

# Structural Analysis of Pesticidal Proteins and Their Interactions



**Lainey Jay Williamson**

School of Biosciences

Cardiff University

A thesis submitted to Cardiff University for the degree of  
*Doctor of Philosophy*

December 2023



## Abstract

Due to their high specificity and biocompatibility, pesticidal proteins produced by *Bacillus thuringiensis* and *Lysinibacillus sphaericus* have been successfully applied as bioinsecticides for use in the control of agricultural pests and disease vectors. Given their significance, it is critical that we understand the mechanisms by which *B. thuringiensis* and *L. sphaericus* exert their pesticidal activity, both to counteract emerging field resistance and develop new bioinsecticides with enhanced potency, stability, and target insect range. In this work, a combination of serial femtosecond crystallography (SFX), single-particle cryogenic electron microscopy (cryoEM), and computational modelling were employed to study the structure and interactions of *B. thuringiensis* and *L. sphaericus* pesticidal proteins.

Chapter 3 uses SFX to elucidate the Cry8Ba2 structure from crystals produced naturally by *B. thuringiensis*. The Cry8Ba2 structure represents the first structure of a full-length Cry protein elucidated from natural crystals and provides insight into crystal packing and the biological function of the pro-toxin domains.

Chapter 4 builds on this work by utilising SFX to elucidate the structure of Tpp49Aa1 from the binary Cry48/Tpp49 mosquitocidal toxin. Complementary pH mixing studies enabled the early structural changes in crystal dissolution to be investigated. In addition, the interaction of Tpp49Aa1 with its partner protein, Cry48Aa1, was predicted, enabling discussion related to models for mode of action. Finally, insect bioassays were performed, leading to the identification of new target mosquito species.

Chapter 5 investigates the interaction of the Tpp2 protein, from the binary Tpp1/Tpp2 mosquitocidal toxin, with its target receptor, Cqm1, using computational modelling. The predicted model was compared with that of the Tpp2Aa2-Cqm1 structure elucidated using single-particle cryoEM.

Directed by the Tpp2Aa2-Cqm1 structure, mutagenesis studies were performed with the aim of engineering the Tpp1Aa2/Tpp2Aa2 pesticidal protein to broaden the target insect range.

## Table of Contents

<b>1.</b>	<b>Introduction.....</b>	<b>1</b>
1.1.	The role of pest insects in agriculture and human disease .....	1
1.2.	<b><i>Bacillus thuringiensis</i> and <i>Lysinibacillus sphaericus</i>.....</b>	<b>2</b>
1.2.1.	Nomenclature for bacterial pesticidal proteins .....	3
1.2.2.	Crystal (Cry) protein class.....	6
1.2.2.1.	Structural analysis of Cry proteins .....	7
1.2.2.2.	Mechanism of action of Cry proteins .....	12
1.2.2.3.	Resistance mechanisms to Cry proteins .....	15
1.2.3.	Toxin-10 Pesticidal Protein (Tpp) class .....	16
1.2.3.1.	Structural analysis of Tpp proteins .....	17
1.2.3.2.	Mechanism of action of Tpp proteins.....	23
1.2.3.3.	Resistance mechanisms to Tpp proteins.....	26
1.3.	<b>Studying pesticidal proteins using structural biology techniques .....</b>	<b>27</b>
1.3.1.	X-ray crystallography .....	29
1.3.1.1.	Conventional crystallography.....	29
1.3.1.2.	Serial femtosecond crystallography (SFX) .....	31
1.3.2.	Single-particle cryogenic electron microscopy (cryoEM) .....	33
1.3.3.	Computational modelling .....	35
1.3.3.1.	Protein structure prediction.....	36
1.3.3.2.	Molecular Docking .....	37
1.3.3.3.	Molecular Dynamics (MD) .....	38
1.4.	<b>Aims and Objectives .....</b>	<b>39</b>
1.4.1.	Using serial femtosecond crystallography (SFX) to elucidate the structures of pesticidal proteins from natural crystals.....	39
1.4.2.	Structural analysis of the interaction of Tpp2Aa2 with its mosquito receptor, <i>Culex quinquefasciatus</i> maltase 1 (Cqm1) .....	39
<b>2.</b>	<b>Materials and Methods .....</b>	<b>41</b>
2.1.	<b>Materials .....</b>	<b>41</b>
2.1.1.	Chemicals and media .....	41
2.1.2.	Antibiotics.....	43
2.1.3.	Bacterial growth media .....	43
2.1.4.	Bacterial strains .....	44
2.1.5.	Vectors for protein expression .....	45
2.2.	<b>Molecular biology.....</b>	<b>45</b>

2.2.1.	DNA oligonucleotides.....	45
2.2.2.	Agarose gel electrophoresis .....	46
2.2.3.	DNA purification and isolation .....	46
2.2.4.	DNA quantification .....	46
2.2.5.	Whole-vector site-directed mutagenesis (SDM) .....	46
2.2.6.	Bacterial transformation .....	48
2.2.7.	Sequencing .....	49
2.2.8.	Preparation of <i>E. coli</i> dimethyl sulphoxide stocks .....	49
<b>2.3.</b>	<b>Protein expression .....</b>	<b>49</b>
2.3.1.	Expression testing of Cqm1 protein in <i>E. coli</i> .....	49
2.3.2.	Expression of Cqm1 protein in <i>E. coli</i> .....	50
2.3.3.	Expression of Tpp1Aa2 and Tpp2Aa2 proteins in <i>E. coli</i> .....	51
2.3.4.	Expression of Cry48Aa1, Tpp49Aa1 and Cry8Ba2 crystal proteins .....	52
<b>2.4.</b>	<b>Recombinant protein purification.....</b>	<b>53</b>
2.4.1.	Nickel affinity chromatography .....	53
2.4.2.	Glutathione-S-transferase (GST) affinity chromatography.....	53
2.4.3.	Size exclusion chromatography (SEC) .....	54
<b>2.5.</b>	<b>Protein analysis methods.....</b>	<b>54</b>
2.5.1.	Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)...	54
2.5.2.	Determination of protein concentration .....	55
2.5.3.	Trypsin activation of proteins .....	55
2.5.4.	Mosquito toxicity bioassays .....	56
2.5.5.	Insect feeding bioassays.....	56
2.5.6.	Alpha-glucosidase assays .....	57
2.5.7.	Static light scattering (RALS) and refractive index (RI) measurements.....	57
2.5.8.	Transmission electron microscopy (TEM) analysis of crystal proteins .....	58
2.5.9.	N-terminal sequencing .....	58
<b>2.6.</b>	<b>Computational modelling and analysis.....</b>	<b>59</b>
2.6.1.	Preparation of starting structures .....	59
2.6.2.	Molecular docking .....	59
2.6.2.1.	ClusPro.....	59
2.6.2.2.	RosettaDock .....	60
2.6.2.3.	AlphaFold-Multimer.....	61
2.6.2.4.	ColabFold.....	61
2.6.3.	Interface analysis .....	61
2.6.4.	Molecular Dynamics (MD) simulations.....	62
2.6.4.1.	GROMACS software.....	62

2.6.4.2. Forcefields .....	62
2.6.4.3. Input structure preparation.....	63
2.6.4.4. Energy minimisation .....	64
2.6.4.5. Two-step equilibration.....	64
2.6.4.6. MD production run .....	64
2.6.4.7. MD analysis .....	65
2.6.4.8. Root-mean-square deviation (RMSD) analysis .....	65
2.6.4.9. Radius of gyration (Rg) analysis.....	66
<b>2.7. Serial femtosecond crystallography (SFX) .....</b>	<b>66</b>
2.7.1. Sample preparation.....	66
2.7.2. Diffraction data collection and processing .....	67
2.7.3. Diffraction data analysis.....	67
2.7.4. pH mixing studies.....	68
<b>2.8. Single-particle cryogenic electron-microscopy (cryoEM).....</b>	<b>68</b>
2.8.1. Sample preparation.....	68
2.8.2. Sample screening and data collection .....	69
2.8.3. Data processing and model building.....	69
<b>3. Structural analysis of the Cry8Ba2 pesticidal protein .....</b>	<b>71</b>
<b>3.1. Introduction .....</b>	<b>71</b>
<b>3.2. Results and Discussion.....</b>	<b>72</b>
3.2.1. Production of the Cry8Ba2 crystal protein .....	72
3.2.1.1. Crystal protein expression and isolation.....	72
3.2.1.2. Transmission electron microscopy (TEM) of nanocrystals .....	74
3.2.2. Serial femtosecond crystallography (SFX).....	75
3.2.2.1. Diffraction data collection and processing .....	76
3.2.2.2. Molecular replacement (MR) and structure solution .....	84
3.2.2.3. General features of the Cry8Ba2 structure .....	86
3.2.2.4. Crystal packing of the Cry8Ba2 structure .....	91
3.2.2.5. Comparison with related protein structures .....	104
3.2.3. Insect feeding bioassays.....	113
<b>3.3. Conclusions .....</b>	<b>114</b>
<b>4. Structural analysis of the Cry48Aa1/Tpp49Aa1 pesticidal protein .....</b>	<b>116</b>
<b>4.1. Introduction .....</b>	<b>116</b>
<b>4.2. Results and Discussion.....</b>	<b>117</b>
4.2.1. Production of the Cry48Aa1 and Tpp49Aa1 crystal proteins .....	117
4.2.1.1. Crystal protein expression and isolation.....	117

4.2.1.2. Transmission electron microscopy (TEM) of nanocrystals .....	118
4.2.2. Serial femtosecond crystallography (SFX) of Tpp49Aa1 .....	120
4.2.2.1. Diffraction data collection and processing .....	120
4.2.2.2. Molecular replacement (MR) and structure solution .....	124
4.2.2.3. General features of the Tpp49Aa1 structure .....	126
4.2.2.4. pH mixing studies of the Tpp49Aa1 native crystals .....	142
4.2.3. Serial femtosecond crystallography (SFX) of Cry48Aa1.....	152
4.2.4. Prediction of the Cry48Aa1-Tpp49Aa1 interaction.....	152
4.2.4.1. Preparation of Cry48Aa1 and Tpp49Aa1 starting structures .....	154
4.2.4.2. Molecular docking studies .....	158
4.2.4.3. Interface analysis .....	161
4.2.4.4. Molecular dynamics (MD) simulations .....	162
4.2.4.5. Root-mean-square deviation (RMSD) trajectory analysis.....	166
4.2.4.6. Radius of gyration (Rg) trajectory analysis .....	170
4.2.4.7. Selection of most likely Cry48Aa1-Tpp49Aa1 model.....	173
4.2.5. Insect bioassays .....	174
<b>4.3. Conclusions .....</b>	<b>175</b>
<b>5. Characterisation of the Tpp1Aa2/Tpp2Aa2 pesticidal protein and its interaction with the receptor protein, Cqm1.....</b>	<b>177</b>
<b>5.1. Introduction .....</b>	<b>177</b>
<b>5.2. Results and Discussion .....</b>	<b>178</b>
5.2.1. Expression of Tpp1Aa2 and Tpp2Aa2 in <i>E. coli</i> .....	178
5.2.2. Purification of Tpp1Aa2 and Tpp2Aa2 .....	179
5.2.3. Protein activation using trypsin treatment .....	181
5.2.4. Mosquito larval bioassays of Tpp1Aa2/Tpp2Aa2.....	183
5.2.5. Expression of Cqm1 in <i>E. coli</i> .....	184
5.2.6. Purification of Cqm1 .....	186
5.2.7. Glucosidase assays .....	187
5.2.8. Prediction of the Tpp2Aa3-Cqm1 interaction .....	189
5.2.8.1. Preparation of Tpp2 and Cqm1 starting structures.....	191
5.2.8.2. Molecular docking studies .....	193
5.2.8.3. Interface analysis .....	196
5.2.8.4. Molecular dynamics (MD) simulations .....	197
5.2.8.5. Root-mean-square deviation (RMSD) trajectory analysis.....	198
5.2.8.6. Radius of gyration (Rg) trajectory analysis .....	201
5.2.8.7. Selection of most likely Tpp2Aa3-Cqm1 model .....	203

5.2.9.	Structural analysis of the Tpp2Aa2-Cqm1 complex using single-particle cryogenic electron-microscopy (cryoEM) .....	209
5.2.9.1.	Data collection and processing .....	209
5.2.9.2.	Comparison of the predicted model with the Tpp2Aa2-Cqm1 structure ....	209
5.2.9.3.	Modelling Tpp2-Cqm1 using Alphafold-Multimer and ColabFold .....	215
5.2.9.4.	General features of the Tpp2Aa2-Cqm1 complex .....	221
5.2.9.5.	Comparison with related structures .....	228
5.2.10.	Engineering Tpp1Aa2/Tpp2Aa2 to expand the target insect range .....	240
5.2.10.1.	Rationale for Tpp2Aa2 mutation .....	240
5.2.10.2.	Site-directed mutagenesis (SDM) of Tpp2Aa2 .....	240
5.2.10.3.	Expression of Tpp2Aa2-E65A mutant .....	241
5.2.10.4.	Mosquito larval bioassays.....	243
<b>5.3.</b>	<b>Conclusions .....</b>	<b>245</b>
<b>6.</b>	<b><i>Discussion</i>.....</b>	<b>248</b>
<b>6.1.</b>	<b>Overview .....</b>	<b>248</b>
<b>6.2.</b>	<b>Application of pesticidal proteins as bioinsecticides and beyond .....</b>	<b>248</b>
<b>6.3.</b>	<b>Importance of molecular structure in protein studies .....</b>	<b>252</b>
<b>6.4.</b>	<b>Future directions and limitations.....</b>	<b>255</b>
<b>6.5.</b>	<b>Summary .....</b>	<b>259</b>
<b>7.</b>	<b><i>References</i> .....</b>	<b>260</b>
<b>8.</b>	<b><i>Appendices</i> .....</b>	<b>290</b>

## Acknowledgements

First, I'd like to offer my sincere thanks to my supervisors Prof. Colin Berry, Dr. Emyr Lloyd-Evans, Dr. Pierre J. Rizkallah, and Dr. Helen Waller-Evans for their continued guidance, advice, and support. A special thanks to Prof. Colin Berry; it has been a pleasure to work in your research group and I am grateful for the abundant opportunities you have provided for my scientific development. Thank you to Dr. Pierre J. Rizkallah for your patience and long sessions spent teaching me protein crystallography. Thank you also to Dr. Hannah L. Best for being an excellent mentor and for all your efforts teaching me new techniques. Thank you to my funding body, BBSRC, and the SWBio Doctoral Training Partnership for providing valuable training to support my PhD and development as a scientist, as well as Cardiff University for the provision of research facilities.

I would also like to extend my thanks to the collaborators who played a part in this project. Dr. Dominik Oberthür, from Deutsches Elektronen-Synchrotron (DESY), who oversaw and gave me the opportunity to be a part of our beamtime at the European XFEL. Thank you for being so supportive of early career researchers. Similarly, thank you to the researchers and beamline scientists at DESY and the European XFEL for their hard work during the long data collection sessions. Thank you to A/Prof. Michelle A. Dunstone, Dr. Bradley A. Spicer, and lab members for welcoming me into your research group at Monash University and for your work on the Tpp2-Cqm1 project. Thank you also to Prof. Brian A. Federici at the University of California, and Dr. Hyun-Woo Park and Dr. Dennis K. Bideshi at the California Baptist University for your work on insect bioassays performed as part of the Cry48/Tpp49 project.

At Cardiff university, thank you to Dr. D. Dafydd Jones for hosting me during my first-year rotation project, Dr. Mark T. Young for your feedback as my assessor and advice surrounding static light scattering experiments, Dr.

Georgina Menzies, Andrew H. Mack, and James Davies for your help with molecular dynamics simulations, and Dr. Magdalena Lipka-Lloyd for your help setting up protein crystal trials. Thank you to members of the CB and ELE labs and many others in the School of Biosciences. A special thanks goes to Dr. Rochelle Ahmed, Dr. Rebecca E. A. Gwyther, and Emily A. Heath for your friendship over the last four-years and for helping me to become accustomed with new techniques.

Finally, I would like to extend my deepest gratitude to all my family and friends for their unconditional encouragement and support.

## List of Publications

**Williamson, L.J.**, Galchenkova, M., Best, H.L., Bean, R.J., Munke, A., Awel, S., Pena, G., Knoska, J., Schubert, R., Dörner, K., Park, H-W., Bideshi, D, K., Henkel, A., Kremling, V., Klopprogge, B., Lloyd-Evans, E., Young, M., Valerio, J., Kloos, M., Sikorski, M., Mills, G., Bielecki, J., Kirkwood, H., Kim, C., Wijn, R., Lorenzen, K., Lourdu, P. X., Rahmani Mashhour, A., Gelisio, L., Yefanov, O., Mancuso, A.P., Federici, B., Chapman, H.N., Crickmore, N., Rizkallah, P.J., Berry, C., Oberthür, D. 2023. Structure of the *Lysinibacillus sphaericus* Tpp49Aa1 pesticidal protein elucidated from natural crystals using MHz-SFX. *Proceedings of the National Academy of Sciences* 120(49), p. e2203241120. doi: [10.1073/pnas.2203241120](https://doi.org/10.1073/pnas.2203241120)

Best, H.L., **Williamson, L.J.**, Heath, E.A., Waller-Evans, H., Lloyd-Evans, E., Berry, C. 2023. The role of glycoconjugates as receptors for insecticidal proteins. *FEMS Microbiology Reviews* 47(4), p. fuad026. doi: [10.1093/femsre/fuad026](https://doi.org/10.1093/femsre/fuad026)

Best, H.L., **Williamson, L.J.**, Lipka-Lloyd, M., Waller-Evans, H., Lloyd-Evans, E., Rizkallah, P.J., Berry, C. 2022. The crystal structure of *Bacillus thuringiensis* Tpp80Aa1 and its interaction with galactose-containing glycolipids. *Toxins* 14(12), p. 863. doi: [10.3390/toxins14120863](https://doi.org/10.3390/toxins14120863)

Worthy, H.L., **Williamson, L.J.**, Auhim, H.S., Leppla, S.H., Sastalla, I., Jones, D.D., Rizkallah, P.J., Berry, C. 2021. The crystal structure of *Bacillus cereus* HblL<sub>1</sub>. *Toxins* 13(4), p. 253. doi: [10.3390/toxins13040253](https://doi.org/10.3390/toxins13040253)

## Abbreviations

Aegerolysin-related pesticidal protein (Gpp)  
Alkaline phosphatase (ALP)  
AlphaFold2 (AF2)  
Aminopeptidase N (APN)  
ATP-binding cassette (ABC)  
Bicinchoninic acid (BCA)  
Biological pesticides (biopesticides)  
Bovine serum albumin (BSA)  
Brush-border membrane vesicle (BBMV)  
Cadherin-like protein (CAD)  
Carbohydrate-binding module (CBM)  
Critical Assessment of Structure Prediction (CASP)  
Cryogenic electron microscopy (CryoEM)  
Crystal pesticidal protein (Cry)  
Cytolytic pesticidal protein (Cyt)  
Deionised water (dH<sub>2</sub>O)  
Deutsches Elektronen-Synchrotron (DESY)  
Dimethyl sulfoxide (DMSO)  
Double flow focused nozzle (DFFN)  
Electron Microscopy Data Bank (EMDB)  
*Escherichia coli* (*E. coli*)  
European X-ray Free Electron Laser (EuXFEL)  
Fast Fourier Transform (FFT)  
Glutathione-S-transferase (GST)  
Glycosylphosphatidylinositol (GPI)  
Isopropyl β-d-1-thiogalactopyranoside (IPTG)  
Lethal concentration 50 (LC<sub>50</sub>)  
Lethal concentration 95 (LC<sub>95</sub>)  
Luria-Bertani (LB)  
Lysis-Equilibration-Wash (LEW)

Lysomyristoylphosphatidylglycerol (LMPG)  
Madin-Darby canine kidney (MDCK)  
Megahertz (MHz)  
Microcrystal electron diffraction (microED)  
Molecular dynamics (MD)  
Molecular replacement (MR)  
Molecular weight (MW)  
Monte Carlo (MC)  
N-acetylgalactosamine (GalNAc)  
Phosphate-buffered saline (PBS)  
Polyvinylidene fluoride (PVDF)  
Potential energy ( $E_{\text{pot}}$ )  
Predicted local-distance difference test (pLDDT)  
Protein Data Bank (PDB)  
Radius of gyration ( $R_g$ )  
Refractive index (RI)  
Root-mean-square deviation (RMSD)  
Rosetta energy units (REU)  
Secreted insecticidal protein (Sip)  
Serial femtosecond crystallography (SFX)  
Shape correlation ( $Sc$ )  
Site-directed mutagenesis (SDM)  
Size exclusion chromatography (SEC)  
Small-angle electron diffraction (SAED)  
Small-angle neutron scattering (SANS)  
Sodium dodecyl sulphate (SDS)  
Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE)  
Solvation free energy gain ( $\Delta^iG$ )  
Static light scattering (RALS)  
Three-dimensional (3D)  
Toxin-10 pesticidal protein (Tpp)  
Transmission electron microscopy (TEM)

Two-dimensional (2D)

Vegetative insecticidal protein (Vip)

X-ray free electron laser (XFEL)

## **1. Introduction**

### **1.1. The role of pest insects in agriculture and human disease**

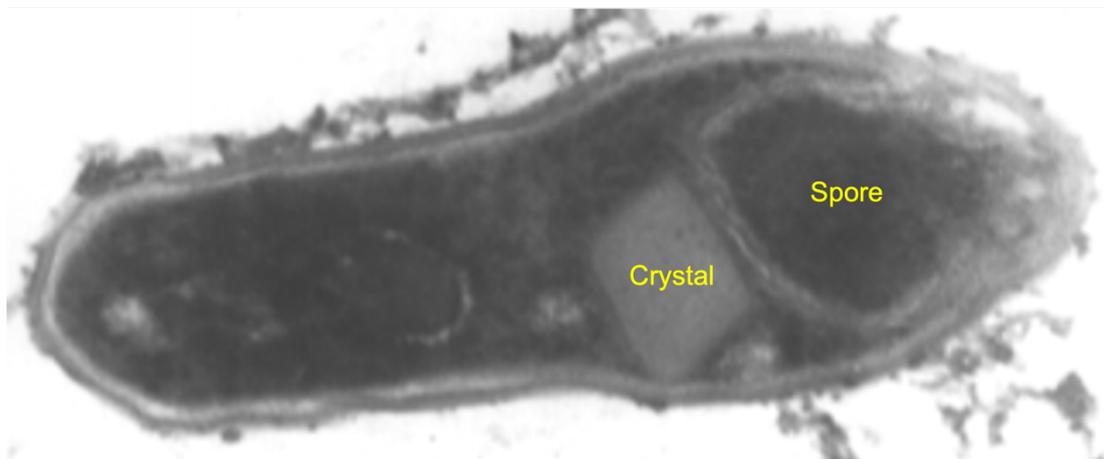
Across the world, pathogens and pests reduce agricultural yield by an estimated 20 – 40%, leading to food insecurity and substantial economic losses (IPPC Secretariat 2021). Insects across several orders (including Coleoptera – beetles, Lepidoptera – butterflies and moths, Diptera – flies, and Hemiptera – aphids and stinkbugs) and nematodes represent major crop pests. Such challenges are exacerbated by an increasing global population and, thus, an increased requirement for food, as well as climate change, which will lead to increased insect activity and, in turn, increased rates of agricultural loss (Skendžić et al. 2021). In addition, some insect pests, such as mosquitoes, act as vectors of human disease (Chala and Hamde 2021). The geographic range of several disease-transmitting mosquito species is expanding due to rising global temperatures (Thomson and Stanberry 2022), highlighting an ever-increasing need for methods of controlling their populations.

Over the years, chemical/synthetic pesticides have been employed to control insect pest populations. Despite their positive effect on agricultural yields and spread of human disease, the use of chemical pesticides is associated with several concerns, including environmental pollution, harm to beneficial insects, detrimental effects on human health and, in the case of agriculture, issues surrounding food safety (Pathak et al. 2022). Due to these disadvantages, biological pesticides (biopesticides), which are cost-effective, sustainable, environmentally friendly, and highly specific, are promising alternatives to chemical pesticides (Ayilara et al. 2023). Biopesticides fall into several classes, including plant extracts and botanicals (originating from plants), microbials (originating from microbes, including bacteria, fungi, viruses, protozoa, and algae), and pheromones and/or semiochemicals (e.g., chemicals emitted by plants) (Ayilara et al. 2023).

Microbials represent the most widely applied form of biopesticide, of which bacterial sources predominate. Within the bacterial genera, *Bacillus thuringiensis* (Palma et al. 2014) and *Lysinibacillus sphaericus* (Silva-Filha et al. 2014) are the most widely used, having been commercially applied with great success.

## **1.2. *Bacillus thuringiensis* and *Lysinibacillus sphaericus***

*B. thuringiensis* and *L. sphaericus* are gram-positive, rod-shaped, sporulating, and ubiquitous bacteria, isolated from water, soil, leaves, dead insects, decaying human tissue, as well as insectivorous mammals (Palma et al. 2014; Silva-Filha et al. 2014). The life cycle of *B. thuringiensis* and *L. sphaericus* can be broken down into vegetative and sporulation cycles. *B. thuringiensis* and *L. sphaericus* produce a wide variety of pesticidal proteins in the form of parasporal crystalline inclusions (**Fig 1.1**) and secreted soluble proteins, recently reclassified according to their structural homology by the Bacterial Pesticidal Protein Resource Center (BPPRC.org - Crickmore et al. 2021; Panneerselvam et al. 2022). These pesticidal proteins are ingested by the target insect larvae where they are solubilised in the gut (which can range from highly acidic (pH 3.6) to highly basic (pH 10.9) depending on the target insect), allowing them to interact specifically with receptors present on the midgut epithelium (Palma et al. 2014; Silva-Filha et al. 2014). This leads to downstream cytopathological effects and larval death (Palma et al. 2014; Silva-Filha et al. 2014). *B. thuringiensis* and *L. sphaericus* pesticidal proteins are active against a diverse range of insect orders and hence, have been commercially applied in genetically modified crops, such as maize, cotton, soya bean and rice (Cheng et al. 1998; Bernardi et al. 2012; Shrestha et al. 2018), and bacterial sprays that aid the control of disease-transmitting mosquitoes (Silva-Filha et al. 2014; Xiao and Wu 2019). Given their significance, it is critical that we understand the mechanisms by which *B. thuringiensis* and *L. sphaericus* exert their pesticidal activity, both to counteract emerging field resistance and to develop new bioinsecticides with enhanced potency, stability, and target insect range.



**Figure 1.1.** Electron micrograph illustrating the spore and crystal protein of *Lysinibacillus sphaericus*. Image courtesy of Dr. Jean-François Charles, Institut Pasteur.

