative ribbon structures of the toxin families discovered in S8-PTS, including nine families of the BECR ribonucleases **(B)**, HicA **(C)**, XHH **(D)**, S8-REase **(E)** and Ntox33 **(F)**. The α-helices and β-sheets of the structures are shown in red and green, respectively, while the loop and other non-conserved regions are shown in grey. The predicted catalytic residues for each family are shown as balls and sticks, with carbon atoms in white, nitrogen atoms in blue and oxygen atoms in red. The NCBI accession numbers of the representative protein sequences used in the AlphaFold2 predictions are shown in parentheses. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Toxin domains related to the VOC (Vicinal oxygen Chelate) families

While the majority of the toxin domains display DNase or RNase activities, we also identified two novel classes of toxin domains which appear to deploy previously unrecognized toxic activities. The first novel class includes two unique domains, S8-VOC1 and S8-VOC2 (Fig. 3A). Despite having different structures, structure similarity searches via DaliLite revealed that both domains are related to enzymes of the VOC (Vicinal Oxygen Chelate) superfamily, including several glyoxalases (with PDB IDs:



S8-VOC1 (WP_008881970.1)

S8-VOC2 (WP_044338763.1)

CdiA-VOC (PDB: 5T86_A)

Glyoxalase I (PDB: 1F9Z_A)

Extradiol dioxygenase (PDB: 1HAN_A)

VOC fold: βαβββ

FxnL-Toxin (WP_206392480.1)

Frataxin (PDB: 2GA5_A)

Nqo15 (PDB: 2FUG_7)

YjbR (PDB: 2KFP_A)

Fig. 3. Representative ribbon structures of the VOC toxins **(A)**, FxnL-Toxin **(B)**, and their distantly-related structural homologs.

3E5D, 4QB5 and 3HPV). The VOC superfamily represents a group of diverse metalloenzymes that catalyze very diverse biochemical reactions on a wide range of molecules, including glyoxalases [89], methylmalonyl-CoA epimerases [90], extradiol dioxygenases [89], α-keto acid oxygenases [91], fosfomycin resistance proteins [92] and bleomycin resistance proteins [89]. These enzymes typically contain multiple copies of a characteristic scaffold with a βαβββ topology (VOC fold) in which a mixed parallel sheet was arranged in an order of 1–4-3–2 [93], such as two tandem VOC copies for the glyoxalase I (PDB: 1F9Z) from *E. coli* [94] and four tandem copies for the extradiol dioxygenase from *Pseudomonas cepacia* (PDB: 1HAN) [95] (Fig. 3A). The catalytic mechanisms of these enzymes, as indicated by the name of the superfamily, Vicinal Oxygen Chelate, usually involve a metal-binding center that is configured by three or four conserved polar residues (usually Glu and His), provided by two structurally adjacent VOC units, together with two additional vicinal oxygen atoms from nearby water molecules and/or their substrates [93].

Structural comparison and conservation analysis of both S8-VOC1 and S8-VOC2 support that they are new members of the VOC superfamily. First, both domains contain the core of the VOC fold, despite S8-VOC1 having an N-terminal extension and S8-VOC2 having a β-hairpin inserted directly after the core α helix. Second, the evolutionarily conserved residues of both domains, namely four His residues in S8-VOC1 and residues of Arg-Ser-Lys-Lys in S8-VOC2, are clustered on the surface of the β sheet, resembling the metal chelating site of other known VOC families (Fig. 3A). Based on both structural and conservation features, we propose that S8-VOC1/2 will utilize metal chelating mechanisms, like those of other VOC families, to catalyze the molecules of target cells. Further, we identified another previously unrecognized VOC family in the polymorphic CDI proteins (CdiA-VOC; PDB: 5T86_A). Its overall structure is more similar to that of S8-VOC2 in having a β-hairpin insert; however, it uses a different set of conserved residues for its potential metal-chelating site. The identification of three distict VOC families in PT systems suggests that the diversification of the VOC superfamily is largely unexplored, and the VOC domains might be another major toxin class in the conflict systems. Further, it needs to be noted that all the canonical VOC families contain two or more copies of the VOC domains, whereas all the new VOC families only have one VOC domain. Therefore, it will be interesting to investigate the hierarchical relationship among VOC domains, and it is possible that some of these one-copy VOC versions represent the ancestral stage of the VOC superfamily [93,96].

The toxin domain with the Frataxin-like fold (FxnL-toxin)

The second class of domains with potential novel toxin activity is from a toxin domain found in the C-terminal region of the protein WP_206392480.1 (Fig. 3B). Structure similarity searches revealed that this domain is significantly related to several members of the Frataxin-like fold, including the typical Frataxin (Dali Z = 6.6, PDB:4HS5) [97], the Nqo15 subunit of respiratory complex I from *Thermus thermophiles* (Dali Z = 6.8, PDB:2FUG_7) [98] and the YjbR domain (Dali Z = 4.5, PDB:2KFP) [99]. These proteins share a structural core composed of αββββββα elements with the leading and ending α-helices arranged on one layer and the central β-sheet on the other [100]. Previous studies have shown that human Frataxin and its yeast and bacterial homologs are functionally involved in regulating the Fe-S cluster assembly [101–103]. This activity appears to be associated with a cluster of several family-specific polar and aromatic residues on the surface of the β-sheet, especially Trp80, which interacts directly with the scaffold protein of the Fe-S cluster complex [102]. Importantly, the FxnL-toxin domain displays a similar cluster configuration of conserved residues, including an aromatic residue Tyr248, as that of Frataxin (Fig. 3B). Therefore, it is possible that the FxnL-Toxin is used to interfere with the Fe-S cluster formation of the rival bacteria since Fe-S clusters are common cofactors involved in energy metabolism pathways [104].

Ntox33

The last toxin family is the Ntox33 domain. It is an α + β domain with a conserved H-K-D triad as the potential catalytic residues (Fig. 2F). However, both profile-based and structure-based similarity searches did not identify any significant hits. Therefore, the potential function of Ntox33 remains unclear.

3.3. Gene neighborhood analysis identified a variety of immunity proteins associated with the BetaH-containing toxins

In addition to the polymorphic toxins, a major feature of the PTSs is the association of the toxins with the immunity proteins and other secretion components on the toxin loci. We thus conducted a systematic analysis on the genomic loci that contain the above-identified BetaH-containing toxins. Both upstream and downstream genes were extracted, and their encoded protein sequences were clustered and annotated using domain architecture. From this analysis, we identified 20 distinct immunity families (Fig. 4A) which are located downstream of the toxin genes [5,7]. We also utilized the same sequence/structure analysis pipeline to predict their structures and conservation patterns and classify them into a superfamily when possible (Figs. 4 and 5). Below we will specifically describe the structure and features of the major immunity families and their interactions.

Six Ferredoxin-like (FdL) immunity families

We found six novel immunity families, categorized into two major types, that adapt the Ferredoxin-like fold. A typical **F**erre**d**-oxin-**L**ike (FdL) fold, belonging to the alpha and beta (α + β) class according to SCOP classification, displays a βαββαβ topology in which the antiparallel sheet is ordered as 4–1-3–2. Four FdL immunity families, namely FdL-Imm1 to FdL-Imm4 (Fig. 4B), are grouped as the first type as they are featured by having two copies of the FdL domain. To be noted, the two copies (or units) are not tandemly arranged; instead, the second FdL domain is inserted into the first domain in the position between the first α-helix and the second β-strand. Two other immunity families, namely FdL-Imm5 (or DUF4279 in Pfam) and FdL-Imm6, are grouped as the second type as they contain only one FdL domain followed by some C-terminal extensions. This type also contains a CdiI immunity protein from *Burkholderia pseudomallei 1026b* (PDB: 4G6V_B) (Fig. 4B) [105], as revealed by the structural similarity. The Ferredoxin-like fold is a pervasive fold shared by numerous domains with different functions, such as the ACT domain for ligand binding [106], the RNA binding domain [107], the archetypal ferredoxin domain found in the iron–sulfur proteins [108], several toxins such as fungal killer toxin (PDB: 1KP6_A) [109] and fungal effector AvrLm4-7 (PDB: 4FPR_A) [110], and CRISPR Cas6 endonuclease (PDB: 3I4H_A) [111]. This is the first report showing that FdL might represent another major fold of immunity families, in addition to SUKH and SuFu, which we have previously identified [5,7].

