



OPEN Prediction of secreted uncharacterized protein structures from *Beauveria bassiana* ARSEF 2860 unravels novel toxins-like families

Peter F. Farag , Aya A. Elsisy, Esraa W. Elabd, Jana J. Sadek, Nada H. Mousa, Rawan M. Zaky & Sara M. Ahmed

Insecticides are toxic substances used to control a wide variety of agricultural insect pests. Most of these are chemicals in nature, and their increasing residues in soil, water, and fruits contribute to environmental pollution, chronic human illnesses, and the emergence of insecticide resistance phenomenon. In the context of a green environment, bioinsecticide metabolites, including proteins, are a safe alternative that mostly has selective toxicity to insects. Thus, this study aimed to predict and identify new toxin-like families through uncharacterized secreted proteins from one of the most potent entomopathogenic fungi, *Beauveria bassiana* ARSEF 2860, which was selected as a model. In this work, a total of 2483 amino acid sequences of uncharacterized proteins (Ups) were retrieved from the RefSeq database. Among these, 365 Ups were identified as secreted proteins using the SignalP web server. We implemented the integration of well-designed bioinformatic tools to characterize and anticipate their homologous similarities at the sequence (InterPro) and structural (AlphaFold2) levels. The structural function annotation of these proteins was predicted using DeepFRI. With 269 successfully predicted folds, we identified new putative families with pathogenesis functions related to toxins like Janus-faced atracotoxins (insecticidal spider toxin), Cry toxins (commercial insecticide from *Bacillus thuringiensis*), ARTs-like toxins, and other insecticidal toxins. Furthermore, some proteins that are not homologous to any known experimental data were functionally predicted as cation metal ion binding (Zn, Na, and Co) with potential toxicity. Collectively, computational structural genomics can be used to study host–pathogen interactions and predict novel families.

Keywords AlphaFold2, *Beauveria bassiana*, Insecticidal proteins, New families, Structural annotations

Abbreviations

AF2	AlphaFold2
AF3	AlphaFold3
AFDB	AlphaFold database
ART	ADP-ribotransferase
BP	Bubble protein
GO	Gene ontology
pLDDT	Predicted local distance difference test
pTM	Predicted template modeling
SP	Signal peptide
Ups	Uncharacterized proteins

Beauveria bassiana (Balsamo-Crivelli) Vuillemin is one of the most devastating necrotrophic soil-borne entomopathogenic fungus that belongs to the Cordycipitaceae family^{1,2}. It causes white muscardine disease in more than 700 insects and spider mite species across 15 orders and 149 families^{3,4}. This fungus is characterized

Department of Microbiology, Faculty of Science, Ain Shams University, Cairo 11566, Egypt. ✉email: peter_jireo@sci.asu.edu.eg

by secreting several natural selective pigments and toxins (beauvericin, bassiana, bassianolide, tenellin, beauverolides, oosporein, and so forth) with highly virulent effect, making it commercially manufactured as an eco-friendly mycoinsecticide and used in Integrated Pest Management (IPM) programs^{3,6}. *B. bassiana* ARSEF 2860 is a popular strain for pest control⁷. It was isolated from *Schizaphis graminum* (wheat aphid), and its genome (GenBank accession number ASM28067v1) was recently assembled^{18,9}. This enables the study of various genes that encode insecticidal proteins, illuminating the mechanism of action and their virulence against insects.

Many proteins encoded by most microorganisms are known to lack experimental proof of translation for their in vivo expression. These proteins, which constitute between 20 and 50% of the protein-coding regions, are referred to as hypothetical proteins (HPs), and their roles remain unclear^{10–12}. HPs can be categorized as uncharacterized proteins (UPs) and the domain of unknown functions (DUF)¹³. Despite the experimental confirmation of their existence, UPs have not yet been named or linked to a known gene. In contrast, DUFs are proteins identified through experiments but do not have recognized structural or functional domains¹⁴. Most of these proteins are expected to play crucial roles, and their annotation may uncover new domains and motifs, functional pathways, and discoveries of novel pathogenesis-related genes with putative toxin-like family homology^{15,16}.

While many obstacles remain in annotating these types of proteins, numerous bioinformatics tools, including databases and web servers, are available for functional annotation and homology assignment for UPs^{16,17}. Despite the widespread popularity of sequence-based annotation tools like BlastP and InterPro, many protein sequences remain unclassified and functionally unannotated¹⁸. This could be attributed to the rapid divergence between homologous proteins, leading to diminished sequence similarity¹⁹. Structure-based annotation depends on building a protein's three-dimensional (3D) structure, which reveals molecular functions, novel folds, and structural similarities, resulting in enhanced genomic annotations^{20,21}. Experimental structure determination is a costly and time-consuming process, making computational structure prediction an appealing alternative for reducing the effort needed to obtain a structural model from months of laboratory work to just a few keystrokes^{22,23}.

The AlphaFold Protein Structure Database (AFDB) is a publicly accessible data collection of protein structures and their confidence metrics (pLDDT 0–100), generated by the AlphaFold2 (AF2) artificial intelligence system^{24,25}. Recently, an AlphaFold 3 (AF3) webserver was launched to predict structures molecular interactions²⁶. The homologous structural similarity of proteins can be measured using predicted template modeling (pTM) scores, which range from 0 to 1. A TM-align score greater than 0.5 indicates the evolutionary relatedness between two structures adopting the same fold. Thus, structural comparison can reveal this startling similarity, which remains elusive for BLAST and other sequence-based methods such as HHblits²⁷.

Considering the significance of studying the fungal secretome (the proteins secreted outside the plasma membrane) that play a vital role in host–pathogen interactions, knowledge about these proteins remains limited in many fungal species^{28–31}. So, the purpose of this study is to describe the secreted UPs encoded by *Beauveria bassiana* ARSF 2680, aiming to explore potential novel toxin-like families based on structural annotations that enhance the understanding of the mechanism of action. This will be achieved by constructing 3D structures for these proteins and then utilizing the homology model to predict their functions (structure-based functions) as well as paralogue and orthologue similarities.

Materials and methods

Sequence information and retrieval

The information data of *Beauveria bassiana* ARSEF 2860 (Genbank accession no. ADAH00000000.1) was submitted to the NCBI database by Zhejiang University, China⁹. The genome of this strain has a length of 33.7 Mb and contains 10,364 genes encoding 10,364 proteins. Out of these, 2483 (24%) of the proteins were classified as UPs, whereas 7881 (76%) were fully characterized proteins. The proteome of this strain was retrieved from the NCBI RefSeq database (https://www.ncbi.nlm.nih.gov/datasets/gene/GCF_000280675.1/, accessed on 25 August 2024) that was submitted recently in January 2024.

Screening of secreted proteins

Secreted proteins that carry a signal peptide (SP) were predicted using SignalP v5.0³². DeepTMHMM V1.0.24 and TMHMM 2.0 web servers were used to detect the transmembrane helix proteins³³. The candidates were excluded if they contained any transmembrane helices.

Domain prediction, homologous similarity, and clustering

The domains of 365-secreted protein sequences were predicted by InterPro 98³⁴. Based on structure prediction, the domains were first screened through AFDB v2 and then re-predicted by AF2 using ColabFold v1.5.5 implementation after removing the SP sequence³⁵. Five models were created for every protein, and we selected the top model based on its best average pLDDT score. RUPEE^{36,37} was used for searching homologous structural similarity against SCOPe v2.08, CATH v4.3, and PDB chain databases downloaded on 16 July 2022 (Top aligned, Full length), while BlastP was used for sequence similarity searches. Two structures were deemed similar if their TM scores exceeded 0.5. A similar network was established based on structural similarity through all-against-all comparisons using the DALI server³⁸. Cytoscape v3.10 was employed to construct and visualize the protein network, and ChimeraX v1.6.1 was used for visualizing 3D protein structures^{39,40}.

Prediction of protein functions

The putative functional annotation to secreted UPs was performed using a DeepFRI (cut-off score ≥ 0.5) to predict enriched Gene Ontology (GO) terms of biological process (BP) and molecular function (MF) based on protein structures⁴¹. Argot^{2.5} web server (cut-off score > 200) was used to predict enriched GO terms according

to protein sequences⁴², where these sequences were downloaded from the NCBI database in FASTA format. ToxinPred 2.0 (cut-off score > 0.2) was used for the anticipated toxicity of some sequences⁴³.