### 1.2.1. Nomenclature for bacterial pesticidal proteins

At the onset of sporulation and during the stationary growth phase, *B. thuringiensis* and *L. sphaericus* produce parasporal crystalline inclusions that, once ingested by the target larvae, are solubilised to release pro-proteins that are proteolytically activated prior to receptor interaction (Palma et al. 2014; Silva-Filha et al. 2014). In the past, *B. thuringiensis* proteins were ordered into Crystal (Cry) and Cytolytic (Cyt) classes (Hofte and Whiteley 1989), with Bin and Cry names for *L. sphaericus* crystal toxins. In addition, secreted soluble proteins may be produced during the vegetative growth phase, ordered for *B. thuringiensis* into vegetative insecticidal proteins (Vip - Estruch et al. 1996) and secreted insecticidal proteins (Sip - Donovan et al. 2006) classes, and Mtx proteins for *L. sphaericus*. Within the initial nomenclature system proposed in 1988, *B. thuringiensis* proteins were classified using roman numerals according to their activity against distinct insect orders (Hofte and Whiteley 1989). More recently, a revised nomenclature (Crickmore et al. 1998) in which proteins were classified

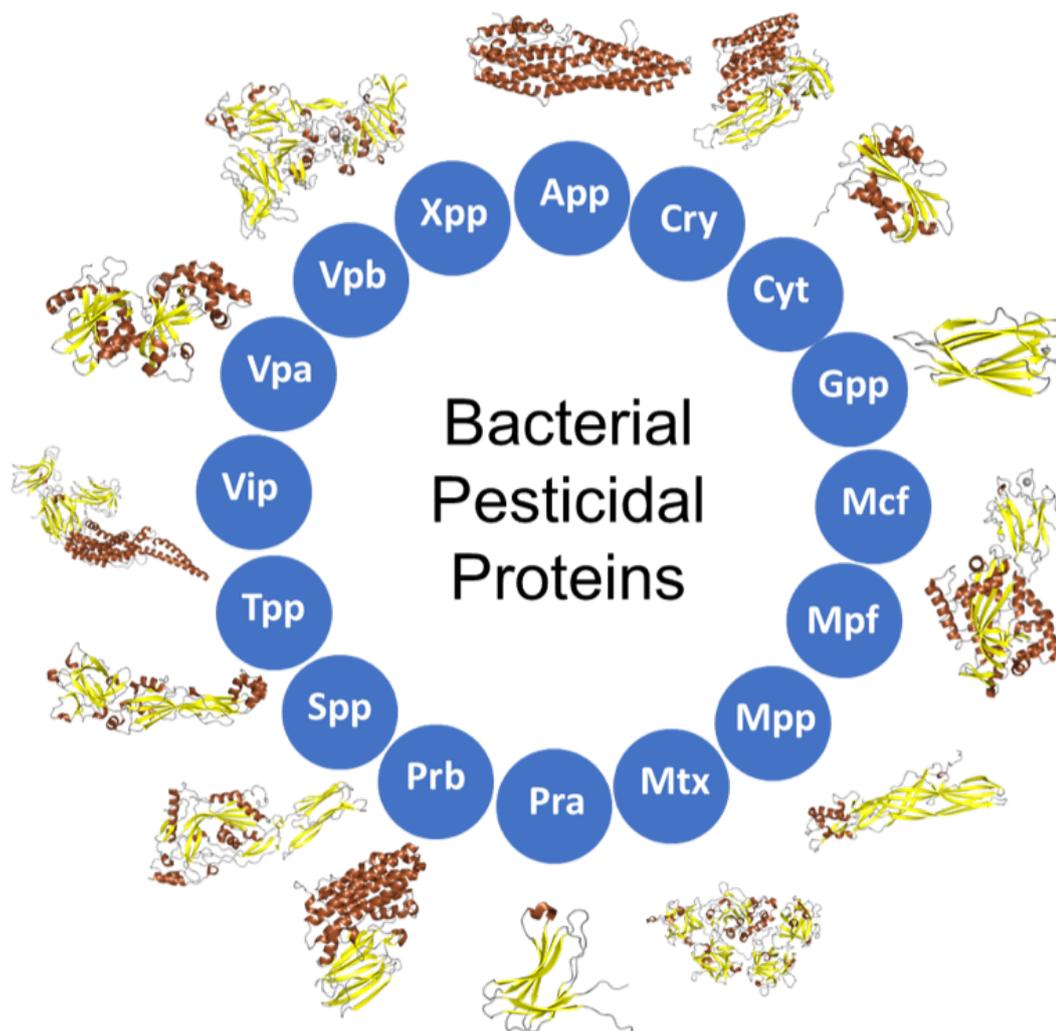
according to their amino acid sequence identity using a four-level system was released. In this system, upper- and lower-case letters are used for the second and third rank, whilst Arabic numbers are used for the first and fourth rank (Crickmore et al. 1998). Proteins with less than 45% sequence identity are assigned different primary ranks (e.g., Cry1/Cry2). At the secondary level, proteins exhibiting less than 78% are assigned different ranks (e.g., Cry1A/Cry1B) and at tertiary less than 95% identity (e.g., Cry1Aa/Cry1Ab). At the quaternary rank, each separate accession to the database exhibiting more than 95% sequence identity is given a new designation (e.g., Cry1Aa1/Cry1Aa2) (Crickmore et al. 1998).

# Cry1Ac1

Pesticidal protein class mnemonic	Primary rank < 45%	Secondary rank < 76%	Tertiary rank < 95%	Quaternary rank > 95%
-----------------------------------	-----------------------	-------------------------	------------------------	--------------------------

**Figure 1.2. Current nomenclature for bacterial pesticidal proteins.** Pesticidal proteins were recently reclassified according to their structural homology and sequence identity by the Bacterial Pesticidal Protein Resource Center (BPPRC.org - Crickmore et al. 2021; Panneerselvam et al. 2022). In this revised nomenclature, pesticidal proteins are arranged into classes (e.g., 3-domain “Cry” class) according to their structural homology. A four-level system is used to illustrate sequence identity within these classes. Upper- and lower-case letters are used for the second and third rank, whilst Arabic numbers are used for the first and fourth rank.

More recently, the nomenclature system has been further revised to include structural homology (Crickmore et al. 2020; Panneerselvam et al. 2022). In the current nomenclature (**Fig 1.2**), the four-level system of sequence identity has been retained (except the 78% sequence identity cut-off at the tertiary level was changed to 76%), but the original classes have been separated into a total of 17 groups (16 structural classes plus the Xpp mnemonic as a temporary name for those proteins without clear structures – **Fig 1.3**). The use of the term ‘pesticidal protein’ is also preferred over ‘toxin’, given the fewer negative connotations that can impact the successful commercialisation and uptake of pesticidal proteins as bioinsecticides and in transgenic crops (Crickmore et al. 2020). The new nomenclature allows for greater appreciation of the diversity of pesticidal proteins produced by *B. thuringiensis*, *L. sphaericus*, and other bacteria, which fall into several pfam groups, including Bin-like Toxin\_10 family (pfam05431), Aegerolysin family (pfam06355), and ETX\_MTX2 family (pfam03318). Of the 17 classes, 7 (Cry, Tpp, App, Mpp, Gpp, Cyt, Xpp) may be produced as crystalline inclusions during the sporulation and stationary growth phases. The Cry and Tpp classes, the focus of this thesis, will be introduced in detail below.



**Figure 1.3. Representative structures of the different pesticidal protein classes.** This diagram does not include the 17<sup>th</sup> (Txp) class. Obtained from the Bacterial Pesticidal Protein Resource Center (BPPRC.org - Crickmore et al. 2021; Panneerselvam et al. 2022).

### 1.2.2. Crystal (Cry) protein class

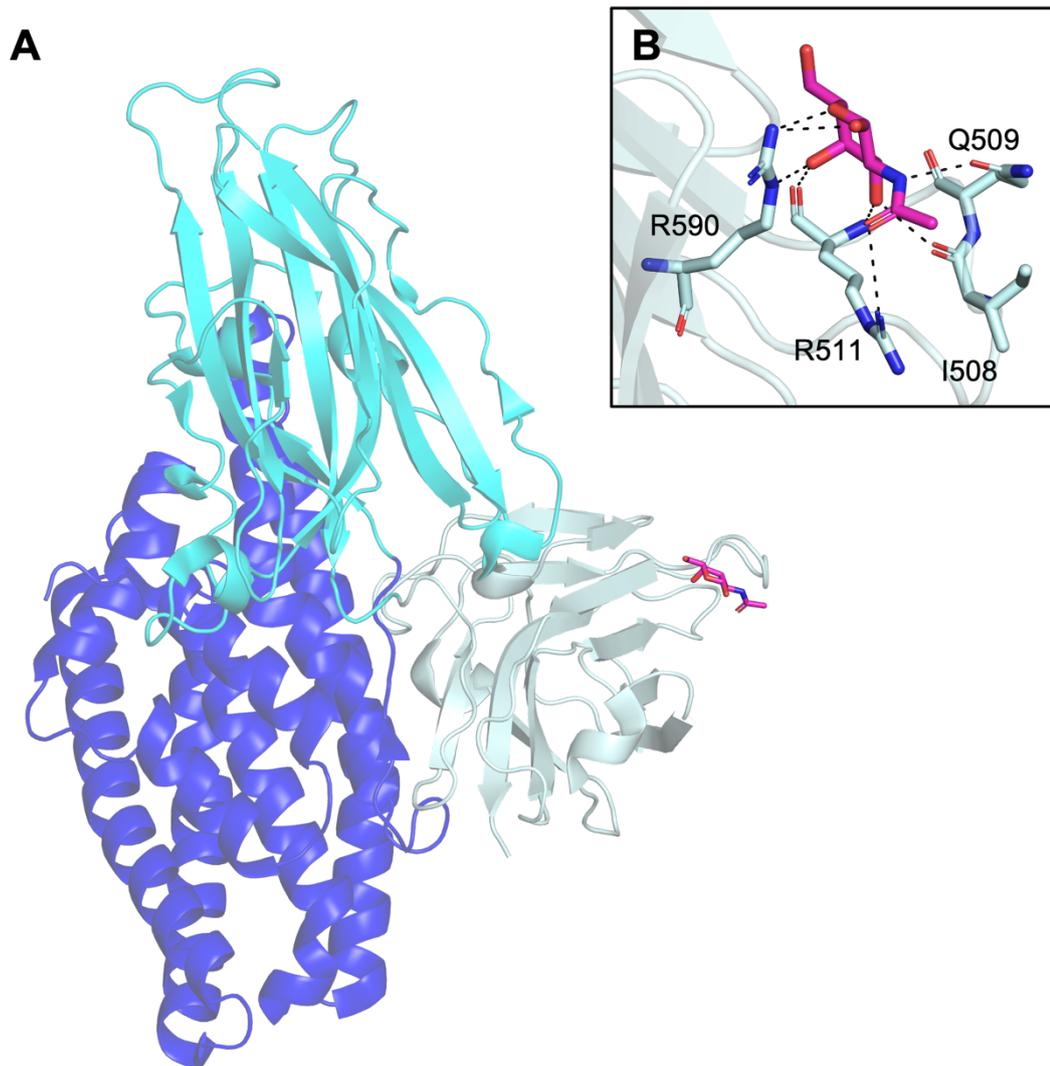
The three-domain Cry proteins represent the largest and most mechanistically well-characterised class of bacterial pesticidal proteins. Individual Cry proteins can display activity against species in several insect orders, including Lepidoptera and Coleoptera, as well as nematodes, and hence, are of great commercial importance. An example of a lepidopteran-protected GM crop is MON810, a maize variety expressing Cry1Ab (Álvarez

et al. 2022). Two models for their mechanism of action have been proposed, namely the sequential binding pore formation model (Bravo et al. 2004; Pigott and Ellar 2007; Pacheco et al. 2009; Soberón et al. 2009) and the G-protein mediated apoptotic signalling pathway model (Zhang et al. 2005; Zhang et al. 2006; Soberón et al. 2009). Binding events leading to toxicity in this family are complex and the relevance of the sequential binding pore-formation model is questioned for many Cry proteins and their insect targets. Despite this, the binding pore-formation model is favoured in the field (Soberón et al. 2009). In addition, a diverse variety of resistance mechanisms, many of which relate to receptor mutation and increased cell and/or gut recovery, have been identified (Pardo-López et al. 2013).

#### **1.2.2.1. Structural analysis of Cry proteins**

Despite displaying distinct differences in their amino acid sequences and target insect range, the activated Cry proteins exhibit a conserved three-domain architecture (**Fig 1.4** – DI, DII, DIII), indicating a similar mechanism of action. Phylogenetic analysis suggests that the ability of Cry proteins to target several insect orders is due to the independent evolution of DI – DIII, as well as swapping of DIII amongst Cry protein variants (De Maagd et al. 2001). Several crystal structures of solubilised and activated Cry proteins have been elucidated using conventional X-ray crystallography techniques and include Cry1Aa (PDB 1CIY - Grochulski et al. 1995), Cry1Ac (PDB 4ARX - Derbyshire et al. 2001), Cry2Aa (PDB 1I5P - Morse et al. 2001), Cry3Aa (PDB 6LFP - Heater et al. 2020), Cry3Bb1 (PDB 1JI6 - Galitsky et al. 2001), Cry4Aa (PDB 2C9K - Boonserm et al. 2006), Cry4Ba (PDB 1W99 - Boonserm et al. 2005), Cry5Ba (4D8M - Hui et al. 2012), Cry7Ca1 (PDB 5ZI1 - Jing et al. 2019), and Cry8Ea1 (PDB 3EB7 - Guo et al. 2009). At the N-terminus, DI, which has been linked to membrane insertion and pore formation (Li et al. 1991; Schwartz et al. 1997; Gazit et al. 1998), is composed of a helical bundle, usually containing six amphipathic  $\alpha$ -helices surrounding a single hydrophobic  $\alpha$ -helix thought to act as the membrane spanning component of the pore. In all Cry proteins, a short N-terminal region is cleaved during proteolytic activation. DII is composed of three

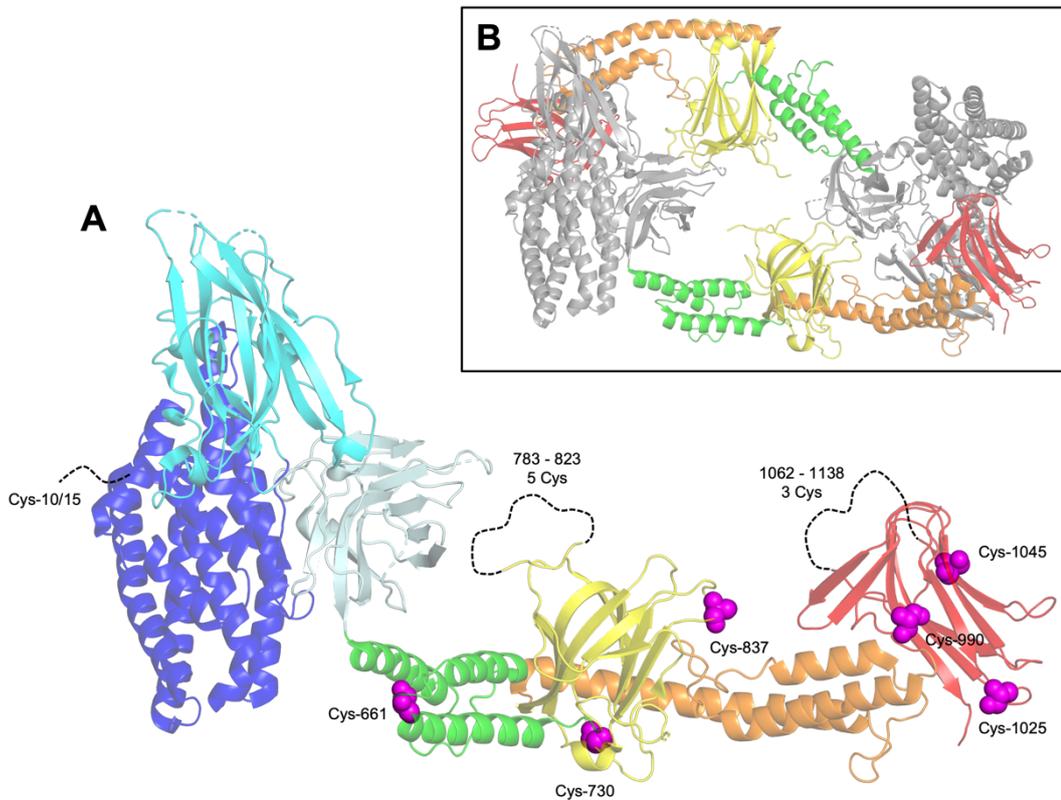
antiparallel  $\beta$ -sheets, in which greatest variation is seen in the apical loops that have been linked to receptor binding and hence, confer target specificity (Rajamohan et al. 1996; Smedley and Ellar 1996; Tuntitippawan et al. 2005; Gómez et al. 2006). DIII has also been linked to receptor binding (Burton et al. 1999; Lee et al. 1999; Gómez et al. 2006) and is composed of a  $\beta$ -sheet sandwich. DII and DIII resemble jacalin and sialidase, respectively, both of which are carbohydrate binding proteins, suggesting that these domains can bind carbohydrate moieties. Indeed, glycan-mediated receptor binding has been implicated in the mechanism of action of several Cry proteins (Best et al. 2023) and a structure for Cry1Ac1 in complex with N-acetylgalactosamine (GalNAc) exists (**Fig 1.4**) (Derbyshire et al. 2001).



**Figure 1.4. Members of the Crystal (Cry) class display a conserved three-domain architecture.** (A) Crystal structure of solubilised and proteolytically activated Cry1Ac1 (PDB 4ARY) in complex with N-acetylgalactosamine (GalNAc – shown as sticks. Carbon – magenta, oxygen – red, nitrogen – blue) illustrating the conserved three-domain (DI – DIII) architecture in the activated toxin core. DI – dark blue, DII – cyan, DIII – light cyan. (B) Binding site of GalNAc with Cry1Ac1 DIII. Cry1Ac1 residues involved in forming polar contacts (black dashed lines) with GalNAc are shown as sticks.

Cry proteins are produced as either ~65 kDa or ~130 kDa pro-proteins, the latter of which exhibit an extended C-terminus containing four pro-domains (DIV, DV, DVI, DVII) cleaved during proteolytic activation (Evdokimov et al. 2014). One structure of a full-length ~130 kDa Cry protein

has been reported to date and that is of a highly-mutated form of Cry1Ac, Cry1Ac- $\Delta$ 14C (Evdokimov et al. 2014). To minimise protein aggregation prior to crystallisation, Evdokimov et al. (2014) mutated 14 out of the 16 cysteine residues (by serine substitution), in addition to an F462V mutation (Evdokimov et al. 2014). The Cry1Ac- $\Delta$ 14C structure (**Fig 1.5**) revealed the presence of two  $\alpha$ -helical bundles – DIV and DVI – with structural similarity to spectrin (Grum et al. 1999) and a fibrinogen-binding complement inhibitor (Hammel et al. 2007), respectively. DV and DVII are composed of  $\beta$ -rolls with structural similarity to the carbohydrate binding modules of sugar hydrolases (Bae et al. 2008; Cid et al. 2010). In addition, Cry1Ac- $\Delta$ 14C was found to form a head-to-tail dimer which favoured contacts between the pro-domains, suggesting a role for the extended C-terminus in crystal packing (Evdokimov et al. 2014). Indeed, several studies have indicated that the pro-domains are dispensable for pesticidal activity and instead are required for optimised crystal packing, stability, and selective solubilisation (Lüthy and Ebersold 1981; Hofte and Whiteley 1989; Evdokimov et al. 2014). Of note, is the presence of a high number of cysteine residues within the pro-domains of Cry1 variants, amongst others, thought to be involved in intermolecular disulphide bonding (Dastidar and Nickerson 1979; Couche et al. 1987; Bietlot et al. 1990; Evdokimov et al. 2014). In addition, conservation of the pro-domains across Cry subclasses may allow packing of different Cry protein variants into the same crystal, ensuring their codelivery to the target insect. Co-delivery of Cry protein variants would be advantageous for those with demonstrated synergistic activity, including Cry1Ab / Cry1Ac, which show up to 5-fold higher activity against *Chilo partellus* larvae when treated in combination (Sharma et al. 2010), and Cry1Aa / Cry1Ac, which show up to 4.9-fold higher activity against *Lymantria dispar* larvae when treated in combination (Lee et al. 1996). More recently, investigations carried out by Peña-Carenda et al. (2018) indicated a role for the pro-domains in pesticidal activity. Specifically, the C-terminal region of the Cry1Ab pro-domain was shown to bind to alkaline phosphatase (ALP) and aminopeptidase N (APN) insect receptors, correlating with increased toxicity in comparison to the activated form of Cry1Ab (Peña-Cardena et al. 2018). Additional work investigating the functional role of Cry pro-domains is required.



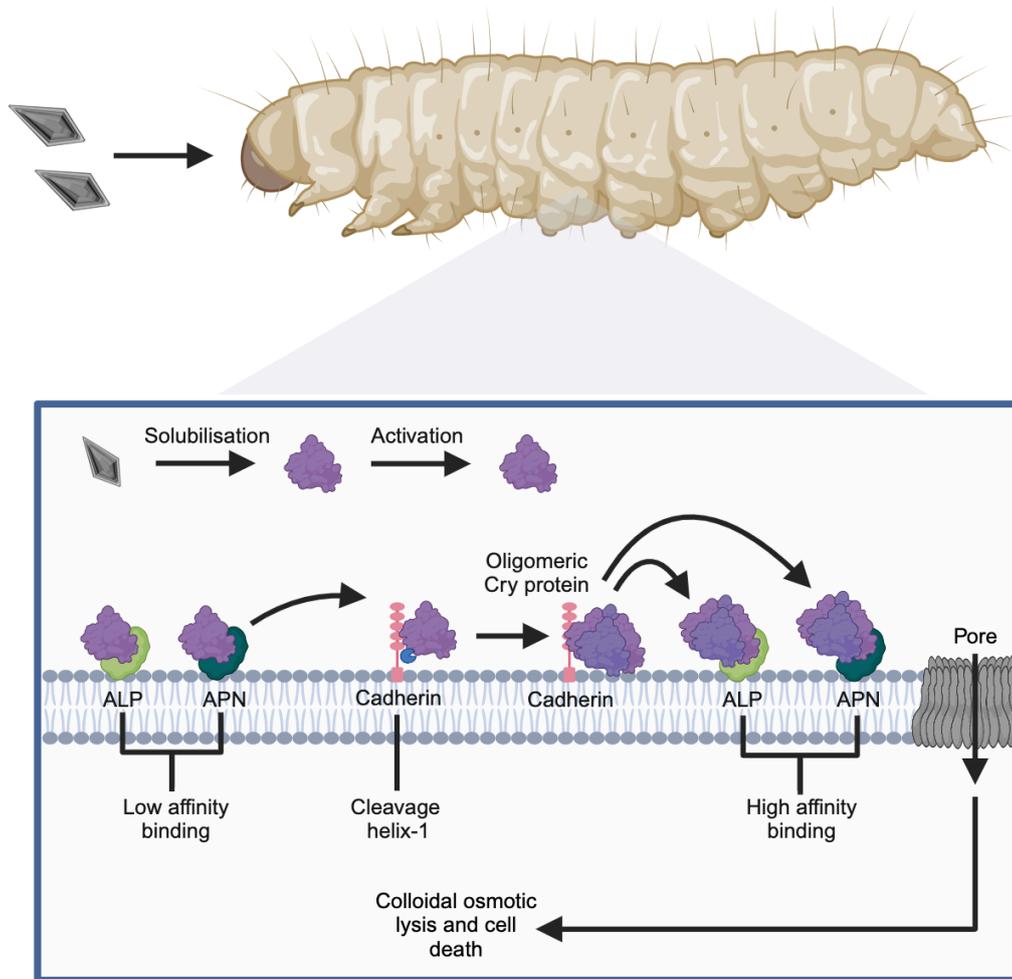
**Figure 1.5. The Cry1Ac- $\Delta$ 14C crystal structure.** (A) The protein core contains three domains (DI – dark blue, DII – cyan, DIII – light cyan). The C-terminal half is composed of an additional four pro-domains (DIV – green, DV – yellow, DVI – orange, DVII – red). To minimise protein aggregation, 14 out of the 16 cysteine residues were mutated to serine. Positions of the cysteine mutations are illustrated as magenta spheres. Some residues are present within regions of the structure which were not seen in the electron density map (illustrated as dashed lines), suggesting that these regions are disordered. (B) Cry1Ac- $\Delta$ 14C forms a head-to-tail dimer which favours contacts between the pro-domains (DIV – DVII – green, yellow, orange, red). Cry1Ac1 core domains (DI – DIII) shown in grey.

More recently, serial femtosecond crystallography (SFX), an emerging form of X-ray crystallography introduced in detail in **section 1.3.1.2**, has been used to solve the structure of Cry3Aa (a short ~73 kDa protoxin) and structures of other crystal protein classes from their natural *in vivo* grown

crystals (Sawaya et al. 2014; Colletier et al. 2016; Tetreau et al. 2020; Tetreau et al. 2022; Williamson et al. 2023). The structure of Cry3Aa revealed that the natural crystals exhibited a high solvent content, demonstrated by large solvent channels (approx. 3-5 nm) running between the Cry3Aa monomers (Sawaya et al. 2014; Tetreau 2021b). These solvent channels have been exploited in subsequent studies aimed at encapsulating other proteins, such as enzymes, within the crystals, with the downstream goal of improving their stability and recyclability for efficient cytosolic delivery in therapeutic products (Heater et al. 2020; Yang et al. 2021; Sun et al. 2022), an additional application of *B. thuringiensis* and *L. sphaericus* crystal proteins.

#### **1.2.2.2. Mechanism of action of Cry proteins**

Given their significance within the agriculture sector, the mechanism of action of Cry proteins has largely been studied using lepidopteran active Cry1 proteins. Two models for the mechanism of action of Cry proteins have been proposed – the sequential binding pore formation model (Bravo et al. 2004; Pigott and Ellar 2007; Pacheco et al. 2009; Soberón et al. 2009) and the G-protein mediated apoptotic signalling pathway model (Zhang et al. 2005; Zhang et al. 2006; Soberón et al. 2009). The sequential binding pore formation model (**Fig 1.6**) has been generally accepted but may only apply to limited toxin/target interactions. It proposes that, following ingestion, crystal proteins are solubilised by the specific pH and midgut environment of the target insect, releasing monomeric protoxins which are then activated by midgut proteases. The activated Cry protein binds to specific target receptors leading to further cleavage of an N-terminal region within DI, which triggers protein oligomerisation to form the pre-pore structure. The central hydrophobic  $\alpha$ -helix of the DI  $\alpha$ -helical bundle is thought to insert into the membrane to form the final pore structure, with cell death hypothesized to occur via colloidal osmotic lysis (Knowles and Ellar 1987).



**Figure 1.6. Sequential binding pore formation model for Cry proteins.** Crystal proteins are solubilised by the specific pH of the target insect midgut, releasing monomeric protoxins which are then activated by midgut proteases. The activated Cry protein is localised to the cell surface by low-affinity interactions with alkaline phosphatase (ALP) and aminopeptidase N (APN) receptors, leading to high-affinity interaction with cadherin-like proteins (CAD) receptors. This interaction induces further cleavage of an N-terminal region within DI, which triggers protein oligomerisation to form the pre-pore structure. The oligomeric pre-pore structure binds ALP and APN receptors with high affinity, leading to insertion of the central hydrophobic  $\alpha$ -helix of domain I and finally, pore-formation. Cell death is hypothesized to occur via colloidal osmotic lysis. The sequential binding pore formation model has been generally accepted but may only apply to limited toxin/target interactions and while receptor binding/pore-formation remain the most common modes for toxicity, the steps involved remain unclear. Created with BioRender.com.

Several binding proteins and receptors have been identified for lepidopteran active Cry1 proteins, including ALP, APN, cadherin-like proteins (CAD), and ATP-binding cassette (ABC) transporters (Pigott and Ellar 2007). In lepidopteran larvae, the sequential binding pore formation model first proposes a low-affinity interaction of activated Cry1A proteins with glycosyl phosphatidylinositol (GPI)-anchored ALP (via Cry1A D-II exposed loop 3) and GPI-anchored APN (via Cry1A D-III  $\beta$ -strand 16) receptors (Masson et al. 1995; Pacheco et al. 2009; Arenas et al. 2010). This low-affinity interaction is hypothesized to increase the abundance of Cry1 at the cell membrane, allowing a subsequent high-affinity interaction with CAD receptors, involving extracellular regions CR7, CR11, and CR12, which bind Cry1A D-II exposed loops 2, 3, and  $\alpha$ -helix 8 (Vadlamudi et al. 1995; Gómez et al. 2006; Atsumi et al. 2008; Pacheco et al. 2009; Arenas et al. 2010). This high-affinity interaction is thought to promote cleavage at the N-terminal site, triggering oligomerisation (Gómez et al. 2002; Pacheco et al. 2009; Arenas et al. 2010). The oligomeric form of Cry1A displays an increased binding affinity to ALP and APN receptors and in Cry1Ab, D-II loop 2 is involved in this interaction (Arenas et al. 2010). Interaction with ALP and APN receptors after oligomerisation leads to membrane insertion, pore formation and cell lysis (Pardo-López et al. 2006; Arenas et al. 2010). Across other insect orders, similar receptor proteins have been identified but the increased range of receptors (including ABC proteins, which often seem to have significant roles in binding) have cast some doubt on the serial binding mechanism. While receptor binding/pore formation (**Fig 1.6**) remains the most common hypothesis for the mechanism of action, the steps involved remain unclear.