Two sPC4L immunity families

Two immunity families, sPC4L-Imm1 and sPC4L-Imm2, are found to belong to the sPC4L superfamily (Pfam Clan: CL0609). The sPC4L fold is featured by a ββββα-ββββα topology which forms two separate β-sheets with two α-helices packed against each other [112]. Both sPC4L-Imm1 and sPC4L-Imm2 share the typical structure, but they differ from other canonical families in that their first β-sheet is three-stranded due to the missing first strand of the typical fold (Fig. 4C). Further, we found that two other CdiI immunity proteins (PDB: 4NTQ_B; PDB: 5FFP_A) also belong to this superfamily (Fig. 4C), suggesting that the sPC4L fold is another versatile scaffold in binding to toxin proteins, in addition to its known ability for nucleic acid-binding as observed in many families such as PurA [112], MRP [113] and Whirly [114].

Four TPR (<u>T</u>etratrico<u>P</u>eptide <u>R</u>epeat) immunity families

We found four novel immunity families that belong to the TPR repeats, namely TPR-Imm1 to TPR-Imm4 (Fig. 5A). The TPR repeats typically adapt a solenoid structure formed by multiple tandem repeats of a hairpin of antiparallel α-helices [115]. They are well-known for the ability to bind with short or long peptides via their concave grooves, or to bind with or block other globular domains via different sections of their solenoid structures [115]. Interestingly, three novel TPR-immunity families (TPR-Imm1 to TPR-Imm3) have at least one conserved negatively-charged residue located on the loop between helices (Fig. 5A). This situation is like that of the CdiI protein (PDB: 6CP8_C) in which a conserved Asp64 on the loop region is used to directly interact with a Arg299 of its cognate CdiA toxin [26]. This might indicate that the novel TPR families use these conserved negatively-charged residues to form polar contacts with their cognate toxins.

Other immunity families

In addition to the above immunity families, there are eight other immunity families that display unique structures. Specifically, 4Cys-Imm is featured by containing four conserved cysteines that are packed against each other like two perpendicular knuckles (Fig. 5B). No disulfide bond was formed based on the AlphaFold2 prediction, suggesting that they might be involved in binding a metal, i.e. zinc. Although DALI returned no reliable structural homologs, 4Cys-Imm shows a local structural similarity to the zinc finger of human transcriptional elongation factor TFIIS (PDB: 1TFI_A) (Fig. 5B). Therefore, we believe 4Cys-Imm might represent a new type of the zinc fingers [85]. Second, DSL-Imm (**D**ouble **S**heet **L**ayers) represent another novel immunity family, which is featured by its composition of two layers of β-sheets (Fig. 5C). No report was found on this fold, but a hypothetical protein (5V77_A) was revealed as a highly significant hit (Z-score = 7.0) in the DaliLite search (Supplementary File S3). Further, like DSL-Imm, some of the homologs of 5V77_A (i.e. WP_172763861.1) are also coupled with the toxin genes including the MafB toxins, supporting the role of the fold as an immunity protein in PTSs [5]. Third, FxnL-Imm (Imm54) is an immunity family with a frataxin-like fold, a fold also observed in the above-mentioned FxnL-Toxin domain (Fig. 5D). However, unlike FxnL-Toxin which displays a striking conservation of multiple polar and aromatic residues on the surface of the β sheet (Fig. 3B), FxnL-Imm has no

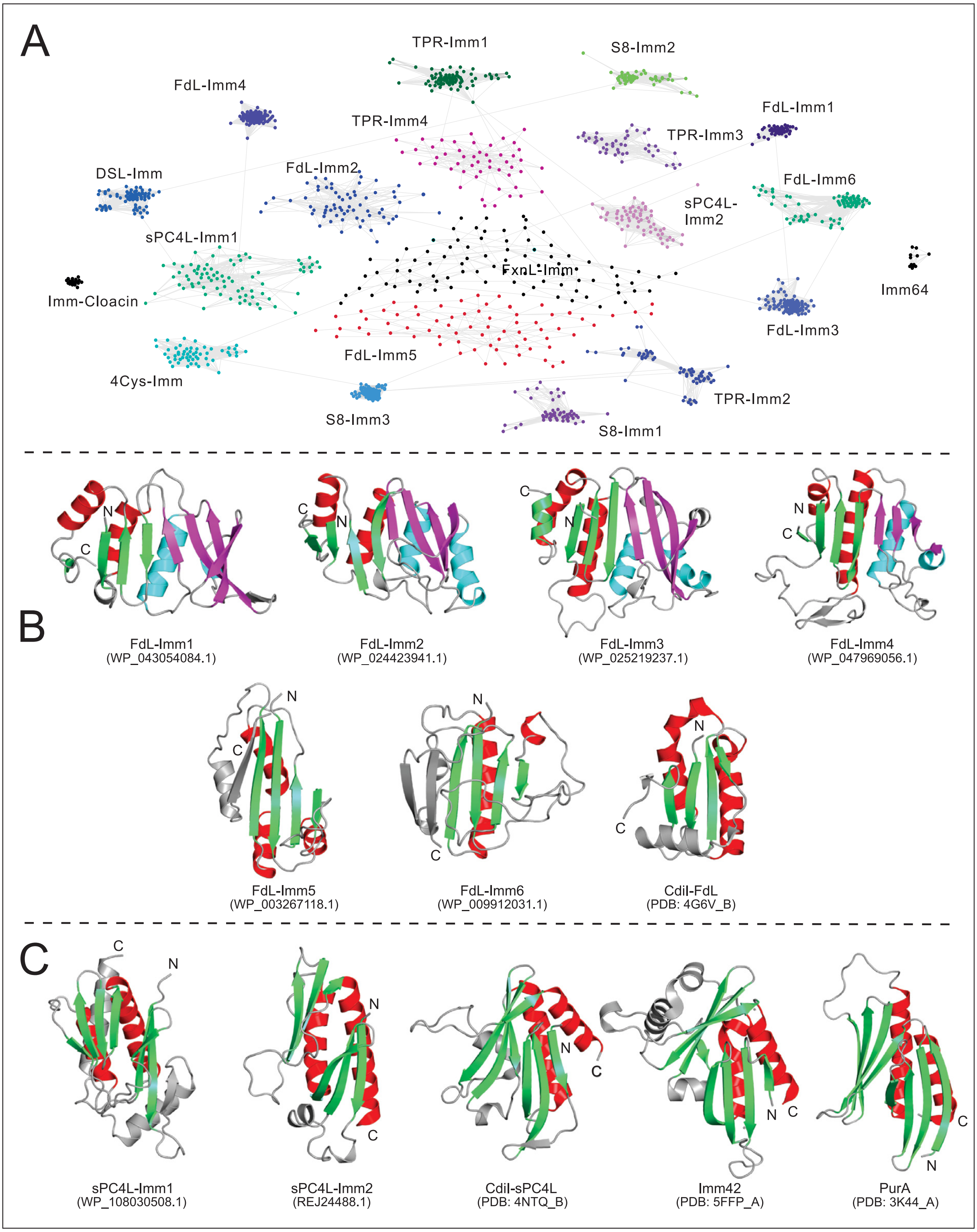


Fig. 4. A network clustering analysis of the immunity families that are associated with the BetaH toxins. The known immunity families are shown in black while the newly-identified ones are shown in various colors. **(B-C)** Representative ribbon structures of FdL **(B)** and sPC4L **(C)** immunity families and their structural homologs.

such feature (Supplementary File S1), highlighting their difference in function. This is an interesting case in which different families adopting the same fold can function as either toxin or immunity proteins in PTSs. Further, Imm64 was found to be structurally related to Imm52 (PDB: 6JDP_A) (Fig. 5E) and a CdiI protein (PDB: 6D7Y_B), whereas S8-Imm1 was found to be structurally related to another CdiI protein (PDB: 5T86_I) (Fig. 5F). Finally, both S8-Imm2 (Fig. 5G) and S8-Imm3 (Fig. 5H) are unique structures, and no structural homologs were found in DaliLite searches.