Biomolecular interactions and molecular docking analysis

The interactions between a subset of UPs and other molecules, including ions, nucleic acids, small molecules, and modified residues, were predicted using the AF3 server²⁶. Protein–ligand docking was performed using CB-Dock2⁴⁴, whilst protein–protein docking was conducted between the receptor and target proteins using the HDock server⁴⁵.

Results and discussion

Statistical insight into describing the uncharacterized secreted proteins

In this study, we used sequence and structure predictions on the putatively secreted UPs ($n = 365$). The structural folds of 269 proteins (74%) were predicted using AF2 with high confidence scores (pLDDT > 70), resulting in a significant value ($p < 0.0001$) compared to projected low confidence scores (Fig. 1A, Supplementary Table S1). Only 68 proteins could be predicted from primary sequences, with more than 80% lacking InterPro annotation (Fig. 1B). Additionally, most of the detected results involved in this study were annotated according to the structure of these proteins. These results are not surprising; other recent research utilized structural annotations to identify novel families through AI machine-learning modeling^{18,19,22,25}. Using Argot2.5 and DeepFRI for functional annotation, most toxin-like clusters exhibit a putative pathogenesis biological process (GO:0009405), highlighting the potential role of these proteins in the fungus' pathogenicity against its host. To reveal the different putative toxin families, the high pLDDT-predicted UP structures were clustered based on a structural alignment against other toxin proteins with known structures. Finally, we summarized the various protein clusters as follows:

Identification of predicted toxin-like proteins with a knottin fold

We investigated structural similarity within two distinct clusters, including the knottin fold. We discovered that these clusters might represent new families: the first resembles a spider toxin family, while the second is similar to bubble proteins. Knottins, or inhibitor cystine knots (ICKs), are structural cysteine-rich protein families (30–50 amino acids) that are classified in the SCOPe database^{46,47}. This family is found in many living organisms, including the toxins of venomous animals like spiders and scorpions. Additionally, it features a distinctive knotted structure formed by three intramolecular disulfide linkages, which offer great chemical, thermal, and proteolytic stability⁴⁸.

After matching with the SCOPe fold database v2.08, we discovered several genes, including BBA_06834, BBA_01324, BBA_03436, BBA_08673, BBA_09303, and BBA_09080, that encoded proteins similar to spider toxin structures. Spider toxin proteins belong to a diverse family of knottins that include various insecticidal peptides targeting neuronal ion channels and receptors. The outstanding specificity, efficacy, and stability of these peptides have attracted significant interest as potential eco-friendly insecticides⁴⁹. From these genes, BBA_01324 (pLDDT 93.8) was similar to delta-theraphotoxin with a TM score of about 0.65. This gene is

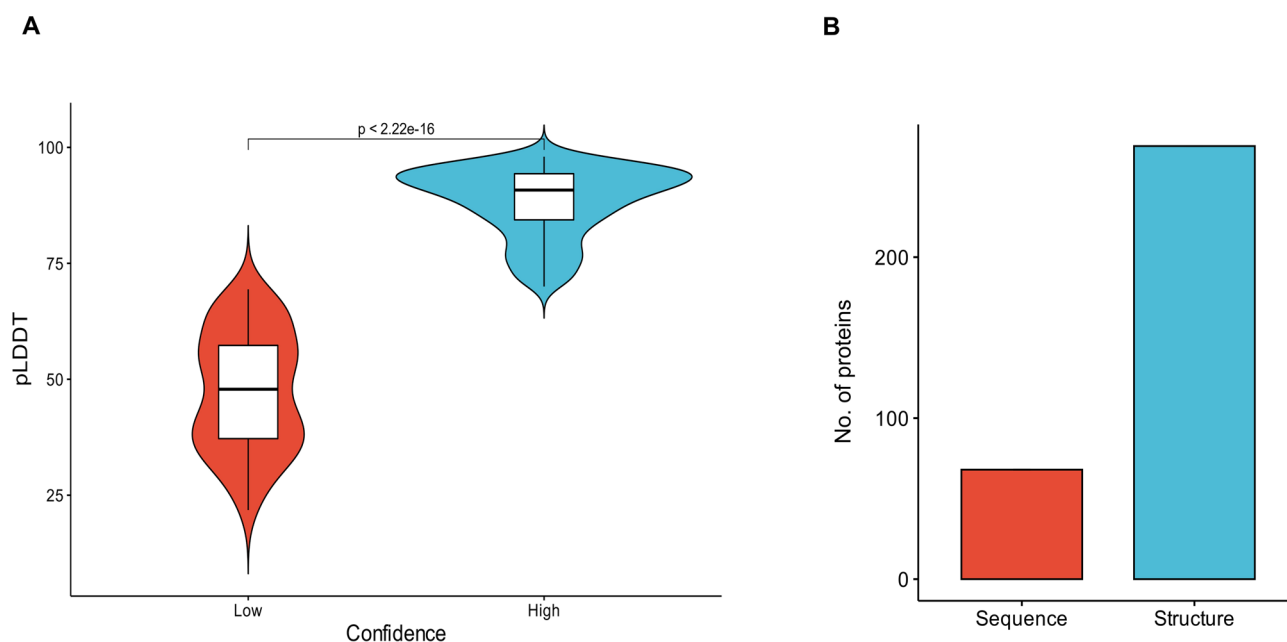


Fig. 1. (A) The violin box plot showed successful screening for folds with high confidence scores (pLDDT > 70). (B) The bar plot showed the comparison between the number of annotated proteins by InterPro (sequence-based annotation, $n = 68$) and AF2 (structure-based annotation, $n = 269$).

species-specific to *Beauveria* based on sequence (BLASTp) and structure (AFDB) levels (Fig. 2A). Moreover, we investigated a vital gene (BBA_08673), exclusive only to this strain, similar to insect-selective neurotoxin Janus-faced atracotoxins (J-ACTXs) from the venom of the Australian funnel-web spider (*Hadronyche versuta*) that seems to be a promising target for insects⁵⁰. Figure 2B demonstrates a very high confidence score (pLDDT 90.3) of this protein structure and the superposed score (TM 0.71), indicating an acceptable similarity to J-ACTX (PDB 1DL0). This toxin is a specific blocker of insect K(Ca) channels⁵¹. Thus, the molecular docking analysis between the original toxin and our protein against the *Drosophila* K (ca) channel (PDB 7PXF) indicated a similar active site and binding affinity (Fig. 2C). While J-ACTX shares a disulfide connection pattern akin to other ICKs, it exhibits limited sequence homology to any protein and DNA sequence databases⁵². The structural similarity between the fungal and spider toxins gives the fungus an edge in mass production for a scalable industry.

Another knottin family had similarities to bubble protein (BP), which was initially identified in *Penicillium brevicompactum* exudate and may act as a toxin against fungi⁵³. This cluster contains seven BP-like proteins, was the average TM-score of representatives aligned with the experimental BP (PDB 1UOY) being 0.65, as illustrated in Fig. 3. These proteins are structurally similar to other antifungal proteins, including the *P. chrysogenum* antifungal protein (PAF) family. BP is categorized as a member of defensins, which consist of five beta sheets and structural classification places BP into knottin fold-containing proteins⁵⁴. Thus, this investigation suggests that *B. bassiana* might be a good biopesticide against insects and fungi.

Clustering of Cry toxin-like proteins

With structure-based clustering, we were able to capture a new family specific to entomopathogenic fungi that are related to *Bacillus thuringiensis* (Bt) Cry toxins (Fig. 4). Crystal (Cry) proteins are selective, pore-forming toxins that specially target midgut invertebrates and are generally innocuous to mammals. They are widely employed as agricultural pesticides to eliminate insects and nematodes⁵⁵. A cluster of five proteins encoded by BBA_06207, BBA_01385, BBA_09344, BBA_10262, and BBA_07997 genes in *B. bassiana* exhibited varied degrees of resemblance to Cry51Aa1 Cry toxin, an insecticidal aerolysin-type β -pore-forming toxin composed of 309 amino acid sequences⁵⁶. Compared with the members in this cluster, BBA_06207 from *B. bassiana* exhibits higher folding similarity (E-value $< 10^{-7}$) than other proteins from the same fungus or other entomopathogenic fungi. Concurrently, BBA_09344, BBA_07997, BBA_10262, and BBA_01385 have structural alignment with TM prediction scores greater than 0.7 (Fig. 4A). The BBA_09344 gene exhibited the most common fold sharing among other genes, making it suitable for clustering with proteins identified in other entomopathogenic fungi through AFDB clusters and aligned using the DALI server, resulting in Z-score outputs. The dendrogram is generated by average linkage clustering of the structural similarity matrix (DALI Z-scores), with a similarity cutoff at $Z = 2$ ⁵⁷. The results reveal a structural similarity dendrogram between them, giving a high Z DALI score ($Z = 19.3$) and a strong similarity between a query protein in *B. bassiana* and *Cordyceps javanica* protein (IF1G_09556) (Fig. 4B).

