bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.476343; this version posted January 15, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Structure of the *Lysinibacillus sphaericus* Tpp49Aa1 pesticidal protein elucidated from natural crystals using MHz-SFX

Lainey J. Williamson [1], Marina Galchenkova [2], Hannah L. Best [1], Richard J. Bean [3], Anna Munke [2], Salah Awel [2], Gisel Pena [2], Juraj Knoska [2], Robin Schubert [3], Katerina Doerner [3], Emyr Lloyd-Evans [1], Mark Young [1], Joana Valerio [3], Marco Kloos [3], Marcin Sikorski [3], Grant Mills [3], Johan Bieleki [3], Henry Kirkwood [3], Chan Kim [3], Raphael de Wijn [3], Kristina Lorenzen [3], P. Lourdu Xavier [2,4], Aida Rahmani [2], Luca Gelisio [2], Oleksandr Yefanov [2], Adrian P. Mancuso [3,5], Henry N. Chapman [2,6,7], Neil Crickmore [8], Pierre J. Rizkallah [9], Colin Berry [1*] and Dominik Oberthür [2*]

[1] School of Biosciences, Cardiff University, UK; [2] Center for Free Electron Laser Science CFEL, Deutsches Elektronen-Synchrotron DESY, Notkestr. 85, 22607 Hamburg, Germany; [3] European XFEL GmbH, Schenefeld, Germany; [4] Max-Planck Institute for the Structure and Dynamics of Matter, 22761 Hamburg, Germany; [5] Department of Chemistry and Physics, La Trobe Institute for Molecular Science, La Trobe University, Melbourne, Victoria 3086, Australia; [6] Centre for Ultrafast Imaging, Universität Hamburg, Hamburg, Germany; [7] Department of Physics, Universität Hamburg, Hamburg, Germany; [8] School of Life Sciences, University of Sussex, Falmer, UK; [9] School of Medicine, Cardiff University, UK

Correspondence: berry@cardiff.ac.uk, dominik.oberthuer@cfel.de

Key words: Xfel; Lysinibacillus sphaericus; Insecticidal protein; Tpp49Aa1; Cry48Aa1.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.476343; this version posted January 15, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Tpp49Aa1 from *Lysinibacillus sphaericus* is a Toxin_10 family protein that must interact with Cry48Aa1, a 3-domain crystal protein, to produce potent mosquitocidal activity, specifically against *Culex quinquefasciatus* mosquitoes. We use *Culex* cell lines to demonstrate for the first time transient detrimental effects of individual toxin components and widen the known target range of the proteins.

MHz serial femtosecond crystallography at a nano-focused X-ray free electron laser allowed rapid and high-quality data collection to determine the Tpp49Aa1 structure at 2.2 Å resolution from the merged X-ray diffraction data. The structure revealed the packing of Cry49Aa1 within the natural nanocrystals isolated from sporulated bacteria, as a homodimer with a large intermolecular interface. We then modelled the potential interaction between Tpp49Aa1 and Cry48Aa1. The structure sheds light on natural crystallisation and, along with cell-based assays broadens our understanding of this two-component system.

1. Introduction

Insecticidal proteins produced by *Bacillus thuringiensis* and *Lysinibacillus sphaericus* constitute the major factors in bioinsecticides and transgenic crops. The Cry48Aa1/Tpp49Aa1 pair is produced by *L. sphaericus* and is composed of Cry48Aa1, a 135 kDa protein belonging to the 3-domain family of Crystal (Cry) proteins, and Tpp49Aa1 (formerly Cry49Aa1), a 53 kDa protein belonging to the family of Toxin_10 pesticidal proteins (Tpp) [1,2]. Both proteins are deposited in the form of natural, parasporal crystals in association with the bacterial spore. Mosquito toxicity bioassays have revealed that both components in combination are required for activity against *Culex quinquefasciatus* larvae, with optimal activity at a ratio of ~1:1 [2,3].

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.476343; this version posted January 15, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The Cry48Aa1/Tpp49Aa1 toxin pair does not show toxicity to any other insect species investigated, including the Coleopteran, *Anthonomus grandis*, the Lepidoptera, *Anticarsia gemmatalis*, *Spodoptera frugiperda* and *Plutella xylostella*, and the Diptera, *Aedes aegypti* and *Anopheles gambiae* [3].

In common with other insecticidal proteins, the mode of action of the Cry48Aa1/Tpp49Aa1 toxin pair begins with ingestion of crystal inclusion protoxins, followed by solubilization and proteolytic cleavage in the alkaline environment of the gut [2]. Following activation, both Cry48Aa1 and Tpp49Aa1 bind specifically and with high affinity to *C. quinquefasciatus* brush border membrane [4]. Although functional receptors are yet to be elucidated, several classes of Cry48Aa1/Tpp49Aa1 binding proteins have been identified, including maltases, aminopeptidases, alkaline phosphatases, and metalloproteases [5]. Following membrane interaction, subsequent cytopathological effects including cytoplasmic and mitochondrial vacuolation, endoplasmic reticulum breakdown, and microvillus disruption, are seen [6], culminating in insect death.

Dot blot assays show Cry48Aa1 and Tpp49Aa1 are able to form a complex [4]. Moreover, competition assays have indicated that the N-terminal region of Tpp49Aa1, residues N49 – S149, is responsible for binding Cry48Aa1, while the C-terminal region, residues S349 – N464, is involved in membrane interaction [4]. Considering the combined role of Tpp49Aa1 in both interaction with Cry48Aa1 and the target cellular membrane, as well as studies indicating a significant loss of larvicidal activity from Tpp49Aa1 truncated fragments, it has been suggested that the region of Tpp49Aa1 required for activity corresponds to the protein fragment remaining after proteolytic activation after residue F48 [3,4]. Mutagenesis studies targeting

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.476343; this version posted January 15, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Tpp49Aa1 cysteine residues C70, C91, C183, and C258, have indicated the functional importance of the latter three, which are required for full larvicidal activity against *Culex* larvae, as well as Cry48Aa1 binding [7]. Although homology modelling has been applied to produce structural models of both Cry48Aa1 and Tpp49Aa1 [3,8], their structures are yet to be experimentally resolved.

The development of X-ray free electron lasers (XFELs) has given rise to a new approach in protein crystallography, known as serial femtosecond crystallography (SFX). SFX makes use of a liquid microjet to introduce a stream of crystals into an XFEL beam, where the delivery of intense X-ray pulses of several femtoseconds duration enables diffraction data to be collected at high exposures in a serial fashion before structural information is lost due to radiation damage [9–11]. As such, diffraction data are not limited by the small crystal size of natural crystals. In the field of insecticidal proteins, SFX has previously been applied to solve the structures of the Cyt1Aa, Cry3Aa and Cry11 toxins from *B. thuringiensis* and, the Tpp1Aa2/Tpp2Aa2 toxin (formerly BinA/BinB toxin) from *L. sphaericus* [12–15]. Here, we employed SFX at the SPB/SFX instrument [16] of the European XFEL to determine the Tpp49Aa1 structure from natural crystals isolated from recombinant *B. thuringiensis* expressing *L. sphaericus* Tpp49Aa1 as parasporal inclusions. Elucidation of the Tpp49Aa1 structure will be important for understanding the mechanism of the Cry48Aa1/Tpp49Aa1 two-component system fully and, given that the Cry48Aa1/Tpp49Aa1 pair exhibits toxicity against Tpp1Aa2/Tpp2Aa2-resistant *C. quinquefasciatus* larvae [2,17], also has implications for managing mosquito resistance.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.476343; this version posted January 15, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2. *Materials and Methods*

2.1. Purification of Tpp49Aa1 Protein Crystals

The *B. thuringiensis* recombinant strain 4Q7::pHTP49, encoding the *tpp49Aa1* gene (accession number AJ841948) [2] was grown in 400 mL Embrapa medium [18] containing 5 µg/mL erythromycin at 30°C with shaking (200 rpm) until sporulation reached > 90%, as judged by phase contrast microscopy. Sporulated cultures were harvested and the natural crystal proteins were isolated using stepped sucrose gradients as previously described [2]. Crystal protein was run on SDS-PAGE and blotted onto PVDF membrane for N-terminal sequencing by Alta Bioscience Ltd (Redditch, UK).

2.2. Bioassays

Bioassays were carried out against a range of insects (*C. quinquefasciatus*, *Ae. aegypti* and *An. gambiae*) and insect cell lines including *C. quinquefasciatus* (Hsu, and MRA-918) *Culex tarsalis* (Ct), and *Spodoptera frugiperda* (SF9) [19–21]. It should be noted that the MRA-918 line (also known as 4A3A) was originally reported as an *Anopheles gambiae* line but the cells have been independently verified by a number of laboratories, including our own (amplifying and sequencing mitochondrial large ribosomal subunit, mitochondrial cytochrome C oxidase and maltase genes), to be *C. quinquefasciatus.*

For insect assays, 10–15 second to third instar larvae were placed in 5 mL of dH_2O and maintained in a humidified room at 24°C. Mortality was assessed at 12, 24, and 48 hours post addition of the natural protein crystals or crystals solubilised by overnight incubation at room temperature in 50 mM Na_2CO_3 pH 10.5 + 0.05% β-mercaptoethanol. Solubilised toxin was

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.476343; this version posted January 15, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

treated with immobilized TPCK Trypsin (Thermo Scientific, 20230) overnight at 37°C, followed by separation from the protein sample by centrifugation. Tpp49Aa1 and Cry48Aa1 were assayed independently and together at a 1:1 molar ratio.