Oligomerisation of Cry proteins has been demonstrated across several variants, including Cry1 (Tigue et al. 2001), Cry3 (Rausell et al. 2004), Cry4 (Likitvivanavong et al. 2006), and Cry11 (Pérez et al. 2007). In addition, pore-formation activity has been analysed using several techniques, including calcein leakage from brush border membrane vesicles (BBMVs) (Rausell et al. 2004), light scattering of liposomes or BBMVs (Haider and

Ellar 1989; Carroll and Ellar 1993), and analysis of changes in membrane potential (Muñoz-Garay et al. 2006). Using osmotic protection assays, Carroll and Ellar (1997) estimated a pore diameter of ~2.4 nm under high pH (Carroll and Ellar 1997). It is unclear how many monomers oligomerise to form the final pore structure. Using two-dimensional crystallography, a trimeric assembly has been demonstrated for both Cry1AbMod (lacking the N-terminal region and first  $\alpha$ -helix) (Muñoz-Garay et al. 2009) and Cry4Ba (Ounjai et al. 2007) in their membrane associated forms. In contrast, atomic force microscopy studies have demonstrated a tetrameric assembly for Cry1Aa (Vié et al. 2001) and Cry4Ba (Puntheeranurak et al. 2005).

The G-protein mediated apoptotic signalling pathway model (Zhang et al. 2005; Zhang et al. 2006) differs from the sequential binding pore formation model in that protein oligomerisation and pore formation is excluded. Instead, binding of Cry proteins to CAD receptors is hypothesized to trigger activation of a G-protein, which in turn activates adenylyl cyclase, leading to an increase in cAMP levels, activation of protein kinase A, and finally cell death (Zhang et al. 2005; Zhang et al. 2006). However, several studies suggest that binding to CAD receptors alone is insufficient to induce cell death. For example, Cry1Ab proteins lacking the N-terminal region cleaved during proteolytic activation, as well as  $\alpha$ -helix 1 (Cry1AbMod), were able to form oligomers and induce toxicity in insects lacking the CAD receptor, clearly indicating that toxicity can occur independent of CAD binding (Soberón et al. 2007; Muñoz-Garay et al. 2009).

### **1.2.2.3. Resistance mechanisms to Cry proteins**

Insect resistance remains the greatest challenge to bioinsecticide efficacy, with several resistance mechanisms identified in both laboratory and field studies (Pardo-López et al. 2013). The emergence of insect resistance depends on general factors, including long-term use of larvicides that lead to increased selection pressure, as well as factors relating to the mode of action of the larvicide itself. In the Cry class, impaired binding of Cry proteins to target receptors appears to be the most common mechanism.

Indeed, reduced Cry protein binding to mutated receptors has been linked with resistance across several lepidopteran insects, e.g., ALPs in *Heliothis virescens* (Jurat-Fuentes et al. 2002; Jurat-Fuentes et al. 2011), *Helicoverpa armigera* (Jurat-Fuentes et al. 2011), *Spodoptera frugiperda* (Jurat-Fuentes et al. 2011); APNs in *Spodoptera exigua* (Herrero et al. 2005), *Spodoptera litura* (Rajagopal et al. 2002), *Trichoplusia ni* (Tiewisiri and Wang 2011), *H. armigera* (Zhang et al. 2009); and CADs in *Diatraea saccharalis* (Yang et al. 2011), *Pectinophora gossypiella* (Morin et al. 2003), *H. virescens* (Gahan et al. 2001; Jurat-Fuentes et al. 2004), *H. armigera* (Xu et al. 2005). Alternative resistance mechanisms have been identified. In *Plodia interpunctella* larvae, a reduction in protease activity, and therefore activation of the Cry1Ac protoxin, was linked to insect resistance (Oppert et al. 1997). Increased levels of protease activity, thought to lead to Cry degradation, have also been linked to insect resistance, as demonstrated in *S. littoralis* resistant to Cry1C (Keller et al. 1996). In addition, toxin sequestration by glycolipids which were present in the gut of *H. armigera* larvae, leading to reduced receptor binding, has been linked to resistance against Cry1Aa and Cry2Ab proteins (Ma et al. 2012). Finally, an elevated immune response has also been linked to insect resistance (Hernández-Martínez et al. 2010).

### **1.2.3. Toxin-10 Pesticidal Protein (Tpp) class**

The **Toxin-10 Pesticidal Proteins (Tpp)** are a class of beta pore-forming bacterial pesticidal proteins. Tpps of *L. sphaericus* origin (Silva-Filha et al. 2014) include the mosquitocidal binary pesticidal proteins Tpp1/Tpp2 (previously BinA/B - Colletier et al. 2016) and Tpp49Aa1 (previously Cry49Aa1) which functions alongside its partner protein from the Cry class, Cry48Aa1 (Jones et al. 2007). Tpps of *B. thuringiensis* origin include Tpp78Aa1 (previously Cry78Aa1 - Cao et al. 2022) and Tpp80Aa1 (previously Cry80Aa1 - Best et al. 2022), both of which act alone to exert mosquitocidal activity. Highly toxic *L. sphaericus* strains, such as 1593, 2362 and C3-41, produce Tpp1/Tpp2 and have been widely and successfully applied as mosquitocides for the control of human diseases, including West Nile virus (Regis et al. 2001; Lacey 2007). Other Tpp proteins, including *B.*

*thuringiensis* Tpp35Ab1 (previously Cry35Ab1), which functions alongside its partner protein from the Aegerolysin-related **Pesticidal Proteins (Gpp)** class, Gpp34Ab1 (previously Cry34Ab1), exert insecticidal activity against coleopteran insects, such as *Diabrotica virgifera virgifera* (Western corn rootworm), a major pest of maize in Europe and North America (Schnepf et al. 2005; Kelker et al. 2014).

Like the Cry class, Tpp proteins are produced as crystalline inclusions which, following ingestion, are solubilised in the acidic or alkaline environment of the larval midgut to release pro-proteins in the gut lumen. The pro-proteins are proteolytically cleaved to produce active fragments which bind target receptors present on the brush border membrane of epithelial cells, leading to several cytotoxic effects which ultimately lead to larval death (Silva-Filha et al. 2014). The molecular details underlying receptor interaction, as well as the mechanisms contributing to cell death remain understudied. In addition, several cases of field resistance to currently marketed bioinsecticides have been identified, and thus, further analysis of the Tpp class is required.

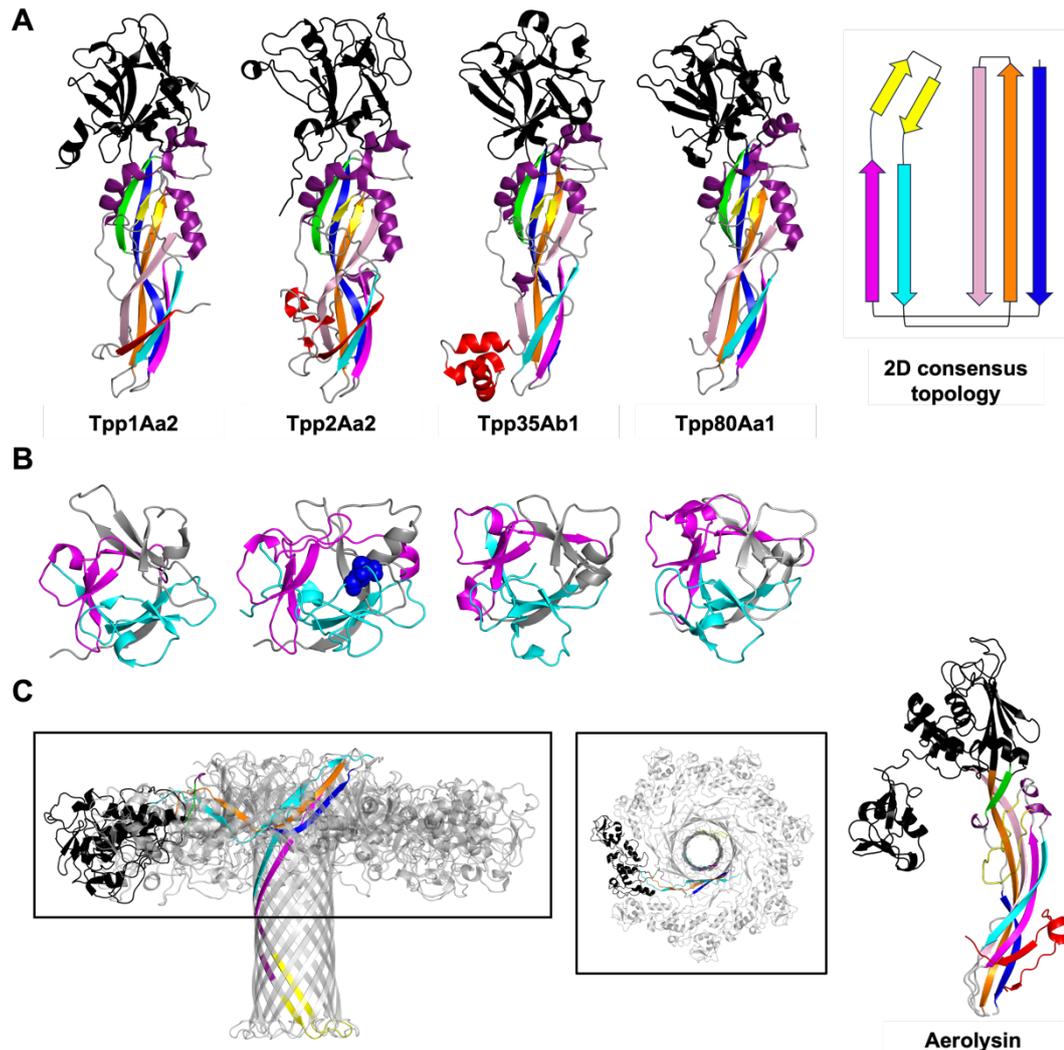
#### **1.2.3.1. Structural analysis of Tpp proteins**

Several crystal structures of solubilised and activated Tpp proteins have been elucidated using conventional X-ray crystallography techniques and include Tpp2Aa3 (PDB 3WA1 – Srisucharitpanit et al. 2014), Tpp35Ab1 (PDB 4JP0 – Kelker et al. 2014), Tpp78Aa1 (PDB 7Y78 – Cao et al. 2022), and Tpp80Aa1 (PDB 8BAD – Best et al. 2022). Structure solution has demonstrated that Tpp proteins are composed of continuous N-terminal head and C-terminal tail domains (**Fig 1.7**). The head domain is composed of a  $\beta$ -trefoil fold (**Fig 1.7**), commonly seen in carbohydrate binding ricin B-like lectins, such as haemolytic lectin from the parasitic mushroom *Laetiporus sulphureus* (Mancheño et al. 2005).  $\beta$ -trefoil folds are usually composed of six two-stranded  $\beta$ -hairpins arranged into three subdomains with pseudo three-fold symmetry, designated  $\alpha$  ( $\beta$ 1- $\beta$ 4),  $\beta$  ( $\beta$ 5- $\beta$ 8), and  $\gamma$  ( $\beta$ 9- $\beta$ 12), each

of which may contain a conserved QxW motif (where X denotes any residue) capable of carbohydrate binding (Hazes 1996). In the Tpp class, binding assays have demonstrated that the N-terminal  $\beta$ -trefoil fold is involved in receptor interaction (Romão et al. 2011). In line with this, the largest structural differences between Tpp family members are seen within the surface exposed loops of the  $\beta$ -trefoil domain, indicating their involvement in target specificity. The presence of a lectin-like  $\beta$ -trefoil fold has led to the hypothesis that receptor binding may be mediated by interaction with glycoconjugates.

The tail domain is  $\beta$ -sheet rich (**Fig 1.7**) and resembles the pore-forming domain of the wider aerolysin, ETX/MTX-2 superfamily of  $\beta$ -pore forming proteins (Parker et al. 1994; Szczesny et al. 2011). Across this superfamily, a core topology consisting of five antiparallel  $\beta$ -sheets and a short  $\beta$ -hairpin, proposed to act as the transmembrane segment following pore formation, is conserved (**Fig 1.7**) (Lacomel et al. 2021). The length and twists of the  $\beta$ -sheets varies across family members and further  $\beta$ -sheets and short  $\alpha$ -helices may also be present. Aerolysin, the virulence factor associated with mammalian diarrhoeal diseases and deep wound infections caused by the gram-negative bacterium *Aeromonas hydrophila*, is the most mechanistically well-characterised. Specifically, Aerolysin exerts its toxicity by forming pores which lead to osmotic lysis and cell death. This pore formation process is well-characterised, with cryo-electron microscopy (cryoEM) structures for both the pre-pore state (PDB 5JZH) and pore state (PDB 5JZT, embedded in lysomyristoylphosphatidylglycerol (LMPG) micelles) having been elucidated (Iacovache et al. 2016). Aerolysin-like proteins display several common structural features, including a conserved  $\beta$ -hairpin thought to act as the transmembrane segment following pore-formation, as well as the presence of surface-exposed hydrophobic patches rich in serine/threonine residues that are thought to drive oligomerization via the formation of intermolecular interactions. The structural similarity of the Tpp C-terminal domain with that of aerolysin (Lacomel et al. 2021), as well as electrophysiology studies demonstrating that Tpp proteins are able to

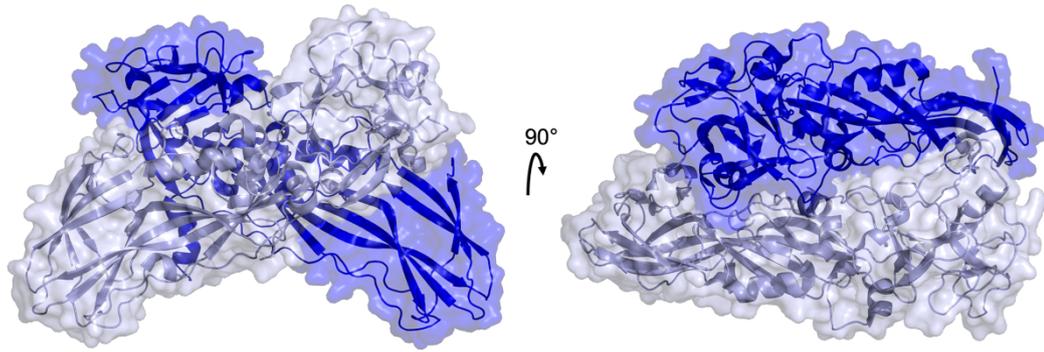
form pores in cell membranes (Cokmus et al. 1997; Schwartz et al. 2001), suggests that the C-terminal domain of the Tpp class functions in pore-formation. The receptor binding head region of aerolysin, which includes two domains (Parker et al. 1994; Iacovache et al. 2016), one of which is discontinuous with the C-terminal pore-forming domain, is structurally distinct to the head domain of the Tpp class, consistent with the distinct target range (mammalian vs invertebrate) between aerolysin and Tpp proteins.



**Figure 1.7. Structural analysis of the Tpp class.** (A) Crystal structures of Tpp1Aa2, Tpp2Aa2 (PDB 5FOY – extracted from the Tpp1Aa2/Tpp2Aa2 co-crystal), Tpp35Ab1 (PDB 4JP0), and Tpp80Aa1 (PDB 8BAD). Structure solution has demonstrated that Tpp proteins are composed of continuous N-terminal head and C-terminal tail domains. The head domain is composed of a  $\beta$ -trefoil fold common in the class of carbohydrate binding Lectin proteins, whilst the tail domain is  $\beta$ -sheet rich and resembles the pore-forming domain of the wider aerolysin, ETX/MTX-2 superfamily of  $\beta$ -pore forming proteins. Across this superfamily, a core topology consisting of five antiparallel  $\beta$ -sheets and a short  $\beta$ -hairpin, proposed to act as the transmembrane segment following pore formation is conserved. (B) Top-down view of the  $\beta$ -trefoil fold of Tpp1Aa2, Tpp2Aa2, Tpp35Ab1, and Tpp80Aa1, showing the common fold composed of six two-stranded  $\beta$ -hairpins arranged into three subdomains with pseudo three-fold symmetry, designated  $\alpha$  ( $\beta$ 1- $\beta$ 4 – cyan ),  $\beta$  ( $\beta$ 5- $\beta$ 8 –

magenta), and  $\gamma$  ( $\beta 9$ - $\beta 12$  – grey), each of which may contain a conserved QxW motif (where X denotes any residue) capable of carbohydrate binding. In Tpp2Aa2, a disulphide bridge (dark blue spheres) appears to distort the three-fold symmetry. (C) Structural comparison of aerolysin, with the conserved core topology in the C-terminal pore-forming domain coloured according to the 2D consensus topology in panel A. The cryoEM structure of the aerolysin pore state (PDB 5JZT, embedded in lysomyristoylphosphatidylglycerol – LMPG – micelles) is also shown.

More recently, SFX has been applied to elucidate the structure of the Tpp1Aa2/Tpp2Aa2 natural co-crystal (**Fig 1.8**) produced by *L. sphaericus* strains (Colletier et al. 2016). The Tpp1Aa2/Tpp2Aa2 structure revealed that Tpp1 and Tpp2 pack into crystals in a 1:1 ratio, forming a large intermolecular interface, approximately 42% of which involves pro-peptide regions. Given this, the authors suggested that proteolytic cleavage may lead to dissociation of the heterodimer, although they also hypothesized that slow release may prevent dissociation prior to cellular internalisation. Indeed, it remains unclear whether the Tpp1Aa2/Tpp2Aa2 heterodimer dissociates into its individual components prior to receptor interaction and internalisation, or if the heterodimeric complex persists.



**Figure 1.8. Tpp1Aa2/Tpp2Aa2 natural co-crystal structure elucidated using serial femtosecond crystallography (SFX).** The Tpp1Aa2/Tpp2Aa2 structure revealed that Tpp1 and Tpp2 pack into crystals in a 1:1 ratio, forming an ‘X’ structure with a large intermolecular interface thought to be involved in crystal stability.

Since data are collected in a serial fashion, SFX can be applied to perform substrate mixing and time-resolved studies, yielding dynamic structural information relating to protein function. In the field of bacterial pesticidal proteins, pH mixing studies have been performed on native *B. thuringiensis* and *L. sphaericus* crystals to probe the early structural events leading to crystal dissolution. Specifically, in their work with Tpp1Aa2/Tpp2Aa2 natural co-crystals, Colletier et al (2016) showed that an elevation of pH from 7 to 10 (mimicking the alkaline environment of the mosquito larval midgut) led to several early structural events, including the unravelling of an  $\alpha$ -helix present in the pro-peptide region of Tpp1Aa2, the loss of intermolecular hydrogen bonds that, at high pH, were deprotonated leading to electrostatic repulsion, as well as the loss of pH-labile intermolecular salt bridges (Colletier et al. 2016).

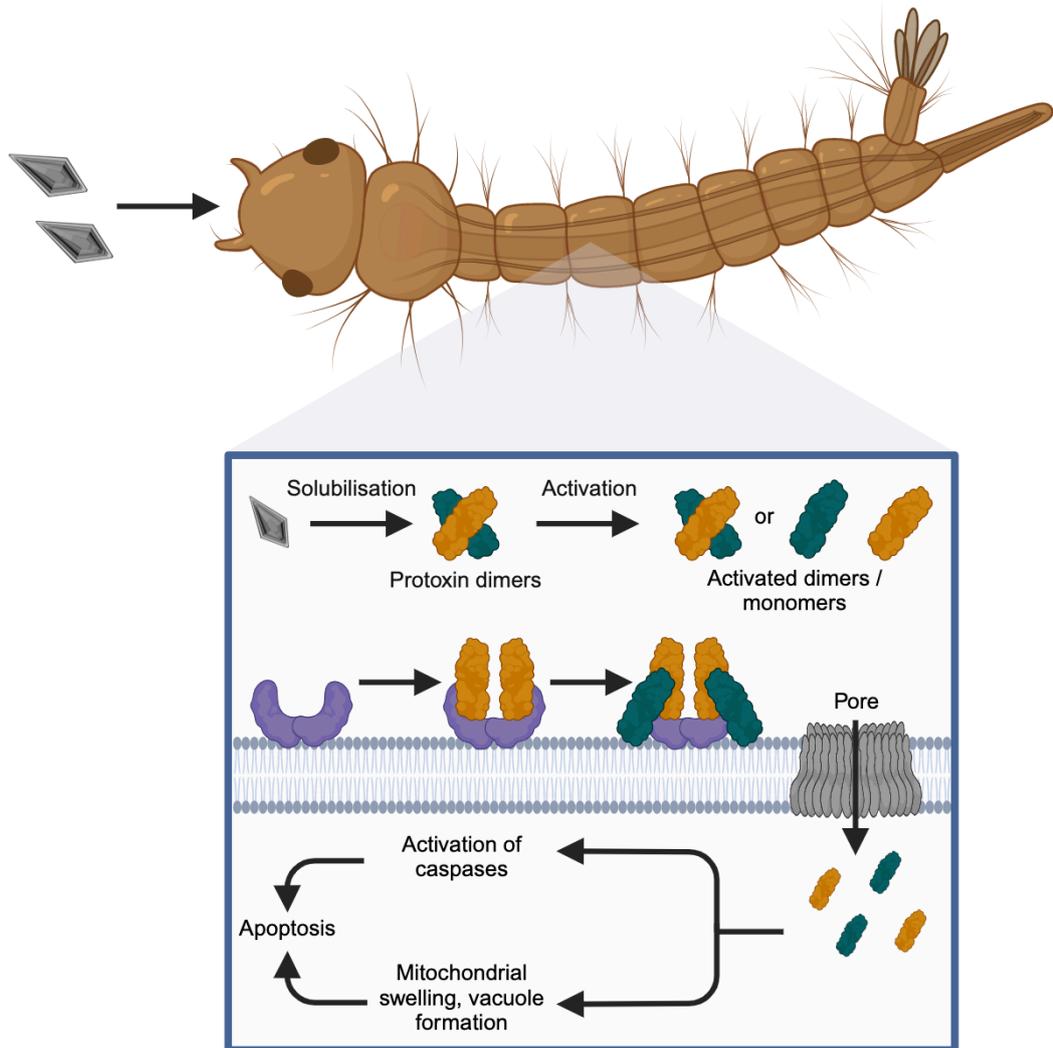
The Tpp1Aa2/Tpp2Aa2 structure also revealed notable differences between the N-terminal  $\beta$ -trefoil domains of the two proteins. In Tpp2Aa2, a disulphide bridge (**Fig 1.7** – Cys67-Cys161) appears to disrupt the three-fold symmetry of the  $\beta$ -trefoil fold, obstructing the  $\alpha$ -carbohydrate binding module. This is in contrast to Tpp1Aa2, where all three carbohydrate binding modules

( $\alpha$ ,  $\beta$ ,  $\gamma$ ) appear capable of carbohydrate interaction, suggesting that members of the Tpp class may confer distinct carbohydrate binding capabilities (Colletier et al. 2016).

### 1.2.3.2. Mechanism of action of Tpp proteins

As previously alluded to, some *B. thuringiensis* and *L. sphaericus* pesticidal proteins act alone to exert their toxicity, whilst others require a partner protein. The former is the case for Tpp78Aa1 and Tpp80Aa1, which act alone to exert mosquitocidal activity. In contrast, some proteins act alongside other members of the same class, as is the case for Tpp1Aa2/Tpp2Aa2, or alongside members from distinct classes, as is the case for Cry48Aa1/Tpp49Aa1 and Gpp34Ab1/Tpp35Ab1. This highlights a complexity in the mechanism of action of *B. thuringiensis* and *L. sphaericus* proteins, including those in the Tpp class.

Within the Tpp class, Tpp1/Tpp2 is the most mechanistically well-characterised (**Fig 1.9**). Tpp1/Tpp2 is active against the larvae of *Culex* and *Anopheles* mosquitoes, with maximum toxicity achieved when both proteins are present in equimolar concentrations. The binary toxin is produced as crystalline inclusions formed by the packing of Tpp1/Tpp2 heterodimers, which are composed of the Tpp1 (~42 kDa) and Tpp2 (~51 kDa) pro-proteins (Baumann et al. 1988; Oei et al. 1992; Charles et al. 1997; Colletier et al. 2016). Once ingested by mosquito larvae, the crystals are solubilised and the pro-protein forms of Tpp1 and Tpp2 are cleaved by midgut proteases (**Fig. 1.9**), resulting in the production of activated Tpp1 (~39 kDa) and Tpp2 (~43 kDa) (Baumann et al. 1988; Oei et al. 1992; Charles et al. 1997). It remains unclear whether the heterodimeric complex persists following proteolytic activation, or whether the activated proteins dissociate into their individual components. However, given that 42% of the large heterodimeric interface of the Tpp1Aa2/Tpp2Aa2 complex (**Fig 1.8**) involves pro-peptide regions, it is likely that the complex dissociates prior to receptor interaction.



**Figure 1.9. Proposed mechanism of action of Tpp1/Tpp2 against *Culex* mosquitoes.** The binary crystal proteins composed of the Tpp1Aa2/Tpp2Aa2 heterodimer are solubilised by the specific pH of the target insect midgut, releasing protoxins which are then activated by midgut proteases. Following activation, it remains unclear whether the heterodimeric complex persists, or whether the activated proteins dissociate into their individual components. In *Culex* mosquitoes, Tpp2 (yellow) acts as the major receptor binding component, which allows Tpp1 (green) – the active component – to bind subsequently. Receptor binding leads to pore-formation, which enables both proteins to be internalised. The cell death pathways remain to be fully elucidated but involve mitochondrial pathway-mediated apoptosis and several cytopathological changes. Created with BioRender.com.

Tpp1/Tpp2 bind a single class of receptor which has been identified in susceptible species as *Culex quinquefasciatus* maltase 1 (Cqm1 - Sharma et al. 2018), for which a structure has been solved (Sharma and Kumar 2019), *Culex pipiens* maltase 1 (Cpm1 - Silva-Filha et al. 1999), and *Anopheles gambiae* maltase 3 (Agm3 - Opota et al. 2008). An orthologous protein, *Aedes aegypti* maltase 1, has been identified in the refractory mosquito species, *Ae. aegypti* (Nielsen-Leroux and Charles 1992; Ferreira et al. 2010). Receptor binding is essential for Tpp1/Tpp2 toxicity and, despite the high sequence identity of Aam1 with Cqm1 (74%), non-susceptibility of *Ae. aegypti* has been linked to a lack of receptor interaction (Ferreira et al. 2010). Given that toxicity is exerted via a single class of receptors, the Tpp1/Tpp2 protein is an ideal target for studies investigating binding interactions, in comparison to other more complex systems, such as members of the Cry class which bind many target receptors. Early studies using combinations of recombinant Tpp1 and Tpp2 proteins against *Culex* mosquitoes demonstrate that Tpp2 acts as the major receptor binding component, which subsequently allows Tpp1 – the active component – to bind (**Fig 1.9**), whereas in *Anopheles* mosquitoes, Tpp1 also appears to be able to bind target receptors (Oei et al. 1992; Charles et al. 1997; Darboux et al. 2001). Subsequently, binding studies investigating deletion fragments demonstrated that the N-terminal region of Tpp2Aa2, Asn33 – Leu158, is sufficient for receptor binding (Singkhamanan et al. 2010; Romão et al. 2011). In *Culex* larvae, receptor binding is regional and localised to the gastric caecum and posterior midgut of mosquito larvae, whereas in *Anopheles* larvae, receptor binding is less defined (Oei et al. 1992). The luminal pH of mosquito larval midgut ranges from ~8 in the gastric caecum (Boudko et al. 2001) to > 10 in the anterior gut and ~7.5 in the posterior gut (Dadd 1975), demonstrating that, in *Culex* mosquitoes, receptor binding occurs at approximately neutral pH.

Electrophysiology studies on large unilamellar phospholipid vesicles and cultured *Cx. quinquefasciatus* cells have demonstrated the ability of Tpp1/Tpp2 to form pores in the cell membrane (Cokmus et al. 1997; Schwartz et al. 2001). In addition, binding and internalisation of fluorescently

labelled Tpp1 and Tpp2 proteins has been shown in the epithelial cells of *Culex* larvae (Oei et al. 1992; Lekakarn et al. 2015). The precise roles of uptake and/or cell-death pathways following pore-formation and internalisation are still being investigated. Early studies investigating the morphological effects of Tpp1/Tpp2 treatment against epithelial cells identified mitochondrial swelling, microvillar disruption, as well as vacuolisation (Charles 1987; Silva-Filha and Peixoto 2003). More recently, it was demonstrated that Tpp1/Tpp2-induced vacuolisation is linked with an autophagic response in mammalian epithelial Madin-Darby canine kidney (MDCK) cells expressing the Cpm1 receptor (Opota et al. 2011). Furthermore, intracellular trafficking of Tpp1/Tpp2 revealed that the toxin components did not localize within any degradative compartment, and instead localised to recycling endosomes, suggesting that Tpp1/Tpp2 is able to avoid degradation (Opota et al. 2011). Finally, morphological changes including mitochondrial swelling, chromatic condensation, and apoptotic cell formation, as well as the activation of caspase-9 and caspase-3 in the epithelial cells of *Culex* larvae suggest that mitochondrial pathway-mediated apoptosis is also involved in cell death (Tangsongcharoen et al. 2015). This is further supported by transcriptomic studies demonstrating a differential expression of transcripts involved in apoptosis between Tpp1/Tpp2-treated vs untreated *Cx. quinquefasciatus* larvae (Tangsongcharoen et al. 2017). Further work is required to understand the molecular mechanisms of Tpp1/Tpp2 fully, as well as other members of the Tpp class.