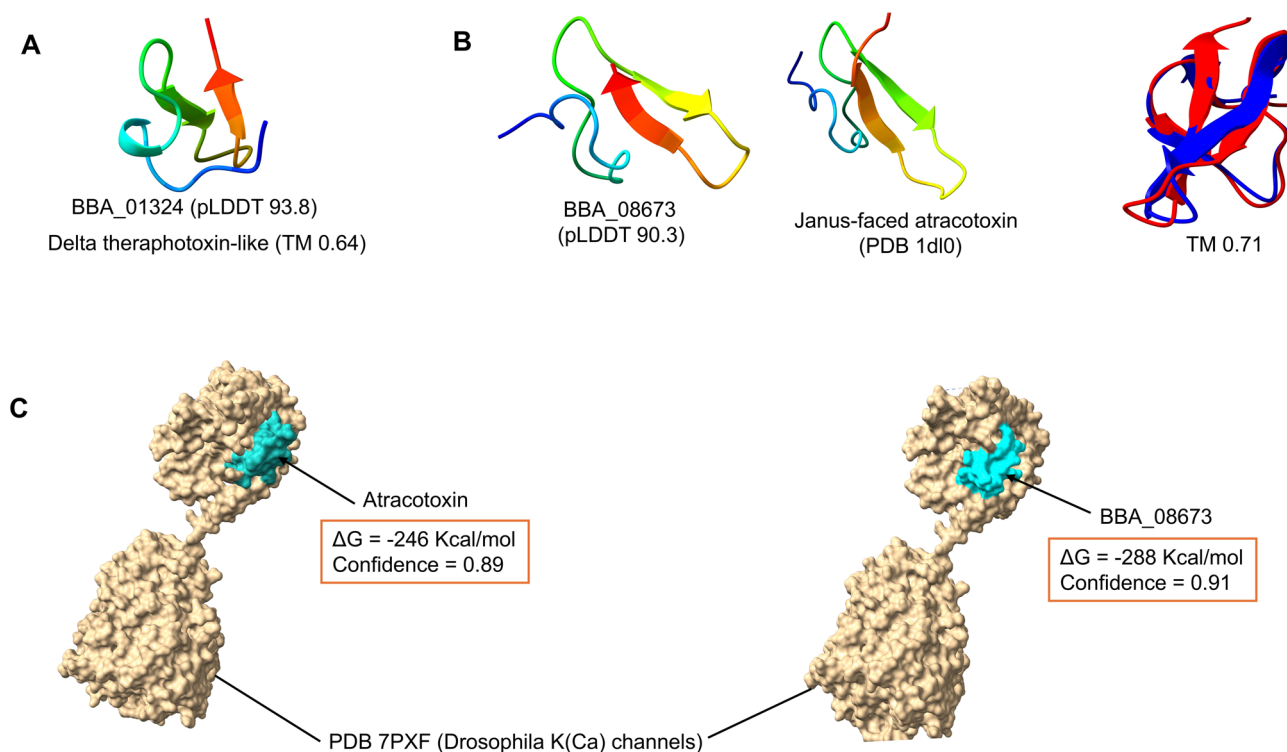


Fig. 2. New spider toxin-like families. (A) delta-theraphotoxin-like protein (BBA_01324) with a TM score equal to 0.64. (B) J-ACTX-like protein (BBA_08673) superposed to Atracotoxin PDB 1dl0 (TM 0.71). (C) Docked complex between J-ACTX and BBA_08673 against *Drosophila* K (Ca) channel.

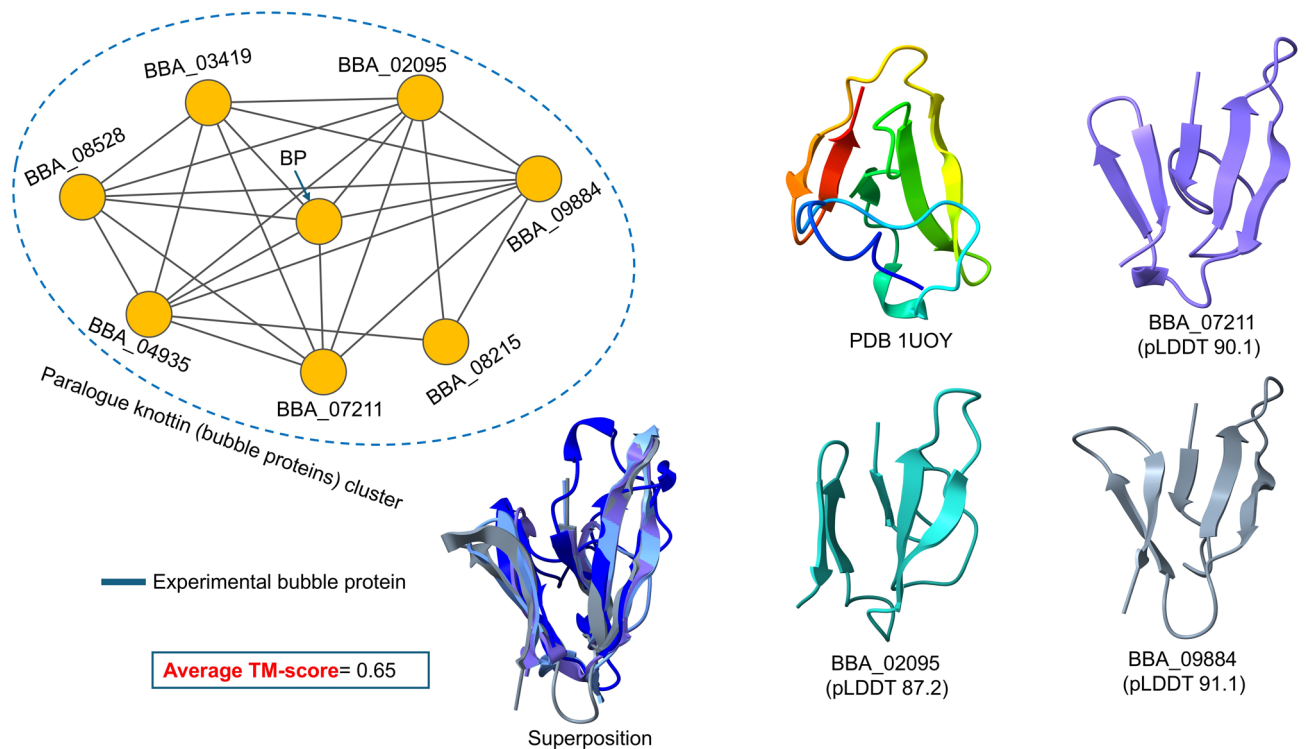


Fig. 3. A cluster of seven BP-like and three high-confidence score models with an average alignment score of 0.65 compared to experimental bubble protein (PDB 1UOY). Grey edges represent structural similarities among these proteins.

Figure 4C,D illustrate the putative 3D structure and active site of BBA_06207, a likeness to Cry51Aa1, which is rich with threonine residues in the middle of the backbone. In previous works, serine or threonine residues were shown to make up about 23% of the different types of Cry toxins, including Cry51Aa1^{56,58}.

Putative insecticidal GNIP1Aa toxin-like proteins

We identified two proteins (BBA_02700 and BBA_09997) that exhibit structural similarities to the insecticidal GNIP1Aa protein, which belongs to the membrane attack complex/PerForin (MACPF) family. BBA_02700 demonstrated a very high confidence score (pLDDT 92.3) and was superposed onto GNIP1Aa (PDB: 6FBM) with a TM value of 0.57 (E-value $< 10^{-5}$) (Fig. 5A). Also, BBA_09997 presented a high confidence score (pLDDT 85.2) and was aligned onto GNIP1Aa (PDB: 6FBM) with a TM value of 0.52 (E-value $< 10^{-5}$) (Fig. 5B). GNIP1Aa is a protein identified from *Chromobacterium piscinae* in 2017 that displays particular toxicity against Western corn rootworm (WCR), one of the most destructive corn pests in the United States. Although GNIP1Aa belongs to the same class of Cry toxins, it is distinct from all insect-control treatments currently available on the market that utilize modern agricultural technologies. Due to its distinctiveness and protein activity, GNIP1Aa is a strong commercial candidate for development into a transgenic product. Such a solution would be highly effective in preventing crop loss in corn and delaying the emergence of pest resistance^{59,60}. Other entomopathogenic fungi, including *Cordyceps javanica* (Gene: IF1G_04403), *Ophiocordyceps camponoti-leonard* (Gene: CP532_0387), and *Metarhizium anisopliae* (Gene: MAN_10237) have comparable structures following structure-based grouping.