Insect cell lines were maintained at 27°C in an appropriate growth medium which was changed every 4 days: Hsu, MRA-918, Ct (Schneider's Insect Medium supplemented with 10% FBS) and SF9 (Grace's Insect Medium supplemented with 10% FBS). For cellular bioassays, cells were plated at 10,000 cells per well of a 96-well plate in 150 µL of media and left to reach ~80% confluency. Tpp49Aa1/Cry48Aa1 proteins were solubilised in 50 mM Na_2CO_3 pH 10.5 + 0.05% β-mercaptoethanol and added either separately, or at a 1:1 molar ratio, in a range of concentrations (20, 10, 5, 2.5 µg/mL total protein) with the equivalent amount of solubilisation buffer added into control wells (at no more than 5% of total well volume). As a measure of cell viability, resazurin (10 µg/ mL) was added into the cell media (10% v/v) at 24 hours post Cry48Aa1 and Tpp49Aa1 addition. Fluorescence was quantified using a Molecular Devices Spectramax Gemini EM plate reader (λ_{ex} = 445 nm: λ_{em} = 585 nm) at 8 and 24 hours after the addition of resazurin. Statistical analyses were performed using GraphPad Prism for mac OS (Ver 8.2.0), using one-way ANOVAs followed by Dunnett's multiple comparisons test to compare individual treatment groups back to the control. Data are presented as mean ± standard deviation. For imaging, MRA-918 cells were seeded in 8-well Ibidi chamber slides (Thistle, IB-80826). Brightfield images were acquired with a Zeiss AX10 inverted microscope with Axiocam Mrm camera and Axiovision 4.5.2 software (Zeiss, Cambridge, UK).

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.476343; this version posted January 15, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2.3. Transmission Electron Microscopy

Purified crystal batches were characterized using transmission electron microscopy (JEM 2100-Plus, JEOL) in the XBI lab of European XFEL [22]. Holey carbon copper grids (Quantifoil R1.2/1.3) were glow discharged (GloQube Plus, Quorum Technologies) freshly before use. Crystal slurry (2 µL) was applied onto the grid and incubated for 30 s and blotted using filter paper (Whatman #1). Samples were negatively stained by placing the grid on a droplet containing 2 % (w/v) uranyl acetate and blotted immediately. Grids were placed on a second uranyl acetate droplet and incubated for 20 seconds before blotting again and left for drying on filter paper. Samples were imaged with the TEM at 200 kV acceleration voltage using an Emsis Xarosa camera in imaging and selected area electron diffraction mode.

2.4. Structure Determination

Crystals were washed with ddH_2O and filtered through a cascade of nylon mesh filters (Sysmex Celltrics), ranging from 100 µm down to 5 µm mesh size. The crystal suspension was centrifuged at 200 x *g* for one minute and the supernatant – containing the Tpp49Aa1 nanocrystals – was subjected to another cascade of filtration and washing before transfer to the high-pressure sample reservoirs for injection into the XFEL beam. Megahertz serial femtosecond crystallography [23,24] diffraction data were collected at the SPB/SFX [16], instrument of the European XFEL facility, Hamburg, Germany, using fast liquid-jet based injection [25] with 3D-printed [26] double flow focusing nozzles (DFFN) [27]. With this set-up, 352 images per X-ray pulse train (with 10 trains/s repetition rate) were recorded with the AGIPD Detector at an intra-train pulse rate of 1.1 MHz. A photon energy of 9.3 keV with an average of 2 mJ / pulse was delivered to the instrument, focused to a spot size of about 300 nm diameter using the Nanoscale-focusing KB optics [28], providing about $6x10^{12}$ photons/μm^2/pulse at the sample.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.476343; this version posted January 15, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The online crystal diffraction 'hit-rate' was monitored using OnDA program [29] with raw data processing largely following the method described by Wiedorn et al. [23]. Hit finding was performed using the program Cheetah [30] with careful optimization of the peak search parameters (hitfinderADC=200, hitfinderMinSNR=5, hitfinderNpeaks=20, hitfinderMinPixCount=1) and masking of bad pixels. Meaningful diffraction patterns were then indexed using CrystFEL [31,32] version 0.9.1 (--int-radius=2,4,6, --multi) using the indexing method XGandalf [33]. For the detector geometry optimization the program Geoptimiser was used [34]. Merging and scaling of the integrated reflection intensities was performed using the Partialator program from the same CrystFEL package (--model=unity --min-res=5 --push-res=1.0). The solvent content and number of molecules in the asymmetric unit was estimated using Matthews 'analysis, available through the MATTPROB web server [35]. The phasing pipeline MRage in Phenix [36,37] was used for initial phasing, using the sequence information and a component stoichiometry of two as input. Molecular replacement was performed using *L. sphaericus* Tpp1Aa2/Tpp2Aa2 (formerly BinA/BinB, PDB 5FOY and PDB 5G37) and *L. sphaericus* Tpp2Aa3 (BinB variant, PDB 3WA1) as starting models. The initial model was optimized using phenix.phase_and_build, followed by another round of Phaser and automatic model building using phenix.autobuild [38]. The resulting model and maps were inspected manually using coot [39], followed by iterative refinement and model building cycles using phenix.refine [40] and coot respectively. In the initial molecular replacement, two NCS-related domains A1 and A2 were switched (A1B2, A2B1), this error was corrected in coot and the residues linking A and B were built manually. This model was subjected to another cycle of phenix.autobuild, followed by iterative refinement and model building cycles using phenix.refine and coot respectively. Final refinement was carried out using Refmac5 [41] in the CCP4i2 package [42] keeping the R_{free}-Flags generated in Phenix.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.476343; this version posted January 15, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2.5. Modelling the Cry48Aa1-Tpp49Aa1 interaction

2.5.1. Structure preparation

The full-length Cry48Aa1 structure was predicted using the AlphaFold2 package [43], as installed at DESY and available through the Maxwell-cluster. For docking, the core toxin domains (residues 53 – 659) were extracted from the predicted structure. The structure of Tpp49Aa1 was obtained by extracting residues 49 – 464 from chain B of the crystal structure elucidated as part of this work.

2.5.2. Molecular docking

Molecular docking was performed using a naïve approach with no presumed Cry48Aa1-Tpp49Aa1 interface. First, a global docking search was performed using the ClusPro web server [44]. Briefly, ClusPro employs a fast Fourier transform (FFT)-based algorithm to perform rigid body docking. The 1,000 lowest energy models are clustered according to their root-mean-square deviation (RMSD) and refined via energy minimisation. ClusPro outputs the central model from each cluster, with the largest clusters ranking highest. Here, some 30 clusters were identified, with the central models from the 5 largest clusters carried forward for local docking refinement.

Local docking refinement was performed using the RosettaDock algorithm [45]. Briefly, RosettaDock employs a Monte Carlo based algorithm to perform rigid body docking and side chain optimisation from a user perturbed starting position. Models are scored using the ref2015 score function [46], which is composed of several weighted terms, enabling the total energy to be calculated in Rosetta Energy Units (REU). To ensure side chains were in their lowest energy

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.476343; this version posted January 15, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

conformations, models were prepacked using the *docking_prepack_protocol.macosclangrelease* executable. Subsequently, docking was performed using the *docking_protocol.macosclangrelease* executable. For each docking search, 1,000 models were generated and the 5 models with the lowest energy scores (a total of 25 models across all 5 RosettaDock searches) were carried forward for further structural analysis.

2.5.3. Molecular dynamics

To assess the structural stability of modelled Cry48Aa1-Tpp49Aa1 complexes, molecular dynamics (MD) simulations were performed using GROMACS (v.2020.1) [47] on Cardiff University's High-Performance Computing cluster, Hawk. Set-up, energy minimisation, equilibration, and production simulations were performed as previously described [48]. Production simulations were performed for 100 ns and the resulting trajectories were visualised. In addition, the RMSD of the position of backbone atoms was analysed as a function of simulation time using GROMACS modules gms rms and gmx gyrate respectively. The Visual Molecular Dynamics (VMD, v.1.9.4) program was used to visualise simulations and the Gnuplot (v.5.2) program was employed to produce the graphics associated with this work.