Therefore, on the genomic loci of the BetaH-containing toxins, we have identified many novel immunity proteins that can be classified into 20 families. Based on the similarity of the folds, we could

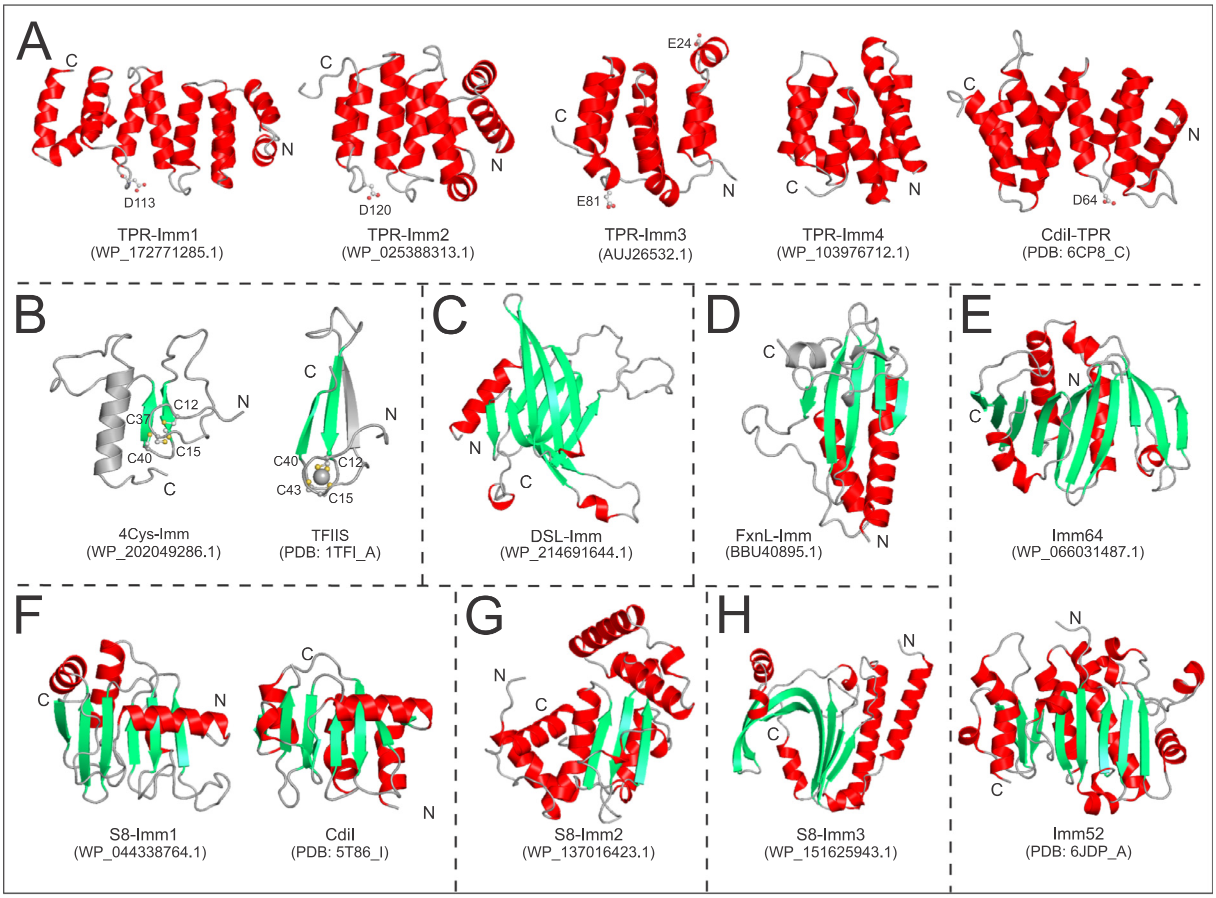


Fig. 5. Representative ribbon structures of the immunity families including TPR **(A)**, 4Cys-Imm **(B)**, DSL-Imm **(C)**, FxnL-Imm **(D)**, Imm64 **(E)**, S8-Imm1 **(F)**, S8-Imm2 **(G)** and S8-Imm3 **(H)**.

further unify some of them into superfamilies. However, the diversity in the overall structures between the families and between the superfamilies are still striking, and provides another typical presentation of the polymorphic toxin systems.

3.4. A peptidase is strongly associated with the toxin-immunity pairs

Importantly, our gene neighborhood analysis also revealed that a group of peptidase genes are frequently coupled with the above-identified toxin and immunity genes (Fig. 6). They are located mostly upstream of the BetaH-containing toxin genes and sometimes downstream of the immunity genes (Fig. 6A). Domain analysis revealed that they are led by a signal peptide, suggesting that they are secreted. The peptidase domain is found to belong to the S8 peptidase family (Pfam ID: PF00082), which is featured by an α/β fold with a 7-stranded parallel β sheet, and an Asp/Ser/His catalytic triad (Fig. 6B). The S8 peptidases, also known as subtilases, represent the second largest family of serine peptidases according to the MEROPS database [116] and are widely distributed in bacteria, archaea and eukaryotes. They are also known as proprotein-processing endopeptidases, as typified by subtilisins in bacteria [117], kexins in yeast [118], cucumisin and SISBT3 in plants [119] and proprotein convertases in mammals [120]. They usually have a multidomain architecture, including a N-terminal signal peptide, a prodomain, a central peptidase domain, and a C-terminal highly variable extension [121–123]. The peptidase remains inactive until the prodomain is autocatalytically cleaved, a process called zymogen maturation [120,124]; then, the peptidase domain will process and facilitate maturation of the downstream substrates, such as neuropeptides and peptide hormones by mammal proprotein convertases [120], the killer toxin precursors by fungal kexins [125], and bacterial cytolysin precursors by lantibiotic leader peptidases [126].

To understand the role of the BetaH-toxin associated S8 peptidase proteins in the system, we collected the bacterial homologs using our custom database (see methods) and conducted a systematic analysis of their domain architectures and gene neighborhoods in a phylogenetic context. We found that the majority of the bacterial S8 peptidases retain multidomain architectures, typically including an N-terminal signal peptide, a prodomain of the I9 inhibitor family, and several C-terminal domains, in addition to the S8 peptidase domain (Fig. 7). This suggests that they will function as the prodomain processing endopeptidases. By contrast, BetaH-associated peptidases and several other clades of S8 peptidases display unique domain architectures. Further, by studying the domain architectures and gene neighborhoods of these bacterial homologs, we noticed that they are involved in different bacterial secretion pathways, respectively. These include a clade of S8 peptidases that are directly fused with the C-terminal β-barrel autotransporter domain, the secretion module for T5SSa [127]; the other clade of peptidases are fused with the Por_Secre_Tail domain of the immunoglobulin-like fold, a marker module for T9SS [128]; a third clade with a solo peptidase domain is operonically associated with the components of the T7SS. Further, several studies have shown that these unique clades of S8 peptidases serve as maturation peptidases in processing the substrates that are secreted by these secretion pathways. For example, the T5SSa-S8 peptidases in *Bordetella pertussis*, after being secreted via the coupled autotransporter [129], are required for the maturation of the filamentous hemagglutinin adhesins [130]. The mycosins, known as T7SS-