Clustering of ARTs-like toxins

One of the largest clusters, including 17 members, was described with ADP-ribosylation fold and NAD⁺-dependent ADP-ribosyltransferase activity according to MF (GO:0003950); six of them possessed predicted structures with estimated TM > 0.6 matching known homologous proteins (Fig. 6). Catalysis of these proteins evolved by a structural superfamily of enzymes, called ADP-ribosyltransferases (ARTs) with NAD⁺ as a co-substrate⁶¹. The paralogue distribution of these proteins was clustered into five groups and one singleton as shown in Fig. 6A, where the orthologue similarity of these groups was exclusive only to entomopathogenic fungi, especially *Metarhizium* and *Cordyceps* species, based on the sequence and structure homology clustering. Cluster 2 contained two proteins (BBA_04708 and BBA_04559) similar to diphtheria toxin (DT), a secreted exotoxin by *Corynebacterium diphtheriae*, with a high confidence score (pLDDT 92.2) and TM-align score about 0.65 (Fig. 6B). Figure 6C illustrated representative of two protein structures (BBA_03706 and BBA_07827) from Cluster 4, which analyzed for structural similarity to the heat-labile enterotoxin. The structure of BBA_03706 was determined with high confidence (pLDDT = 87.9), and structural alignment with the heat-labile enterotoxin revealed high structural similarity with a TM-score of 0.85. Similarly, the structure of BBA_07827 was determined with very high confidence (pLDDT = 92.7), and alignment with the heat-labile enterotoxin also showed high

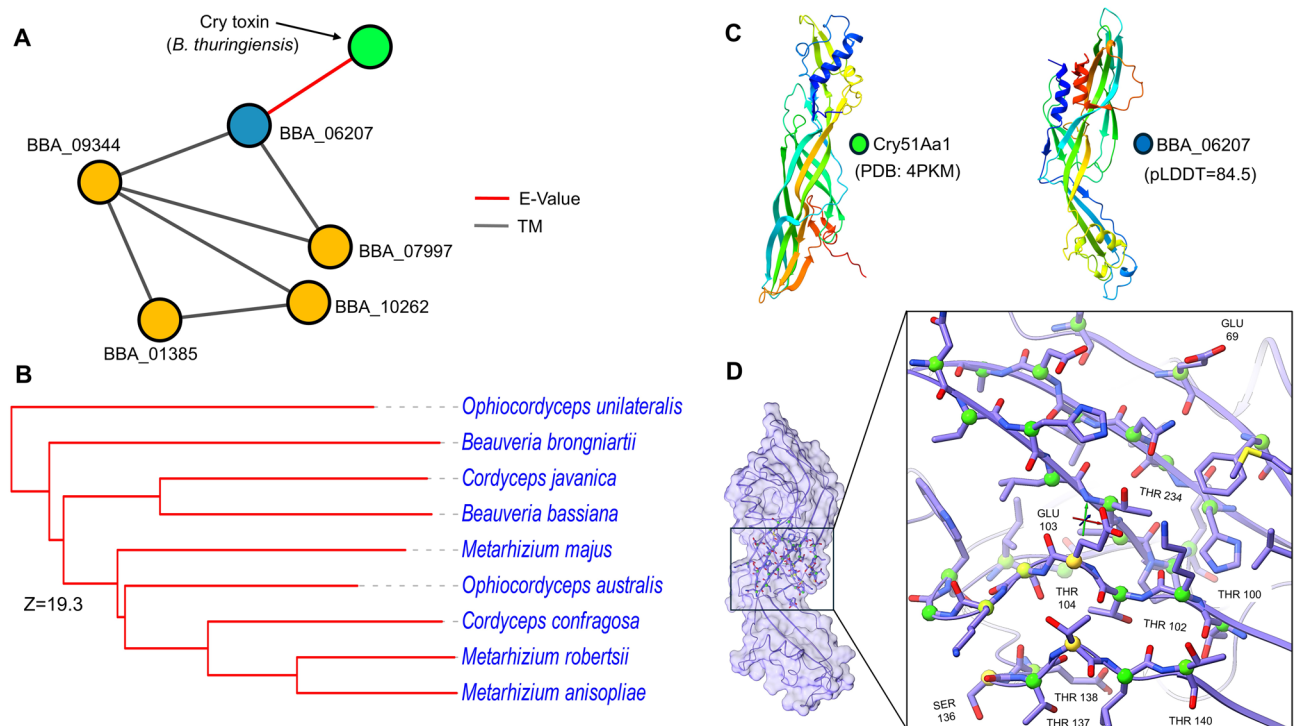


Fig. 4. Putative new Cry toxin-like family based on structural homology. **(A)** The network of members of *B. bassiana* compared to Bt Cry toxin Cry51Aa1. Each orange node is similar to each other by calculating TM-score > 0.7 (Grey edge). In contrast, the green node represents the BBA_06207 protein (Blue node) with a high E-value (Red edge) structural similarity with Bt Cry toxin (Green node). **(B)** Structural similarity dendrogram between highly similar protein from other entomopathogenic fungi to BBA_09344 protein. The average DALI Z-score was 19.3 (cutoff Z = 2). **(C)** 3D structure comparison between Cry51Aa1 (PDB: 4PKM) and BBA_06207 (confidence score = 84.5). **(D)** Active site of BBA_06207 with threonine-rich residues.

similarity, with a TM-score of 0.71. Historically, various classes of enzymes form the ART superfamily, including the diphtheria-toxin-like transferases (ARTDs) and the cholera-toxin-like transferases (ARTCs)⁶². Although it is uncommon for secreting toxins to resemble human-infecting bacterial toxins from entomopathogenic fungi, Aravind et al.⁶³ reported the putative expanded evolution of ARTs throughout eukaryotes by horizontal gene transfer.

Discovery of potential novel families with putative toxicity

Several proteins were identified with no homology to any known experimental proteins (TM < 0.4), forming novel families with putative MFs and biological processes. The selection criteria were critical as they concentrated on anticipated toxicity using ToxinPred 2.0 and were restricted to the entomopathogenic fungus group. Furthermore, a set of proteins appeared to share the same MF: cation metal ion binding (GO:0043169). From these proteins, BBA_01910 (209 amino acids) contains two repeated motifs fused with intrinsically disordered regions (IDRs) detected by the ODINPred server⁶⁴, in which proteins with IDRs are noted to be highly prevalent in diseases⁶⁵ (Fig. 7A,B). ToxinPred 2.0 predicted the toxicity of the motif sequence (ATCEPHEDHWHCPAGVPQPSLNPDPGTPNPKATQ) with a score of approximately 0.75. To predict the type of cation metal ions, AlphaFold3 (AF3) was utilized for biomolecular interaction detection between the query protein and various ions, with results provided by interface-predicted template modeling (ipTM) scores²⁶. Zinc ion interacted with six histidine residues (His-tag) and was the best matching ion to the studied protein giving a high ipTM score equal to 0.92 (Fig. 7C). Other novel folds were detected from genes BBA_09398 (pLDDT 85.4) and BBA_02207 (pLDDT 80.2) (Fig. 8). BBA_09398 (236 a.a.) binds to sodium ion through three residues (alanine, asparagine, and cysteine) with a high ipTM score (0.89) (Fig. 8A), while BBA_02207 (215 a.a.) attaches with cobalt ion through histidine residue (ipTM 0.91) (Fig. 8B). Like humans and plants, insects depend on various metal ions such as zinc, sodium, and calcium for proper physiological functions, where chelating these ions by any compound may block their functions⁶⁶. Zinc metal is a necessary cofactor for many enzymes and is involved in a variety of processes, such as DNA synthesis, oxidation reactions, and cuticle production⁶⁷. In neurons and other excitable cells, sodium ions are necessary for the propagation of the action potential. Numerous synthetic and naturally occurring neurotoxins, including various types of insecticides, target sodium channels due to their crucial functions in electrical signaling⁶⁸. Cobalt has a key role in the synthesis of hemoglobin, which is necessary for insects to transport oxygen, as well as the metabolism of lipids, carbohydrates, and amino acids⁶⁹.