2.5.4. Interface analysis

The PDBePISA web server [49] was used to analyse the interfacial interactions of modelled Cry48-Tpp49 complexes. PISA enabled calculation of the interface area ($Å^2$) and solvation free energy gain (Δ^iG, kcal mol^1), as well as the identification of interfacial hydrogen bonds and salt bridges.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.476343; this version posted January 15, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2.6. Size Exclusion Chromatography

For size exclusion chromatography (SEC), Cry48Aa1 and Tpp49Aa1 protein crystals were solubilised in 50 mM Na_2CO_3 pH 10.5 + 0.05% β-mercaptoethanol overnight at room temperature, with agitation. Insoluble material was removed by centrifugation, and the solubilised proteins buffer exchanged into 20 mM TrisHCl pH 8.5. SEC was performed using a calibrated Hiload™ 16/60 Superdex™ S200 pg column. For column calibration, BioRad standard proteins (of molecular weights 670, 158, 44, 17, and 1.35 kDa) were run on the column at a flow rate of 0.5 mL/min with 50 mM TrisHCl buffer pH 8. All protein samples were run on the column at a flow rate of 1 mL/min with 20 mM TrisHCl pH 8.5. Elution fractions were collected and analysed using SDS-PAGE.

2.7. Dynamic light scattering

To investigate whether Tpp49Aa1 was present as a monomer in solution we used dynamic light scattering (DLS) and refractive index (RI) measurements. The Zetasizer MicroV system (Malvern Instruments Ltd., Malvern, UK) was used to quantify the size distribution of Tpp49Aa1 particles based upon time-dependent fluctuations in the scattered light intensity (at 90° scattering angle), due to the Brownian motion of the protein in solution. Concentration was determined by a refractive index (RI) detector (VE 3580, Viscotek Corp). Protein samples (100 µL at 1 mg/ mL) were prepared and solubilised as above, and separated via SEC using a Superdex 75 Increase 10/300 GL with a flow rate of 0.8 mL/min, coupled to the Zetasizer and the refractive index (RI) detector. Eluted samples were measured every ~3 s at 30°C. Data were collected and analysed in OmniSEC software (Ver 5.12) and calibrated to BSA (1 mg/ mL).

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.476343; this version posted January 15, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

3. Results and Discussion

3.1. Insect Bioassays

A combination of the recombinant Cry48Aa1 and Tpp49Aa1 resulted in 100% mortality of *C. quinquefasciatus* after 24 h (data not shown), confirming observations noted in previous publications [2], and the functionality of the Tpp49Aa1 produced for this study. Mortality was not observed with the individual proteins, nor against the other mosquito species assayed (*Ae. aegypti*, *An. gambiae*).

3.2. Structure Description

Diffraction data were collected from Tpp49Aa1 nanocrystals, which had been visualized by transmission electron microscopy **(Fig S1)**. The sample was injected in the vacuum chamber of the SPB/SFX beamline at European XFEL using 3D-printed microfluidic DFFN. In total, 11278683 diffraction patterns have been collected from which 116387 could be indexed in space group $P2_12_12$, (a~79; b~83; c~157 Å; α=β=ɣ=90). Phasing was performed by molecular replacement using *L. sphaericus* Tpp1Aa2/Tpp2Aa2 (PDB 5FOY and PDB 5G37) and *L. sphaericus* Tpp2Aa3 (PDB 3WA1) as starting models. Manual building and refinement from the ensembled diffraction data led to a model with R_{work}/R_{free} of 0.160/0.197 at 2.2 Å resolution **(Table 1)**. The electron density map **(Fig S2)** showed continuous density for residues 49 – 464 of the protein sequence. The first 48 N-terminal residues are not observed in the map. The structural data reveal that Tpp49Aa1 crystals are formed by the packing of dimers into the crystal lattice.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.476343; this version posted January 15, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 1 – Crystallographic data statistics for Tpp49Aa1

PDB Entry	Tpp49Aa1
Data Collection	
Accession Code	7QA1
Beamline	SPB/SFX at European XFEL
X-ray Energy (keV)	9.3
Wavelength (Å)	1.33
Crystal Data	
a, *b*, *c* (Å)	79.76, 83.15, 157.33
α, β, γ	90.0, 90.0, 90.0
Space group	P 2_1 2_1 2_1
Resolution (Å)	2.22 – 28.57
Outer shell	2.22 – 2.26
Rsplit (%)	12.0 (91.8)
CC1/2	0.990 (0.480)
CC*	0.997 (0.806)
I / σ (I)	8.91 (1.25)
Completeness (%)	100 (100)

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.476343; this version posted January 15, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Multiplicity	485 (28.8)
Total Measurements	26,166,570 (76837)
Unique Reflections	53,943 (2667)
Wilson B-factor (Å^2)	20.54
Refinement Statistics	
Non-H Atoms	7204
R-work reflections	52,112
R-free reflections	1755
R-work/R-free	0.160/0.197
rms deviations	
Bond lengths (Å)	0.0126
Bond Angles (°)	1.848
[1]Coordinate error	0.134
Mean B value (Å^2)	17.24
Ramachandran Statistics	
Favoured/allowed/outliers	809/19/0
Favoured/allowed/outliers (%)	97.71/2.29/0

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.476343; this version posted January 15, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Within each Tpp49Aa1 monomer **(Fig 1)**, two distinct domains exist: An N-terminal lectin-like head domain covering residues 49 – 214, and a C-terminal putative pore-forming domain (PFD) covering residues 215 – 464. The first 48 N-terminal residues, not seen in the structure, but are known to be present in Tpp49Aa1 crystals produced in the native host *L. sphaericus* [2]. To ensure that this region is also present when (as in this work) the protein is expressed in *B. thuringiensis*, N-terminal sequencing was performed for 5 cycles. The resulting sequence MXNQ(E) appears to correspond to the authentic N-terminal sequence MENQI, showing that the N-terminus is intact and, therefore, apparently disordered in the crystals. This region is known to be proteolytically removed from the protein by *C. quinquefasciatus* gut extracts [3] so that the apparent structural flexibility may facilitate this activation of the protein. The N-terminal domain seen in the structure comprises a β-trefoil fold consisting of six two-stranded β-hairpins which form a β-barrel with a triangular cap, as well as one disulphide bond, Cys91-Cys183. The β-hairpins in this, and other β-trefoil folds, are arranged into three subdomains commonly designated α, β, and γ, giving rise to a pseudo three-fold axis. β-trefoil folds are highly conserved in the lectin family, a class of carbohydrate-binding proteins [50]. Moreover, several studies have indicated a role for carbohydrate moieties in eliciting insecticidal action of Cry proteins [51–54]. Hence, it is possible that carbohydrate binding may also be important for insecticidal activity of the Cry48Aa1/Tpp49Aa1 toxin pair. However, in Tpp2Aa2 of the Tpp1Aa2/Tpp2Aa2 complex, the Cys67-Cys161 disulphide bond, which is the equivalent of the Cys91-Cys183 disulphide bond identified in the Tpp49Aa1 structure **(Fig S3)**, has been proposed to obstruct the putative α sugar binding module [13]. This may suggest that carbohydrate binding is less important for Tpp2Aa2 function, and this may also be the case for Tpp49Aa1.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.476343; this version posted January 15, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The C-terminal domain of Tpp49Aa1 comprises a β sheet-rich topology characteristic of the aerolysin family of pore forming toxins [55,56]. Indeed, Lacomel *et al*. speculated that the Tpp family may represent a subclass of the larger aerolysin, ETX/MTX-2 superfamily of β-pore forming proteins [57]. Within this superfamily, the aerolysin toxin is considered the model member and its mechanism of pore formation has been well characterised. Aerolysin, which is secreted as the inactive proaerolysin homodimer, is first activated by proteolytic cleavage of the C-terminal propeptide, allowing dissociation to a monomer [55,58]. The C-terminal PFD interacts with target glycosylphosphatidylinositol (GPI) anchored receptors, whilst the N-terminal receptor binding domain interacts with target glycans [59]. Oligomerization of seven monomers via interaction of the PFDs leads to formation of a pre-pore structure constituting an amphipathic β-barrel which inserts into the membrane, leading to cell death by osmotic lysis [58]. The structural homology seen between Tpp49Aa1 and the aerolysin family may suggest that Tpp49Aa1 is able to form pores by a similar mechanism. In an analysis of the Tpp proteins and the wider aerolysin family [57], a region including a short beta hairpin with predominantly amphipathic structure, tucked under a loop within the PFD, was proposed as transmembrane domains in Tpp1 (residues 256-268) and Tpp2 (residues 302-317) **(Fig 3)**. A similar structure is found in the structure of Tpp35 (formerly Cry35: residues 249-259) and in our Tpp49 structure (residues 322-334) **(Fig 3)**. We speculate that, as in an aerolysin-like mechanism, this region may unfold to form the beta barrel pore in the target cell membrane. However, it also cannot be ruled out that pore formation by Tpp49Aa1 may occur via unique mechanisms and the necessity for Cry48Aa1 in the action of Tpp49Aa1 raises further questions regarding the formation of pores by this protein pair. Additional structural studies are required to investigate the pore-formation process further.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.476343; this version posted January 15, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