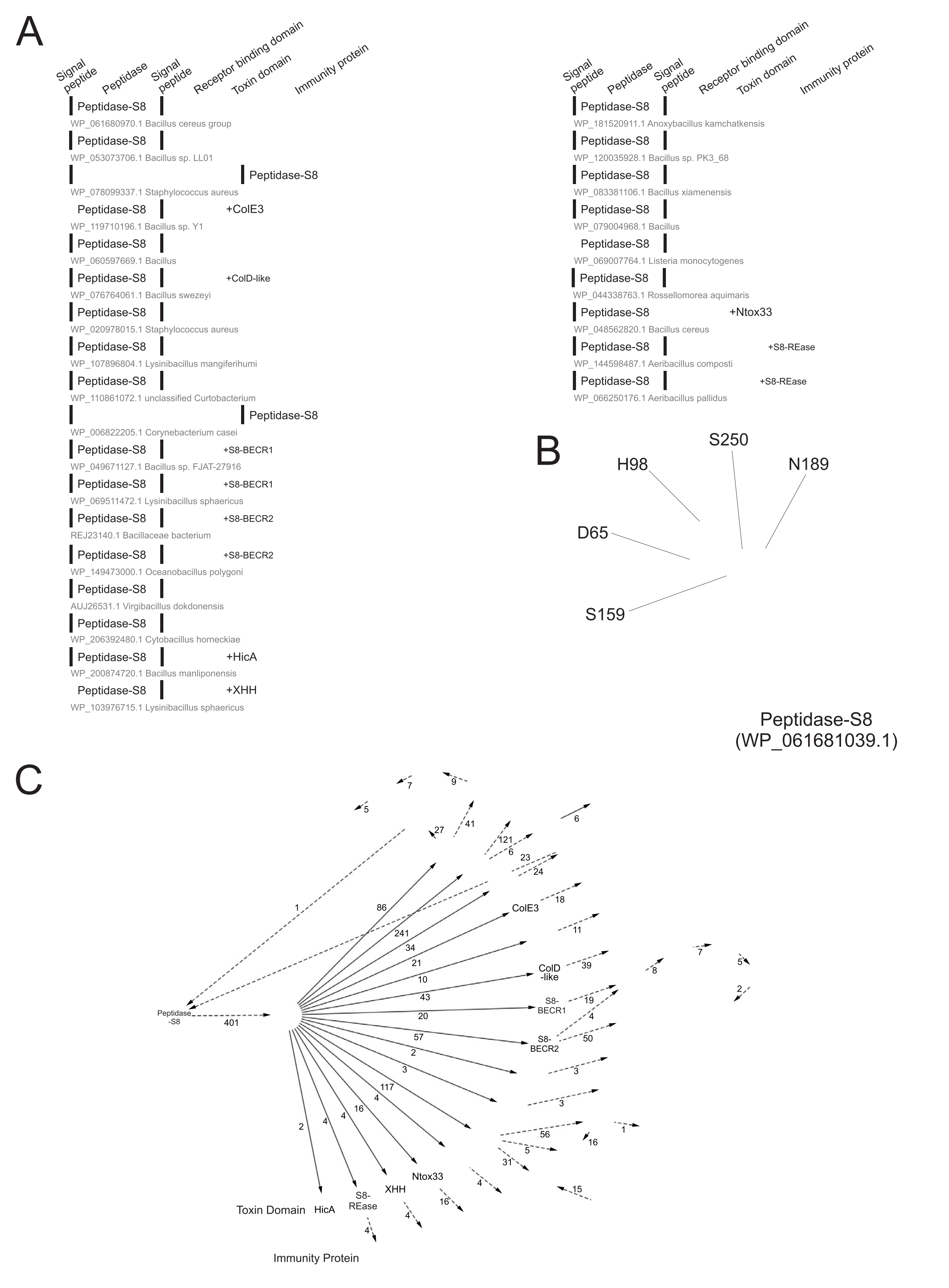

A
Signal peptide
Peptidase
Signal peptide
Receptor binding domain
Toxin domain
Immunity protein
Peptidase-S8
WP_061680970.1 Bacillus cereus group
WP_053073706.1 Bacillus sp. LL01
WP_078099337.1 Staphylococcus aureus
+ColE3
WP_119710196.1 Bacillus sp. Y1
WP_060597669.1 Bacillus
+ColD-like
WP_076764061.1 Bacillus swezeyi
WP_020978015.1 Staphylococcus aureus
WP_107896804.1 Lysinibacillus mangiferihumi
WP_110861072.1 unclassified Curtobacterium
WP_006822205.1 Corynebacterium casei
+S8-BECR1
WP_049671127.1 Bacillus sp. FJAT-27916
WP_069511472.1 Lysinibacillus sphaericus
+S8-BECR2
REJ23140.1 Bacillaceae bacterium
WP_149473000.1 Oceanobacillus polygoni
AUJ26531.1 Virgibacillus dokdonensis
WP_206392480.1 Cytobacillus horneckiae
+HicA
WP_200874720.1 Bacillus manliponensis
+XHH
WP_103976715.1 Lysinibacillus sphaericus
WP_181520911.1 Anoxybacillus kamchatkensis
WP_120035928.1 Bacillus sp. PK3_68
WP_083381106.1 Bacillus xiamenensis
WP_079004968.1 Bacillus
WP_069007764.1 Listeria monocytogenes
WP_044338763.1 Rossellomorea aquimaris
+Ntox33
WP_048562820.1 Bacillus cereus
+S8-REase
WP_144598487.1 Aeribacillus composti
WP_066250176.1 Aeribacillus pallidus
B
S250
H98
N189
D65
S159
Peptidase-S8
(WP_061681039.1)
C
Peptidase-S8
401
1
86
241
34
21
10
43
20
57
2
3
117
4
16
4
4
2
ColE3
ColD-like
S8-BECR1
S8-BECR2
Ntox33
XHH
S8-REase
HicA
Toxin Domain
Immunity Protein
7
9
5
27
41
121
6
23
24
6
18
11
39
8
7
5
2
19
4
50
3
3
56
16
1
5
31
4
15
16
4
4

Fig. 6. (**A**) Representative gene neighborhoods of the S8-PTS. Each gene is presented as a block arrow, in which the major domains of the encoded protein are separately shown as rectangular segments. The direction of the arrow shows the direction of the transcription. The loci are labelled by the NCBI accession numbers of the S8 peptidase genes followed by their species names. (**B**) The ribbon structure of the S8 peptidase that is associated with the BetaH toxin (WP_061680970.1) shown in the first operonic example. The α-helices (shown as cylinders) and β-sheets of the structure are shown in red and blue, respectively, whereas the loop and other non-conserved regions are shown in grey. The predicted catalytic residues are shown as balls and sticks, with carbon atoms in light blue, nitrogen atoms in blue and oxygen atoms in red. (**C**) Domain architecture and operonic association network of S8-PTS. Direct domain association is indicated by the solid line, whereas the operonic association is indicated by the dash line. The occurrence frequency of each association is shown along the line. The coloring theme of each domain is corresponding to the that of the respective domain architectures shown in **(A)**. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