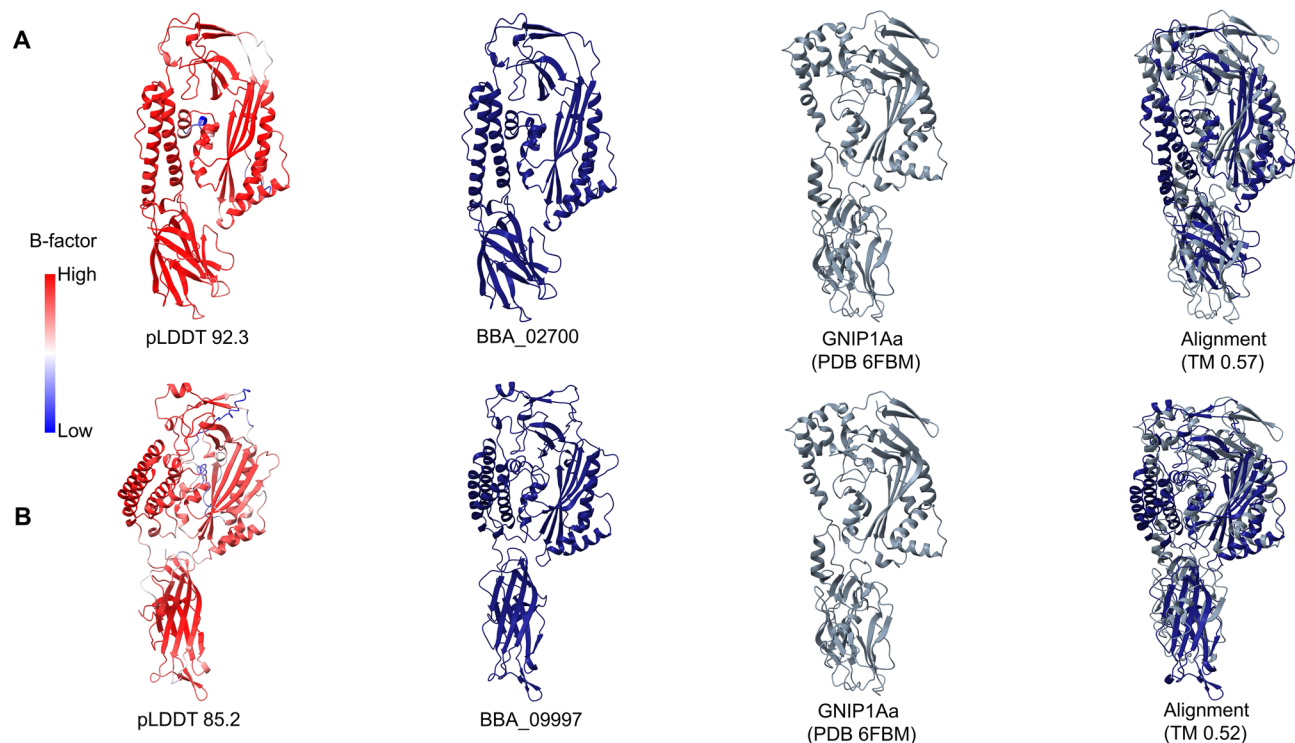


Fig. 5. Structural superposition between the predicted and solved structures. **(A)** The structure of BBA_02700 protein predicted by AF2 was superposed with TM-align against the experimental structure (PDB: 6FBM). **(B)** The structure of BBA_09997 protein was superposed with TM-align against 6FBM.

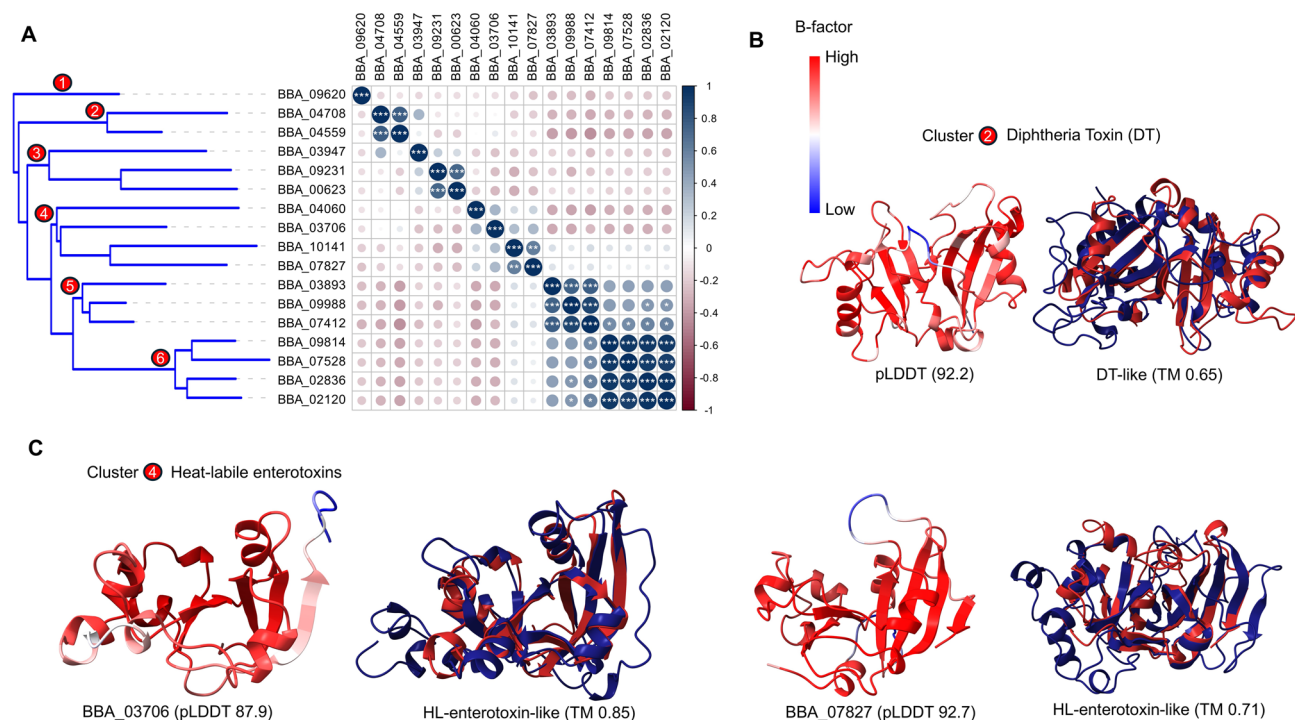


Fig. 6. Structural similarity analysis of ADP-ribosylation fold-containing proteins. **(A)** Dendrogram and similarity matrix of 17 clustering members. **(B)** Homology of Cluster 2 to diphtheria toxin. **(C)** Homology of Cluster 4 to heat-labile enterotoxin. * $P < 0.05$, ** $P < 0.01$, and *** $P < 0.001$.