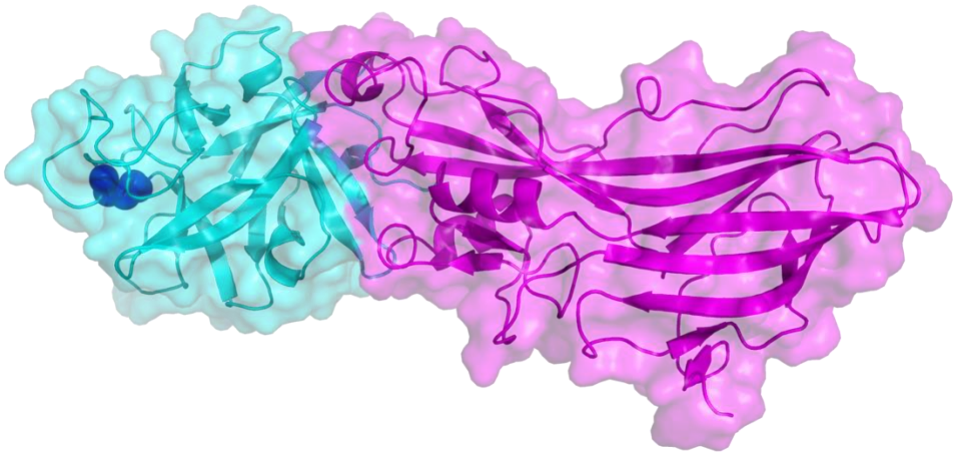

Figure 1. Crystal structure of Tpp49Aa1. Two distinct domains exist within the Tpp49Aa1 monomer. The N-terminal lectin-like head domain (cyan) consisting of six β-hairpins that form a β-trefoil fold containing a disulphide bond Cys91-Cys183 (shown as spheres – dark blue). The C-terminal pore forming domain (magenta) comprises the aerolysin-like domain.

The final model of Tpp49Aa1 revealed the presence of a homodimer forming an "X" structure with a large intermolecular interface **(Fig 2)**, similar to that described for the Tpp1Aa2/Tpp2Aa2 binary pesticidal protein heterodimers in their natural crystals **(Fig S3)** [13]. Superposition of the Tpp49Aa1 monomers shows the two copies to be almost identical, with a RMSD of 0.299 Å estimated by PyMOL **(Fig S4)**. The interface between the two monomers, which involves 42 residues from monomer A and 39 residues from monomer B, including 17 hydrogen bonds, exhibits a shape complementarity (Sc) index of 0.729 [60]. Interface analysis by PISA **(Table S1)** estimates the interface area at 1331.4 $Å^2$ and the binding energy at -13.5 kcal mol^{-1}. In the Tpp1Aa2/Tpp2Aa2 heterodimer **(Fig S3)**, which exhibits an RMSD of 2.295 Å

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.476343; this version posted January 15, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

when compared to the Tpp49Aa1 homodimer, the interface between the Tpp1Aa2 and Tpp2Aa2 monomers involves 49 residues from Tpp1Aa2 and 63 residues from Tpp2Aa2, making 19 hydrogen bonds and 2 salt bridges. Consistent with this, interface analysis by PISA estimates the interface area at 1833.1 $Å^2$ and the binding energy at -22.5 kcal mol^{-1}, indicating a more stable complex for Tpp1Aa2/Tpp2Aa2 heterodimers than for Tpp49Aa1 homodimers. Hence, whilst the Tpp1Aa2/Tpp2Aa2 interaction is required for toxicity, interaction of the Tpp49Aa1 monomers may only be required for crystal packing and stability. Indeed, the N-terminal region of Tpp49Aa1 (residues N49 – S149), which has been shown to interact with Cry48Aa1 [4], is partially buried within the large intermolecular interface of the Tpp49Aa1 dimer **(Fig S5)**. Hence, a weaker interaction may also allow dissociation of the Tpp49Aa1 monomers for subsequent formation of the Cry48Aa1-Tpp49Aa1 complex. These different interactions highlight an interesting feature of the Tpp group of proteins in which some members (Tpp36, Tpp78, Tpp80) are able to exert insecticidal activity acting alone, while Tpp1/Tpp2 requires interaction of two distinct members of the Tpp group and Tpp49Aa1, described here, requires interaction with Cry48Aa1, a member of a distinct (Cry) structural family.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.476343; this version posted January 15, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

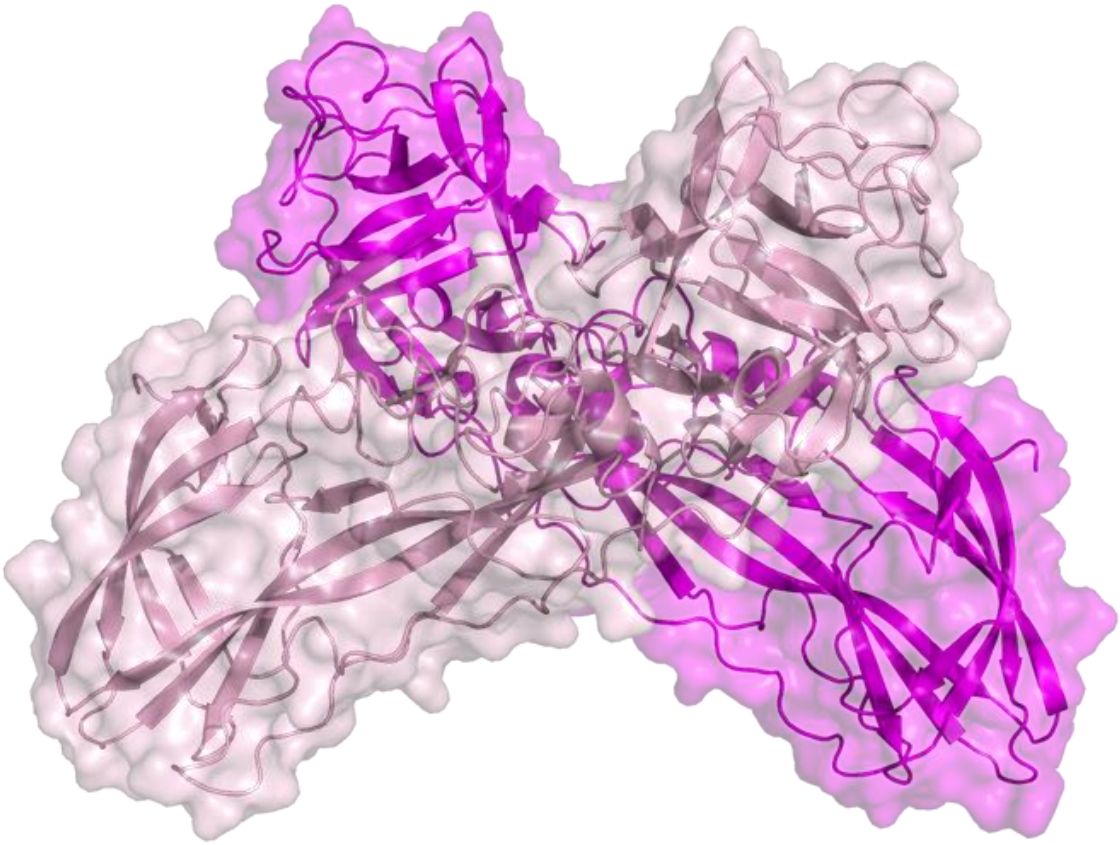

Figure 2. Structure of Tpp49Aa1 homodimer. Tpp49Aa1 homodimer (monomer A – magenta, monomer B – light pink) forming an "X" structure with a large intermolecular interface.

The final structural model of Tpp49Aa1 presented here shows that the protein exists as a homodimer in crystalline form. To investigate whether the dimeric form is maintained in solution, solubilised Tpp49Aa1 was assessed using SEC. Tpp49Aa1 is present as two major bands that run at, and just above, the 48 kDa molecular marker on an SDS-PAGE gel (**Fig S6.A**). Solubilised Tpp49Aa1 crystals have previously been shown to produce similar size bands, one of which starts at residue Asn76 and these may be products that are produced by crystal-associated proteinases during analysis [2]. SEC showed a main peak at 51 kDa, but with a shoulder peak before at 121 kDa indicating that it is predominantly a monomer but that some dimers may persist (**Fig S6.B**). In further analysis, dynamic light scattering (DLS) and

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.476343; this version posted January 15, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

refractive index measurements were used to analyse molecular weight. DLS data, calibrated to BSA, indicated Tpp49Aa1 is present in solution at a single MW of 52.1 kDa, corresponding with the approximate weight of the monomer (**Fig S6.C**).