### **1.2.3.3. Resistance mechanisms to Tpp proteins**

As previously discussed, insect resistance remains the greatest challenge to the long-term use of bioinsecticides. Highly toxic *L. sphaericus* strains produce Tpp1/Tpp2 and hence, these strains have been widely applied as mosquitocides for the control of human diseases (Regis et al. 2001; Lacey 2007). Resistance to Tpp1/Tpp2 has been detected in both field treated populations and lab selected colonies of *Cx. pipiens* and *Cx. quinquefasciatus* mosquitoes (Silva-Filha et al. 2021). In most cases, resistance to Tpp1/Tpp2 is caused by the selection for mutations in the

Cpm1/Cqm1 receptor, which lead to the expression of truncated proteins without the GPI anchor, such that the receptor is no longer presented on the epithelial cell surface (Nielsen-Leroux et al. 1995; Oliveira et al. 2004). Given that receptor interaction is critical for insecticide activity, these mutations completely disrupt the action of Tpp1/Tpp2 and thus, lead to high levels of mosquito resistance. Resistant *Cx. pipiens* colonies expressing functional Cpm1 receptors have been identified, indicating that additional mechanisms of resistance also exist (Nielsen-Leroux et al. 1997; Nielsen-Leroux et al. 2002).

Given that cases of resistance have been identified in several field populations treated with the current bioinsecticides in use, there is a need for the identification of new strains / pesticidal proteins with the ability to overcome insect resistance. Indeed, screening of new *L. sphaericus* isolates gave rise to the identification of the Cry48Aa1/Tpp49Aa1 pesticidal protein (Yuan et al. 2003; Jones et al. 2007; Jones et al. 2008). Cry48Aa1/Tpp49Aa1 exerts activity against *Cx quinquefasciatus* mosquitoes that are resistant to Tpp1Aa2/Tpp2Aa2, thus demonstrating that Cry48Aa1/Tpp49Aa1 acts via unique molecular mechanisms (Jones et al. 2007; Jones et al. 2008). This example represents one case with potential to overcome mosquito resistance to currently marketed bioinsecticides.

### **1.3. Studying pesticidal proteins using structural biology techniques**

It is well known that insight into a protein's three-dimensional (3D) structure is essential for understanding its function and mechanism of action. In the field of pesticidal proteins, knowledge of a protein's structure can be applied to counteract field resistance mechanisms, and for protein engineering efforts aimed at developing new bioinsecticides with enhanced potency, stability, and target insect range. By applying a variety of techniques, including X-ray crystallography, cryoEM, and computational modelling (Carugo and Djinović-Carugo 2023), we may be able to answer some of the key remaining questions within the field such as; How do pesticidal proteins interact with their target receptors? How do pesticidal

proteins oligomerise to form the pre-pore and pore structures? How are *B. thuringiensis* and *L. sphaericus* strains able to package and crystallise such a diverse range of structurally distinct proteins? What role do the pro-domains of *B. thuringiensis* and *L. sphaericus* proteins play in the crystallisation process?

The most common technique applied to study the structures of pesticidal proteins to a high resolution is X-ray crystallography (Wlodawer et al. 2013). Indeed, improved ease of access and beamtime availability at synchrotrons, the major X-ray source used for diffraction data collection, has enabled the structures of many *B. thuringiensis* and *L. sphaericus* proteins to be elucidated. However, despite its success, conventional crystallography requires the production of a well-ordered crystal of sufficient size to withstand radiation damage. Some *B. thuringiensis* and *L. sphaericus* proteins (namely members of the Cry, Tpp, App, Mpp, Gpp, Xpp classes) are produced naturally as crystalline inclusions (Tetreau 2021b). Despite this, the *in vivo* grown crystals cannot be studied by conventional crystallography due to their nano – micro size range. This has limited crystallography studies to those where either the protein has been produced in recombinant and soluble form for crystal trials in the laboratory, or to those which solubilise and activate the *in vivo* grown crystals, and then recrystallize the proteins. By doing so, information surrounding the pro-domains and natural crystal packing is lost. In addition, some proteins do not crystallise well *in vitro* and hence, their structures have remained undetermined. A major breakthrough in the field of crystallography has been the development of SFX (see below) (Chapman et al. 2011). This technique is suitable for crystals in the nano – micro size range, hence enabling direct study of the *in vivo* grown crystals, expanding the possibilities for structural studies of *B. thuringiensis* and *L. sphaericus* proteins.

In addition to studying the structures of individual proteins, it is also critical that we gain insight into their interaction with target receptors and/or partner proteins - in the case of binary pesticidal proteins. Given that receptor interaction confers insect specificity, structural elucidation of

pesticidal proteins in complex with receptors will enable engineering studies aimed at broadening the potency and target insect range of pesticidal proteins. For this, the development of single-particle cryoEM (Chua et al. 2022), which itself has experienced a “resolution revolution” and is applicable to the study of larger macromolecules, including protein-protein complexes, will prevail. Single-particle cryoEM will also enable the study of pre-pore and pore structures, as demonstrated for related pore-forming proteins, including aerolysin (Iacovache et al. 2016), providing insight into the cell death mechanisms of pesticidal proteins.

Technological innovations within the field of structural biology have not been limited to wet-lab techniques, as computational approaches (including protein structure prediction, molecular docking, and molecular dynamics) have also experienced major breakthroughs (Mészáros et al. 2023). Computational approaches can overcome some of the limitations that wet-lab techniques present, including limited instrument access, cost, and time-consumption. In addition, proteins are inherently dynamic, a characteristic which can be difficult to analyse using wet-lab techniques. Computational approaches, including molecular dynamics simulations, can be applied to model both protein dynamics and interactions.

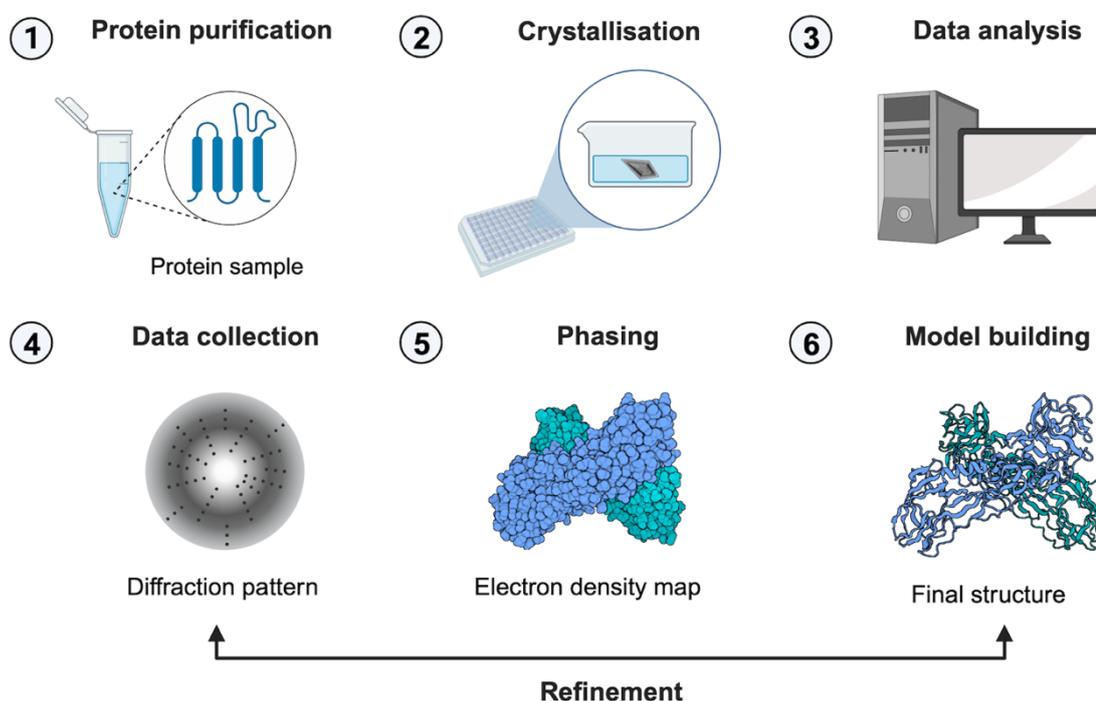
Here, each of these techniques, as well as their advantages and limitations, will be introduced.

### **1.3.1. X-ray crystallography**

#### **1.3.1.1. Conventional crystallography**

X-ray crystallography (**Fig 1.10**), represents one of the predominant techniques applied to elucidate the 3D structure of proteins (Holton and Frankel 2010). Briefly, X-ray crystallography begins with the crystallisation of a highly concentrated and purified protein sample using either hanging-drop or sitting-drop vapour diffusion to produce a single crystal of sufficient size (usually within 50 – 250 µm), purity, and regularity. The successful

crystallisation of a protein sample depends upon several variables, including protein concentration, choice and pH of the buffer solution, temperature, and crystallisation technique, with a trial-and-error approach often required. Hence, elucidation of conditions required for protein crystallisation represents the rate-limiting step of structure solution. Following crystallisation, a single crystal is mounted within an X-ray beam and gradually rotated, allowing diffraction patterns for all orientations of the crystal to be collected. X-ray diffraction experiments may be carried out using laboratory-based X-ray sources, or at a synchrotron facility. The resulting diffraction patterns are processed to identify the size of the unit cell and possible space-groups. In addition, both the amplitude of diffracted waves, which can be calculated directly from the intensity of each diffracted spot, and relative phase of the wave, which must be calculated indirectly by techniques such as molecular replacement or isomorphous replacement, are required to determine a parameter known as the structure factor. Ultimately, determination of the structure factor allows the calculation of an electron density map which may be used to build the atomic structure of the protein using well-established algorithms available in software packages such as CCP4 (Agirre et al. 2023) and Phenix (Adams et al. 2010).

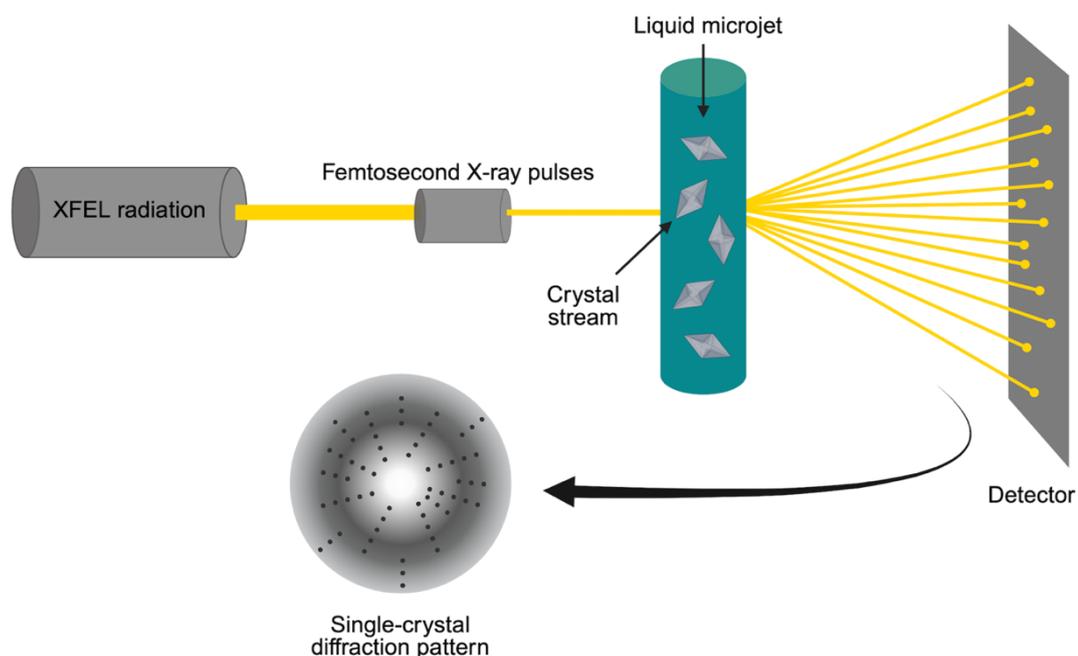


**Figure 1.10. X-ray crystallography to determine the structures of macromolecules.** Created with BioRender.com.

### 1.3.1.2. Serial femtosecond crystallography (SFX)

Although the development of synchrotron radiation has greatly facilitated conventional X-ray crystallography, limitations surrounding crystal quality and radiation damage remain (Holton and Frankel 2010). Specifically, X-ray crystallography relies on the production of large, well-ordered crystals which can produce high resolution diffraction data and at the same time, withstand radiation damage (Holton and Frankel 2010). However, large crystals are often difficult to produce. The invention of X-ray free electron lasers (XFELs) has been shown to mitigate this problem, enabling structure solution around 2 Å or better (Chapman et al. 2011). Specifically, XFELs produce intense pulses of light with femtosecond durations and, therefore, offer the ability to record a diffraction pattern from crystals ranging from micrometres to nanometres in size before radiation damage can occur (Chapman et al. 2007). Indeed, several applications of

XFEL radiation have arisen and include SFX (Chapman et al. 2011). In contrast to conventional crystallography, SFX (**Fig 1.11**) introduces a stream of crystals into an XFEL beam, enabling single pulse diffraction patterns to be collected from individual crystals (Boutet et al. 2012). Several sample delivery systems exist and include liquid jet-based injection, fixed-target chips, and drop-on demand systems, with many research efforts aimed at developing these systems to reduce sample consumption, one of the limitations associated with SFX. More recently, SFX has been applied to perform time-resolved studies of photoactivated proteins, yielding dynamic structure-function information (Aquila et al. 2012; Tenboer et al. 2014; Pandey et al. 2020).



**Figure 1.11. Serial femtosecond crystallography (SFX) at an X-ray free electron laser (XFEL) to determine the structures of macromolecules.** Created with BioRender.com.

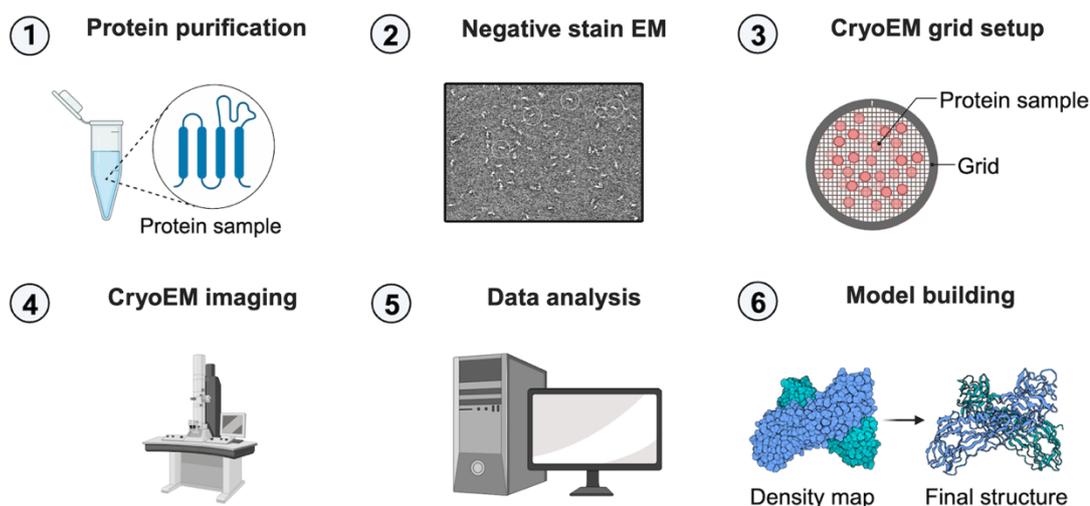
In the field of pesticidal proteins, SFX has been applied to solve the structures of *B. thuringiensis* and *L. sphaericus* proteins in their natural crystal form, which previously have proved too small for X-ray

crystallography at a synchrotron. In 2014, Sawaya et al. used SFX to solve the structure of the Cry3Aa toxin from natural crystals to a final resolution of 2.9 Å (Sawaya et al. 2014). Moreover, diffraction data were collected by directly streaming whole *B. thuringiensis* cells (where the crystal is present on the exterior of dormant cells) through the XFEL beam. Colletier et al. (2016) went on to use SFX to solve the structure of the Tpp1Aa2/Tpp2Aa2 toxin pair from natural crystals isolated from *B. thuringiensis* spores to a final resolution of 2.25 Å (Colletier et al. 2016). More recently, SFX has been applied to study the structures of Cyt1Aa (Tetreau et al. 2020), Cry11Aa, and Cry11Ba (Tetreau et al. 2022). These studies indicate the suitability of SFX for structure solution of pesticidal proteins in their natural crystal form.

### **1.3.2. Single-particle cryogenic electron microscopy (cryoEM)**

An alternative method to study 3D protein structure is single-particle cryoEM (Chua et al. 2022; Chari and Stark 2023). Single-particle cryoEM (**Fig 1.12**) is a powerful tool that involves flash-freezing protein solutions and bombarding the molecules with an electron beam to produce a microscope image, subsequently used to reconstruct the 3D structure. CryoEM itself has experienced a 'resolution revolution' due to developments in electron microscopes, detectors, and data processing software (Chua et al. 2022). This, alongside increased availability and access to instruments throughout the scientific community, has led to a continuing exponential growth in the number of high-resolution protein structures solved by single-particle cryoEM, deposited in both the Protein Data Bank (PDB) and Electron Microscopy Data Bank (EMDB). CryoEM is suitable for the study of large complexes, heterogenous samples, and membrane proteins, the latter of which are particularly challenging to study by X-ray crystallography. Briefly, the cryoEM workflow involves sample preparation and characterisation, which itself includes analysis of the molecular weight, purity and homogeneity, alongside quality screens using negative stain transmission electron microscopy (TEM). Following this, the protein solution is applied to an EM grid, which is blotted to remove excess sample, and flash-frozen by plunging into liquid ethane. Ideally, particles are distributed evenly across

the grid at sufficient concentration and in several orientations, embedded within a uniformly thin layer of vitrified ice. At this stage, time-resolved cryoEM may also be performed by trapping intermediates (e.g., following substrate mixing) at fixed time-points. Images are acquired using a cryogenic transmission electron microscope, which, broadly, includes an electron beam source set up within an ultrahigh vacuum system, a series of magnetic lenses which focus that beam, a cryogenic sample holder used to translate and rotate the grid, and finally, an electron detector. Briefly, image processing involves motion correction, particle picking and extraction, 2D-classification, 3D reconstruction to obtain a cryoEM map, and finally, model building, refinement, and validation using this map. One of the major challenges of cryoEM is beam-induced motion (Kühlbrandt and D'imprima 2021), which may be corrected for during image processing using software that has been the target of much development. Specifically, changes in the ice layer and carbon support film of the EM grid can lead to beam-induced motion, which ultimately results in blurred images, limiting structure resolution.



**Figure 1.12. Single-particle cryogenic electron microscopy (cryoEM) to determine the structures of macromolecules.** Created with BioRender.com.

Within the field of pesticidal proteins, single-particle cryoEM has been applied to elucidate the pro-protein and activated states of lepidopteran-active Vip3Aa16 (Núñez-Ramírez et al. 2020) and Vip3Bc1 (Byrne et al. 2021), both of which are produced during the vegetative growth phase by *B. thuringiensis* strains. Vip3 assembles into a tetramer which, once activated, undergoes a large conformational change to form an extended coiled coil, thought to permeate the membrane (Núñez-Ramírez et al. 2020; Byrne et al. 2021). In addition, pre-pore and pore structures of aerolysin (Iacovache et al. 2016), a pore-forming protein related to the Tpp pesticidal protein class (Lacomel et al. 2021), have also been elucidated using cryoEM, further highlighting its potential for the study of *B. thuringiensis* and *L. sphaericus* proteins. Currently, no toxin-receptor structures exist for any *B. thuringiensis* or *L. sphaericus* pesticidal proteins.

An emerging cryoEM technique is microcrystal electron diffraction (microED) (Nannenga and Gonen 2019). MicroED utilises the diffraction mode of transmission electron microscopes to study crystalline protein structures. In comparison to conventional X-ray crystallography, microED is suitable for much smaller crystals, in the size range of 0.2 – 1  $\mu\text{m}$  (Chua et al. 2022). Briefly, micro-crystals are applied to an EM grid which, as in cryoEM, is blotted to remove excess sample, and flash-frozen by plunging into liquid ethane. Due to its suitability to study crystals in the micro size range, microED may be applicable for the study of *B. thuringiensis* and *L. sphaericus* proteins in their natural crystal form.

### **1.3.3. Computational modelling**

X-ray crystallography and cryoEM represent two of the most predominant wet-lab techniques utilised to study 3D protein structure. However, these techniques can be expensive, time-consuming, and limited by instrument access. Moreover, it can be difficult to routinely investigate protein dynamics. Computational methods, including protein structure

prediction, molecular docking, and molecular dynamics can be complementary to wet-lab techniques, overcoming these limitations.

### 1.3.3.1. Protein structure prediction

The accurate prediction of a protein's 3D structure has been a major goal within the field of computational biology (Bertoline et al. 2023), with a community-wide, independent, and worldwide experiment (Critical Assessment of Structure Prediction – CASP) aimed at objectively testing structure prediction programs/methods occurring on a bi-annual basis (Moult et al. 1995; Kryshtafovych et al. 2023). Protein structure prediction approaches can be divided into templated based (homology modelling and threading/fold recognition) and free modelling (*ab initio*) approaches (Bertoline et al. 2023). As the name suggests, template-based approaches utilise sequence alignments and 3D templates available in the PDB to build template backbones (homology modelling) or identify similar folds (threading/fold recognition) for protein structure prediction (Bertoline et al. 2023). Examples of homology modelling and threading/fold recognition programs include SWISS-MODEL (Waterhouse et al. 2018) and GenTHREADER (Jones 1999), respectively. *Ab initio* approaches, such as QUARK (Mortuza et al. 2021), are based on the thermodynamic hypothesis, which states that the native protein structure will exist at the lowest free energy (Anfinsen 1973), and thus, consider physiochemical properties to predict the protein structure.

More recently, DeepMind, a start-up of Google, released AlphaFold (AF), followed by AlphaFold2 (AF2), a deep-learning based prediction tool which uses a convolutional neural network trained on structures available in the PDB (Jumper et al. 2021b). AF performed exceptionally well in the most recent (14<sup>th</sup>) edition of CASP, which demonstrated a median backbone accuracy of 0.96 Å root-mean-square deviation (RMSD) between predicted and experimental backbone structures (Jumper et al. 2021a; Kryshtafovych et al. 2021). This was followed by the next best-performing program which exhibited an RMSD of 2.8 Å between predicted and experimental structures

(Jumper et al. 2021a; Kryshtafovych et al. 2021). In addition to the open-source code for AF2 being freely available on github (<https://github.com/deepmind/alphafold>), DeepMind has also partnered with the EMBL-European Bioinformatics Institute (EMBL-EBI) to produce the AF Protein Structure Database (AF-DB), which now has over 200 million entries for the human proteome and proteomes of 47 other key organisms, including *B. thuringiensis* and *L. sphaericus* (Varadi et al. 2022). The release of AF and AF2 is a ground-breaking development within the scientific community and has led to spin-off algorithms such as AF-multimer (for prediction of macromolecular complexes) (Evans et al. 2022) and AF-pulldown (for screening of protein-protein interactions) (Yu et al. 2023). Some limitations and areas for development remain (Bertoline et al. 2023) and include the inability of AF2 to predict intrinsically disordered proteins or regions accurately, as well as flexible loop regions. In addition, AF2 does not accurately predict the effect of mutations on protein structure and/or folding stability (Pak et al. 2023), and further to this, does not consider the impact of external factors, such as the effect of environment pH on protonation state, temperature, and conformational dynamics. Despite these limitations, the release of AF2 has been critical for generating hypotheses about protein function and for investigating systems which have remained challenging to study using wet-lab techniques.

#### **1.3.3.2. Molecular Docking**

Molecular docking predicts the 3D structure of protein-protein complexes using unbound protein structures determined by either wet-lab techniques or protein structure prediction (Vakser 2014). A number of docking software programs exist and include ClusPro (Kozakov et al. 2017), RosettaDock (Chaudhury et al. 2011), HADDOCK (Van Zundert et al. 2016), FRODOCK (Ramírez-Aportela et al. 2016), ZDOCK (Pierce et al. 2014), and ATTRACT (De Vries et al. 2015), as well as deep-learning techniques, AlphaFold-Multimer (Evans et al. 2022) and ColabFold (Mirdita et al. 2022). In general, computational docking consists of a docking stage, during which billions of possible conformations are sampled, and a scoring stage, during

which docked conformations are evaluated and ranked according to a weighted scoring function (Zhang et al. 2016). Docking may be carried out either globally or locally, thus enabling suspected binding sites to be explored.

One limitation of the standard computational docking protocol is its inability to account for protein flexibility and binding-induced conformational changes (Lensink et al. 2017). Nonetheless, recent advances include the ability to model sidechain and backbone flexibility better in RosettaDock (Wang et al. 2005; Marze et al. 2018). This is achieved by including alternative and unbound native rotamers, and by producing large conformational ensembles of the unbound proteins prior to docking (Wang et al. 2005; Marze et al. 2018). Despite this progress, the ability to account for protein flexibility remains the primary challenge of computational docking (Lensink et al. 2017). Furthermore, output models often have similarly good scores and hence, discriminating the near-native structure from similarly scored, non-native structures, remains a challenge.

#### **1.3.3.3. Molecular Dynamics (MD)**

Molecular dynamics (MD) is an approach that utilises Newton's equations of motion to analyse the dynamics of atoms and molecules with respect to time (Hollingsworth and Dror 2018). Hence, MD can be used to study protein dynamics (Sakano et al. 2016; Radom et al. 2018). In the late 1950s, the first MD simulation experiment was carried out by Alder and Wainwright to study the interactions of hard spheres (Alder and Wainwright 1957; Alder and Wainwright 1959). This was followed by the simulation of liquid water by Stillinger and Rahman in 1974 (Stillinger and Rahman 1974). In 1977, the first simulation of proteins, namely – Bovine pancreatic trypsin inhibitor, was performed (McCammon et al. 1977). Nowadays, MD simulations are routinely performed to study several system types, including solvated proteins, protein-protein and protein-DNA complexes, proteins within membrane systems. The development of better forcefields, including AMBER (Tian et al. 2020) and CHARMM (Huang et al. 2016), alongside

more powerful computers has enabled longer and more complex simulations to be performed. The resulting trajectory, which provides information about the total energy of the system and atom coordinates, can be analysed to gain biological insight into the system. In this thesis, MD has been applied to evaluate the stability and likelihood of docked protein-protein complexes persisting in solution.

#### **1.4. Aims and Objectives**

##### **1.4.1. Using serial femtosecond crystallography (SFX) to elucidate the structures of pesticidal proteins from natural crystals**

A key objective of this PhD was to utilise SFX to elucidate the structures of pesticidal proteins from natural crystals. As part of this work, the structures of 7 new pesticidal proteins (Tpp49Aa1, Cry8Ba2, Cry1Ac1, Cry1Ca1, Mpp60Aa1, App6Aa2, MD32, as well as variants exposed to different pH or bound to carbohydrate molecules) were solved across three separate beamtimes in collaboration with researchers from Cardiff University, the University of Sussex, the European X-ray Free Electron Laser (EuXFEL) facility, and the Deutsches Elektronen-Synchrotron (DESY) facility. For each of these structures, I was involved in various stages and timepoints of the experiment, including crystal protein production, diffraction data collection, and/or diffraction data analysis. This thesis presents the structure solution of the Cry8Ba2 and Tpp49Aa1 pesticidal proteins.