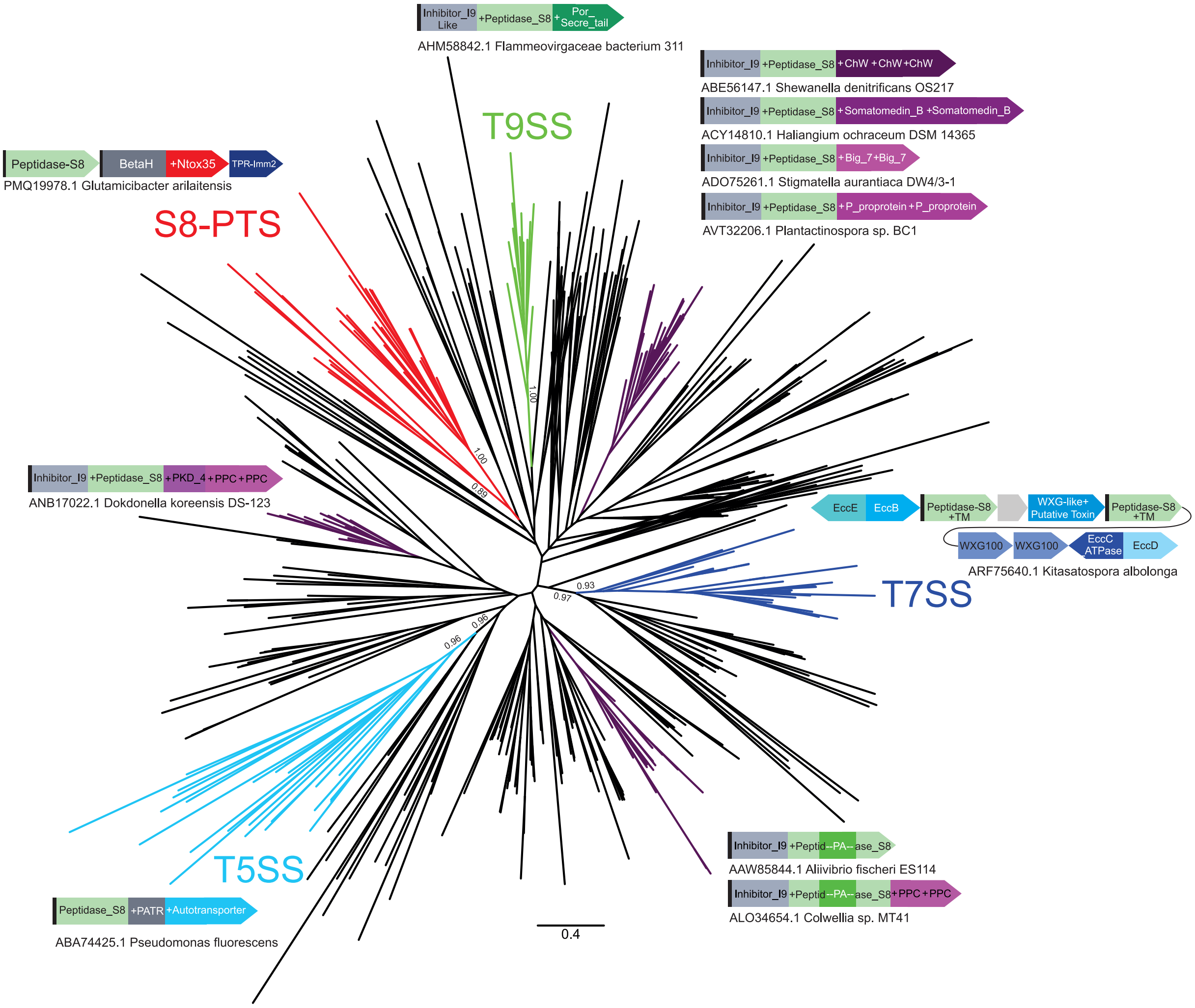

Inhibitor_I9 Like
+Peptidase_S8
+Por_ Secre_tail
AHM58842.1 Flammeovirgaceae bacterium 311
T9SS
Inhibitor_I9 +Peptidase_S8 +ChW +ChW +ChW
ABE56147.1 Shewanella denitrificans OS217
Inhibitor_I9 +Peptidase_S8 +Somatomedin_B +Somatomedin_B
ACY14810.1 Haliangium ochraceum DSM 14365
Inhibitor_I9 +Peptidase_S8 +Big_7 +Big_7
ADO75261.1 Stigmatella aurantiaca DW4/3-1
Inhibitor_I9 +Peptidase_S8 +P_proprotein +P_proprotein
AVT32206.1 Plantactinospora sp. BC1
Peptidase-S8
BetaH
+Ntox35
TPR-Imm2
PMQ19978.1 Glutamicibacter arilaitensis
S8-PTS
1.00
1.00
0.99
Inhibitor_I9 +Peptidase_S8 +PKD_4 +PPC +PPC
ANB17022.1 Dokdonella koreensis DS-123
EccE
EccB
Peptidase-S8 +TM
WXG-like+ Putative Toxin
Peptidase-S8 +TM
WXG100
WXG100
EccC ATPase
EccD
ARF75640.1 Kitasatospora albolonga
0.93
0.97
T7SS
0.96
0.96
T5SS
Peptidase_S8 +PATR +Autotransporter
ABA74425.1 Pseudomonas fluorescens
0.4
Inhibitor_I9 +Peptid --PA-- ase_S8
AAW85844.1 Aliivibrio fischeri ES114
Inhibitor_I9 +Peptid --PA-- ase_S8 +PPC +PPC
ALO34654.1 Colwellia sp. MT41

Fig. 7. Phylogenetic relationship of bacterial S8 peptidases. The phylogeny was constructed using the FastTree method, with the major branches supported by the SH-like local support values. The major clades that display either specific domain architectures or genomic context are highlighted in different colors. Representative domain architectures and gene neighborhoods are also shown along the clades.

specific S8 peptidases [131], are found to constitute the trimer dome-like structure that caps the cavity of the T7SS complex with catalytic sites facing towards the central lumen of the cavity, supporting the possibility that they process the substrates of T7SS, such as the toxins, that pass through the central channel [132]. Therefore, across the bacterial S8 peptidases, the function of processing the associated substrates appears to be a common theme, including both prodomain processing peptidases and those associated with the secretion pathways, such as T5SSa, T7SS and T9SS. We therefore propose that the association of S8 peptidases with different secretion pathways is derived during evolution and that the peptidases function as a processing peptidase to facilitate the release of the associated components, including the above-identified polymorphic toxins. Importantly, S8 peptidases, such as proprotein convertases, typically cleave precursors at the motif [R/K]-X_n-[R/K]↓, where X can be any amino acids and n = 0, 2, 4 or 6 [133]. Indeed, we have found that the C-terminal regions of BetaH domains are featured by a patch of positive residues, suggesting that they are potential cleavage sites of these peptidases (Fig. 1).

3.5. S8-PTS, a novel PT system with three components

In conclusion, with our analysis of the genomic loci and protein components, we have identified a new type of polymorphic toxin system. The system has two defining features. The first is the strong association of the S8 peptidase with the toxin and immunity components typical of other PTS. Therefore, we term it S8-PTS (S8 peptidase-associated PTS) (Fig. 6). Our previous studies have identified several peptidase families in the PT systems. However, the majority of them are present as a component of the toxin proteins, such as HINT, ZU5, PrsW, and the so-called PVC-metallopeptidase [5]. They are typically used to release the C-terminal toxin domains during toxin secretion. S8-PTS is the first PTS that uses operonically-associated peptidase to release the toxins. The other defining feature is the presence of a new N-terminal domain, BetaH, in their toxins. We found that BetaH is structurally related to the receptor-binding domain of the colicin N toxin and that both toxins (ColN and BetaH-containing ones) share a comparable organization, suggesting that BetaH might function similarly in facilitating entry of the toxin into the target cell cytoplasm.

In addition to the above-discussed defining features, we observed a striking level of variation in the composition of both the toxins and immunity proteins (Fig. 6), despite the S8-PTS only being found in a small set of bacterial species. The toxins, sharing a common BetaH domain, have employed 16 mechanically distinct toxin domains. Likewise, their immunity proteins can be classified into 20 families. Further, the domain shuffling appears to be mainly between the BetaH domain and the C-terminal toxin

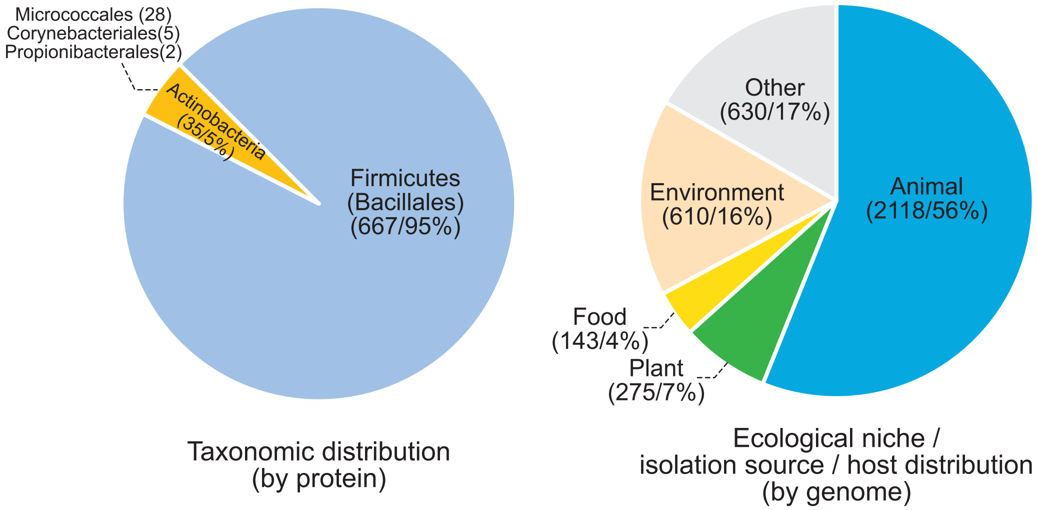


Fig. 8. Taxonomy distribution of the S8-PTS based on the characteristic BetaH domain sequences (left) and classification of host/isolation source of these bacteria based on the retrieved genomes that contain the S8-PTS (right). The number of instances (either protein sequences or genomes) and their proportions associated with each taxonomy or host/source category are shown in the parentheses.

domains, whereas the toxin-immunity domain pairs appear to be stable for most of the cases (Fig. 6B). Additionally, we did not see a strong association between toxin and immunity domains in term of their folds. In other words, the toxin domains with the same fold can be neutralized by the immunity families that have different folds (Fig. 6). For example, different BECR domains are found to be coupled with FdL immunity domains, TPR-like families, 4Cys-Imm and DSL-Imm. Similarly, all six FdL immunity families are found to block at least three classes of toxin domains, including two BECR toxins (Ntox21 and Ntox21-like), S8-VOC1, and Ntox33, respectively.

Not only did elucidate the organizational principle of the S8-PTS system, but we have also provided confident structure models of the components based on the AlphaFold2 algorithm and predicted the potential catalytic residues for toxin domains and interacting residues for immunity families. The newly discovered toxin types, including VOC and FxnL domains, have extended our understanding of the mechanisms of toxin action. Further, the unification of the immunity families, for the first time, suggests that an extensive evolutionary diversification of immunity proteins from a relatively small set of ancestors has paralleled with the expansion of the toxins.