3.3. Comparison with Related Structures

Related proteins were identified in the Protein Data Bank using the DALI server [61] to perform a structural similarity search with Tpp49Aa1. The best matches were other insecticidal proteins belonging to the Tpp family (PF05431), sharing both the common PFD and β-trefoil domains **(Fig 3)**: *L. sphaericus* Tpp2Aa2 (formerly BinB, PDB 5FOY – chain B), *L. sphaericus* Tpp2Aa3 (BinB variant, PDB 3WA1 – chain A), *L. sphaericus* Tpp1Aa2 (formerly BinA, PDB 5FOY – chain A), and *B. thuringiensis* Tpp35Ab1 (formerly Cry35Ab1, PDB 4JP0 – chain A), with greatest structural similarity between Tpp49Aa1 and Tpp2Aa2 **(Table 2)**. For these structurally similar proteins the Z-scores ranged from 29.2 to 41.4. Scores greater than 2 are considered significant [61].

Table 2 – Structural similarity of Tpp49Aa1 with other β-sheet toxins

PDB ID	Name (formally)	Z-score	RMSD (Å)	Sequence identity (%)
5FOY_B	Tpp2Aa2 (BinB)	41.4	2.1	31
3WA1_A	Tpp2Aa3 (BinB)	38.0	2.3	30
5FOY_A	Tpp1Aa2 (BinA)	37.8	2.5	27
4JP0_A	Tpp35Ab2 (Cry35)	29.2	3.5	17

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.476343; this version posted January 15, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

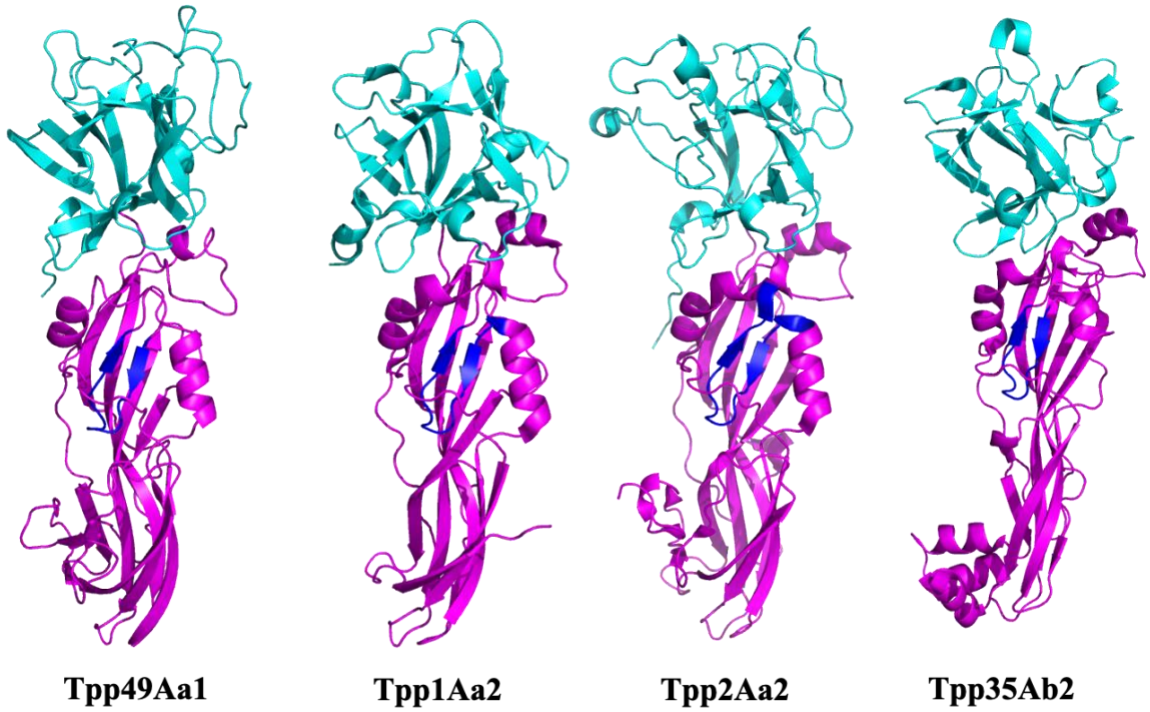

Figure 3. Comparative structures of β-sheet toxins. The structures of insecticidal proteins: Tpp49Aa1 (PDB 7QA1) from *L. sphaericus*, Tpp1Aa1 from *L. sphaericus* (PDB 5FOY – chain A), Tpp2Aa1 from *L. sphaericus* (PDB 5FOY – chain B), Tpp35Ab1 from *B. thuringiensis* (PDB 4JP0 – chain A). Lectin-like domains are shown in cyan and pore-forming domains are shown in magenta. Regions including a short beta hairpin with predominantly amphipathic structure (shown in dark blue) were proposed as transmembrane domains in Tpp1 (residues 256-268) and Tpp2 (residues 302-317). A similar structure is found in the structure of Tpp35 (residues 249-259) and in our Tpp49 structure (residues 322-334).

Another important feature that Tpp49Aa1 shares with a range of other insecticidal proteins in the taxonomic class Bacilli, is its propensity to form natural crystals when expressed at high level (in its source organism, *L. sphaericus*, and in this study, in recombinant form in *B. thuringiensis*). Tetreau *et al*, have recently reviewed factors within proteins that may contribute to the ability of bacteria to sequester them into natural crystals [62]. Natural crystals of Cry3Aa (PDB 4QX0) [12] were found to have a high solvent content: 60.4% by Matthews' analysis compared to a mean of ~47% in a previous analysis of *in vitro* grown crystals of

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.476343; this version posted January 15, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

proteins in the PDB [35]. Tpp49Aa1 has an estimated 49.8% solvent content - closer to the mean of *in vitro* grown crystals - but, like natural crystals of Cry3Aa, the Tpp49Aa1 crystals are permeated by wide solvent accessible channels **(Fig. S7)**. The ability of natural crystals to remain stable in the environment but to be solubilised by insect gut pH is thought to rely, in part, on intermolecular salt bridges that may be pH labile. PISA predicts that the packing of dimers into crystals introduces eight possible interfaces between neighbouring dimers. A network of hydrogen bonds and salt bridges is also predicted, including the interaction between Asp88 residues from monomers within one dimer and Arg92 residues in neighbouring dimers. We can speculate that the solvent channels within the crystals may facilitate dissolution in the appropriate environment.

3.4. Modelling the Cry48Aa1-Tpp49Aa1 interaction

Mosquito toxicity bioassays have revealed that both Cry48Aa1 and Tpp49Aa1 in combination are required for activity against *C. quinquefasciatus*, with optimal activity arising from a ratio of ~1:1 [2,3]. In addition, dot blot assays have revealed the ability of Cry48Aa1 and Tpp49Aa1 to form a complex [4]. Given these studies, it is interesting to speculate how Tpp49Aa1 and Cry48Aa1 might interact.

To model the Cry48Aa1-Tpp49Aa1 complex, we used a combined ClusPro and RosettaDock approach. ClusPro docking produced 30 clusters, of which the central models from the top 5 clusters were carried forward for local docking refinement using RosettaDock. From each of the 5 local docking searches, 1,000 models were produced. Within each of these searches, output models were ranked according to their total energy scores and the 5 lowest scoring models were then identified. Models are referred to according to their ClusPro cluster (c1 –

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.476343; this version posted January 15, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

c5) followed by their Rosetta energy rank (r1 – r5, with r1 being the lowest energy score and r5 being the highest energy score). Energy scores of the generated Cry48Aa1-Tpp49Aa1 models ranged from -1580.96 to -1730.57 REU **(Table S2)**.

To assess the structural stability of the Cry48Aa1-Tpp49Aa1 models, 100 ns MD simulations were performed. MD simulations were visualised and the RMSD of the position of backbone atoms relative to the starting structure was calculated as a function of time. Low RMSD values indicate stable structures which remain close to their starting positions, whilst high RMSD values indicate unstable structures which deviate away from their starting positions. All 5 models originating from ClusPro cluster 3 were found to exhibit RMSD values that remained relatively constant throughout the 100 ns MD simulations performed **(Fig S8.C)**. In addition, several models originating from ClusPro clusters 1 and 4 also exhibited low RMSD values **(Fig S8.A, S8.D)**. To further assess the Cry48Aa1-Tpp49Aa1 models, the orientation of docking was analysed. It has been shown that the N-terminal region of Tpp49Aa1 is responsible for binding Cry48Aa1 [4]. Since models originating from ClusPro cluster 1 were found to bind via the C-terminal domain of Tpp49Aa1, these models were excluded from any further analyses. Of the remaining stable models originating from ClusPro clusters 3 and 4, models c3r3 **(Fig 4.A)** and c4r3 **(Fig 4.B)** were found to exhibit the lowest RMSD values **(Fig S8.C, S8.D)** and hence, may represent possible structures of the Cry48Aa1-Tpp49Aa1 complex. Clearly, further analyses will be necessary to validate these models.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.476343; this version posted January 15, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