##### **1.4.2. Structural analysis of the interaction of Tpp2Aa2 with its mosquito receptor, *Culex quinquefasciatus* maltase 1 (Cqm1)**

As previously discussed, it is critical that we develop our knowledge of toxin-receptor interactions to gain a better understanding of mechanism of action, overcome emerging field resistance, and for downstream engineering efforts aimed at broadening target insect range. Across all *B. thuringiensis* and *L. sphaericus* pesticidal protein classes, no structure of a toxin-receptor complex had been elucidated, and hence, a second objective of this PhD

was to characterise the interaction of the mosquitocidal protein, Tpp2Aa2 (of the binary protein – Tpp1Aa2/Tpp2Aa2), with its target receptor, Cqm1. This thesis presents the prediction of the Tpp2Aa3-Cqm1 complex using computational modelling techniques, which is compared with the experimentally resolved Tpp2Aa2-Cqm1 structure, the latter of which was subsequently elucidated using single-particle cryoEM in collaboration with researchers from Monash University. In addition, knowledge of the Tpp2Aa2-Cqm1 structure was applied in an attempt to broaden target insect range to the refractory mosquito species, *Aedes aegypti*, by using mutagenesis techniques with the goal of inducing binding of Tpp2Aa2 to Aam1, the Cqm1 ortholog.

## 2. Materials and Methods

### 2.1. Materials

#### 2.1.1. Chemicals and media

Routine laboratory chemicals and media were used in the following methodologies. Product sources (manufacturers) and specifications have been detailed in **Table 2.1**. Where stock solutions were required, chemicals were dissolved in deionised water (dH<sub>2</sub>O). Compositions of buffers and solutions have been detailed throughout the relevant methodologies.

**Table 2.1. Chemicals and media.**

<b>Chemical</b>	<b>Specification / preparation</b>	<b>Manufacturer</b>
Acetic acid	≥ 99.5% purity	Scientific Laboratory Supplies
Acrylamide: Bis-Acrylamide	40% (w/v) Acrylamide: Bis-Acrylamide 29:1	Geneflow
Agarose	Molecular biology grade	Appleton Woods
Ammonium persulphate (APS)	Prepared as 10% (w/v) stock	Melford
β-mercaptoethanol	≥ 99% purity	Sigma-Aldrich
Bromophenol blue	Technical grade	Sigma-Aldrich
Calcium carbonate (CaCO <sub>3</sub> )	≥ 98.5% purity	Duchefa Biochemie
Coomassie blue	Coomassie Brilliant Blue R250	Thermo Fisher Scientific
Deoxyadenosine triphosphate (dATP)	Prepared as 10 mM stock	Promega
Deoxycytidine triphosphate (dCTP)	Prepared as 10 mM stock	Promega
Deoxyguanosine triphosphate (dGTP)	Prepared as 10 mM stock	Promega
Deoxythymidine triphosphate (dTTP)	Prepared as 10 mM stock	Promega
Dimethyl sulfoxide (DMSO)	≥ 99.5% purity	Sigma-Aldrich

Ethanol	≥ 99.8% purity	VWR
Ethidium bromide	Prepared stock as 10 mg/mL stock	Sigma-Aldrich
Ethylenediamine-tetraacetic acid (EDTA)	EDTA disodium salt dihydrate	Apollo Scientific
Glycerol	Analytical grade	Thermo Fisher Scientific
Glycine	Glycine free base	Melford
Imidazole	≥ 99% purity	Sigma-Aldrich
Iron sulphate (FeSO <sub>4</sub> )	≥ 97% purity	Acros Organics
Isopropyl β-d-1-thiogalactopyranoside (IPTG)	Prepared as 1 M stock	Melford
Isopropanol	≥ 99.5% purity	Thermo Fisher Scientific
Luria-Bertani (LB)	Molecular biology grade	Melford
Luria-Bertani agar (LB-agar)	Molecular biology grade	Miller
Magnesium sulphate (MgSO <sub>4</sub> )	≥ 99.5% purity	Sigma-Aldrich
Manganese sulphate (MnSO <sub>4</sub> )	≥ 99% purity	VWR Chemicals BDH
Methanol	≥ 99.8% purity	Thermo Fisher Scientific
Nutrient broth	Molecular biology grade	Sigma-Aldrich
Potassium hydrogen phosphate (K <sub>2</sub> HPO <sub>4</sub> )	≥ 98% purity	VWR Chemicals BDH
Reduced glutathione	≥ 98% purity	Sigma-Aldrich
Sodium azide (NaN <sub>3</sub> )	≥ 99.5% purity	Sigma-Aldrich
Sodium carbonate (Na <sub>2</sub> CO <sub>3</sub> )	≥ 99.95% purity	Fisher Scientific
Sodium chloride (NaCl)	≥ 99.5% purity	Thermo Fisher Scientific
Sodium citrate	≥ 99.9% purity	Sigma-Aldrich
Sodium dodecyl sulphate (SDS)	Prepared as 10% (w/v) stock	Melford
Sucrose	Ultra-high purity	Melford

Tetramethylethylenediamine (TEMED)	≥ 99% purity	Melford
Tris(hydroxymethyl)aminomethane acetate (Tris-acetate)	≥ 99% purity	Thermo Fisher Scientific
Tris(hydroxymethyl)aminomethane hydrochloride (Tris-HCl)	≥ 99% purity	Duchefa Biochemie
Tris(hydroxymethyl)aminomethane base (Tris-base)	≥ 99.5% purity	Melford
Yeast extract	Molecular biology grade	Oxoid
Zinc sulphate (ZnSO <sub>4</sub> )	≥ 99.9% purity	VWR Chemicals BDH
2x YT	Molecular biology grade	Melford

### 2.1.2. Antibiotics

All antibiotics were prepared as 1000x stocks and stored at -20°C. The final stock concentrations are provided in **Table 2.2**.

**Table 2.2. Antibiotic preparations.**

Antibiotic	Stock concentration	Solvent	Manufacturer
<b>Ampicillin</b>	100 mg/mL	dH <sub>2</sub> O	Melford
<b>Erythromycin</b>	5 – 10 mg/mL	Ethanol	Sigma-Aldrich
<b>Kanamycin</b>	50 mg/mL	dH <sub>2</sub> O	Sigma-Aldrich

### 2.1.3. Bacterial growth media

Embrapa medium was prepared by dissolving 3.2 g nutrient broth (Sigma-Aldrich), 0.4 g yeast extract (Oxoid), and 0.4 g K<sub>2</sub>HPO<sub>4</sub> (VDW Chemicals BDH) in 400 mL dH<sub>2</sub>O supplemented with 4 mL salt solution

containing 100 mM CaCO<sub>3</sub>, 40 mM MgSO<sub>4</sub>, 3.6 mM FeSO<sub>4</sub>, 3.6 mM MnSO<sub>4</sub>, 3.5 mM ZnSO<sub>4</sub>. Embrapa medium was adjusted to pH 7.0. LB broth and solid medium LB-agar plates were prepared by dissolving 20 g of LB-broth powder (Melford) and 40 g of LB-agar (Miller) in 1 L of dH<sub>2</sub>O. 2xYT medium was prepared by dissolving 31 g of granulated 2xYT (Melford) in 1 L of dH<sub>2</sub>O. Embrapa medium, LB-broth, LB-agar and 2xYT were sterilised by autoclaving at 121°C.

#### 2.1.4. Bacterial strains

For DNA plasmid amplification and recombinant protein expression, two *Escherichia coli* (*E. coli*) strains were used. The genotypes and source have been detailed in **Table 2.3**.

**Table 2.3. Bacterial strains and genotypes.**

Bacterial strain	Genotype	Manufacturer
5-alpha <i>E. coli</i>	fhuA2Δ(argF-lacZ)U169 phoA glnV44 Φ80Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17	New England Biolabs
BL21 (DE3) <i>E. coli</i>	fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5	New England Biolabs

### 2.1.5. Vectors for protein expression

Plasmids used throughout this study for protein production are detailed in **Table 2.4** along with their parental expression plasmids.

**Table 2.4. Expression plasmids.**

<b>Parental vectors</b>	<b>Gene</b>	<b>Source</b>
<b>pGEX (Ampicillin<sup>R</sup>)</b>	Encoding Tpp1Aa2 and Tpp2Aa2 proteins containing N-terminal glutathione-S-transferase (GST) tag	Professor Colin Berry, School of Biosciences, Cardiff University, Cardiff, UK
<b>pET28a (Kanamycin<sup>R</sup>)</b>	Encoding Cqm1 protein containing N-terminal 6x histidine tag	Associate Professor Michelle Dunstone, Biomedicine Discovery Institute, Department of Biochemistry and Molecular Biology, Monash University, Melbourne, Australia
<b>pSTAB (Erythromycin<sup>R</sup>)</b>	Encoding Cry48Aa1 crystal protein	Professor Colin Berry, School of Biosciences, Cardiff University, Cardiff, UK
<b>pHT304 (Erythromycin<sup>R</sup>)</b>	Encoding Tpp49Aa1 crystal protein	Professor Colin Berry, School of Biosciences, Cardiff University, Cardiff, UK

## 2.2. Molecular biology

### 2.2.1. DNA oligonucleotides

SnapGene software (available at: [snapgene.com](http://snapgene.com)) was used for visualisation of DNA, annotation of features, and design of primers. Primers were synthesized using the Eurofins Genomics oligonucleotide synthesis service. Annealing temperatures were determined using the T<sub>m</sub> calculator available online from NEB (accessed at: <https://tmcalculator.neb.com>).

### **2.2.2. Agarose gel electrophoresis**

DNA samples were analysed and separated using agarose gel electrophoresis. Gels containing 1% agarose were prepared by dissolving 0.7 g agarose powder (Melford) in 70 mL of 1x TAE (40 mM Tris-acetate, 1 mM EDTA, pH 8.8). Ethidium bromide was added to a final concentration of 0.5 µg/mL. DNA samples (5 mL) were mixed with 1mL 6x gel loading dye (NEB) and run against 2 mL 1 kB plus DNA ladder (NEB) at 120 V for 40 minutes. A GelDoc-It UV-Transilluminator (Ultra-Violet Products Ltd) was used to visualise the DNA bands.

### **2.2.3. DNA purification and isolation**

Plasmid DNA was isolated and purified from bacterial cells using the QIAprep Spin Miniprep Kit (Qiagen, 27104) as described by the manufacturer's procedures. Briefly, cells were harvested from overnight bacterial cultures by centrifugation at 4000 x g for 15 minutes. Subsequently, the cell pellet was resuspended, lysed under alkaline conditions, and plasmid DNA extracted in TE buffer (10 mM TrisHCl, 1 mM EDTA, pH 8.0).

### **2.2.4. DNA quantification**

DNA concentration was estimated using a NanoDrop ND1000 Spectrophotometer (Thermo Fisher Scientific). The spectrophotometer was blanked using TE buffer and the concentration of samples was determined at 260 nm according to the Beer-Lambert law. Contaminants (organic compounds, RNA, or proteins) were detected by absorbance at 230 and 280 nm.

### **2.2.5. Whole-vector site-directed mutagenesis (SDM)**

To generate the Tpp2Aa2-E65A mutant, whole-vector site-directed mutagenesis (SDM) was carried out using the Q5 Site-Directed Mutagenesis

Kit (NEB, E0554S). Briefly, plasmid DNA containing the desired gene (pGEX-Tpp2Aa2) was amplified using Q5 Hot Start High-Fidelity DNA polymerase and custom primers, one of which contained the desired mutation (**Table 2.5**). An annealing temperature of 64°C and extension time of 3 minutes were used (**Table 2.6**). Following PCR, the expression plasmid was re-ligated using the Quick Ligation Kit (NEB, M2200S) and transformed into 5-alpha competent *E. coli* cells (NEB, C2987H). Cells were plated onto LB-agar plates containing ampicillin (100 µg/mL) and incubated at 37°C overnight. An individual Tpp2Aa2 positive colony was utilised to inoculate 5 mL LB containing ampicillin (100 µg/mL) and incubated overnight at 37°C with shaking (200 rpm). The pGEX-Tpp2Aa2 expression plasmid was purified and sent for sequencing. Following identification of a colony with the expected mutation, the pGEX-Tpp2Aa2 expression plasmid was re-transformed into BL21 (DE3) chemically competent *E.coli* cells (NEB, C2527).

**Table 2.5. Custom primers used for site-directed mutagenesis.**

Primer	Sequence (5' to 3')	Tm (°C)
<b>Tpp2Aa2 E65A mutation</b>		
Forward	CTT CAA GTA TCG CTA ATT GCC CAT CTA AC	65
Reverse	GAA ATT CGG TTT TTG ATA AAC CAT AAC C	63

\* Mutations introduced are highlighted in red

**Table 2.6. DNA amplification using Q5 DNA polymerase.**

Reagent		Final concentration				
Q5 reaction buffer		1x				
Q5 DNA polymerase		1 unit				
Forward primer		0.5 mM				
Reverse primer		0.5 mM				
dNTPs		0.2 mM				
Template DNA		1 ng/mL				
Total volume *		50				
PCR conditions						
Condition	Initial denaturation	30 cycles			Final extension	Hold
		Denaturation	Annealing	Extension		
Time (s)	30	10	30	30 / kB product	120	∞
Temp (°C)	98	98	64	72	72	4

\* Made up with nuclease-free water

### 2.2.6. Bacterial transformation

For plasmid amplification, plasmids were transformed into 5-alpha chemically competent *E. coli* cells (NEB, C2987). For protein production, expression plasmids were transformed into BL21 (DE3) chemically competent *E. coli* cells (NEB, C2527). In both cases, transformation was performed according to the NEB heat shock protocol. Briefly, competent cells were thawed on ice for 10 minutes before being incubated with 1 mL plasmid DNA (50 – 100 ng of DNA) for 30 minutes. Cells were incubated at 42°C in a water bath for exactly 30 seconds before being placed on ice for an additional 5 minutes. Subsequently, the cell mixture was added to 950 mL Super Optimal Broth (SOB, NEB) and incubated with shaking at 37°C for 1 hour. Following incubation, cells were spread onto LB-agar plates containing the appropriate selection antibiotic and incubated at 37°C overnight.

### 2.2.7. Sequencing

The Eurofins Genomics DNA TubeSeq service and standard primers (Table 2.7) were used to sequence expression plasmids, enabling verification of mutations and gene inserts.

**Table 2.7. Standard primers used for Sanger sequencing.**

Primer	Sequence (5' to 3')
<b>Verification of Tpp1Aa2 and Tpp2Aa2</b>	
pGEX-5'	CTG GCA AGC CAC GTT TGG
pGEX-3'	GGA GCT GCA TGT GTC AGA GG
<b>Verification of Cqm1</b>	
T7	TAA TAC GAC TCA CTA TAG GG
T7term	CTA GTT ATT GCT CAG CGG T

### 2.2.8. Preparation of *E. coli* dimethyl sulphoxide stocks

For long-term storage, *E. coli* cells transformed with the desired expression plasmid were added to 8% DMSO and stored at -80°C. DMSO stocks were prepared by adding 920 µL bacterial culture (grown in LB containing appropriate selection antibiotic at 37°C overnight) to 80 µL DMSO and mixing thoroughly.

## 2.3. Protein expression

### 2.3.1. Expression testing of Cqm1 protein in *E. coli*

The histidine (6x His) tagged *cqm1* gene lacking the GPI anchor and cloned into the pET28a expression plasmid under the control of the lac operator was donated by A/Prof. Michelle Dunstone (Monash University, Melbourne, Australia). The pET28a-Cqm1 expression plasmid was verified

by Sanger sequencing and transformed into BL21 (DE3) chemically competent *E. coli* cells (NEB, C2527) and stored as a DMSO stock at -80°C. From these DMSO stocks, pET28a-Cqm1 transformed cells were plated onto LB-agar plates containing kanamycin (50 µg/mL) and incubated at 37°C overnight. Individual colonies were utilised to inoculate 5 mL LB containing kanamycin (50 µg/mL) and this culture was incubated at 37°C overnight with shaking (200 rpm). The following day, 1 mL culture was utilised to inoculate 20 mL fresh LB containing kanamycin (50 µg/mL) and incubated for 8 h at 37°C with shaking (200 rpm). Subsequently, IPTG was added at a range of concentrations (0.1 mM, 0.5 mM, 1 mM) to the cultures, which were incubated overnight at two different temperatures (16°C and 37°C) with shaking (200 rpm). The following day, 1 mL samples were obtained, pelleted at 13,000 rpm for 3 minutes in a table-top microcentrifuge and resuspended to an OD<sub>600</sub> (degree of light scattering / attenuation at 600 nm) of 10. Lysozyme (10 µL 10 mg/mL) was added, and the samples were incubated at 37°C for 15 minutes, and freeze-thawed (-80°C, 37°C) twice. Following this, the samples were sonicated in a water bath for 3 minutes and spun at 13,000 rpm for 3 minutes in a table-top microcentrifuge. The supernatant (soluble fraction) was obtained, and the pellet was resuspended to an OD<sub>600</sub> of 10 (insoluble fraction). All samples were stored at 4°C before analysis by sodium dodecyl-sulphate polyacrylamide gel electrophoresis (SDS-PAGE).

### **2.3.2. Expression of Cqm1 protein in *E. coli***

Following expression testing and for all subsequent protein production, pET28a-Cqm1 transformed BL21 cells were plated onto LB-agar plates containing kanamycin (50 µg/mL) and incubated at 37°C overnight. An individual colony was utilised to inoculate 5 mL LB containing kanamycin (50 µg/mL) and this culture was incubated at 37°C overnight with shaking (200 rpm). The following day, this 5 mL culture was used to inoculate 500 mL 2xYT medium containing kanamycin (50 µg/mL). Subsequently, this culture was incubated at 37°C until an OD<sub>600</sub> of 0.6 was obtained, at which point IPTG was added to a final concentration of 1 mM. Cultures were grown at 16°C overnight. The following day, the culture was harvested at 4,000 x g

for 20 minutes at 4°C using the Fibrelite™ F9-6x1000 LEX rotor and Sorvall Lynx 6000 Centrifuge (Thermo Fisher Scientific). The pellet was resuspended in 20 mL 1 x Lysis-Equilibration-Wash (LEW) buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, pH 8.0) supplemented with one tablet cOmplete™, EDTA-free Protease Inhibitor Cocktail (Roche) and stored at -20°C until affinity purification. Immediately prior to purification, the cell suspension was pulse-sonicated (4 x 1 min). The lysed cell suspension was then centrifuged at 25,000 x g for 20 minutes at 4°C using the Fibrelite™ F21-8x50y rotor and Sorvall Lynx 6000 Centrifuge (Thermo Fisher Scientific). The 6xHis-Cqm1 fusion protein containing supernatant was obtained, passed sequentially through 0.45 and 0.2 mM syringe filters (Sartorius) and purified using nickel affinity purification (**section 2.4.1**).

### **2.3.3. Expression of Tpp1Aa2 and Tpp2Aa2 proteins in *E. coli***

The glutathione-S-transferase (GST) tagged *tpp1Aa2* and *tpp2Aa2* genes were previously cloned into separate pGEX expression plasmids under the control of the lac operator, transformed into BL21 (DE3) chemically competent *E. coli* cells (NEB, C2527) and stored as DMSO stocks at -80°C. From these DMSO stocks, pGEX-Tpp1Aa2 and pGEX-Tpp2Aa2 transformed cells were plated onto separate LB-agar plates containing ampicillin (100 µg/mL) and incubated at 37°C overnight. Individual colonies were utilised to inoculate 5 mL LB containing ampicillin (100 µg/mL) and these cultures were incubated at 37°C overnight with shaking (200 rpm). The following day, 5 mL cultures were used to inoculate 500 mL 2xYT media containing ampicillin (100 µg/mL). Subsequently, these cultures were incubated at 37°C until an OD<sub>600</sub> of 0.6 was obtained, at which point IPTG was added to a final concentration of 1 mM. Growth at 37°C ensued for 4 hours. Following this, the cultures were harvested at 9,000 g for 7 minutes at 4°C using the Fibrelite™ F9-6x1000 LEX rotor and Sorvall Lynx 6000 Centrifuge (Thermo Fisher Scientific). The pellets were resuspended in 20 mL 1x Phosphate-Buffered Saline (PBS) and stored at -20°C until affinity purification. Immediately prior to purification, the cell suspensions were frozen and thawed (-20°C, 37°C) twice and pulse-sonicated (3 x 20 sec). The lysed cell

suspension was then centrifuged at 31,000 g for 15 minutes at 4°C using the Fibrelite™ F21-8x50y rotor and Sorvall Lynx 6000 Centrifuge (Thermo Fisher Scientific). The GST-Tpp1Aa2 and GST-Tpp2Aa2 fusion protein containing supernatant was obtained, passed through a 0.2 µm filter (Sartorius) and purified individually using GST affinity purification (**section 2.4.2**).

#### **2.3.4. Expression of Cry48Aa1, Tpp49Aa1 and Cry8Ba2 crystal proteins**

The *cry48Aa1* and *tpp49Aa1* genes (accession number AJ841948) were previously cloned into pSTAB and pHT304 expression vectors, respectively. Expression plasmids contained an erythromycin resistance gene and were separately transformed into *B. thuringiensis* subsp. *israelensis* 4Q7 (Jones et al. 2007) and donated by Prof. Colin Berry (Cardiff University, Cardiff, UK). The native *B. thuringiensis* strain Na210, encoding the *cry8Ba2* gene (accession number MZ355710), was donated by Prof. Primitivo Caballero (The Public University of Navarra, Pamplona, Spain). For all strains, sporulated cells were immobilized on sterilised filter paper and stored at room temperature. From these stocks, an individual spore paper was used to inoculate 5 mL LB and incubated at 30°C overnight with shaking (200 rpm). The following day, this preculture was used to inoculate 400 mL Embrapa medium. For strains expressing recombinant protein, growth medium was supplemented with erythromycin to a final concentration of 5 µg/mL. Cultures were grown at 30°C with shaking (200 rpm) until sporulation reached 90%, as judged by phase contrast microscopy. Sporulated cultures were subsequently harvested by centrifugation for 15 minutes at 12,000 g, 4°C using the Fibrelite™ F9-6x1000 LEX rotor and Sorvall Lynx 6000 Centrifuge (Thermo Fisher Scientific). The pellet was resuspended and washed once in 250 mL ice-cold 1 M NaCl containing 10 mM EDTA and twice in 100 mL ice-cold 10 mM EDTA before final resuspension in 7.5 mL dH<sub>2</sub>O. The final spore suspension was lysed by sonication using the MSE Soniprep 150 at 50% amplitude in 4 x 30 second pulses. The lysed spore suspension was separated on a discontinuous sucrose gradient (67% / 72% / 79% / 84%) using an SW 28 Ti swinging-bucket rotor and Beckman Coulter

Optima L 90K Ultracentrifuge for 16 hours at 24,000 rpm, 15°C. Purified crystal protein bands were extracted from the sucrose gradient by piercing the ultracentrifuge tubes using a 16-G needle and syringe. Extracted crystals were washed twice in dH<sub>2</sub>O using the Fibrelite™ F21-8x50y rotor and Sorvall Lynx 6000 Centrifuge (Thermo Fisher Scientific) before resuspension in 1 mL dH<sub>2</sub>O. Crystal protein suspensions were maintained using 10 mM sodium azide and stored at 4°C before SDS-PAGE analysis.

## **2.4. Recombinant protein purification**

### **2.4.1. Nickel affinity chromatography**

For purification of His-tagged proteins, gravity-flow affinity purification was performed using the Protino Ni-TED 2000 packed columns kit (Macherey-nagel, Düren, 745120) according to the manufacturer's instructions. Briefly, columns were equilibrated with 4 mL 1 x LEW buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, pH 8.0). Following this, the clarified lysate containing the 6xHis-Cqm1 fusion protein was added to the capped column and allowed to drain. The column was washed with 2 x 4 mL LEW buffer, after which the 6xHis-Cqm1 fusion protein was eluted in 3 x 3 mL LEW-imidazole elution buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, 250 mM imidazole, pH 8.0). Elution fractions were stored at -20°C before SDS-PAGE analysis.

### **2.4.2. Glutathione-S-transferase (GST) affinity chromatography**

For purification of GST-tagged proteins, gravity-flow affinity purification was used. Purification columns were washed with 10 mL 50 mM TrisHCl, 10 mM reduced glutathione (Sigma-Aldrich), pH 8.0, capped and equilibrated with 50% glutathione sepharose beads (GE Healthcare, 17075601) followed by 30 mL PBS, pH 7.4. The clarified lysate containing the GST fusion protein was added to the capped column and incubated for 1 hour at room temperature to allow bead binding. The column was then drained and washed with 30 mL PBS, pH 7.4. The column was drained, and the GST fusion protein was eluted in 10 x 1 mL fractions of 50 mM TrisHCl,

10 mM reduced glutathione, pH 8.0. Fractions were stored at -20°C before SDS-PAGE analysis.

### **2.4.3. Size exclusion chromatography (SEC)**

For SEC, proteins were buffer exchanged into 20 mM TrisHCl, pH 8.5. SEC was performed using a calibrated HiLoad™ 16/60 Superdex™ S200 pg column. For calibration, BioRad standard proteins (molecular weight 670, 158, 44, 17 and 1.35 kDa) were run on the column at a flow rate of 0.5 mL/min in 50 mM TrisHCl, pH 8.0. Protein samples were run on the column at a flow rate of 1.0 mL/min in 20 mM TrisHCl, pH 8.5. Elution fractions were stored at -20°C before SDS-PAGE analysis.

## **2.5. Protein analysis methods**

### **2.5.1. Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)**

Protein samples (8 µL) were added to 4 µL 5x SDS-loading buffer (0.2 M TrisHCl pH 6.8, 10% (v/v) β-mercaptoethanol, 8% (w/v) glycerol, 2% (w/v) SDS, 0.04% (w/v) bromophenol blue) and 8 µL distilled H<sub>2</sub>O and heated at 98°C for 5 min. Heating and treating with β-mercaptoethanol caused the protein to denature, whilst SDS coated the polypeptides with a negative charge (Laemmli 1970).

Following this, protein samples were loaded into the wells of pre-cast gels composed of a 12.5% resolving gel (42.2% (v/v) dH<sub>2</sub>O, 31.3% (v/v) acrylamide: Bis acrylamide, 24.8% (v/v) 1.5 M TrisHCl pH 8.8, 1.0% (v/v) SDS, 0.5% (v/v) APS, and 0.2% (v/v) TEMED) and stacking gel (73.8% (v/v) dH<sub>2</sub>O, 13.0% (v/v) acrylamide: Bis acrylamide, 10.0% (v/v) 0.5 M TrisHCl pH 6.8, 2.0% (v/v) SDS, 1.0% (v/v) APS, 0.2% (v/v) TEMED. BLUeye Pre-Stained Protein Ladder (Geneflow, S6-0024), Color Prestained Protein Standard (NEB, P7719S), and Precision Plus Protein Standards Dual Color (Bio-Rad, 1610374) were used as molecular markers. Gels were run in Tris-

glycine SDS running buffer (192 mM glycine, 25 mM TrisBase, 0.1% (w/v) SDS, pH 8.3) at 200 V for a total of ~ 45 minutes before being stained with Coomassie Blue for 30 minutes. Finally, the gel was destained (10% acetic acid, 40% methanol, 50% distilled H<sub>2</sub>O) overnight before visualisation using the Bio-Rad Gel Doc EZ Imager.

### **2.5.2. Determination of protein concentration**

The bicinchoninic acid (BCA) protein assay was used to determine protein concentration. Diluted protein samples, 10 µL of 1:5, 1:10, and 1:20 dilutions, were plated in triplicate into a 96-well microplate. A standard curve of bovine serum albumin (BSA) with a range of concentrations (0 – 75 µg/mL) plated in triplicate was used to calculate the protein sample concentration. Cu<sub>2</sub>SO<sub>4</sub> (200 µL) was added to 10 mL BCA reagent and 90 µL of this BCA reagent mixture was added to each well of the 96-well microplate containing the protein sample or BSA standard. The reaction was incubated with shaking for approximately 30 minutes at 37°C, or until the colour of the reaction changed from green to light grey / purple. Absorbance was measured at 570 nm using a microplate reader (Infinite F50, Tecan, Switzerland).

### **2.5.3. Trypsin activation of proteins**

Trypsin cleaves at the carboxyl side of arginine and lysine residues. Immobilized TPCK-Trypsin agarose gel (Thermo Fisher Scientific, 20230) was used to trypsin activate Cry48Aa1, Tpp49Aa1 and Cry8Ba2, all of which are usually proteolytically cleaved in the target insect gut. Briefly, ~1 mg of protein sample was incubated with 100 mL of Immobilized TPCK Trypsin for 18 hours at room temperature with shaking. Cleavage was confirmed by SDS-PAGE and trypsin agarose gel was removed from the protein sample by centrifugation.