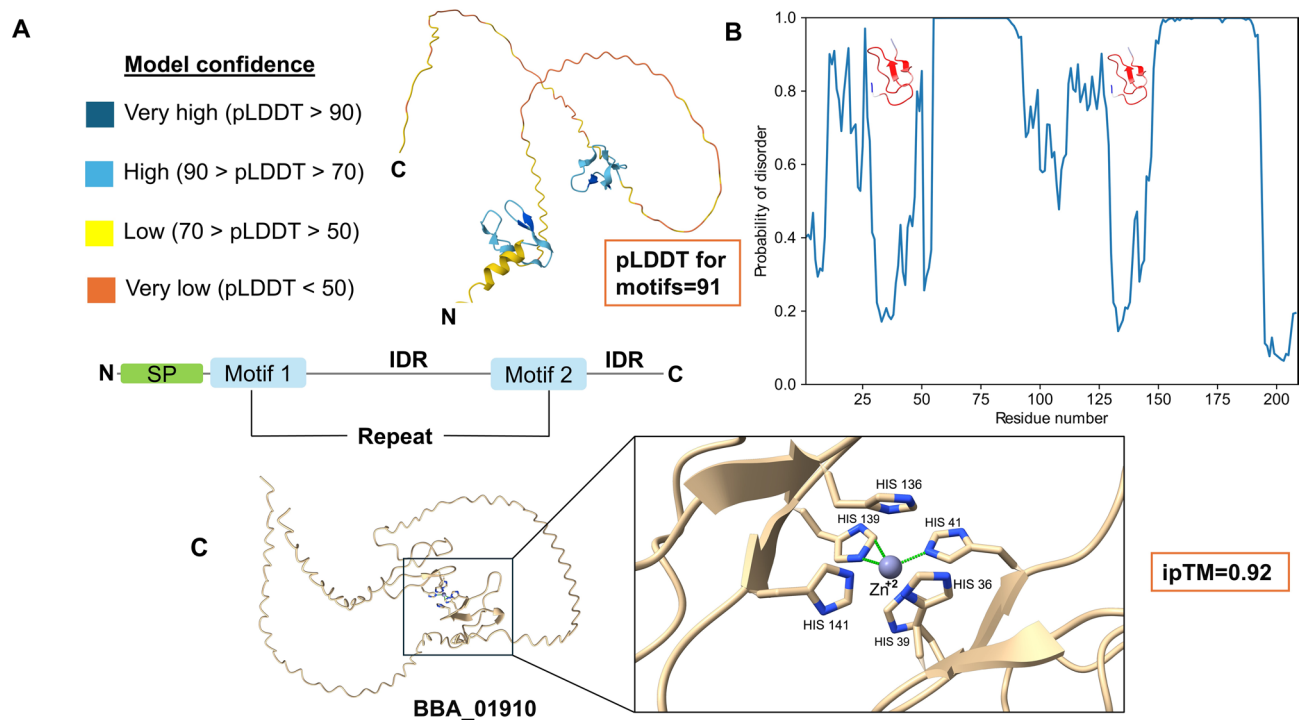


Fig. 7. Characterization of BBA_01910 protein structure. **(A)** Prediction of the protein and folding of its repeated motifs with a confidence scale. **(B)** ODINPred was used to predict the intrinsically disordered regions for the protein (cutoff > 0.5). **(C)** Prediction of the binding sites between the BBA_01910 protein and zinc ion using AF3.

In conclusion, the prediction of secreted uncharacterized protein structures from *Beauveria bassiana* ARSEF 2860 uncovered: (i) new pathogenesis-related proteins belong to putative toxin-like families, most of which exhibit potential pesticides for controlling insects and fungi, (ii) the evolution of expanded putative ADP-ribose transferases (ARTs-like family), (iii) mechanisms and functions of nonhomologous novel folds.

Limitations and future perspectives

While computational structural genomics has proven to be an excellent supplement to the costly and time-consuming wet lab setting, some limits in our work must be acknowledged. Firstly, AF2 was unable to predict approximately 25% of protein structures. Secondly, several proteins did not fit with any known annotation category. Lastly, the structural prediction is insufficient to predict the putative functions. Despite these limitations, in silico structural-based annotation is a first step toward future studies that will focus on in vitro and in vivo validations of such proteins to assess their insecticidal efficacy and potential for agricultural uses. Regarding the non-homology protein structures, future advances might lead to their complete annotation, and these structures might be useful for the scientific community.

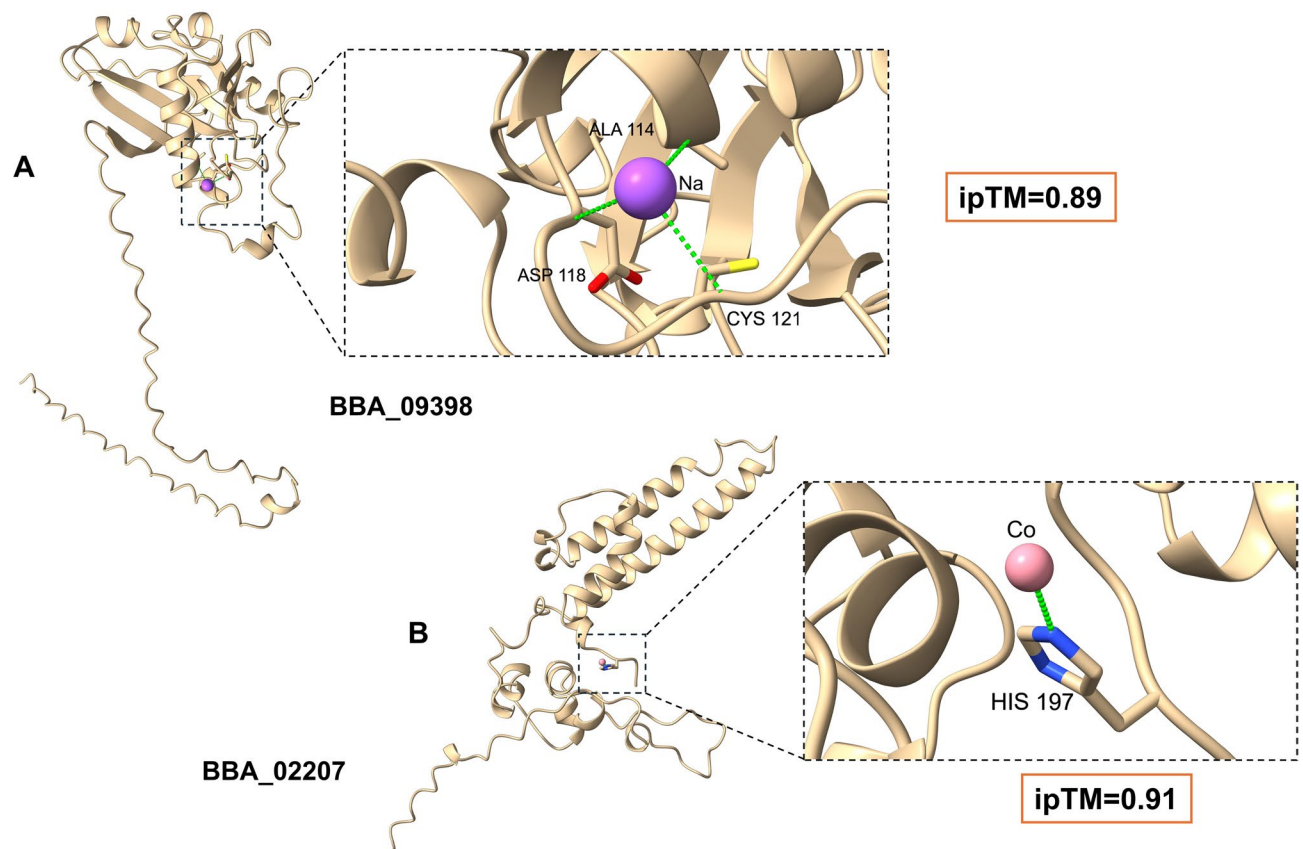


Fig. 8. The interactions between (A) the BBA_09398 protein and sodium ion. (B) the BBA_02207 protein and cobalt ion.

Data availability

The datasets generated and/or analysed during the current study are available in the NCBI RefSeq database under the Bioproject accession number PRJNA225503.

Received: 13 February 2025; Accepted: 14 May 2025

Published online: 22 May 2025

References

- Solano-González, S., Castro-Vásquez, R. & Molina-Bravo, R. Genomic characterization and functional description of *Beauveria bassiana* isolates from Latin America. *J. Fungi* **9**(7), 711. <https://doi.org/10.3390/jof9070711> (2023).
- Ranesi, M. et al. Field isolates of *Beauveria bassiana* exhibit biological heterogeneity in multitrophic interactions of agricultural importance. *Microbiol. Res.* **286**, 127819. <https://doi.org/10.1016/j.micres.2024.127819> (2024).
- Mascarin, G. M. & Jaronski, S. T. The production and uses of *Beauveria bassiana* as a microbial insecticide. *World J. Microbiol. Biotechnol.* **32**(11), 177. <https://doi.org/10.1007/s11274-016-2131-3> (2016).
- Gu, Z. et al. A sensitive method for detecting *Beauveria bassiana*, an insecticidal biocontrol agent, population dynamics, and stability in different substrates. *Can. J. Infect. Dis. Med. Microbiol.* **2023**, 9933783. <https://doi.org/10.1155/2023/9933783> (2023).
- Wang, H., Peng, H., Li, W., Cheng, P. & Gong, M. The toxins of *Beauveria bassiana* and the strategies to improve their virulence to insects. *Front. Microbiol.* **12**, 705343. <https://doi.org/10.3389/fmicb.2021.705343> (2021).
- Muthabathula, P. & Biruduganti, S. Analysis of biodegradation of the Synthetic pyrethroid cypermethrin by *Beauveria bassiana*. *Curr. Microbiol.* **79**(2), 46. <https://doi.org/10.1007/s00284-021-02744-x> (2022).
- Litwin, A., Bernat, P., Nowak, M., Slaba, M. & Różalska, S. Lipidomic response of the entomopathogenic fungus *Beauveria bassiana* to pyrethroids. *Sci. Rep.* **11**(1), 21319. <https://doi.org/10.1038/s41598-021-00702-y> (2021).
- Feng, M. G., Johnson, J. B. & Kish, L. P. Survey of entomopathogenic fungi naturally infecting cereal aphids (Homoptera: Aphididae) of irrigated grain crops in Southwestern Idaho. *Environ. Entomol.* **19**(5), 1534–1542. <https://doi.org/10.1093/ee/19.5.1534> (1990).
- Xiao, G. et al. Genomic perspectives on the evolution of fungal entomopathogenicity in *Beauveria bassiana*. *Sci. Rep.* **2**, 483. <https://doi.org/10.1038/srep00483> (2012).
- Mazumder, L., Hasan, M., Rusd, A. A. & Islam, M. A. In-silico characterization and structure-based functional annotation of a hypothetical protein from *Campylobacter jejuni* involved in propionate catabolism. *Genom. Inform.* **19**(4), e43. <https://doi.org/10.5808/gi.21043> (2021).
- Rahman, M. A., Heme, U. H. & Parvez, M. A. K. In silico functional annotation of hypothetical proteins from the *Bacillus paralicheniformis* strain Bac84 reveals proteins with biotechnological potentials and adaptational functions to extreme environments. *PLoS ONE* **17**(10), e0276085. <https://doi.org/10.1371/journal.pone.0276085> (2022).
- Ang'angò, L. M., Herren, J. K. & Tastan Bishop, Ö. Structural and functional annotation of hypothetical proteins from the microsporidia species *Vittaforma corneae* ATCC 50505 using *in silico* approaches. *Int. J. Mol. Sci.* **24**(4), 3507. <https://doi.org/10.3390/ijms24043507> (2023).