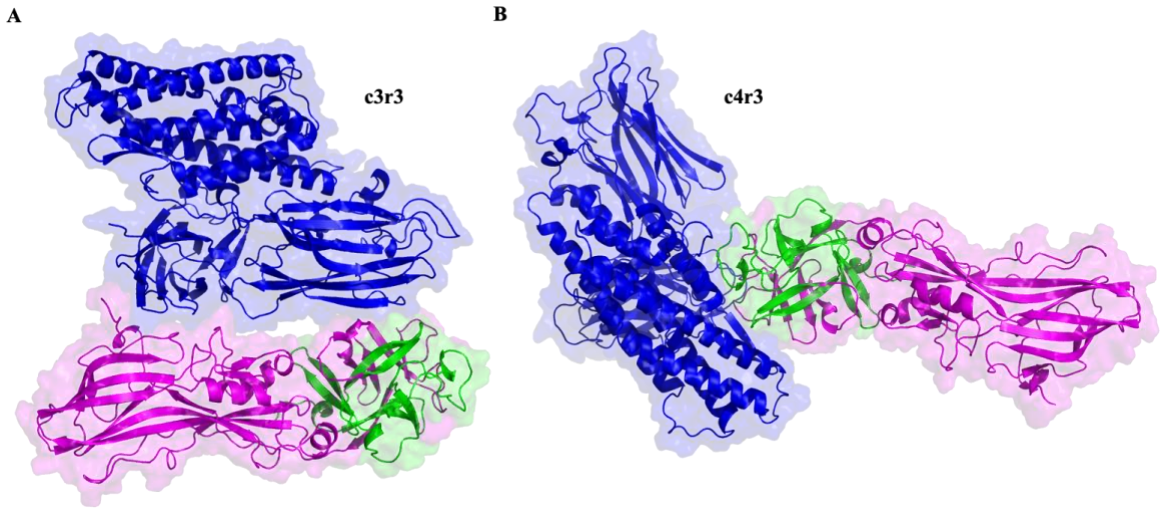

Figure 4. Cry48Aa1-Tpp49Aa1 models. The predicted dimer structures are illustrated showing Cry48Aa1 (dark blue) and Tpp49Aa1 (magenta with the region of Tpp49Aa1 that has been experimentally shown to interact with Cry48Aa1 shown in green). (A) Model c3r3. (B) Model c4r3.

Subsequent analysis of the interfacial interactions of model c3r3 **(Fig 4.A)** identified a large interface area of 1066.6 $Å^2$ involving 7 hydrogen bonds and a binding energy of -7.4 kcal mol^{-1}. In comparison, model c4r3 **(Fig 4.B)** contains an interface area of 802.1 $Å^2$ involving 4 hydrogen bonds and a binding energy of -9.1 kcal mol^{-1}. In model c3r3, Cry48Aa1 interacts via domain II (beta-sheet domain) and domain III (antiparallel beta-sandwich domain) with both the N- and C-terminal domains of Tpp49Aa1. This contrasts with model c4r3, where Cry48Aa1 interacts via domain I (alpha-helical bundle domain) with the Tpp49Aa1 N-terminal domain alone.

3.5. Cellular models of Cry48Aa1 and Tpp49Aa1

The cellular toxicity of Cry48Aa1 and Tpp49Aa1 was investigated in two cell lines derived from the known target species *C. quinquefasciatus* (MRA-918 and the ovary-derived Hsu), and

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.476343; this version posted January 15, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

a cell line from the same genus - but unconfirmed target - *C. tarsalis* (Ct). To investigate cytotoxicity, resazurin was used as an indicator of metabolic activity and cellular viability [63]. Cry48Aa1-Tpp49Aa1 treatment reduced cellular metabolism in all three cell lines confirming the functional effect of the proteins generated in this study and indicating *C. tarsalis* as a new potential target organism for Cry48Aa1-Tpp49Aa1 (Fig 5).

In the known target species, *C. quinquefasciatus*, higher concentrations (20, 10, & 5 µg/mL) of the Cry48Aa1-Tpp49Aa1 combination reduced cellular metabolic activity after 30 h in both cell lines (**Fig 5.A**). By 48 h, the cells subjected to the lower concentration (5 µg/mL) of combined Cry48Aa1 and Tpp49Aa1 recovered with no observable difference between treated and controls (data not shown), while higher concentrations (20 & 10 µg/mL) still affect viability (**Fig 5.B**). When added as individual components, both Cry48Aa1 and Tpp49Aa1 caused reduced viability, as quantified at 30 h (**Fig 5.A**). However, by 48 h cells treated with an individual protein appeared to have largely recovered when compared to the combination. This suggests Cry48Aa1 and Tpp49Aa1 have an adverse cellular effect as individual components, but this effect is not sufficient to elicit long-term toxicity and cell death. These noted decreases on cell viability were similar in the *C. tarsalis* derived line (Ct), suggesting this species may also be susceptible to Cry48Aa1/Tpp49Aa1. No effects on the metabolic activity or cellular morphology were apparent from addition to SF9 cells, either in combination or as individual proteins, consistent with the fact that Cry48Aa1-Tpp49Aa1 have no effect on the source insect *S. frugiperda* (data not shown).

Light microscopy of the MRA-918 cell line shows the Cry48Aa1-Tpp49Aa1 combination - and to a lesser degree Tpp49Aa1 alone – induces vacuolisation (**Fig 5.C**). This supports

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.476343; this version posted January 15, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

previous studies which show Cry48Aa1-Tpp49Aa1, in combination, to induce cytoplasmic and mitochondrial vacuolisation in midgut epithelial cells isolated from *C. quinquefasciatus* larvae [6]. Similar phenotypes are also reported with Tpp49Aa-related Toxin_10 family members Tpp1Aa2/Tpp2Aa2 in both larvae [64,65] and cells expressing the relevant Cpm1 receptor [66]. There is also a slight rounding of cellular morphology that is present in Cry48Aa1, Tpp49Aa1, and a combination of both, at 24 h post addition. At 48 h post addition, there is no apparent difference between control cells and those treated with Cry48Aa1 or Tpp49Aa1 alone, yet the combination shows a retraction of cellular processes and a rounded morphology, with many cells detached from the plate. Complementary to the cellular viability assays, this indicates a small effect of Cry48Aa1 or Tpp49Aa1 alone on cellular health. Whether these adverse effects on viability are mediated through the same cellular pathways as the Cry48Aa1-Tpp49Aa1 combination is yet to be determined. All three *Culex* cell lines may be useful tools for investigating the cellular pathways through which Cry48Aa1-Tpp49Aa1 toxicity is elicited.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.476343; this version posted January 15, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

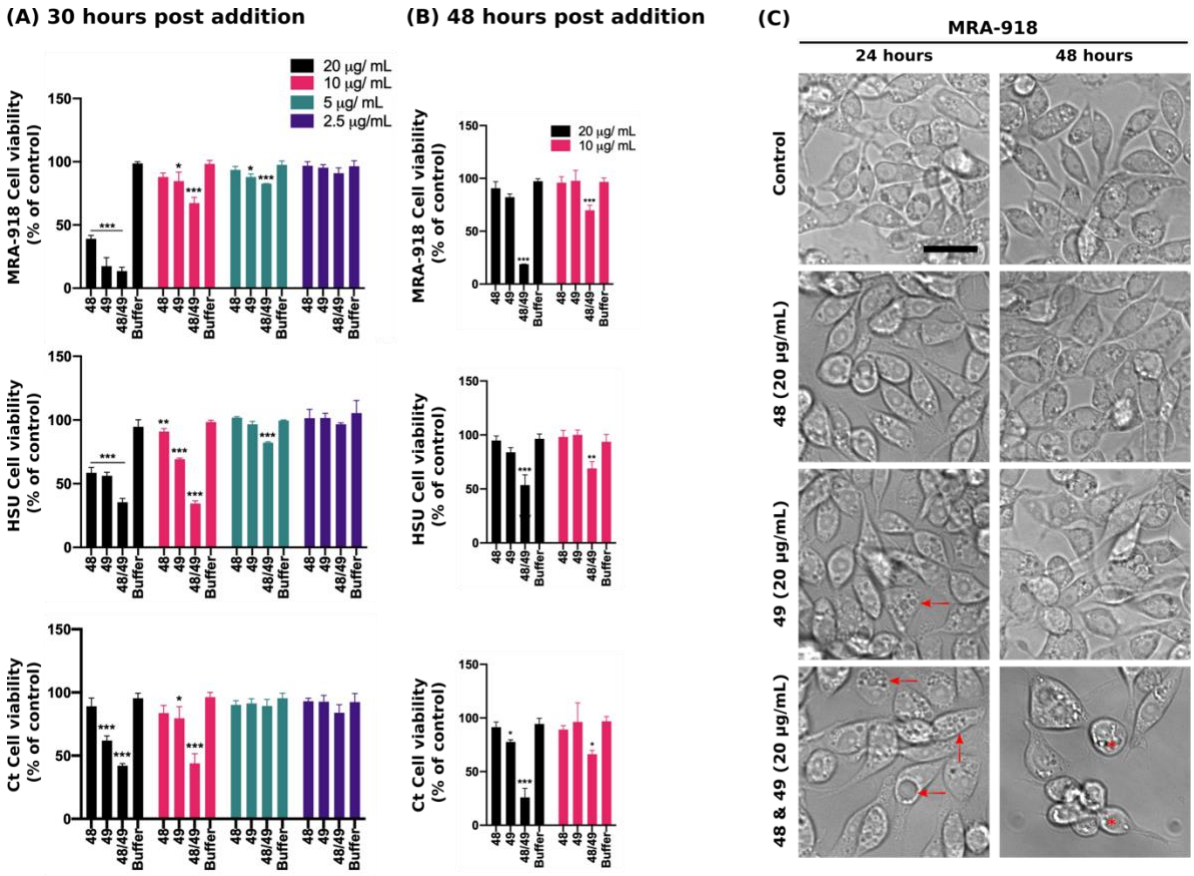

Figure 5. Cry48Aa1 and Tpp49Aa1 reduce cell viability in *C. quinquefasciatus* and *C. tarsalis* cell lines.