#### **2.5.4. Mosquito toxicity bioassays**

To confirm known activity of Tpp1Aa2/Tpp2Aa2 and Cry48Aa1/Tpp49Aa1, high doses of toxins (in combination or alone) were added to 1 mL of water containing 5 third instar larvae (*Cx. quinquefasciatus*, *Ae. aegypti*, and *An. gambiae*), with three replicates per species. Mortality was observed after 24 hours.

To identify new targets for Cry48Aa1/Tpp49Aa1, 100 mL dH<sub>2</sub>O containing 10 fourth instar larvae (*Ae. albopictus*, *An. stephensi*, and *C. tarsalis*) were placed in 350 mL cups. A 1:1 ratio of Cry48Aa1:Tpp49Aa1 at a range of concentrations was tested, with three replicates per concentration. Mortality was observed at 24- and 48-hour time periods. Finney's probit analysis was used to determine LC<sub>50</sub> and LC<sub>95</sub> values. Mosquito bioassays to identify new targets were performed by researchers at the Department of Entomology and Institute for Integrative Genome Biology, University of California, USA and Department of Biological Sciences, California Baptist University, USA.

#### **2.5.5. Insect feeding bioassays**

To investigate the target insect range of Cry8Ba2, insect feeding bioassays were performed against *Tenebrio molitor* (mealworms) and *Zophobas morio* (superworm) larvae. Portions of sweet potato or courgette (1 cm<sup>3</sup>) were soaked with 100 µL Cry8Ba2 crystal samples (~ 1 mg/mL). As a negative control, sweet potato or courgette were soaked with 100 µL Cry48Aa1/Tpp49Aa1 (50 µL each of ~ 1 mg/mL). Cry48Aa1/Tpp49Aa1 is a mosquitocidal toxin with no known coleopteran activity. The larvae and toxin soaked food were placed in a petri dish and incubated in the dark for 14 days to record mortality daily. Three replicates were performed, and the diet was replenished every two days.

### 2.5.6. Alpha-glucosidase assays

Alpha-glucosidase activity assays were performed using the Alpha-Glucosidase Activity Assay Kit (Abcam, ab174093) according to the manufacturer's instructions. Briefly,  $\alpha$ -glucosidase protein was mixed with the substrate (*p*-nitrophenol- $\alpha$ -D-glucopyranoside) to release *p*-nitrophenol, which can be measured colorimetrically at 410 nm. Here, 10  $\mu$ L of a 1:10 dilution of Cqm1 protein sample (~ 500  $\mu$ g/mL – estimated by BCA assay) were plated in triplicate into a 96-well microplate and made up to 50  $\mu$ L with  $\alpha$ -glucosidase assay buffer. An  $\alpha$ -glucosidase positive control was also plated in triplicate. The  $\alpha$ -glucosidase substrate mix (3  $\mu$ L) was added to 47  $\mu$ L  $\alpha$ -glucosidase assay buffer and this reaction mix was subsequently added to each well of the 96-well microplate containing the Cqm1 protein sample or positive control. Absorbance was measured immediately at 410 nm using a microplate reader (CLARIOstar, BMG LABTECH, UK) in kinetic mode for 60 minutes. The average absorbance was plotted against time using GraphPad Prism (version 10.1.0).

### 2.5.7. Static light scattering (RALS) and refractive index (RI) measurements

RALS and RI measurements were performed in collaboration with Dr. Hannah Best (Cardiff University, Wales) to identify the oligomeric state of Tpp49Aa1 in solution. Crystalline protein samples (100  $\mu$ L at 1 mg/mL) were prepared as described in **section 2.3.4** and solubilised in 50 mM Na<sub>2</sub>CO<sub>3</sub> pH 10.5 + 0.05%  $\beta$ -mercaptoethanol overnight at room temperature, with shaking. The soluble sample was obtained by centrifugation and buffer exchanged into 20 mM TrisHCl pH 8.5. SEC was performed as described in **section 2.5.1**. Protein-containing fractions were identified using SDS-PAGE, pooled and concentrated. For subsequent RALS and RI measurements, the sample was subjected to a further round of SEC using a Superdex 75 Increase 10/300 GL column with a flow rate of 0.8 mL/min. The column was attached to a Zetasizer MicroV system (Malvern Instruments Ltd., Malvern, UK). Due to Brownian motion of the protein in solution, the size distribution

of Tpp49Aa1 could be quantified based upon time-dependent fluctuations in the scattered light intensity. In addition, protein concentration was determined by a RI detector (VE 3580, Viscotek Corp). Eluted samples were measured every 3 secs at 30°C. OmniSEC software (v.5.12) was used to analyse data calibrated to BSA (1 mg/mL).

### **2.5.8. Transmission electron microscopy (TEM) analysis of crystal proteins**

TEM analysis (JEM 2100-Plus, JEOL) of crystal proteins was performed by Dr. Robin Schubert from the XBI laboratory at the EuXFEL facility. Purified crystal protein preparations (2  $\mu$ L) were applied to glow discharged (GloQube Plus, Quorum Technologies) holey carbon copper grids (Quantifoil R1.2/1.3) and incubated for 30 seconds before blotting with filter paper (Whatman #1). For negative staining, grids were placed on a droplet containing 2% (w/v) uranyl acetate and blotted immediately. This step was repeated, and grids were left to dry on filter paper before TEM imaging at 200 kV acceleration voltage using an Emsis Xarosa camera.

### **2.5.9. N-terminal sequencing**

For N-terminal sequencing, crystal proteins or trypsin-treated soluble protein were run on SDS-PAGE. Following SDS-PAGE, gels were transferred to CAPS buffer (10 mM 3-(cyclohexylamino)-1-propanesulfonic acid, 10% methanol, pH 11.0) and assembled alongside 3MM paper and polyvinylidene fluoride (PVDF) transfer membrane (Thermo Fisher Scientific – activated with methanol before use) to produce the membrane sandwich. Semi-dry transfers were run at 25 V, 1.0 A for 30 min using the Bio-Rad Trans-Blot Turbo Transfer system. Subsequently, PVDF membranes were stained with Coomassie Blue for 30 minutes and destained (10% acetic acid, 40% methanol, 50% distilled H<sub>2</sub>O) overnight. The PVDF membrane slice containing the desired protein band was extracted and sent to Alta Bioscience Limited (Redditch, UK) for N-terminal sequencing.

## **2.6. Computational modelling and analysis**

Computational modelling studies were performed to predict the structures of the Cry48Aa1-Tpp49Aa1 and Tpp2Aa2-Cqm1 complexes. A general overview of the methods used to perform computational modelling studies has been provided below, with specific details included in the relevant results chapters.

### **2.6.1. Preparation of starting structures**

For modelling of the Tpp2Aa3-Cqm1 complex, the crystal structures of trypsin-treated Tpp2Aa3 (PDB 3WA1) and monomeric Cqm1 (PDB 6K5P) were used. For modelling of the Cry48Aa1-Tpp49Aa1 complex, the crystal structure of Tpp49Aa1 elucidated as part of this work (PDB 8BEY) was used. A model of the full-length Cry48Aa1 structure was predicted using the AF2 program (Jumper et al. 2021b) installed on the DESY Maxwell-cluster. Starting structures were prepared using the PyMOL Molecular Graphics System (version 2.5.3).

### **2.6.2. Molecular docking**

Molecular docking was performed using ClusPro and RosettaDock. Both docking programs score well in CAPRI, a communitywide experiment on the comparative evaluation of docking programs for structure prediction of protein-protein complexes (Lensink et al. 2018; Lensink et al. 2019). To prevent any bias from the outset of these studies, a naïve docking approach was used, meaning that the results of mutagenesis studies available in the literature were utilised to assess the output of docking (and more specifically, identify the most likely model), rather than direct it.

#### **2.6.2.1. ClusPro**

First, the ClusPro online server (available at: <https://cluspro.bu.edu/>) was used to perform a global docking search (Kozakov et al. 2017). ClusPro

utilises a direct docking approach based on the laws of thermodynamics, which states that the native (most thermodynamically stable) structure is located at the global energy minimum (Anfinsen 1973). ClusPro employs a three-step protocol which begins with the sampling of billions of possible conformations produced by rigid body docking using a Fast Fourier Transform (FFT) correlation algorithm available in the PIPER program (Kozakov et al. 2006). The 1000 lowest energy structures are then clustered according to RMSD. These clusters are ranked according to size (number of conformations present in those clusters), with the largest clusters ranking highest. Finally, an energy minimisation step is employed to refine these structures. ClusPro outputs the central model (according to the energy scores) of the 30 most populated clusters. As part of this work, docking in ClusPro was performed using standard settings and the central models from the 5 most populated clusters were carried forward for local docking refinement.

#### **2.6.2.2. RosettaDock**

Local docking refinement was performed on the command-line interface version of RosettaDock (version 3.2, available at: <https://www.rosettacommons.org/>) installed locally. RosettaDock (Chaudhury et al. 2011) employs a Monte Carlo (MC) based algorithm across a two-step protocol involving low-resolution and high-resolution phases. At the low-resolution phase, sidechains are represented as coarse-grained 'pseudo-atoms' and the ligand is translated and rotated from the starting pose with standard deviations of 3 Å and 8 Å, respectively. The lowest energy complex generated enters the high-resolution phase, which involves sidechain packing and energy minimisation. Since packing of sidechains is only performed for interfacial residues, a prepacking step prior to docking is required. To further improve sidechain modelling, unbound rotamer conformations were provided.

For local docking, it is recommended to generate 1,000 decoys (output models). Output models are scored using the Rosetta *ref2015*

function (Alford et al. 2017), which considers several weighted energy terms to calculate a final total energy. Since this calculation involves both physical-based and statistical-based potentials, Rosetta energies exist on an arbitrary scale, referred to as Rosetta Energy Units (REU).

According to the thermodynamic hypothesis of protein folding, the native protein state occurs at the global energy minimum of the free energy landscape (Anfinsen 1973). Therefore, the model with the lowest energy score from each local docking search, assumed to be the most likely native state, was carried forward for further structural analysis, including interface analysis and stability analysis using MD simulations.

#### **2.6.2.3. AlphaFold-Multimer**

In parallel to the above studies, molecular docking was performed using AlphaFold-Multimer (Evans et al. 2022), based on the AF2 program (Jumper et al. 2021b). Protein-protein complexes were predicted by providing the amino acid sequence and stoichiometry of binding on the command-line interface version of AlphaFold-Multimer installed on the Supercomputing Wales high-performance cluster, Hawk.

#### **2.6.2.4. ColabFold**

In addition, molecular docking was performed using ColabFold (Mirdita et al. 2022), a program which combines multiple sequence alignments with AF2 to predict 3D protein structures. Again, protein-protein complexes were predicted by providing the amino acid sequence and stoichiometry of binding on the command-line interface version of ColabFold installed on the Supercomputing Wales high-performance cluster, Hawk.

#### **2.6.3. Interface analysis**

The interfaces of modelled complexes were analysed using the PISA (Krissinel 2015). PISA identifies real and crystallographic interfaces. Across

these interfaces, PISA calculates the interface area ( $\text{\AA}^2$ ), number of residues from each molecule involved, as well as the number of interfacial interactions (hydrogen bonds, salt bridges, disulphide bridges). In addition, PISA predicts the solvation and binding energies ( $\text{kcal mol}^{-1}$ ) of that interface.

## **2.6.4. Molecular Dynamics (MD) simulations**

### **2.6.4.1. GROMACS software**

MD simulations were performed using the open-source GROMACS (*GROningen MAachine for Chemical Solutions*) software (Abraham et al. 2015). All stages of the MD simulation, including energy minimisation, two-step equilibration, and the final MD run (discussed in detail below), were run on GROMACS (version 2020.1) installed on the Supercomputing Wales high-performance cluster, Hawk. Simulations were controlled by molecular dynamics parameters (.mdp) files. Each MD simulation was repeated three times.

### **2.6.4.2. Forcefields**

Forcefields are composed of a set of equations and constants that describe the potential energy ( $E_{\text{pot}}$ ) function. The energy function can be broken down into bonded (bond stretching, bond angles, bond torsion) and non-bonded (electrostatic and van der Waals) terms. The parameters that make up these terms are typically obtained (and constantly developed) from experimental data (e.g., nuclear magnetic resonance, infrared, raman, and neutron spectroscopy, and neutron, X-ray, and electron diffraction) or quantum mechanical calculations. As new data becomes available, forcefields are developed to improve the accuracy in which they replicate the natural behaviour of atoms and molecules. The AMBER (Assisted Model Building with Energy Refinement) forcefield (Lindorff-Larsen et al. 2010) was selected for simulations of Tpp2Aa3-Cqm1 models. Subsequently, the CHARMM (Chemistry at HARvard Molecular Mechanics) forcefield (Foloppe

and MacKerell 2000) was chosen for simulations of Cry48Aa1-Tpp49Aa1 models due to its compatibility with performing simulations at varied pH.

#### **2.6.4.3. Input structure preparation**

Starting structural coordinates for MD simulations were derived from the output models produced in molecular docking studies (**section 2.6.2**). Where MD simulations were performed at varied pH, the protonation states of titratable residues were first assigned by PROPKA in the PDB2PQR server using the PARSE forcefield and CHARMM output naming system (Dolinsky et al. 2007).

Using the pdb2gmx module available in GROMACS, the input pdb file was converted to a GROMACS-formatted structure file, Gromos87 format (.gro) file. Hydrogens were added where required. As default, pdb2gmx selects the protonated forms of lysine and arginine and the deprotonated forms of aspartic acid and glutamic acid, whilst optimising the protonation state of histidine based on optimal hydrogen bonding conformation, which itself is based upon maximum donor-acceptor distances and angles. These protonation states reflect the charges that are likely to occur at neutral pH. For simulations modelled at varied pH, the PROPKA-assigned protonation states were interactively selected. The pdb2gmx module allows interactive selection of the titration states of lysine, arginine, aspartic acid, and glutamic acid.

Following generation of the .gro file, the models were input to a cubic box, with a minimum distance of 1.0 nm between the model and box edge. Using the solvate command and the 3-point solvent (spc216.gro) model, the box was filled with water. The net charge of the system was calculated and neutralised via the addition of sodium ions.

#### **2.6.4.4. Energy minimisation**

Energy minimisation enabled the structure to be relaxed, ensuring correct geometry and removal of steric clashes (e.g., between hydrogens added in the previous step) within the system. Here, the  $E_{\text{pot}}$  of the system was lowered by applying the steepest descent algorithm using a maximum number of 50,000 steps in 0.01 step sizes, and a maximum force (Fmax) of 1000 kJ mol<sup>-1</sup> nm<sup>-1</sup>. A successful energy minimisation was identified by a negative  $E_{\text{pot}}$  with an order of 10<sup>5</sup> / 10<sup>6</sup>. In addition, the maximum force (Fmax) was expected to be no greater than the target value indicated above. Satisfaction of these variables indicated a stable system.

#### **2.6.4.5. Two-step equilibration**

Following energy minimisation, the temperature and pressure of the system were equilibrated. To do so, short (100 ps) simulations were performed. These simulations were performed prior to the final MD run and were disregarded from any analysis. Inclusion of the -DPOSRES flag instructs GROMACS to use position restraints during the simulation. This allowed the solvent around the molecule to be equilibrated without any structural changes to the model. To stabilize the temperature of the system, an isothermal-isochoric ensemble was applied, without pressure-coupling. To stabilize the pressure and density of the system, an isothermal-isobaric ensemble was applied. Pressure-coupling was performed using a Parrinello-Rahman barostat. Both the isothermal-isochoric and isothermal-isobaric equilibration stages were performed for 100 ps with 2 fs time steps for integration.

#### **2.6.4.6. MD production run**

Following energy minimisation and system equilibration, the final 100 ns MD simulation (MD production run) was performed. Here, a timestep of 2 fs was applied to integrate Newton's equations of motion. To calculate non-

bonded long-range electrostatic interactions, the Particle-Mesh Ewald method using a 10 Å cut-off (AMBER99SB) or 12 Å cut-off (CHARMM 27) was applied. Van der Waals interactions were limited to a 10 Å cut-off (AMBER99SB) or 12 Å cut-off (CHARMM 27), whilst the LINCS constraint algorithm was applied to constrain all bonds involving hydrogen atoms. Temperature coupling was performed using a velocity rescaling thermostat with a relaxation constant of 0.1 ps. Proteins and non-proteins were coupled separately, and a reference temperature of 300 K was applied. Pressure coupling was performed using a Parrinello-Rahman barostat with an isothermal compressibility of  $4.5 \times 10^{-5} \text{ bar}^{-1}$ . The pressure was coupled isotropically using a coupling constant of 2.0 ps with a reference pressure of 1.0 bar.

#### **2.6.4.7. MD analysis**

Following completion of the MD simulation, the trajectory file was extracted and converted using the `gmx_trjconv` module to centre the mass of molecules in the box and remove any jumps caused by the periodic boundary conditions. The resulting trajectory files were visualised using the Visual Molecular Dynamics (VMD, version 1.9.4) program (Humphrey et al. 1996). Root-mean-square deviation (RMSD) and radius of gyration (Rg) analyses were performed as detailed below. GraphPad Prism (version 10.1.0) was used to produce the graphics associated with this work.

#### **2.6.4.8. Root-mean-square deviation (RMSD) analysis**

Analysis of the RMSD of backbone atoms was applied to assess the overall stability of the modelled protein-protein complexes. For each simulation repeat, RMSD analysis was performed using the GROMACS `gmx_rms` module and the starting coordinates as the reference structure. The extracted RMSD values were plotted against time.

#### **2.6.4.9. Radius of gyration (Rg) analysis**

Rg analyses the mass and position of atoms in relation to the centre of mass of the molecule and, thus, indicates the overall compactness of the structure. For each simulation repeat, the Rg of protein atoms was analysed using the GROMACS `gmx_gyrate` module and the starting coordinates as the reference structure. The extracted Rg values were plotted against time.

### **2.7. Serial femtosecond crystallography (SFX)**

Megahertz (MHz) SFX diffraction data (Wiedorn et al. 2018a; Yefanov et al. 2019) were collected for the Cry8Ba2 and Tpp49Aa1 crystal proteins in collaboration with researchers from DESY and beamline scientists at the EuXFEL facility on the SPB/SFX instrument (Mancuso et al. 2019). Sample preparation, diffraction data analysis, and structural analysis were performed by me. Diffraction data collection and processing were performed by researchers from DESY and the EuXFEL. A general overview of the methods used to perform SFX studies has been provided below, with specific details included in the relevant results chapters.

#### **2.7.1. Sample preparation**

Protein nanocrystals were prepared as described in **section 2.3.4**. To remove any large contaminating material, including remaining cellular fragments, the protein nanocrystals were filtered and washed in dH<sub>2</sub>O using nylon mesh filters (PluriSelect) ranging from 100 µm down to 5 µm mesh size. To do so, crystal suspensions were resuspended in dH<sub>2</sub>O and centrifuged at 200 x g for one minute and the resulting supernatant was filtered. This process was repeated three times, using filters with progressively smaller mesh size. Following filtration, the crystal suspensions were transferred to high-pressure sample reservoirs.

### 2.7.2. Diffraction data collection and processing

A fast liquid-jet based injection (Wiedorn et al. 2018b) with 3D-printed (Knoška et al. 2020) double flow focused nozzles (DFFN) (Oberthuer et al. 2017) was used to inject crystals into the beam. With this configuration, the AGIPD Detector was able to record 202 pictures per X-ray pulse train (with a repetition rate of 10 trains/s) at an intra-train pulse rate of 0.564 MHz. A total of 9.3 keV photons with an average energy of 4 mJ per pulse were delivered to the instrument and focused to a spot size of about 300 nm diameter using the nanoscale-focusing KB optics (Bean et al. 2016), providing the sample with approximately  $6 \times 10^{12}$  photons/ $\mu\text{m}^2$ /pulse. The online crystal diffraction 'hit-rate' was tracked using the OnDA program (Mariani et al. 2016), with the raw data being processed as previously described (Wiedorn et al. 2018a). Hit-finding was carried out using the program Cheetah (Barty et al. 2014). The peak search parameters were optimised, enabling masking of bad pixels. Using CrystFEL (version 0.10.1) (White et al. 2012; White et al. 2016) and the XGandalf (Gevorkov et al. 2019) indexing method, significant diffraction patterns were then indexed. Detector geometry was optimised using the program Geoptimiser (Yefanov et al. 2015), whilst the Partialator program from the same CrystFEL package was used to merge and scale the integrated reflection intensities.

### 2.7.3. Diffraction data analysis

Matthews' analysis, available via the MATTPROB (Kantardjieff and Rupp 2003) program with the CCP4i2 software suite, was used to estimate the solvent content and number of molecules in the asymmetric unit. Initial phasing was performed by molecular replacement using Phaser (McCoy et al. 2007) in CCP4i2 (Potterton et al. 2018), with the sequence data and an estimated component stoichiometry provided as input. MR requires a structurally similar model usually with a sequence identity of >25%. Blast searches were performed to identify homologous protein structures used as models for molecular replacement. Model building was performed using a combination of automated and manual model building programs available in

the CCP4i2 software suite. Coot (Emsley et al. 2010) was used to inspect the final model and maps manually, whilst Refmac5 (Murshudov et al. 2011) was used for the final refinement.

#### **2.7.4. pH mixing studies**

To probe the early structural changes which occur upon exposure to extreme pH, mixing studies were performed. For these mixing studies, 0.1 M sodium citrate (pH 3.0) and 0.1 M sodium carbonate (pH 11.0) buffers were transferred to the high-pressure sample reservoirs for injection. The protein crystals were mixed with the pH 3 or pH 11 buffers approximately 1 minute before probing with X-rays. Diffraction data collection and analysis were performed as described above (**sections 2.7.2 and 2.7.3**). To identify changes in interactions at the crystallographic interfaces, the structures were compared using PISA (Krissinel 2015).

### **2.8. Single-particle cryogenic electron-microscopy (cryoEM)**

Single-particle cryoEM analysis of the Tpp2Aa2-Cqm1 complex was performed in collaboration with researchers from the Biomedicine Discovery Institute at Monash University. Sample preparation, data collection and model building were carried out by Dr. Bradley Spicer, Hari Venugopal, and Dr. Christopher Lupton. Structural analysis was carried out by me.

#### **2.8.1. Sample preparation**

To prepare Tpp2Aa2-Cqm1 samples for cryoEM and ensure the highest amount of occupancy, Tpp2Aa2 and Cqm1 were incubated at a 1.5:1 (Tpp2:Cqm1) in 50 mM Tris, 150 mM NaCl pH 8.0 and the complex was subsequently purified by SEC using a Superose® 6 10 300 increase column. Copper R1.2/1.3 Quantifoil Holey Carbon Grids (Quantifoil Micro Tools GmbH) were glow-discharged at ambient atmospheric conditions (38 mbar) for 30 seconds prior to blotting. Purified Tpp2Aa2-Cqm1 complexes were

concentrated at 2.3 mg/mL and applied to the glow-discharged grids prior to vitrification using the Vitrobot™ Mark IV (FEI). Briefly, vitrification involved blotting the sample at specified a specified blot force for a certain length of time followed by a drain time (e.g., blot force -3, for 4 sec and 1 sec drain). Following this, the grids were plunged into liquid ethane that had been cooled using liquid nitrogen. The grids were stored in liquid nitrogen prior to sample screening and data collection.

### **2.8.2. Sample screening and data collection**

Grids were prepared in duplicate to be used individually for sample screening and data collection. Sample screening using a Tecnai G2 T20 Electron Twin Transmission Electron Microscope (TEM, FEI) enabled protein conditions to be optimised for data collection. Subsequently, data collection was performed using a Double-corrected Arctica 3 80-300 Field Emission Gun Transmission Electron Microscope (FEGTEM, FEI) or the Krios Titan equipped with a K3 detector. The optimised collection included 5040 movies with a pixel size of 0.82 Å an acceleration of 300 kV, and a total dose of 60 e/Å<sup>2</sup>. The resulting raw movies were stored for single-particle analysis.

### **2.8.3. Data processing and model building**

CryoSPARC (version 4.1.2) available on Monash University's high-performance cluster, MASSIVE, was used for single-particle cryoEM analysis. Briefly, a total of 5040 movies were motion corrected using full frame motion correction. Contrast transfer function (CTF) estimation was performed using CTFFIND4 (Rohou and Grigorieff 2015), followed by particle picking with maximum and minimum diameters of 45 Å and 130 Å. A box size of 256 pixels was used for particle extraction with particles subsequently classified into 50 2D classes. *Ab initio* reconstruction generated 3 distinct 3D classes. The class with the highest number of particles was used for subsequent classification jobs to obtain the most uniform subset of particles (i.e., 3D classification). The best subset of particles was then used to refine

a high-resolution structure using homogenous refinement and C2 symmetry. It was observed that the distal region of Tpp2Aa2 was of poorer quality and was, therefore, subjected to local refinement experiments using a mask to subtract the Cqm1 region from the particles (i.e., particle subtraction). Following this, the subtracted particles were refined using localised refinement with a separate mask specific to Tpp2Aa2. The final maps were sharpened using EM Ready (He et al. 2023), which uses machine learning to apply B-factors across the map to sharpen the quality. The atomic models for Tpp2Aa2 and Cqm1 were derived from PDB 5FOY (chain B) and PDB 6K5P, respectively, and manually adjusted in Coot (Emsley et al. 2010). After manual positioning, the model was refined using Isolde (Croll 2018) in ChimeraX (Meng et al. 2023), which was used to simulate areas in the model that had poor Ramachandran geometry. For real-space refinement, Refmac5 (Murshudov et al. 2011) available in the Phenix software suite (Adams et al. 2010) was employed leading to a model with final global resolution of 2.42 Å.

### 3. Structural analysis of the Cry8Ba2 pesticidal protein

#### 3.1. Introduction

The largest class of bacterial pesticidal proteins is the Cry protein family. Cry proteins contain a conserved three-domain core (DI, DII, DIII) and are produced as either ~65 kDa or ~130 kDa pro-proteins, the latter of which exhibit an extended C-terminus containing four pro-domains (DIV, DV, DVI, DVII) cleaved during proteolytic activation (Evdokimov et al. 2014). Cry proteins display activity against several insect orders, including Lepidoptera and Coleoptera, as well as nematodes, and hence, are of great commercial importance. Several crystal structures of solubilised and activated Cry proteins have been elucidated using conventional X-ray crystallography and include Cry1Aa (PDB 1CIY - Grochulski et al. 1995), Cry1Ac (PDB 4ARX - Derbyshire et al. 2001), Cry2Aa (PDB 1I5P - Morse et al. 2001), Cry3Aa (PDB 6LFP - Heater et al. 2020), Cry3Bb1 (PDB 1JI6 - Galitsky et al. 2001), Cry4Aa (PDB 2C9K – Boonserm et al. 2006), Cry4Ba (PDB 1W99 - Boonserm et al. 2005), Cry5Ba1 (4D8M - Hui et al. 2012), Cry5Ba2 (PDB 8HHE – Li et al. 2022), Cry7Ca1 (PDB 5ZI1 - Jing et al. 2019), and Cry8Ea1 (PDB 3EB7 - Guo et al. 2009). More recently, SFX has been applied to elucidate the protoxin structures of Cry3Aa (PDB 4QX0 - Sawaya et al. 2014), Cry11Aa and Cry11Ba (PDB 7QX4 and 7QYD – Tetreau et al. 2021), all of which are examples of Cry proteins in the short (~65 kDa) form. To date, only one structure of a full-length ~130 kDa Cry protein has been reported and that is of a highly-mutated form of Cry1Ac, Cry1Ac- $\Delta$ 14C (PDB 4W8J - Evdokimov et al. 2014).