13. Rahman, A., Susmi, T. F., Yasmin, F., Karim, M. E. & Hossain, M. U. Functional annotation of an ecologically important protein from *Chloroflexus aurantiacus* involved in polyhydroxyalkanoates (PHA) biosynthetic pathway. *SN Appl. Sci.* **2**(11), 2020. <https://doi.org/10.1007/s42452-020-03598-x> (2020).
14. Mazumder, L., Hasan, M. R., Fatema, K., Islam, M. Z. & Tamanna, S. K. Structural and functional annotation and molecular docking analysis of a hypothetical protein from *Neisseria gonorrhoeae*: An in-silico approach. *Biomed. Res. Int.* **2022**, 4302625. <https://doi.org/10.1155/2022/4302625> (2022).
15. Ashrafi, H. et al. Structure to function analysis with antigenic characterization of a hypothetical protein, HPAG1_0576 from *Helicobacter pylori* HPAG1. *Bioinformatics* **15**(7), 456–466. <https://doi.org/10.6026/97320630015456> (2019).
16. Abbasi, B. A. et al. In silico characterization of uncharacterized proteins from multiple strains of *Clostridium difficile*. *Front. Genet.* **13**, 878012. <https://doi.org/10.3389/fgene.2022.878012> (2022).
17. Pranavathiyan, G., Prava, J., Rajeev, A. C. & Pan, A. Novel target exploration from hypothetical proteins of *Klebsiella pneumoniae* MGH 78578 reveals a protein involved in host-pathogen interaction. *Front. Cell. Infect. Microbiol.* **10**, 109. <https://doi.org/10.3389/fcimb.2020.00109> (2020).
18. Durairaj, J. et al. Uncovering new families and folds in the natural protein universe. *Nature* **622**(7983), 646–653. <https://doi.org/10.1038/s41586-023-06622-3> (2023).
19. Seong, K. & Krasileva, K. V. Prediction of effector protein structures from fungal phytopathogens enables evolutionary analyses. *Nat. Microbiol.* **8**(1), 174–187. <https://doi.org/10.1038/s41564-022-01287-6> (2023).
20. Humphreys, I. R. et al. Computed structures of core eukaryotic protein complexes. *Science* **374**(6573), eabm4805. <https://doi.org/10.1126/science.abm4805> (2021).
21. Akdel, M. et al. A structural biology community assessment of AlphaFold2 applications. *Nat. Struct. Mol. Biol.* **29**(11), 1056–1067. <https://doi.org/10.1038/s41594-022-00849-w> (2022).
22. Seong, K. & Krasileva, K. V. Computational structural genomics unravels common folds and novel families in the secretome of fungal phytopathogen *Magnaporthe oryzae*. *Mol. Plant Microbe Interact.* **34**(11), 1267–1280. <https://doi.org/10.1094/MPMI-03-21-0071-R> (2021).
23. Lane, T. J. Protein structure prediction has reached the single-structure frontier. *Nat. Methods* **20**(2), 170–173. <https://doi.org/10.1038/s41592-022-01760-4> (2023).
24. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589. <https://doi.org/10.1038/s41586-021-03819-2> (2021).
25. Barrio-Hernandez, I. et al. Clustering predicted structures at the scale of the known protein universe. *Nature* **622**(7983), 637–645. <https://doi.org/10.1038/s41586-023-06510-w> (2023).
26. Abramson, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500. <https://doi.org/10.1038/s41586-024-07487-w> (2024).
27. Al-Fatlawi, A., Menzel, M. & Schroeder, M. Is protein BLAST a thing of the past?. *Nat. Commun.* **14**(1), 8195. <https://doi.org/10.1038/s41467-023-44082-5> (2023).
28. Schwenbacher, M. et al. Analysis of the major proteins secreted by the human opportunistic pathogen *Aspergillus fumigatus* under in vitro conditions. *Med. Mycol.* **43**(7), 623–630. <https://doi.org/10.1080/13693780500089216> (2005).
29. Ranganathan, S. & Garg, G. Secretome: Clues into pathogen infection and clinical applications. *Genome Med.* **1**(11), 113. <https://doi.org/10.1186/gm113> (2009).
30. Dionisio, G., Kryger, P. & Steenberg, T. Label-free differential proteomics and quantification of exoenzymes from isolates of the entomopathogenic fungus *Beauveria bassiana*. *Insects* **7**(4), 54. <https://doi.org/10.3390/insects7040054> (2016).
31. Bouquellah, N. A., Elkady, N. A. & Farag, P. F. Secretome analysis for a new strain of the blackleg fungus *Plenodomus lingam* reveals candidate proteins for effectors and virulence factors. *J. Fungi* **9**(7), 740. <https://doi.org/10.3390/jof9070740> (2023).
32. Almagro Armenteros, J. J. et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* **37**(4), 420–423. <https://doi.org/10.1038/s41587-019-0036-z> (2019).
33. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J. Mol. Biol.* **305**(3), 567–580. <https://doi.org/10.1006/jmbi.2000.4315> (2001).
34. Paysan-Lafosse, T. et al. InterPro in 2022. *Nucleic Acids Res.* **51**(D1), D418–D427. <https://doi.org/10.1093/nar/gkac993> (2023).
35. Mirdita, M. et al. ColabFold: Making protein folding accessible to all. *Nat. Methods* **19**(6), 679–682. <https://doi.org/10.1038/s41592-022-01488-1> (2022).
36. Ayoub, R. & Lee, Y. RUPEE: A fast and accurate purely geometric protein structure search. *PLoS ONE* **14**(3), e0213712. <https://doi.org/10.1371/journal.pone.0213712> (2019).
37. Ayoub, R. & Lee, Y. Protein structure search to support the development of protein structure prediction methods. *Proteins* **89**(6), 648–658. <https://doi.org/10.1002/prot.26048> (2021).
38. Holm, L., Laiho, A., Törönen, P. & Salgado, M. DALI shines a light on remote homologs: One hundred discoveries. *Protein Sci.* **32**(1), e4519. <https://doi.org/10.1002/pro.4519> (2023).
39. Otasek, D., Morris, J. H., Bouças, J., Pico, A. R. & Demchak, B. Cytoscape automation: Empowering workflow-based network analysis. *Genome Biol.* **20**, 185. <https://doi.org/10.1186/s13059-019-1758-4> (2019).
40. Meng, E. C. et al. UCSF ChimeraX: Tools for structure building and analysis. *Protein Sci.* **32**(11), e4792. <https://doi.org/10.1002/pro.4792> (2023).
41. Gligorijević, V. et al. Structure-based protein function prediction using graph convolutional networks. *Nat. Commun.* **12**(1), 3168. <https://doi.org/10.1038/s41467-021-23303-9> (2021).
42. Lavezzo, E., Falda, M., Fontana, P., Bianco, L. & Toppo, S. Enhancing protein function prediction with taxonomic constraints—The Argot2.5 web server. *Methods* **93**, 15–23. <https://doi.org/10.1016/j.jymeth.2015.08.021> (2016).
43. Sharma, N., Naorem, L. D., Jain, S. & Raghava, G. P. S. ToxinPred2: An improved method for predicting toxicity of proteins. *Brief Bioinform.* **23**(5), bbac174. <https://doi.org/10.1093/bib/bbac174> (2022).
44. Liu, Y. et al. CB-Dock2: Improved protein-ligand blind docking by integrating cavity detection, docking and homologous template fitting. *Nucleic Acids Res.* **50**(W1), W159–W164. <https://doi.org/10.1093/nar/gkac394> (2022).
45. Yan, Y., Tao, H., He, J. & Huang, S. Y. The HDock server for integrated protein–protein docking. *Nat. Protoc.* **15**, 1829–1852. <https://doi.org/10.1038/s41596-020-0312-x> (2020).
46. Postic, G., Gracy, J., Périn, C., Chiche, L. & Gelly, J. C. KNOTTIN: The database of inhibitor cystine knot scaffold after 10 years, toward a systematic structure modeling. *Nucleic Acids Res.* **46**(D1), D454–D458. <https://doi.org/10.1093/nar/gkx1084> (2018).
47. Chandonia, J. M. et al. SCOPe: Improvements to the structural classification of proteins - Extended database to facilitate variant interpretation and machine learning. *Nucleic Acids Res.* **50**(D1), D553–D559. <https://doi.org/10.1093/nar/gkab1054> (2022).
48. Li, Y. et al. Cystine-knot peptide inhibitors of HTRA1 bind to a cryptic pocket within the active site region. *Nat. Commun.* **15**(1), 4359. <https://doi.org/10.1038/s41467-024-48655-w> (2024).
49. King, G. F. & Hardy, M. C. Spider-venom peptides: Structure, pharmacology, and potential for control of insect pests. *Annu. Rev. Entomol.* **58**, 475–496. <https://doi.org/10.1146/annurev-ento-120811-153650> (2013).
50. Wang, X. et al. Discovery and characterization of a family of insecticidal neurotoxins with a rare vicinal disulfide bridge. *Nat. Struct. Biol.* **7**(6), 505–513. <https://doi.org/10.1038/75921> (2000).
51. Gunning, S. J. et al. The Janus-faced atracotoxins are specific blockers of invertebrate K(Ca) channels. *FEBS J.* **275**(16), 4045–4059. <https://doi.org/10.1111/j.1742-4658.2008.06545.x> (2008).