To assess the use of various insect cell lines for modelling Cry48Aa1 and Tpp49Aa1 toxicity, resazurin was used to quantify the effect of these proteins on cell viability. *C. quinquefasciatus* (MRA-918, Hsu) and *C. tarsalis* (Ct) cells were treated with a range of concentrations of either Cry48Aa1 ("48"), Tpp49Aa1 ("49"), or equimolar amounts in combination (48/49). Resazurin was added to the cells 24 h after the insecticidal proteins. **(A)** Reduced cell viability (% of control) is observable at 8 h post resazurin addition, in all cell lines, with either 48/49 in combination or alone. **(B)** Twenty-four hours post resazurin addition (48 h post protein addition), reduced cell viability is still seen with both insecticidal proteins (alone and in combination), however, the decreased viability between the single protein treatments is reduced compared to the combination, indicating recovery. All data are presented as the mean ± SD and all statistical analysis was performed using a one-way ANOVA, n = 3 (***$p < 0.001$, **$p < 0.01$, * $p < 0.05$). **(C)** Brightfield images of MRA-918 cells at 24 h, post insecticidal protein addition, show a slightly rounded cellular morphology with either 48 or 49, alongside the presence of some small vacuoles in 49 alone (red arrows). Larger vacuoles are apparent in 48/49 combination (red arrows). At 48 h post 48/49 addition, cells have rounded and detached from the plate (red asterix). Representative scale bar = 10 µm.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.476343; this version posted January 15, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

4. Conclusions

We have used megahertz SFX at the European XFEL to determine the Tpp49Aa1 structure to a final resolution of 2.2 Å. It has been shown that the nano-focus option at the SPB/SFX instrument, in combination with megahertz repetition rates, can be used for rapid and high-quality data collection from natively grown nanocrystals and this paves the way for further investigations on structure and dynamics of bacterial insecticides. In addition, we have suggested *C. tarsalis* as a new target organism of Cry48Aa1-Tpp49Aa1, and identified MRA-918, Hsu, and Ct cell lines as useful models for investigating the cellular pathways by which toxicity is elicited. In the future, structural investigations of Cry48Aa1, both in the crystalline protoxin and activated forms, and the Cry48Aa1-Tpp49Aa1 complex, will shed further light on the mode of action of these toxins and aid the design of optimized and safe insecticides to combat mosquitoes as vectors of emerging diseases such as Dengue, West Nile Virus and Zika.

5. Acknowledgements

We acknowledge European XFEL in Schenefeld, Germany, for provision of X-ray free-electron laser beamtime at the SPB/SFX Instrument and use of the XBI laboratories and would like to thank the staff for their assistance. We acknowledge the support of the Supercomputing Wales project, which is part-funded by the European Regional Development Fund (ERDF) via the Welsh Government. This research was supported in part through the Maxwell computational resources operated at Deutsches Elektronen-Synchrotron DESY, Hamburg, Germany. We would like to thank Frank Schluenzen (DESY) for installation and testing of Alphafold2 on the Maxwell-cluster. We would like to thank Dr. Esther Schnettler for kindly providing us with the *C. tarsalis* (Ct) cell line.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.476343; this version posted January 15, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

6. *Funding*

This work was supported by the Biotechnology and Biological Sciences Research Council (BBSRC) [grant reference BB/S002774/1] and BBSRC-funded South West Biosciences Doctoral Training Partnership [training grant reference BB/M009122/1]. PLX thanks Joachim Herz Stiftung for a fellowship.

7. *Author contributions*

Experimental design: CB, HNC, NC and DO; Sample production: LJW, HLB, NC, ELE and CB; Protein analysis and cell-based assays were carried out by HLB and MY; Experimental lead at European XFEL: DO; Sample delivery at European XFEL: MK, JV, GP, SA, JK, JB; European XFEL data collection was executed by: AM, AR, KD, KL, DO, RJB, RdW, CK, HK, GM, MS, APM, KLP and RS. Experimental settings at European XFEL were coordinated by RJB and APM; Data processing was carried out by LG, OY, and MG; TEM imaging was carried out by PLX and RS. Data were analysed by DO, PJR, CB and LJW. The manuscript was written by LJW, CB, HLB and DO with RJB, OY, MG, NC, and input from all the authors.

8. *References*

1. Crickmore, N. *et al.* A structure-based nomenclature for *Bacillus thuringiensis* and other bacteria-derived pesticidal proteins. *J. Invertebr. Pathol.* 107438 (2020) doi:10.1016/j.jip.2020.107438.

2. Jones, G. W. *et al.* A new Cry toxin with a unique two-component dependency from *Bacillus sphaericus*. *FASEB J.* **21**, 4112–4120 (2007).

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.476343; this version posted January 15, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

3. Jones, G. W., Wirth, M. C., Monnerat, R. G. & Berry, C. The Cry48Aa-Cry49Aa binary toxin from *Bacillus sphaericus* exhibits highly restricted target specificity. *Environ. Microbiol.* **10**, 2418–2424 (2008).

4. Guo, Q. Y., Hu, X. M., Cai, Q. X., Yan, J. P. & Yuan, Z. M. Interaction of *Lysinibacillus sphaericus* Cry48Aa/Cry49Aa toxin with midgut brush-border membrane fractions from *Culex quinquefasciatus* larvae. *Insect Mol. Biol.* **25**, 163–170 (2016).

5. Rezende, T. M. T. *et al.* Identification of Cry48Aa/Cry49Aa toxin ligands in the midgut of *Culex quinquefasciatus* larvae. *Insect Biochem. Mol. Biol.* **88**, 63–70 (2017).

6. De Melo, J. V. *et al.* Cytopathological effects of *Bacillus sphaericus* Cry48Aa/Cry49Aa toxin on binary toxin-susceptible and -resistant *Culex quinquefasciatus* larvae. *Appl. Environ. Microbiol.* **75**, 4782–4789 (2009).

7. Guo, Q., Ding, L., Gao, Y., Niu, Y. & Dai, X. Cys183 and Cys258 in Cry49Aa toxin from *Lysinibacillus sphaericus* are essential for toxicity to *Culex quinquefasciatus* larvae. *Arch. Microbiol.* (2021) doi:10.1007/s00203-021-02436-x.

8. Kelker, M. S. *et al.* Structural and biophysical characterization of *Bacillus thuringiensis* insecticidal proteins Cry34Ab1 and Cry35Ab1. *PLoS One* **9**, 112555 (2014).

9. Chapman, H. N. *et al.* Femtosecond X-ray protein nanocrystallography. *Nature* **470**, 73–78 (2011).

10. Chapman, H. N. *et al.* Femtosecond time-delay X-ray holography. *Nature* **448**, 676–679 (2007).

11. Boutet, S. *et al.* High-Resolution Protein Structure Determination by Serial

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.476343; this version posted January 15, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Femtosecond Crystallography. *Science (80-.).* **337**, 362–364 (2012).

12. Sawaya, M. R. *et al.* Protein crystal structure obtained at 2.9 Å resolution from injecting bacterial cells into an X-ray free-electron laser beam. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 12769–12774 (2014).

13. Colletier, J. P. *et al.* De novo phasing with X-ray laser reveals mosquito larvicide BinAB structure. *Nature* **539**, 43–47 (2016).

14. Tetreau, G. *et al.* Serial femtosecond crystallography on in vivo-grown crystals drives elucidation of mosquitocidal Cyt1Aa bioactivation cascade. *Nat. Commun.* **11**, 1–16 (2020).

15. Tetreau, G. *et al.* De novo determination of mosquitocidal Cry11Aa and Cry11Ba structures from naturally-occurring nanocrystals. *bioRxiv* 2021.12.15.472578 (2021) doi:10.1101/2021.12.15.472578.

16. Mancuso, A. P. *et al.* The single particles, clusters and biomolecules and serial femtosecond crystallography instrument of the european XFEL: Initial installation. *J. Synchrotron Radiat.* **26**, 660–676 (2019).