Proteins in the Cry8 subclass are specifically active against members of the Coleoptera insect order (Asano et al. 2003; Shu et al. 2007; Yamaguchi et al. 2008; Guo et al. 2009; Zhang et al. 2013; Park et al. 2014; Shu et al. 2020). One case of low toxicity against *Anticarsia gemmatalis*, a member of the Lepidoptera insect order, has also been reported (Amadio et al. 2013). One structure of a Cry8 family member exists and that is of chymotrypsin-activated Cry8Ea1, which demonstrates a 3-domain

architecture consistent with other Cry protein families (Guo et al. 2009). This chapter focuses on the production of Cry8Ba2 crystal proteins for subsequent structure solution using SFX at the EuXFEL. The Cry8Ba2 structure represents the first structure of a full-length 130 kDa Cry protein from natural crystals and sheds light on crystal packing. During indexing, it became clear that the native *B. thuringiensis* strain Na210, encoding the *cry8Ba2* gene, produced two forms of Cry8Ba2 crystals with different unit-cell dimensions and, therefore, datasets for each one were integrated and extracted separately to produce two datasets which could be analysed individually. Given that diffraction data were collected for both crystals in one data collection, this work highlights the applicability for SFX in the analysis of native *B. thuringiensis* / *L. sphaericus* strains that produce multiple crystal proteins. The asymmetric unit of both cell-types contained one molecule. Using PISA, different symmetry mate assemblies were identified, with a dimer interface being the most significant. The Cry8Ba2 structure was compared to the structures of related pesticidal proteins elucidated using conventional X-ray crystallography, including Cry1Ac- $\Delta$ 14C and the activated form of Cry8Ea1. Finally, insect bioassays were performed with the aim of identifying novel Cry8Ba2 targets. Throughout this chapter, procedures that were carried out by other researchers have been indented and italicised.

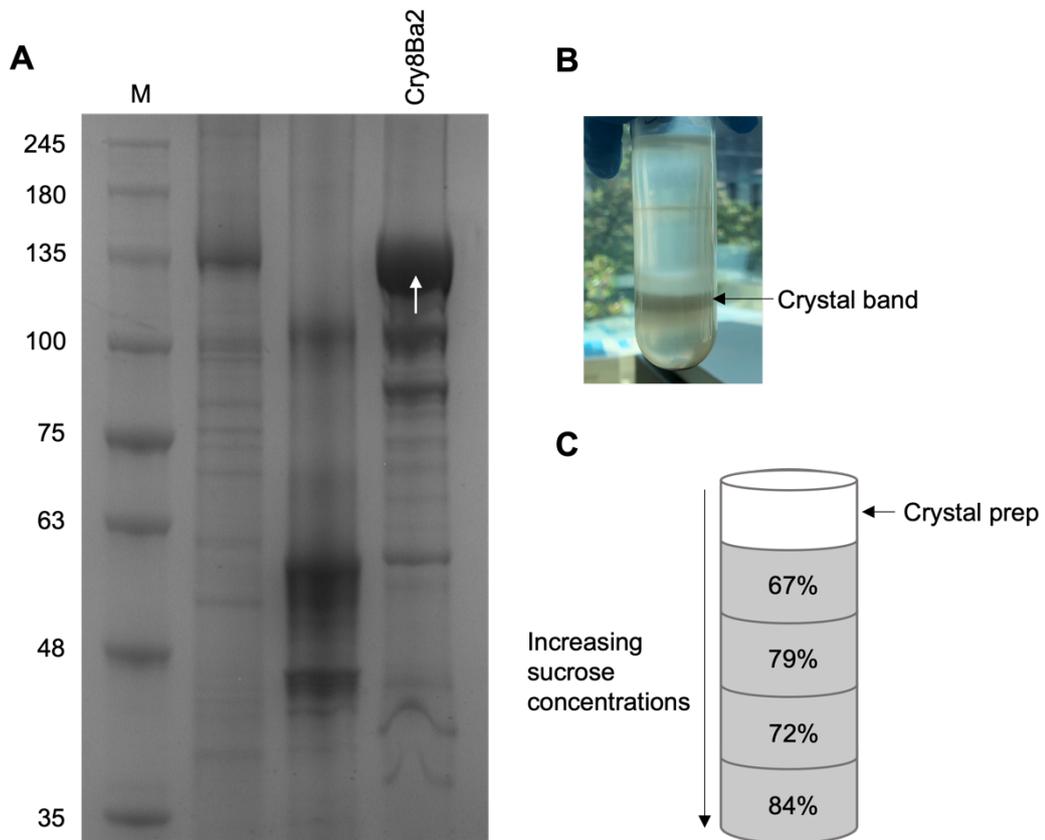
## **3.2. Results and Discussion**

### **3.2.1. Production of the Cry8Ba2 crystal protein**

#### **3.2.1.1. Crystal protein expression and isolation**

The native *B. thuringiensis* bacterial strain Na210 encoding the *cry8Ba2* gene was obtained from Prof. Primitivo Caballero (The Public University of Navarra, Pamplona, Spain). Whole genome sequencing had previously confirmed that Cry8Ba2 was the only crystal protein (**Appendix 1**) encoded by this strain. The *cry8Ba2* gene was expressed under the control of its own promoter during growth in Embrapa medium by incubating at 30 °C until sporulation reached 90% (typically between 48 - 72 h). Subsequently,

the lysed sporulated culture was separated on a discontinuous sucrose gradient. Crystal enriched bands were extracted and analysed by SDS-PAGE (Fig 3.1). SDS-PAGE analysis revealed the enrichment of an ~135 kDa protein.

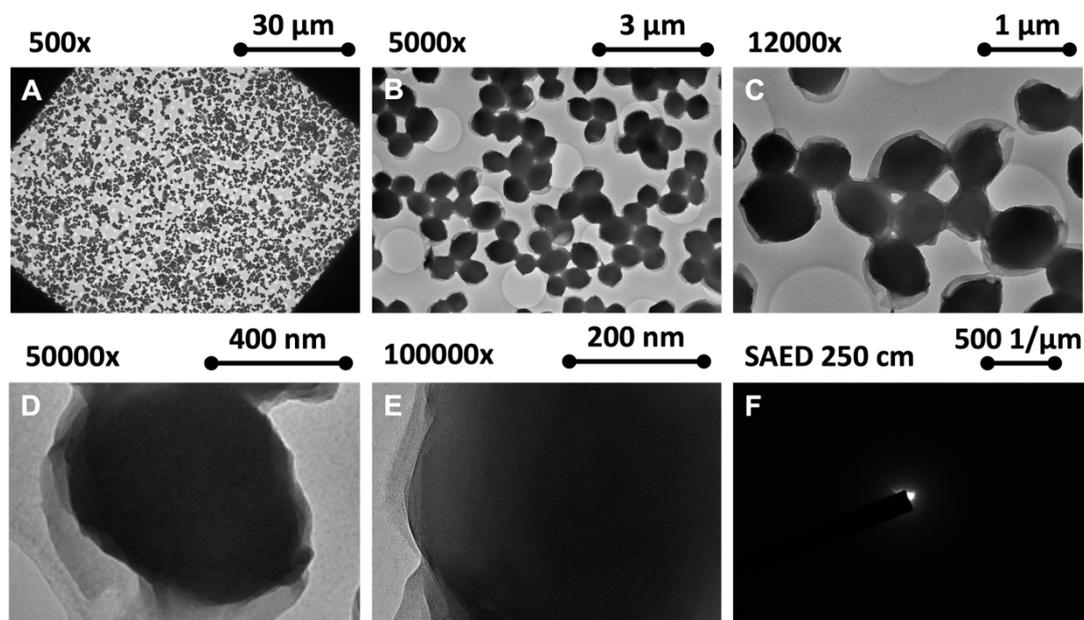


**Figure 3.1. Production of Cry8Ba2 nanocrystals.** (A) Coomassie stained SDS-PAGE showing Cry8Ba2 (lane 4, ~ 135 kDa) crystal samples. M (marker): BLUeye Pre-Stained Protein Ladder (kDa). (B) Sucrose density gradient of lysed sporulated culture showing example crystal band. (C) Diagram of sucrose gradient set up.

### 3.2.1.2. Transmission electron microscopy (TEM) of nanocrystals

In the past, well-ordered crystal lattices and the presence of electron diffraction has been used to identify diffraction-quality crystals (Stevenson et al. 2016). Here, TEM was employed to assess crystal quality prior to data collection. Since access to the SPB/SFX instrument at the EuXFEL was limited, such quality checks enabled efficient use of the beamtime allocation.

*TEM experiments were carried out by Dr. Robin Schubert in the XBI lab at the EuXFEL (Han et al. 2021). Briefly, samples were negatively stained using 2% (w/v) uranyl acetate and imaged using an Emsis Xarosa camera. TEM revealed high concentrations of spherical crystals with a diameter of approximately 500 nm (Fig 3.2). In addition, imaging in selected area electron diffraction (SAED) mode showed the presence of weak electron diffraction, indicated by the presence of streaks in Fig 3.2F. Taken together, TEM confirmed the presence of high concentrations of crystals which were suitable for SFX studies.*



**Figure 3.2. Transmission electron microscopy on negative stained Cry8Ba2 nanocrystals.** (A - E) Cry8Ba2 native nanocrystals (F) Weak electron diffraction was detected by selected area electron diffraction (SAED) imaging. Images obtained from Dr. Robin Schubert (EuXFEL – Schenefeld, Germany).

### 3.2.2. Serial femtosecond crystallography (SFX)

Despite the success of conventional X-ray crystallography, discussed in detail in **section 1.3.1.1**, significant challenges have remained surrounding radiation damage and the ability to obtain large, well-ordered crystals, often the rate-limiting step of crystallography (Holton and Frankel 2010). Third-generation synchrotrons, such as the Diamond Light Source (Harwell, UK), have enabled data collection from smaller crystals, but this often leads to increased absorbed X-ray dose and consequent radiation induced structural changes (Holton and Frankel 2010). In addition, the divergent X-ray beam from synchrotrons requires focussing, which can produce a smearing effect of the diffraction spots, particularly at high resolution. In the past decade, the invention of XFELs has been shown to mitigate this problem, enabling structures of crystal proteins to be determined at a resolution of around 2 Å or better (Chapman et al. 2011). XFELs produce intense pulses of parallel

light with femtosecond durations and, therefore, offer the ability to record a diffraction pattern from much smaller crystals (in the micrometre to nanometre size range), before radiation damage can occur (Chapman et al. 2007). SFX is an emerging crystallography technique that introduces a stream of crystals into an XFEL beam, enabling single pulse diffraction patterns to be collected from single crystals. Due to the high radiation density, crystals are destroyed with each pulse but, due to the femtosecond timescale, diffraction data are collected before radiation damage occurs. *B. thuringiensis* and *L. sphaericus* pesticidal proteins are naturally produced in high quantities as crystalline inclusions of nano – micro size range, thus making them highly applicable for structural analysis using SFX.

To investigate the structure of Cry8Ba2 in natural crystals, SFX at the EuXFEL was performed.

### **3.2.2.1. Diffraction data collection and processing**

MHz serial femtosecond crystallography (Wiedorn et al. 2018a; Yefanov et al. 2019) diffraction data were collected for Cry8Ba2 in one beamtime (p2442) on the SPB/SFX beamline at the EuXFEL using the novel nano-focus option to match the beam size to that of the crystals. Crystal samples were prepared by filtration and centrifugation before being transferred to high-pressure reservoirs that were subsequently loaded onto the fast liquid-jet based injection system (Wiedorn et al. 2018b) with 3D-printed (Knoška et al. 2020) DFFN (Oberthuer et al. 2017). Liquid jet-based injection systems enable a constant flow of crystals to be delivered to the XFEL beam. Use of the OnDA program (Mariani et al. 2016) enabled the crystal diffraction ‘hit-rate’ to be monitored in real-time. In SFX, a ‘hit’ denotes an image (frame) with detectable Bragg diffraction. The ‘hit-rate’ is, therefore, calculated by dividing the number of hit images by the total number of images collected. Real-time monitoring of the hit-rate informed data collection, enabling fast decision-making regarding experimental parameters and sample consumption.

*The early stages of data processing were performed by Dr. Marina Galchenkova and Dr. Oleksandr Yefanov (DESY – Hamburg, Germany) following the method described by Wiedorn et al. (Wiedorn et al. 2018a) and involved several stages, including data preparation, indexing and integration, resolution of indexing ambiguities, and finally, merging. The Cheetah program (Barty et al. 2014) was employed to perform hit finding. From this, the strongest patterns were indexed and integrated using the XGandalf (Gevorkov et al. 2019) indexing method in CrystFEL (White et al. 2012; White et al. 2016) version 0.10.1. Briefly, indexing assigns the 3D reflection indices ( $hkl$ ) to the peaks in the diffraction data, whilst integration refers to the measurement of the intensity of diffracted patterns. Here, diffraction patterns (**Fig 3.3**) were indexed in the space group  $p4_12_12$  of the tetragonal crystal system. The space group describes the symmetry properties of the crystal. The unit cell was inspected using the `cell_explorer` tool. The `cell_explorer` tool outputs a series of histograms for each of the six unit-cell parameters, comprised of cell edges ( $a$ ,  $b$ ,  $c$ ) and angles ( $\alpha$ ,  $\beta$ ,  $\gamma$ ), where sharp peaks coincide with the correct solution.*

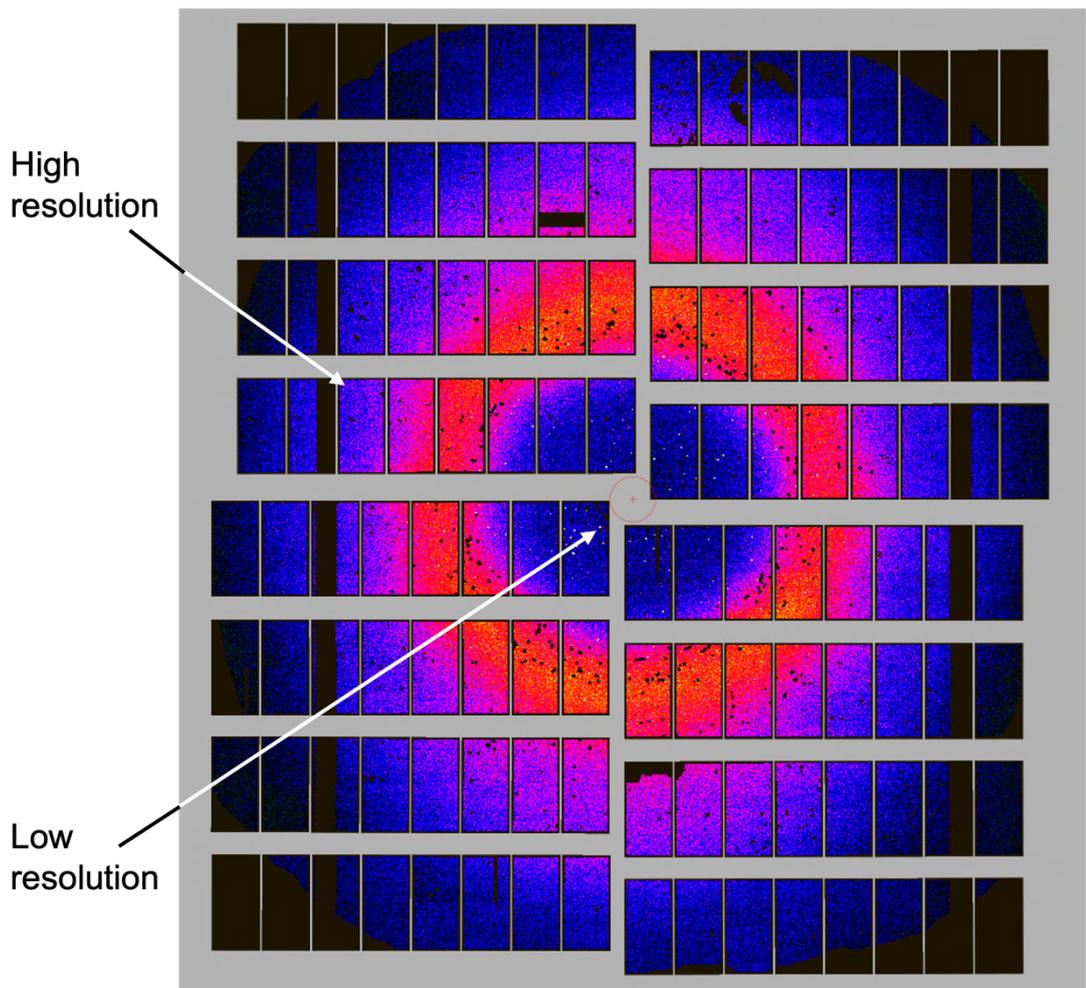
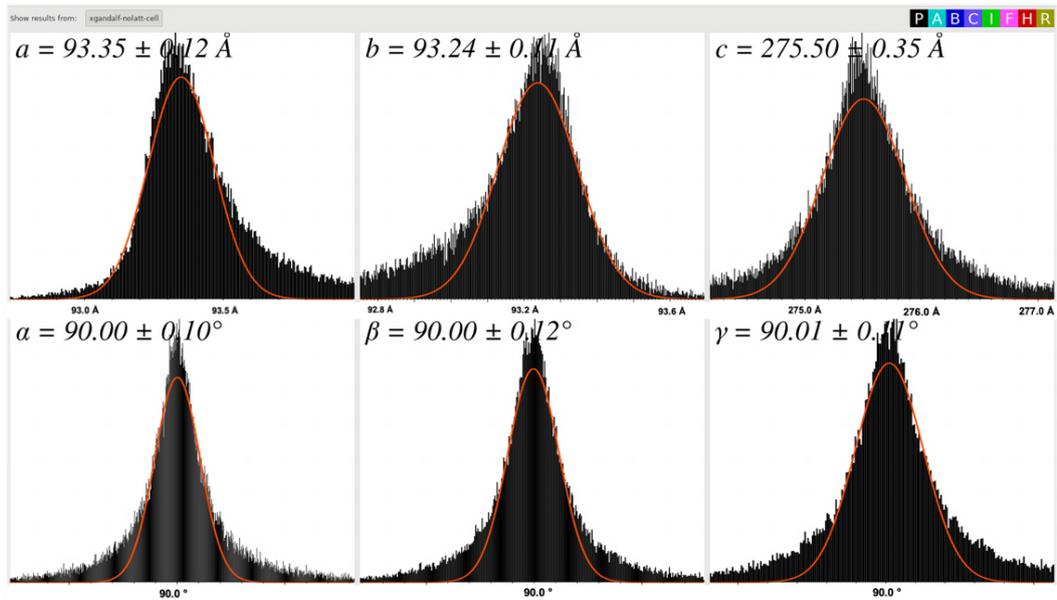
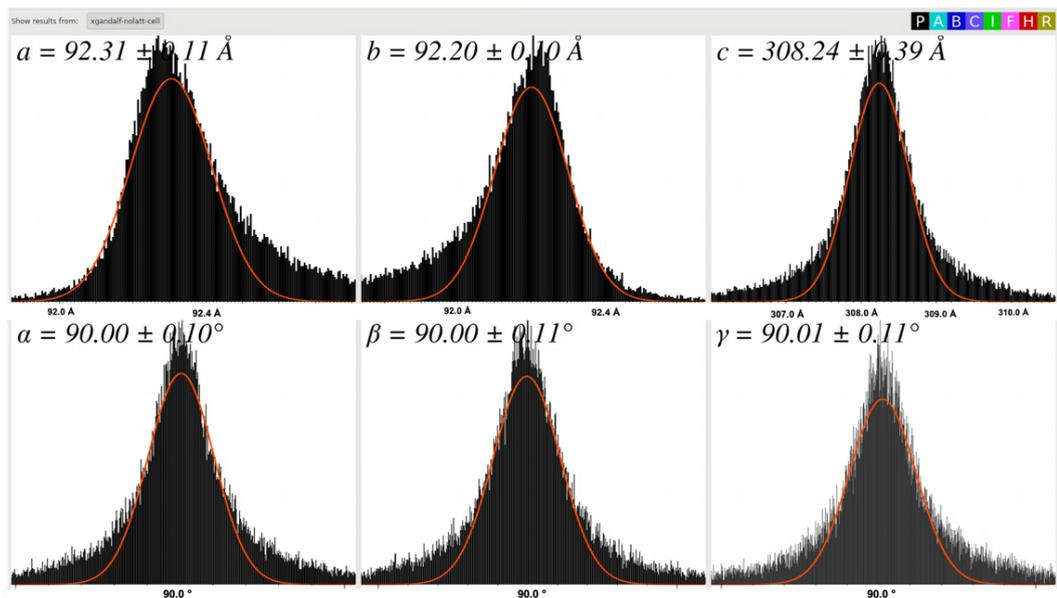


Figure 3.3. X-ray diffraction pattern (frame) of Cry8Ba2.

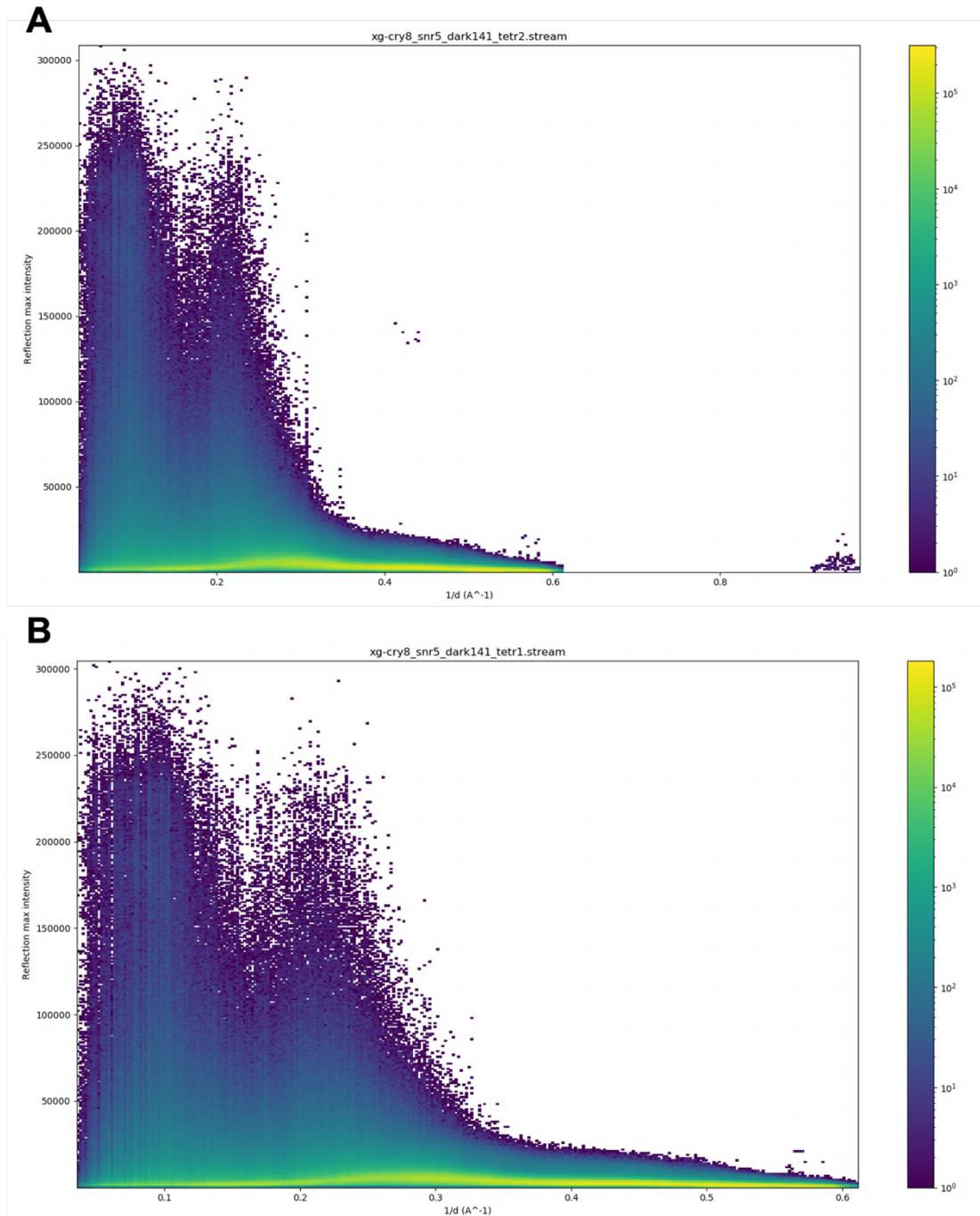
*Inspecting the unit cell parameters using the cell\_explorer tool, it became clear that there were two forms of Cry8Ba2 crystal proteins within the sample, depicted by the bimodal distribution present in the output histograms. Diffraction data for each crystal form were separated and extracted to produce two datasets (Fig 3.4) which could be analysed individually. Both crystals exhibited tetragonal symmetry and could be indexed in space group  $P4_12_12$ , but displayed differences in their unit cell dimensions, with the largest difference occurring in the length of the c-axis (the unique axis in tetragonal crystal systems). Hence, crystals and their resulting structures are referred to as the short (Cry8Ba2-S) and long (Cry8Ba2-L) forms.*

**A****B**

**Figure 3.4. Histograms output from cell-explorer tool.** The unit cell parameters ( $a$ ,  $b$ ,  $c$  and  $\alpha$ ,  $\beta$ ,  $\gamma$ ) for each frame (images with detectable Bragg diffraction) have been plotted. (A) Cry8Ba2-S (B) Cry8Ba2-L.

The strongest patterns were indexed and integrated using the XGandalf (Gevorkov et al. 2019) indexing method in CrystFEL (White et al. 2012; White et al. 2016) version 0.10.1. The unit cell dimensions of Cry8Ba2-S were defined as:  $a = 93.32 \text{ \AA}$ ,  $b = 93.32 \text{ \AA}$ ,  $c = 275.51 \text{ \AA}$  and  $\alpha = \beta = \gamma = 90.0^\circ$  (**Table 3.1**). In contrast, the unit cell dimensions of Cry8Ba2-L were defined as:  $a = 92.23 \text{ \AA}$ ,  $b = 92.23 \text{ \AA}$ ,  $c = 308.23 \text{ \AA}$  and  $\alpha = \beta = \gamma = 90.0^\circ$  (**Table 3.1**).

Following this, the Geoptimiser program (Yefanov et al. 2015) was employed to optimise detector geometry and indexing ambiguity was resolved using the Ambigator program. Prior to merging, the saturation value of the detector was identified using the peakogram-stream script, which outputs an intensity plot showing maximum intensity of pixels plotted against the resolution (**Fig 3.5**). Reflections which were too strong to be recorded by the detector appear as a cloud of points and can be excluded from the merging step. Finally, the indexed diffraction patterns were merged using the Partialator program in CrystFEL to produce a reflection (.hkl) file which was converted and exported as an MTZ file for data analysis and structure solution.



**Figure 3.5. Intensity plot output from CrystFEL peakogram-stream script.** The intensity of each reflection (diffraction spot) was plotted against the reciprocal resolution ( $1/d$ ). Reflections that were too strong to be recorded by the detector appear as a cloud of points at low resolution and were excluded from data merging. The graph is coloured according to the density of diffraction spots. (A) Cry8Ba2-S (B) Cry8Ba2-L.