52. Nakasu, E. Y. et al. Novel biopesticide based on a spider venom peptide shows no adverse effects on honeybees. *Proc. Biol. Sci.* **281**(1787), 20140619. <https://doi.org/10.1098/rspb.2014.0619> (2014).
53. Olsen, J. G., Flensburg, C., Olsen, O., Bricogne, G. & Henriksen, A. Solving the structure of the bubble protein using the anomalous sulfur signal from single-crystal in-house Cu K α diffraction data only. *Acta Crystallogr. D Biol. Crystallogr.* **60**(Pt 2), 250–255. <https://doi.org/10.1107/S0907444903025927> (2004).
54. Seibold, M., Wolschann, P., Bodevin, S. & Olsen, O. Properties of the bubble protein, a defensin and an abundant component of a fungal exudate. *Peptides* **32**(10), 1989–1995. <https://doi.org/10.1016/j.peptides.2011.08.022> (2011).
55. Torres-Quintero, M. C. et al. Engineering *Bacillus thuringiensis* Cyt1Aa toxin specificity from dipteran to lepidopteran toxicity. *Sci. Rep.* **8**(1), 4989. <https://doi.org/10.1038/s41598-018-22740-9> (2018).
56. Xu, C. et al. Crystal structure of Cry51Aa1: A potential novel insecticidal aerolysin-type β -pore-forming toxin from *Bacillus thuringiensis*. *Biochem. Biophys. Res. Commun.* **462**(3), 184–189. <https://doi.org/10.1016/j.bbrc.2015.04.068> (2015).
57. Holm, L. DALI and the persistence of protein shape. *Protein Sci.* **29**(1), 128–140. <https://doi.org/10.1002/pro.3749> (2020).
58. Cao, B. et al. The crystal structure of Cry78Aa from *Bacillus thuringiensis* provides insights into its insecticidal activity. *Commun. Biol.* **5**(1), 801. <https://doi.org/10.1038/s42003-022-03754-6> (2022).
59. Sampson, K. et al. Discovery of a novel insecticidal protein from *Chromobacterium piscinae*, with activity against Western Corn Rootworm, *Diabrotica virgifera virgifera*. *J. Invertebr. Pathol.* **142**, 34–43. <https://doi.org/10.1016/j.jip.2016.10.004> (2017).
60. Zaitseva, J. et al. Structure-function characterization of an insecticidal protein GNIP1Aa, a member of an MACPF and β -tripod families. *Proc. Natl. Acad. Sci.* **116**(8), 2897–2906. <https://doi.org/10.1073/pnas.1815547116> (2019).
61. Zhang, X. N. et al. A ribose-functionalized NAD⁺ with unexpected high activity and selectivity for protein poly-ADP-ribosylation. *Nat. Commun.* **10**(1), 4196. <https://doi.org/10.1038/s41467-019-12215-4> (2019).
62. Poltronieri, P., Celetti, A. & Palazzo, L. Mono(ADP-ribosyl)ation enzymes and NAD⁺ metabolism: A focus on diseases and therapeutic perspectives. *Cells* **10**(1), 128. <https://doi.org/10.3390/cells10010128> (2021).
63. Aravind, L., Zhang, D., de Souza, R. F., Anand, S. & Iyer, L. M. The natural history of ADP-ribosyltransferases and the ADP-ribosylation system. *Curr. Top. Microbiol. Immunol.* **384**, 3–32. https://doi.org/10.1007/82_2014_414 (2015).
64. Dass, R., Mulder, F. A. A. & Nielsen, J. T. ODiNPred: Comprehensive prediction of protein order and disorder. *Sci. Rep.* **10**(1), 14780. <https://doi.org/10.1038/s41598-020-71716-1> (2020).
65. Darling, A. L. & Uversky, V. N. Intrinsic disorder in proteins with pathogenic repeat expansions. *Molecules* **22**(12), 2017. <https://doi.org/10.3390/molecules22122027> (2017).
66. Khan, S. & Lang, M. A comprehensive review on the roles of metals Mediating insect-microbial pathogen interactions. *Metabolites* **13**(7), 839. <https://doi.org/10.3390/metabo13070839> (2023).
67. Nawaz, A. et al. Nanobiotechnology in crop stress management: An overview of novel applications. *Discov. Nano.* **18**, 74. <https://doi.org/10.1186/s11671-023-03845-1> (2023).
68. Catterall, W. A. et al. Voltage-gated ion channels and gating modifier toxins. *Toxicon* **49**(2), 124–141. <https://doi.org/10.1016/j.toxicon.2006.09.022> (2007).
69. Bretscher, H. & O'Connor, M. B. The role of muscle in insect energy homeostasis. *Front. Physiol.* **11**, 580687. <https://doi.org/10.3389/fphys.2020.580687> (2020).

Acknowledgements

We thank Prof. Mohamed Khaled Ibrahim and Prof. Sahar Tolba, coordinators of the Applied and Analytical Microbiology Program, for their technical support, as this work is part of a graduation project for students supervised by Dr. Peter Farag

Author contributions

P.F.F conceptualized the study, developed the methodology, performed the software, conducted the investigation, performed formal analysis, curated the data, supervised the study, validated the findings, and wrote, reviewed the original draft. A.A.E developed the methodology, performed formal analysis, and wrote, reviewed the original draft. J.S.S developed the methodology, conducted the investigation, and wrote the original draft. S.M.A developed the methodology, conducted the investigation, and wrote the original draft. E.W.E performed software, performed formal analysis, and wrote the original draft. N.H.M performed software, performed formal analysis, reviewed, and edited the manuscript. R.M.Z performed software, reviewed, and edited the manuscript. All authors have read and agreed to publish the current version of the manuscript.

Funding

Open access funding provided by The Science, Technology & Innovation Funding Authority (STDF) in cooperation with The Egyptian Knowledge Bank (EKB).

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-02618-3>.

Correspondence and requests for materials should be addressed to P.F.F.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025