Finally, the S8-PTS was present mainly in firmicutes (Bacillales) and actinobacteria, two specific groups of gram-positive bacteria (Fig. 8 and Supplementary File S4). Gram-positive bacteria have only a single cytoplasmic membrane and their protein secretion is typically mediated by the Sec secretion pathway, which is consistent with the presence of the signal peptide on these BetaH-containing toxins. Analysis of these bacterial species suggest that they are mostly animal- and plant-associated bacteria (Fig. 8 and Supplementary File S4). Among them, many are pathogens, such as different species of the genus *Staphylococcus*, including *Staphylococcus aureus* and *Staphylococcus pseudintermedius*, both of which cause severe human and animal diseases [134], as well as *Listeria monocytogenes*, that causes the infection listeriosis [135], *Mycobacteroides abscessus*, responsible for a wide spectrum of skin and soft tissue diseases as well as central nervous system infections [136], and *Curtobacterium flaccumfaciens*, that leads to bacterial wilt and tan spot of dry beans in plants [137]. There are also many bacteria which appear to be beneficial to the host, such as *Bacillus velezensis*, a plant growth-promoting bacteria [138], and *Terribacillus goriensi*, which has antagonistic activity against several phytopathogenic fungi [139]. The majority of the genomes have only one S8-PTS locus, but a few have two to four PTS loci, with each one having distinct sets of toxin and immunity pairs. We propose that these PTSs might facilitate the competition of these bacteria against other microbes or contribute to the pathogen-host interactions.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the Saint Louis University Start-up fund. The authors would like to thank Theresa Schneider for editing the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2022.08.036.

References

[1] Ghoul M, Mitri S. The Ecology and Evolution of Microbial Competition. Trends Microbiol 2016;24(10):833–45.
[2] Munita, J.M. and C.A. Arias. *Mechanisms of Antibiotic Resistance.* Microbiol Spectr. 2016. **4**(2).
[3] Harms A et al. Toxins, Targets, and Triggers: An Overview of Toxin-Antitoxin Biology. Mol Cell 2018;70(5):768–84.
[4] Koonin EV, Makarova KS, Zhang F. Diversity, classification and evolution of CRISPR-Cas systems. Curr Opin Microbiol 2017;37:67–78.
[5] Zhang D et al. Polymorphic toxin systems: Comprehensive characterization of trafficking modes, processing, mechanisms of action, immunity and ecology using comparative genomics. Biol Direct 2012;7:18.

[6] Iyer LM et al. Evolution of the deaminase fold and multiple origins of eukaryotic editing and mutagenic nucleic acid deaminases from bacterial toxin systems. Nucleic Acids Res 2011;39(22):9473–97.

[7] Zhang D, Iyer LM, Aravind L. A novel immunity system for bacterial nucleic acid degrading toxins and its recruitment in various eukaryotic and DNA viral systems. Nucleic Acids Res 2011;39(11):4532–52.

[8] Jamet A, Nassif X. New players in the toxin field: polymorphic toxin systems in bacteria. mBio 2015;6(3):e00285–315.

[9] Ruhe, Z.C., D.A. Low, and C.S. Hayes, *Polymorphic Toxins and Their Immunity Proteins: Diversity, Evolution, and Mechanisms of Delivery.* Annual Review of Microbiology, Vol 74, 2020, 2020. **74**: p. 497-520.

[10] Korotkov KV, Sandkvist M, Hol WG. The type II secretion system: biogenesis, molecular architecture and mechanism. Nat Rev Microbiol 2012;10(5):336–51.

[11] Desvaux M et al. The unusual extended signal peptide region of the type V secretion system is phylogenetically restricted. FEMS Microbiol Lett 2006;264(1):22–30.

[12] Shneider MM et al. PAAR-repeat proteins sharpen and diversify the type VI secretion system spike. Nature 2013;500(7462):350–3.

[13] Pallen MJ. The ESAT-6/WXG100 superfamily – and a new Gram-positive secretion system? Trends Microbiol 2002;10(5):209–12.

[14] Whitney JC et al. A broadly distributed toxin family mediates contact-dependent antagonism between gram-positive bacteria. Elife 2017;6:e26938.

[15] Burroughs AM, Iyer LM, Aravind L. Comparative genomics and evolutionary trajectories of viral ATP dependent DNA-packaging systems. Genome Dyn 2007;3:48–65.

[16] Jamet A et al. A widespread family of polymorphic toxins encoded by temperate phages. BMC Biol 2017;15(1):75.

[17] Ellermeier CD, Losick R. Evidence for a novel protease governing regulated intramembrane proteolysis and resistance to antimicrobial peptides in Bacillus subtilis. Genes Dev 2006;20(14):1911–22.

[18] Jamet A et al. A new family of secreted toxins in pathogenic Neisseria species. PLoS Pathog 2015;11(1):e1004592.

[19] Jana B, Salomon D, Bosis E. A novel class of polymorphic toxins in Bacteroidetes. Life Sci Alliance 2020;3(4).

[20] Aravind L et al. The natural history of ADP-ribosyltransferases and the ADP-ribosylation system. Curr Top Microbiol Immunol 2015;384:3–32.

[21] Ruhe ZC, Low DA, Hayes CS. Bacterial contact-dependent growth inhibition. Trends Microbiol 2013;21(5):230–7.

[22] Poole SJ et al. Identification of functional toxin/immunity genes linked to contact-dependent growth inhibition (CDI) and rearrangement hotspot (Rhs) systems. PLoS Genet 2011;7(8):e1002217.

[23] Jana B et al. A modular effector with a DNase domain and a marker for T6SS substrates. Nat Commun 2019;10(1):3595.

[24] Makarova KS et al. Antimicrobial peptides, polymorphic toxins, and self-nonself recognition systems in archaea: an untapped armory for intermicrobial conflicts. MBio 2019;10(3):e00715. e00719.

[25] Fitzsimons TC et al. Identification of Novel Acinetobacter baumannii Type VI Secretion System Antibacterial Effector and Immunity Pairs. Infect Immun 2018;86(8).

[26] Gucinski GC et al. Convergent Evolution of the Barnase/EndoU/Colicin/RelE (BECR) Fold in Antibacterial tRNase Toxins. Structure 2019;27(11):1660–1674 e5.

[27] Michalska K et al. Structure of a novel antibacterial toxin that exploits elongation factor Tu to cleave specific transfer RNAs. Nucleic Acids Res 2017;45(17):10306–20.

[28] Michalska K et al. Functional plasticity of antibacterial EndoU toxins. Mol Microbiol 2018;109(4):509–27.

[29] Batot G et al. The CDI toxin of Yersinia kristensenii is a novel bacterial member of the RNase A superfamily. Nucleic Acids Res 2017;45(9):5013–25.

[30] Ohr RJ et al. EssD, a Nuclease Effector of the Staphylococcus aureus ESS Pathway. J Bacteriol 2017;199(1).

[31] Pissaridou P et al. The Pseudomonas aeruginosa T6SS-VgrG1b spike is topped by a PAAR protein eliciting DNA damage to bacterial competitors. Proc Natl Acad Sci U S A 2018;115(49):12519–24.

[32] Sibinelli-Sousa S et al. A Family of T6SS Antibacterial Effectors Related to l, d-Transpeptidases Targets the Peptidoglycan. Cell Rep 2020;31(12):107813.

[33] Tang JY et al. Diverse NADase effector families mediate interbacterial antagonism via the type VI secretion system. J Biol Chem 2018;293(5):1504–14.

[34] Basso P et al. Pseudomonas aeruginosa Pore-Forming Exolysin and Type IV Pili Cooperate To Induce Host Cell Lysis. mBio 2017;8(1).

[35] Alcoforado Diniz J, Coulthurst SJ. Intraspecies Competition in Serratia marcescens Is Mediated by Type VI-Secreted Rhs Effectors and a Conserved Effector-Associated Accessory Protein. J Bacteriol 2015;197(14):2350–60.

[36] Allen JP et al. A comparative genomics approach identifies contact-dependent growth inhibition as a virulence determinant. Proc Natl Acad Sci U S A 2020;117(12):6811–21.

[37] Gong Y et al. A nuclease-toxin and immunity system for kin discrimination in Myxococcus xanthus. Environ Microbiol 2018;20(7):2552–67.

[38] Ruhe, Z.C., et al., Programmed secretion arrest and receptor-triggered toxin export during antibacterial contact-dependent growth inhibition. Cell, 2018. **175**(4): p. 921-933. e14.