17. Yuan, Z. M., Pei, G. F., Regis, L., Nielsen-Leroux, C. & Cai, Q. X. Cross-resistance between strains of *Bacillus sphaericus* but not *B. thuringiensis israelensis* in colonies of the mosquito *Culex quinquefasciatus*. *Med. Vet. Entomol.* **17**, 251–256 (2003).

18. Monnerat, R. G. *et al.* Screening of Brazilian *Bacillus thuringiensis* isolates active against *Spodoptera frugiperda*, *Plutella xylostella* and *Anticarsia gemmatalis*. *Biol. Control* **41**, 291–295 (2007).

19. Hsu, S. H., Mao, W. H. & Cross, J. H. Establishment of a line of cells derived from ovarian tissue of *Culex quinquefasciatus* say. *J. Med. Entomol.* **7**, 703–707 (1970).

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.476343; this version posted January 15, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

20. Chao, J. & Ball, G. H. A comparison of amino acid utilization by cell lines of *Culex tarsalis* and of *Culex pipiens*. in *Invertebrate Tissue Culture* 263–266 (Academic Press, 1976). doi:10.1016/b978-0-12-429740-1.50028-x.

21. Summers, M. & Smith, G. *A manual of methods for baculovirus vectors and insect cell culture procedures*. (1987).

22. Han, H. *et al.* The XBI BioLab for life science experiments at the European XFEL. *J. Appl. Crystallogr.* **54**, 7–21 (2021).

23. Wiedorn, M. O. *et al.* Megahertz serial crystallography. *Nat. Commun.* **9**, 1–11 (2018).

24. Yefanov, O. *et al.* Evaluation of serial crystallographic structure determination within megahertz pulse trains. *Struct. Dyn.* **6**, 064702 (2019).

25. Wiedorn, M. O. *et al.* Rapid sample delivery for megahertz serial crystallography at X-ray FELs. *IUCrJ* **5**, 574–584 (2018).

26. Knoška, J. *et al.* Ultracompact 3D microfluidics for time-resolved structural biology. *Nat. Commun.* **11**, 1–12 (2020).

27. Oberthuer, D. *et al.* Double-flow focused liquid injector for efficient serial femtosecond crystallography. *Sci. Rep.* **7**, 1–12 (2017).

28. Bean, R. J., Aquila, A., Samoylova, L. & Mancuso, A. P. Design of the mirror optical systems for coherent diffractive imaging at the SPB/SFX instrument of the European XFEL. *J. Opt.* **18**, 074011 (2016).

29. Mariani, V. *et al.* OnDA: Online data analysis and feedback for serial X-ray imaging. *J. Appl. Crystallogr.* **49**, 1073–1080 (2016).

30. Barty, A. *et al.* Cheetah: Software for high-throughput reduction and analysis of serial

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.476343; this version posted January 15, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

femtosecond X-ray diffraction data. *J. Appl. Crystallogr.* **47**, 1118–1131 (2014).

31. White, T. A. *et al.* Recent developments in CrystFEL. *J. Appl. Crystallogr.* **49**, 680–689 (2016).

32. White, T. A. *et al.* CrystFEL: A software suite for snapshot serial crystallography. *J. Appl. Crystallogr.* **45**, 335–341 (2012).

33. Gevorkov, Y. *et al.* XGANDALF - Extended gradient descent algorithm for lattice finding. *Acta Crystallogr. Sect. A Found. Adv.* **75**, 694–704 (2019).

34. Yefanov, O. *et al.* Accurate determination of segmented X-ray detector geometry. *Opt. Express* **23**, 28459 (2015).

35. Kantardjieff, K. A. & Rupp, B. Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Sci.* **12**, 1865–1871 (2003).

36. Adams, P. D. *et al.* PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 213–221 (2010).

37. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).

38. Terwilliger, T. C. *et al.* Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **64**, 61–69 (2007).

39. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010).

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.476343; this version posted January 15, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

40. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **68**, 352–367 (2012).

41. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 355–367 (2011).

42. Potterton, L. *et al.* CCP4i2: The new graphical user interface to the CCP4 program suite. *Acta Crystallogr. Sect. D Struct. Biol.* **74**, 68–84 (2018).

43. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

44. Kozakov, D. *et al.* The ClusPro web server for protein-protein docking. *Nat. Protoc.* **12**, 255–278 (2017).

45. Chaudhury, S. *et al.* Benchmarking and analysis of protein docking performance in Rosetta v3.2. *PLoS One* **6**, e22477 (2011).

46. Alford, R. F. *et al.* The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J. Chem. Theory Comput* **13**, 3031–3048 (2017).

47. Abraham, M. J. *et al.* Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1**–**2**, 19–25 (2015).

48. Worthy, H. L. *et al.* The Crystal Structure of *Bacillus cereus* HblL1. *Toxins (Basel).* **13**, 253 (2021).

49. Krissinel, E. & Henrick, K. Inference of Macromolecular Assemblies from Crystalline State. *J. Mol. Biol.* **372**, 774–797 (2007).

50. Hazes, B. The (QxW)3 domain: A flexible lectin scaffold. *Protein Sci.* **5**, 1490–1501 (1996).

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.476343; this version posted January 15, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

51. Knowles, B. H., Thomas, W. E. & Ellar, D. J. Lectin-like binding of *Bacillus thuringiensis* var, kurstaki lepidopteran-specific toxin is an initial step in insecticidal action. *FEBS Lett.* **168**, 197–202 (1984).

52. Dennis, R. D., Wiegandt, H., Haustein, D., Knowles, B. H. & Ellar, D. J. Thin layer chromatography overlay technique in the analysis of the binding of the solubilized protoxin of *Bacillus thuringiensis* var. kurstaki to an insect glycosphingolipid of known structure. *Biomed. Chromatogr.* **1**, 31–37 (1986).

53. Sangadala, S., Azadi, P., Carlson, R. & Adang, M. J. Carbohydrate analyses of *manduca sexta* aminopeptidase n, co-purifying neutral lipids and their functional interactions with *Bacillus thuringiensis* cry1ac toxin. *Insect Biochem. Mol. Biol.* **32**, 97–107 (2001).

54. Jurat-Fuentes, J. L., Gould, F. L. & Adang, M. J. Altered glycosylation of 63- and 68-kilodalton microvillar proteins in *Heliothis virescens* correlates with reduced Cry1 toxin binding, decreased pore formation, and increased resistance to *Bacillus thuringiensis* Cry1 toxins. *Appl. Environ. Microbiol.* **68**, 5711–5717 (2002).

55. Parker, M. W. *et al.* Structure of the Aeromonas toxin proaerolysin in its water-soluble and membrane-channel states. *Nature* **367**, 292–295 (1994).

56. Szczesny, P. *et al.* Extending the aerolysin family: From bacteria to vertebrates. *PLoS One* **6**, e20349 (2011).

57. Lacomel, C. J., Dunstone, M. A. & Spicer, B. A. Branching out the aerolysin, ETX/MTX-2 and Toxin_10 family of pore forming proteins. *J. Invertebr. Pathol.* 107570 (2021) doi:10.1016/j.jip.2021.107570.

58. Iacovache, I. *et al.* Cryo-EM structure of aerolysin variants reveals a novel protein fold

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.14.476343; this version posted January 15, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and the pore-formation process. *Nat. Commun.* **7**, 1–8 (2016).

59. Iacovache, I. *et al.* Dual chaperone role of the c-terminal propeptide in folding and oligomerization of the pore-forming toxin aerolysin. *PLoS Pathog.* **7**, 1002135 (2011).

60. Lawrence, M. C. & Colman, P. M. Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234**, 946–950 (1993).

61. Holm, L. Using Dali for Protein Structure Comparison. *Methods Mol. Biol.* **2112**, 29–42 (2020).

62. Tetreau, G. How does *Bacillus thuringiensis* crystallize such a large diversity of toxins? *Toxins (Basel).* **13**, 443 (2021).

63. Präbst, K., Engelhardt, H., Ringgeler, S. & Hübner, H. Basic colorimetric proliferation assays: MTT, WST, and resazurin. in *Methods in Molecular Biology* vol. 1601 1–17 (2017).

64. De Melo, J. V., Vasconcelos, R. H. T., Furtado, A. F., Peixoto, C. A. & Silva-Filha, M. H. N. L. Ultrastructural analysis of midgut cells from *Culex quinquefasciatus* (Diptera: Culicidae) larvae resistant to *Bacillus sphaericus*. *Micron* **39**, 1342–1350 (2008).

65. Singh, G. J. P. & Gill, S. S. An electron microscope study of the toxic action of *Bacillus sphaericus* in *Culex quinquefasciatus* larvae. *J. Invertebr. Pathol.* **52**, 237–247 (1988).

66. Opota, O. *et al. Bacillus sphaericus* binary toxin elicits host cell autophagy as a response to intoxication. *PLoS One* **6**, (2011).