[39] Quentin D et al. Mechanism of loading and translocation of type VI secretion system effector Tse6. Nat Microbiol 2018;3(10):1142–52.

[40] Jurėnas D et al. Mounting, structure and autocleavage of a type VI secretion-associated Rhs polymorphic toxin. Nat Commun 2021;12(1):1–11.

[41] Vassallo CN et al. Infectious polymorphic toxins delivered by outer membrane exchange discriminate kin in myxobacteria. Elife 2017;6.

[42] Geller AM et al. *The extracellular contractile injection system is enriched in environmental microbes and associates with numerous toxins.* Nature. Communications 2021;12(1).

[43] Ma LS et al. Agrobacterium tumefaciens deploys a superfamily of type VI secretion DNase effectors as weapons for interbacterial competition in planta. Cell Host Microbe 2014;16(1):94–104.

[44] Tan Y et al. Comparative Phylogenomic Analysis Reveals Evolutionary Genomic Changes and Novel Toxin Families in Endophytic Liberibacter Pathogens. Microbiol Spectr 2021;9(2):e0050921.

[45] Taylor JC et al. A type VII secretion system of Streptococcus gallolyticus subsp. gallolyticus contributes to gut colonization and the development of colon tumors. PLoS Pathog 2021;17(1):e1009182.

[46] Kobayashi K. Diverse LXG toxin and antitoxin systems specifically mediate intraspecies competition in Bacillus subtilis biofilms. PLoS Genet 2021;17(7):e1009682.

[47] Nachmias, N., et al., *Systematic Discovery of Antimicrobial Polymorphic Toxins.* bioRxiv, 2021: p. 2021.10.19.465003.

[48] Altschul SF et al. Basic local alignment search tool. J Mol Biol 1990;215(3):403–10.

[49] Altschul SF et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25(17):3389–402.

[50] Eddy SR. Accelerated Profile HMM Searches. PLoS Comput Biol 2011;7(10):e1002195.

[51] Doolittle RF. The multiplicity of domains in proteins. Annu Rev Biochem 1995;64:287–314.

[52] Vogel C et al. Structure, function and evolution of multidomain proteins. Curr Opin Struct Biol 2004;14(2):208–16.

[53] Jumper J et al. Highly accurate protein structure prediction with AlphaFold. Nature 2021;596(7873):583–9.

[54] Frickey T, Lupas A. CLANS: a Java application for visualizing protein families based on pairwise similarity. Bioinformatics 2004;20(18):3702–4.

[55] Kans J, Entrez direct,. *E-utilities on the UNIX command line*, in *Entrez Programming Utilities Help [Internet].* National Center for Biotechnology Information (US); 2022.

[56] Lassmann T, Sonnhammer EL. Kalign–an accurate and fast multiple sequence alignment algorithm. BMC Bioinf 2005;6:298.

[57] Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 2004;32(5):1792–7.

[58] Pei J, Grishin NV. PROMALS3D: multiple protein sequence alignment enhanced with evolutionary and three-dimensional structural information. Methods Mol Biol 2014;1079:263–71.

[59] Taylor WR. The classification of amino acid conservation. J Theor Biol 1986;119(2):205–18.

[60] Goodstadt L, Ponting CP. CHROMA: consensus-based colouring of multiple alignments for publication. Bioinformatics 2001;17(9):845–6.

[61] Mistry J et al. Pfam: The protein families database in 2021. Nucleic Acids Res 2021;49(D1):D412. D419.

[62] Kall, L., A. Krogh, and E.L. Sonnhammer, *Advantages of combined transmembrane topology and signal peptide prediction–the Phobius web server.* Nucleic Acids Res, 2007. **35**(Web Server issue): p. W429-32.

[63] Muntjes K et al. Establishing Polycistronic Expression in the Model Microorganism Ustilago maydis. Front Microbiol 2020;11:1384.

[64] Sadava, D.E., et al., *Life: The Science of Biology*. 2011: W. H. Freeman.

[65] Aravind L. Guilt by association: contextual information in genome analysis. Genome Res 2000;10(8):1074–7.

[66] Fruchterman TM, Reingold EM. Graph drawing by force-directed placement. Software: Practice and Experience 1991;21(11):1129–64.

[67] Mirdita M et al. ColabFold: making protein folding accessible to all. Nat Methods 2022;19(6):679–82.

[68] Holm L, Laakso LM. Dali server update. Nucleic Acids Res 2016;44(W1):W351. W355.

[69] Berman HM et al. The Protein Data Bank. Nucleic Acids Res 2000;28(1):235–42.

[70] Illergård K, Ardell DH, Elofsson A. Structure is three to ten times more conserved than sequence—a study of structural response in protein cores. Proteins Struct Funct Bioinf 2009;77(3):499–508.

[71] DeLano WL, Pymol,. An open-source molecular graphics tool. CCP4 Newsletter on protein crystallography 2002;40(1):82–92.

[72] Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol 2009;26(7):1641–50.

[73] Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol 2016;33(7):1870–4.

[74] Vetter IR et al. Crystal structure of a colicin N fragment suggests a model for toxicity. Structure 1998;6(7):863–74.

[75] Klein TA et al. Molecular Basis for Immunity Protein Recognition of a Type VII Secretion System Exported Antibacterial Toxin. J Mol Biol 2018;430(21):4344–58.

[76] Baboolal TG et al. Colicin N binds to the periphery of its receptor and translocator, outer membrane protein F. Structure 2008;16(3):371–9.

[77] Yang W. Nucleases: diversity of structure, function and mechanism. Q Rev Biophys 2011;44(1):1–93.
[78] Mossakowska DE, Nyberg K, Fersht AR. Kinetic characterization of the recombinant ribonuclease from Bacillus amyloliquefaciens (barnase) and investigation of key residues in catalysis by site-directed mutagenesis. Biochemistry 1989;28(9):3843–50.
[79] Mushegian A et al. An ancient evolutionary connection between Ribonuclease A and EndoU families. RNA 2020;26(7):803–13.
[80] Neubauer C et al. The structural basis for mRNA recognition and cleavage by the ribosome-dependent endonuclease RelE. Cell 2009;139(6):1084–95.
[81] Makarova KS, Grishin NV, Koonin EV. The HicAB cassette, a putative novel, RNA-targeting toxin-antitoxin system in archaea and bacteria. Bioinformatics 2006;22(21):2581–4.
[82] Bibi-Triki S et al. Functional and structural analysis of HicA3-HicB3, a novel toxin-antitoxin system of Yersinia pestis. J Bacteriol 2014;196(21):3712–23.
[83] Turnbull KJ, Gerdes K. HicA toxin of Escherichia coli derepresses hicAB transcription to selectively produce HicB antitoxin. Mol Microbiol 2017;104 (5):781–92.
[84] Zhang D et al. Resilience of biochemical activity in protein domains in the face of structural divergence. Curr Opin Struct Biol 2014;26:92–103.
[85] Krishna SS, Majumdar I, Grishin NV. Structural classification of zinc fingers: survey and summary. Nucleic Acids Res 2003;31(2):532–50.
[86] Roberts RJ et al. REBASE–a database for DNA restriction and modification: enzymes, genes and genomes. Nucleic Acids Res 2015;43(Database issue): D298. D299.
[87] Laganeckas M, Margelevicius M, Venclovas C. Identification of new homologs of PD-(D/E)XK nucleases by support vector machines trained on data derived from profile-profile alignments. Nucleic Acids Res 2011;39(4):1187–96.
[88] Steczkiewicz K et al. Sequence, structure and functional diversity of PD-(D/E) XK phosphodiesterase superfamily. Nucleic Acids Res 2012;40(15):7016–45.
[89] Bergdoll M et al. All in the family: structural and evolutionary relationships among three modular proteins with diverse functions and variable assembly. Protein Sci 1998;7(8):1661–70.
[90] McCarthy AA et al. Crystal structure of methylmalonyl-coenzyme A epimerase from P. shermanii: a novel enzymatic function on an ancient metal binding scaffold. Structure 2001;9(7):637–46.
[91] Serre L et al. Crystal structure of Pseudomonas fluorescens 4-hydroxyphenylpyruvate dioxygenase: an enzyme involved in the tyrosine degradation pathway. Structure 1999;7(8):977–88.
[92] Bernat BA, Laughlin LT, Armstrong RN. Fosfomycin resistance protein (FosA) is a manganese metalloglutathione transferase related to glyoxalase I and the extradiol dioxygenases. Biochemistry 1997;36(11):3050–5.
[93] He P, Moran GR. Structural and mechanistic comparisons of the metal-binding members of the vicinal oxygen chelate (VOC) superfamily. J Inorg Biochem 2011;105(10):1259–72.
[94] He MM et al. Determination of the structure of Escherichia coli glyoxalase I suggests a structural basis for differential metal activation. Biochemistry 2000;39(30):8719–27.
[95] Han S et al. Crystal structure of the biphenyl-cleaving extradiol dioxygenase from a PCB-degrading pseudomonad. Science 1995;270(5238):976–80.
[96] Armstrong RN. Mechanistic diversity in a metalloenzyme superfamily. Biochemistry 2000;39(45):13625–32.
[97] Roman EA et al. Frataxin from Psychromonas ingrahamii as a model to study stability modulation within the CyaY protein family. Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics 2013;1834(6):1168–80.
[98] Sazanov LA, Hinchliffe P. Structure of the hydrophilic domain of respiratory complex I from Thermus thermophilus. Science 2006;311(5766):1430–6.
[99] Anantharaman V, Iyer LM, Aravind L. Ter-dependent stress response systems: novel pathways related to metal sensing, production of a nucleoside-like metabolite, and DNA-processing. Mol BioSyst 2012;8(12):3142–65.
[100] Andreeva A et al. The SCOP database in 2020: expanded classification of representative family and superfamily domains of known protein structures. Nucleic Acids Res 2020;48(D1):D376. D382.
[101] Fox NG et al. Structure of the human frataxin-bound iron-sulfur cluster assembly complex provides insight into its activation mechanism. Nat Commun 2019;10(1):2210.
[102] Leidgens S, De Smet S, Foury F. Frataxin interacts with Isu1 through a conserved tryptophan in its beta-sheet. Hum Mol Genet 2010;19(2):276–86.
[103] Adinolfi S et al. Bacterial frataxin CyaY is the gatekeeper of iron-sulfur cluster formation catalyzed by IscS. Nat Struct Mol Biol 2009;16(4):390–6.
[104] Ayala-Castro, C., A. Saini, and F.W. Outten, *Fe-S cluster assembly pathways in bacteria.* Microbiol Mol Biol Rev, 2008. **72**(1): p. 110-25, table of contents.
[105] Morse RP et al. Structural basis of toxicity and immunity in contact-dependent growth inhibition (CDI) systems. Proc Natl Acad Sci U S A 2012;109(52):21480–5.
[106] Grant GA. The ACT domain: a small molecule binding domain and its role as a common regulatory element. J Biol Chem 2006;281(45):33825–9.
[107] Maris C, Dominguez C, Allain FH. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. FEBS J 2005;272(9):2118–31.
[108] Nielsen MS et al. The 1.5 A resolution crystal structure of [Fe3S4]-ferredoxin from the hyperthermophilic archaeon Pyrococcus furiosus. Biochemistry 2004;43(18):5188–94.
[109] Li N et al. Structure of Ustilago maydis killer toxin KP6 alpha-subunit. A multimeric assembly with a central pore. J Biol Chem 1999;274 (29):20425–31.
[110] Blondeau K et al. Crystal structure of the effector AvrLm4-7 of Leptosphaeria maculans reveals insights into its translocation into plant cells and recognition by resistance proteins. Plant J 2015;83(4):610–24.
[111] Carte J et al. Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. Genes Dev 2008;22(24):3489–96.
[112] Graebsch A, Roche S, Niessing D. X-ray structure of Pur-alpha reveals a Whirly-like fold and an unusual nucleic-acid binding surface. Proc Natl Acad Sci U S A 2009;106(44):18521–6.
[113] Schumacher MA et al. Crystal structures of T. brucei MRP1/MRP2 guide-RNA binding complex reveal RNA matchmaking mechanism. Cell 2006;126 (4):701–11.
[114] Desveaux D et al. A new family of plant transcription factors displays a novel ssDNA-binding surface. Nat Struct Biol 2002;9(7):512–7.
[115] Perez-Riba A, Itzhaki LS. The tetratricopeptide-repeat motif is a versatile platform that enables diverse modes of molecular recognition. Curr Opin Struct Biol 2019;54:43–9.
[116] Rawlings ND, Barrett AJ, Bateman A, Merops,. the peptidase database. Nucleic Acids Res 2010;38(Database issue):D227. D233.
[117] Smith EL et al. The complete amino acid sequence of two types of subtilisin. *BPN' and Carlsberg* J Biol Chem 1966;241(24):5974–6.
[118] Wheatley JL, Holyoak T. Differential P1 arginine and lysine recognition in the prototypical proprotein convertase Kex2. Proc Natl Acad Sci U S A 2007;104 (16):6626–31.
[119] Schaller A et al. From structure to function – A family portrait of plant subtilases. New Phytol 2018;218(3):901–15.
[120] Artenstein AW, Opal SM. Proprotein convertases in health and disease. N Engl J Med 2011;365(26):2507–18.
[121] Tangrea MA et al. Solution structure of the pro-hormone convertase 1 pro-domain from Mus musculus. J Mol Biol 2002;320(4):801–12.
[122] Foophow T et al. Crystal structure of a subtilisin homologue, Tk-SP, from Thermococcus kodakaraensis: requirement of a C-terminal beta-jelly roll domain for hyperstability. J Mol Biol 2010;400(4):865–77.
[123] Murayama K et al. Crystal structure of cucumisin, a subtilisin-like endoprotease from Cucumis melo L. J Mol Biol 2012;423(3):386–96.
[124] Kojima S, Minagawa T, Miura K. The propeptide of subtilisin BPN' as a temporary inhibitor and effect of an amino acid replacement on its inhibitory activity. FEBS Lett 1997;411(1):128–32.
[125] Wickner RB. Chromosomal and nonchromosomal mutations affecting the" killer character" of Saccharomyces cerevisiae. Genetics 1974;76(3):423–32.
[126] Booth MC et al. Structural analysis and proteolytic activation of Enterococcus faecalis cytolysin, a novel lantibiotic. Mol Microbiol 1996;21(6):1175–84.
[127] Rojas-Lopez M et al. Identification of the Autochaperone Domain in the Type Va Secretion System (T5aSS): Prevalent Feature of Autotransporters with a beta-Helical Passenger. Front Microbiol 2017;8:2607.
[128] Veith PD et al. Type IX secretion: the generation of bacterial cell surface coatings involved in virulence, gliding motility and the degradation of complex biopolymers. Mol Microbiol 2017;106(1):35–53.
[129] Leyton DL, Rossiter AE, Henderson IR. From self sufficiency to dependence: mechanisms and factors important for autotransporter biogenesis. Nat Rev Microbiol 2012;10(3):213–25.
[130] Coutte L et al. Subtilisin-like autotransporter serves as maturation protease in a bacterial secretion pathway. The EMBO journal 2001;20(18):5040–8.
[131] Chen JM. Mycosins of the Mycobacterial Type VII ESX Secretion System: the Glue That Holds the Party Together. mBio 2016;7(6).
[132] Bunduc CM et al. Structure and dynamics of a mycobacterial type VII secretion system. Nature 2021;593(7859):p. 445-+.
[133] Duckert P, Brunak S, Blom N. Prediction of proprotein convertase cleavage sites. Protein Eng Des Sel 2004;17(1):107–12.
[134] Bhooshan S, Negi V, Khatri PK. *Staphylococcus pseudintermedius: an undocumented, emerging pathogen in humans.* GMS Hyg. Infect Control 2020;15:p. Doc32.
[135] Todd ECD, Notermans S. Surveillance of listeriosis and its causative pathogen. *Listeria monocytogenes* Food Control 2011;22(9):1484–90.
[136] Lee MR et al. Mycobacterium abscessus Complex Infections in Humans. Emerg Infect Dis 2015;21(9):1638–46.
[137] Osdaghi E, Young AJ, Harveson RM. Bacterial wilt of dry beans caused by Curtobacterium flaccumfaciens pv. flaccumfaciens: A new threat from an old enemy. Mol Plant Pathol 2020;21(5):605–21.
[138] Fan B, Bacillus velezensis FZB42 in,, et al. The Gram-Positive Model Strain for Plant Growth Promotion and Biocontrol. Front Microbiol 2018;2018:9.
[139] Lu P et al. Complete Genome Sequence of Terribacillus aidingensis Strain MP602, a Moderately Halophilic Bacterium Isolated from Cryptomeria fortunei in Tianmu Mountain in China. Genome Announc 2015;3(2).