**Table 3.1. Data collection and refinement statistics for Cry8Ba2.**

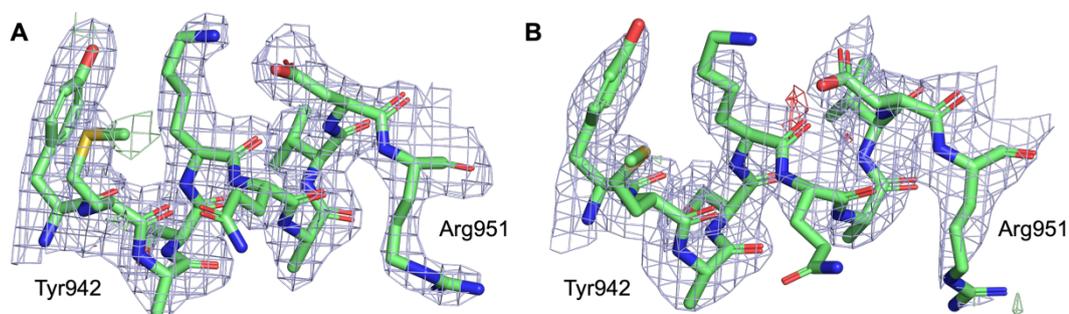
<b>Data Collection</b>		
PDB ID	To be deposited	To be deposited
Beamline	SPB/SFX at European XFEL	SPB/SFX at European XFEL
X-ray Energy (keV)	9.3	9.3
Wavelength (Å)	1.33	1.33
<b>Crystal Data (figures in brackets refer to outer resolution shell)</b>		
Crystal form	Cry8Ba2-S	Cry8Ba2-L
a,b,c (Å)	93.32, 93.32, 275.51	92.23, 92.23, 308.23
$\alpha,\beta,\gamma$ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Space group	P 4 <sub>1</sub> 2 <sub>1</sub> 2	P 4 <sub>1</sub> 2 <sub>1</sub> 2
Resolution (Å)	28.11 – 2.20	28.17 – 2.20
Outer shell	2.23 – 2.20	2.23 – 2.20
R-split (%)	14.54 (302.32)	15.20 (320.07)
CC*	0.997 (0.401)	0.996 (0.344)
I / $\sigma$ (I)	6.00 (0.74)	5.70 (0.19)
Completeness (%)	98.0 (71.61)	97.9 (68.81)
Multiplicity	226.21 (3.7)	196.1 (3.6)
Unique Reflections	62,899 (4075)	68,748 (4502)
Wilson B-factor(Å <sup>2</sup> )	27.4	28.4
<b>Refinement Statistics</b>		
Refined atoms	9358	9349
Protein atoms	9185	9176
Non-protein atoms	0	0
Water molecules	173	173
R-work reflections	59,670	65,214
R-free reflections	3140	3426
R-work/R-free (%)	18.3 / 24.4	18.6 / 23.2
<b>rms deviations (target in brackets)</b>		
Bond lengths (Å)	0.008 (0.011)	0.008 (0.011)
Bond Angles (°)	1.598 (1.670)	1.609 (1.670)
<sup>1</sup> Coordinate error (Å)	0.209	0.193
Mean B value (Å <sup>2</sup> )	53.165	51.965
<b>Ramachandran Statistics (PDB Validation)</b>		
Favoured/allowed/Outliers	1062 / 54 / 17	1076 / 47 / 10
%	93.7 / 4.8 / 1.5	95.0 / 4.1 / 0.9

<sup>1</sup> Coordinate estimated standard uncertainty in (Å), calculated based on maximum likelihood statistics.

### 3.2.2.2. Molecular replacement (MR) and structure solution

MR requires a structurally similar model, usually with a sequence identity of >25%. Here, a blast search showed 71% sequence identity with the full-length Cry8Ea1 protein, for which the structure of the activated form has been solved (PDB 3EB7 – Guo et al. 2009) and 40% sequence identity with unactivated Cry1Ac1, of which a structure of the mutated form, Cry1Ac- $\Delta$ 14C, has been solved (PDB 4W8J – Evdokimov et al. 2014). MR was performed using the phasing pipeline Phaser (McCoy et al. 2007) in the CCP4i2 software suite (Potterton et al. 2018) and Cry8Ea1 and Cry1Ac- $\Delta$ 14C broken down into their component domains as templates. Cry8Ea1 was used as the template for DI – DIII and Cry1Ac- $\Delta$ 14C was used as the template for DIV – DVII. Matthews' analysis using the MATTPROB web server (Kantardjieff and Rupp 2003) was employed to predict the solvent content and number of molecules in the asymmetric unit. As per this prediction, phasing was performed using a stoichiometry of one copy as input. The initial model was incomplete and, therefore, subjected to several rounds of automated model building in Buccaneer (Cowtan 2006). The output model from these steps showed that, rather than modelling a contiguous strand for DI – DVII as seen in the mutant Cry1Ac- $\Delta$ 14C structure, fragments from one monomer (DIV – DVII) had been assigned as belonging to another monomer (DI – DIII) and vice versa (due to non-crystallographic symmetry-related fragments). To correct this error, Coot (Emsley et al. 2010) was used to switch the fragments and the residues linking the DI - DIII and DIV - DVII fragments were built manually. This was followed by iterative rounds of refinement and manual model building cycles in Refmac5 (Murshudov et al. 2011) and Coot, respectively, until the R/Rfree values reached convergence. Cry8Ba2-S was solved first using this approach, producing a final model with R/Rfree 0.183/0.244 at a resolution of 2.2 Å (**Table 3.1**). Subsequently, Cry8Ba2-L was solved using Cry8Ba2-S as a template for phasing. A complete model was produced by molecular replacement and again, refined iteratively in Refmac5 and Coot, producing a final model with R/Rfree 0.186/0.232 at a resolution of 2.2 Å (**Table 3.1**). In both models, the electron density map (**Fig 3.6**) showed continuous density

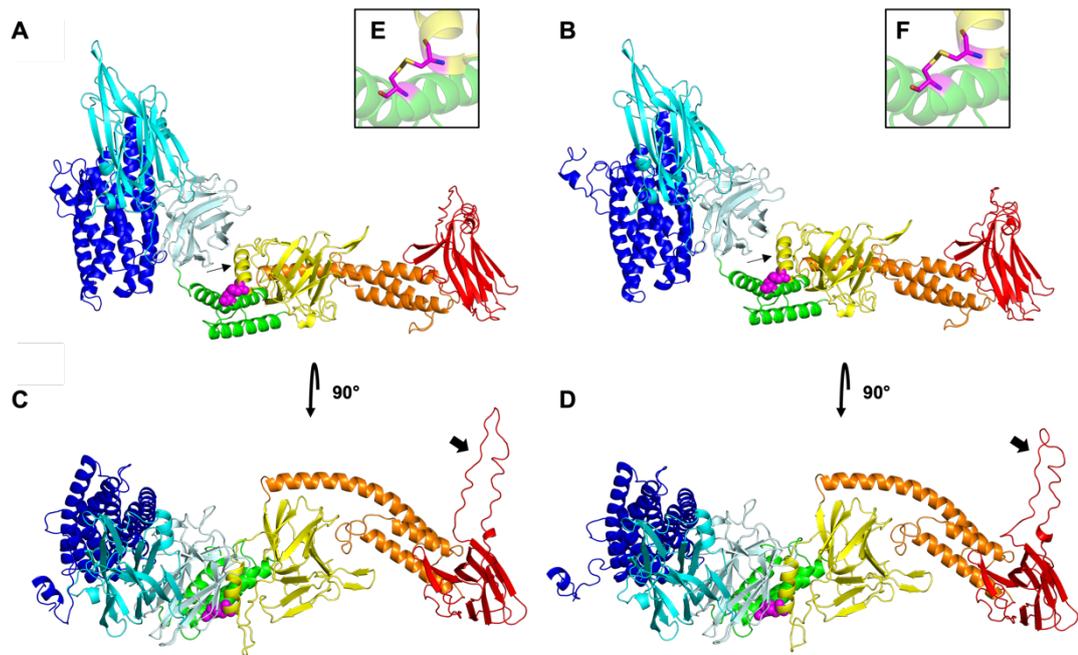
for residues 35 – 1169 of the protein sequence, with regions of weak density identified between residues 50 – 66 / 1107 – 1129 in Cry8Ba2-S and 51 – 63 / 1107 – 1130 in Cry8Ba2-L, respectively. Model building of these residues was guided by electron density contoured at the lower 0.5 rms level, ensuring that there were no clashes with atoms of neighbouring models and that correct coordinate geometry existed. The first 34 residues were not observed at all in the map. To confirm whether these residues were also present in the Cry8Ba2 crystals, we performed N-terminal sequencing as described in **section 2.5.9**. The sequence was identified as MSAGN, which is a possible match for the authentic N-terminal Cry8Ba2 sequence, MSPNN, and, given that no internal MSAGN sequence exists in the protein, suggests that the N-terminal region is present in the crystals. Given that this region is not seen in the electron density map, it is, therefore, apparently disordered. Indeed, disordered regions are often difficult to resolve using X-ray crystallography due to their flexible nature.



**Figure 3.6. Partial electron density map and model of the Cry8Ba2 structures.** 2Fo-Fc electron density map contoured at the 1.0  $\sigma$  level (coloured blue) and Fo-Fc electron density map contoured at the +/-3.0  $\sigma$  level (positive coloured green, negative coloured red). Protein backbone is coloured green, nitrogen atoms are coloured blue, oxygen atoms are coloured red. Residues Tyr942 – Arg951 are shown. (A) Cry8Ba2-S. (B) Cry8Ba2-L.

### 3.2.2.3. General features of the Cry8Ba2 structure

Structure solution revealed the presence of a linear string of seven distinct domains in the Cry8Ba2 monomer (**Fig 3.7**), which packed into a head-to-tail dimer with unit-cell repeats in the c-direction. In common with other Cry proteins, the toxin core contains three domains (DI – DIII). At the N-terminus, the core of DI, which is usually described as a seven-helix bundle (although in several cases, helix-2 is split into two separate helices), contains eight helices (**Fig 3.7**), where the usual helix 2 is split into helix 2a and 2b. Cry8Ba2 DI also reveals the presence of an extended N-terminal region, which, excluding a short helix, is entirely disordered. DII is composed of three antiparallel  $\beta$ -sheets and DIII is composed of a  $\beta$ -sandwich (**Fig 3.7**).

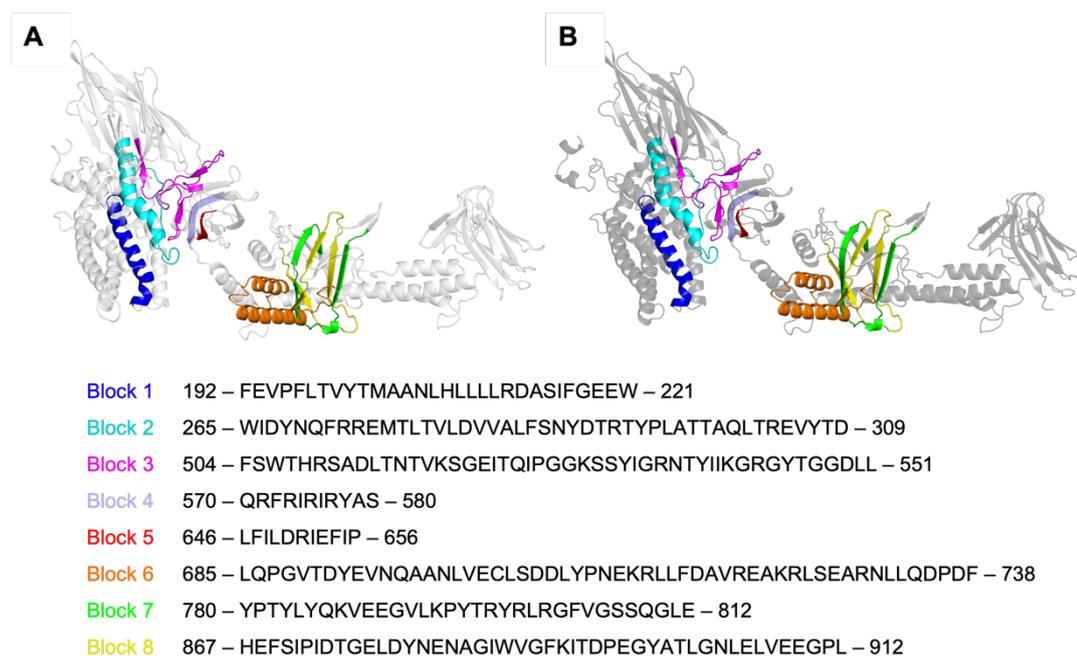


**Figure 3.7. Overall structure of the Cry8Ba2 monomers.** (A, C) Cry8Ba2-S. (B, D) Cry8Ba2-L. In common with other Cry proteins, the toxin core contains three domains (DI – DIII). The C-terminal half of Cry8Ba2 protoxin is composed of an additional four protoxin domains (DIV – DVII). DI = blue, DII = cyan, DIII = light cyan, DIV = green, DV = yellow, DVI = orange, DVII = red. DV contains a single helix (arrow), which interrupts the  $\beta$ -roll structure, and forms a disulphide bridge (Cys703-Cys847, shown as spheres - magenta) with a helix of DIV. (C, D) DVII contains an extended loop (arrow) which projects out perpendicularly from the Cry8Ba2 monomer. (E, F) Cys703-Cys847 shown as sticks and coloured by atom. Carbon = magenta, oxygen = red, nitrogen = blue, sulphur = yellow.

The C-terminal half of the Cry8Ba2 protoxin is composed of four protoxin domains (**Fig 3.7**, DIV – DVII). Like the Cry1Ac- $\Delta$ 14C protoxin structure (Evdokimov et al. 2014), DIV and DVI are composed of helical bundles, whilst the other two domains, DV and DVII, are composed of  $\beta$ -rolls (**Fig 3.7**). Cry8Ba2 DV also contains a single helix, which interrupts the  $\beta$ -roll structure, and forms a disulphide bridge (Cys703-Cys847) with a helix of DIV (**Fig 3.7**). In addition, DVII contains an extended loop (within a region of weak density between residues 1107 – 1129 and 1107 – 1130 for Cry8Ba2-S

and Cry8Ba2-L, respectively), which projects out perpendicular to the rest of the structure (**Fig 3.7**). The Cry8Ba2 DV helix and DVII loop correspond to regions which could not be solved in the Cry1Ac- $\Delta$ 14C due to a lack of electron density.

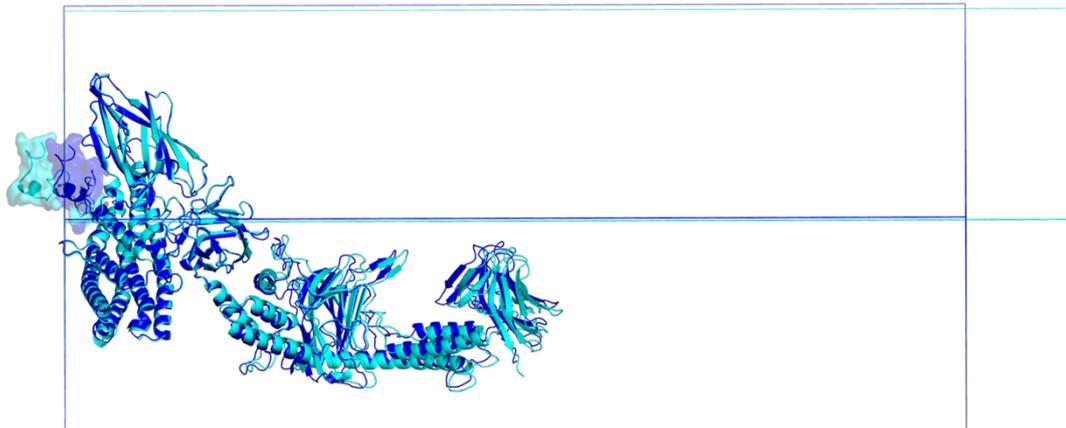
A total of eight blocks of highly conserved amino acids (**Fig 3.8**) have been identified in the Cry class, with blocks 1 – 5 occurring within the activated core (DI – DIII) and blocks 6 – 8 occurring within the extended C-terminal region (Hofte and Whiteley 1989; Schnepf et al. 1998). In an analysis of Cry1Aa and Cry3A, Schnepf et al (1998) found that block 1 encompasses DI helix 5 (implicated in pore-formation), block 2 encompasses DI helix 7 and the first  $\beta$ -strand of DII (the DI-DII contact region), whilst blocks 3, 4, and 5 encompass the  $\beta$ -strands of DII and DIII (Schnepf et al. 1998). Blocks 2 – 5 are thought to be involved in intramolecular domain contacts that, themselves, may be significant for maintaining the structural integrity of the toxin core and for domain movements required for pore-formation. The secondary structure and positioning of these features are nearly all conserved across blocks 1 – 5 in the Cry8Ba2 structure, excluding block 2, which encompasses helix 5 followed by an extended loop region (with no  $\beta$ -strand) linking DI and DII (**Fig 3.8**). Given that blocks 1 – 5 constitute structural features within the toxin core, they are likely significant for the insecticidal activity of the Cry class, as indicated above. Due to a lack of structural knowledge surrounding the C-terminal region, the structural features of blocks 6 – 8 have not been investigated. To identify the structural features of blocks 6 – 8, the conserved amino acid sequences were mapped onto the Cry8Ba2 structure (**Fig 3.8**). Here, we find that block 6 encompasses helix 2 and 3 of DIV, whilst block 7 and 8 encompass  $\beta$ -strands of DV (**Fig 3.8**). Interface analysis of the Cry8Ba2 dimer (discussed in detail in **section 3.2.2.4**) indicated that blocks 6 – 8 are either directly involved in, or reside near to the Cry8Ba2 dimer interface, suggesting that the conserved sequences in the pro-domain are significant for crystal packing.



**Figure 3.8. Conserved amino acid blocks in the Cry8Ba2 structure.** Eight blocks of highly conserved amino acids have been identified in sequence analysis of Cry proteins. Blocks 1 – 8 have been mapped onto (A) the Cry8Ba2-S (shown in light grey) and (B) the Cry8Ba2-L structure (shown in dark grey). The corresponding Cry8Ba2 sequences are also shown. Block 1 = blue, block 2 = cyan, block 3 = magenta, block 4 = light blue, block 5 = red, block 6 = orange, block 7 = green, block 8 = yellow.

During data processing, two forms of Cry8Ba2 crystals were identified, as mentioned above. The Cry8Ba2 crystals exhibited significant differences in their c-axis unit cell dimensions. Hence, the crystals were named the short (Cry8Ba2-S) and long (Cry8Ba2-L) forms (**Fig 3.9**). To investigate the structural differences between the two models, a structural alignment was performed (**Fig 3.9**). The Cry8Ba2 models display an all-atom RMSD of 2.980 Å and almost identical secondary structure features. Alignment of the N-terminal half (residues 35 – 654) produced an all-atom RMSD of 2.387 Å, whilst alignment of the C-terminal half (residues 655 – 1169) produced an all-atom RMSD of 1.495 Å, indicating that the largest structural differences occur in the N-terminal half. Within the structures, the placement of the extended N-terminal region (the peptide thought to be cleaved during

proteolytic activation) represents one of the key differences between the two Cry8Ba2 crystal forms (**Fig 3.9**). In the short model (Cry8Ba2-S), the N-terminal region (Asn35 – Ser49) is tucked into the rest of the molecule, whilst in the long model (Cry8Ba2-L), the N-terminal region (Asn35 – Ser49) is projected out by a hinge motion around region Ala50 to Ser65 and docks into a neighbouring molecule (**Fig 3.9**). The placement of the C-terminal half also differs between the two crystal forms. In Cry8Ba2-L, the C-terminal half appears to have hinged into a straighter conformation (**Fig 3.9**). The biological relevance of two distinct populations of Cry8Ba2 crystals will require further investigation, but it is possible that Cry8Ba2-L represents a more readily solubilised crystal form. Alternatively, the presence of two crystal forms may simply be caused by environmental factors and their impact on crystal production and growth.

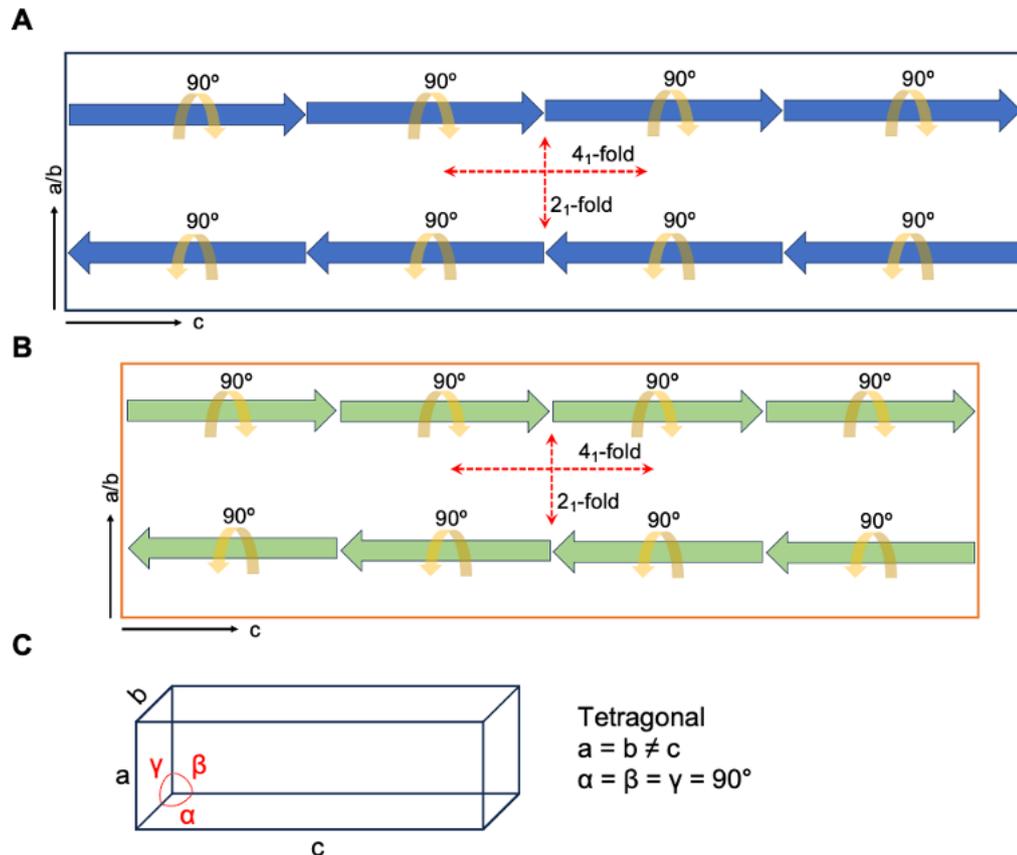


**Figure 3.9. Alignment of Cry8Ba2-S and Cry8Ba2-L.** In Cry8Ba2-S (blue), the N-terminal region (shown as surface) is tucked into the rest of the molecule. In Cry8Ba2-L (cyan), the N-terminal region (shown as surface) is projected out by a hinge region composed of a flexible loop. The C-terminal half of Cry8Ba2-L also appears to have hinged into a straighter conformation. The unit cell is shown as a box.

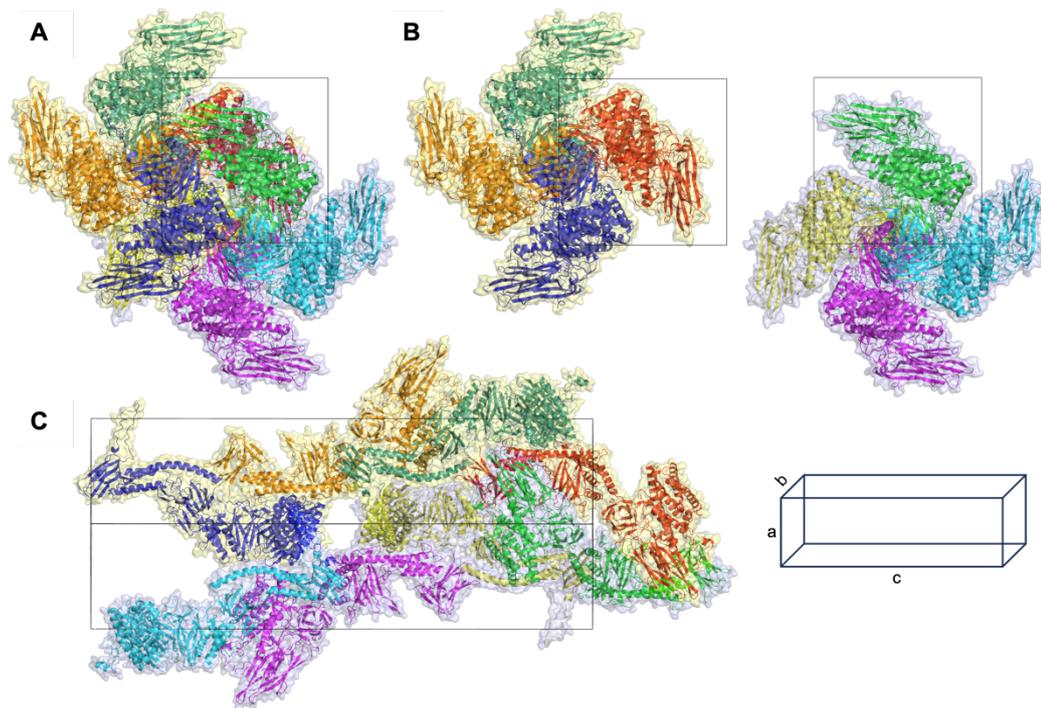
#### 3.2.2.4. Crystal packing of the Cry8Ba2 structure

The packing of pesticidal crystal proteins has biological relevance for understanding the process by which natural *B. thuringiensis* crystals are formed, as well as the multifaceted roles that they can play in pesticidal activity. Like the Cry1Ac- $\Delta$ 14C protoxin crystals (Evdokimov et al. 2014), and the *B. thuringiensis* crystals (long Cry form) studied using powder X-ray diffraction and electron microscopy (Holmes and Monro 1965), the Cry8Ba2 crystals belong to the tetragonal space group  $P4_12_12$ . Hence, this symmetry of packing may be favoured by the long form of Cry proteins. This may be biologically important, as high conservation of the C-terminal pro-domains may enable Cry protein variants to co-crystallise, ultimately ensuring their co-delivery to the target insect, which, itself, would be advantageous for those Cry protein subclasses that display synergistic activity (Lee et al. 1996; Sharma et al. 2010). It is important to note that Cry crystal proteins in the short (~65 kDa) form appear to pack differently to Cry proteins in the long (~130 kDa) form. For example, Cry3Aa, Cry11Aa and Cry11Ba (examples of short Cry proteins the structures of which have been solved directly from native crystals using SFX) belong to distinct space groups ( $C222_1$ ,  $I222$  and  $P2_12_12$ , respectively).

To understand crystal packing further, the Cry8Ba2 crystal interfaces were analysed using PISA. In both Cry8Ba2-S and Cry8Ba2-L, crystal packing has successive molecules lined up head-to-toe rectilinearly, parallel to the 4-fold screw ( $4_1$ ) axis (**Fig 3.10, 3.11**). The molecules fall on either side of a 2-fold screw ( $2_1$ ) axis normal to the 4-fold screw axis. The two strands of molecules run in opposite directions to one another (**Fig 3.10, 3.11**). Interface analysis using PISA identifies 10 interfaces in the Cry8Ba2-S crystal (**Table 3.2**) and 8 interfaces in the Cry8Ba2-L crystal (**Table 3.3**). Crystal contact regions total 8222.6 Å<sup>2</sup> in Cry8Ba2-S and 7598.9 Å<sup>2</sup> in Cry8Ba2-L, with many of these interfaces involving several hydrogen bonds and salt bridges (**Tables 3.2, 3.3**).



**Figure 3.10. Schematic of Cry8Ba2 packing in the unit cell.** (A) Cry8Ba2-L unit cell. Cry8Ba2 monomers depicted as blue arrows. (B) Cry8Ba2-S unit cell. Cry8Ba2 monomers depicted as green arrows. In both models, crystal packing has the successive molecules lined up head-to-toe along a 4-fold screw ( $4_1$ ) axis, which is parallel to the c-axis. With each successive molecule, there is a  $90^\circ$  turn, and the molecules fall either side of a 2-fold ( $2_1$ ) axis. The two strands of molecules run in opposite directions to one another. The Cry8Ba2-S and Cry8Ba2-L unit cells exhibit significant differences in the length of the c-axis. (C) Schematic of a tetragonal unit cell showing lengths a, b, c, where lengths  $a = b \neq c$  and angles  $\alpha = \beta = \gamma = 90^\circ$ .



**Figure 3.11. Packing of Cry8Ba2-L monomers in the unit cell.** (A) 2D-projection looking down the c-axis showing the full contents of the unit cell. (B) 2D-projection looking down the c-axis showing each strand of Cry8Ba2 molecules independently. (C) 2D-projection looking down the diagonal between the a/b-axis showing the full contents of the unit cell. In all figures, Cry8Ba2 monomers are shown in cartoon and coloured by chain. The surface view of each strand is also shown in pale yellow and pale blue. The unit cell is shown as a black box. A schematic of the unit cell axes (*a*, *b*, *c*) has been included.

Table 3.2. Calculated structural properties of the molecular interfaces in the Cry8Ba2-S crystal.

Interface	Monomer 1			Monomer 2			Interface area (Å <sup>2</sup> )	$\Delta^i G^*$ (kcal mol <sup>-1</sup> )	H-bonds	Salt bridges	Binding energy (kcal mol <sup>-1</sup> )
	Chain	Symmetry	AA	Chain	Symmetry	AA					
1	A	x,y,z	103	A	-Y+1,-X+1,-Z+1/2	104	3243.6	-5.1	44	8	-27.6
2	A	x,y,z	35	A	-Y,-X+1,-Z+1/2	41	1138.2	-4.0	17	1	-11.9
3	A	x,y,z	35	A	-Y+1/2,X-1/2,Z+1/4	36	1124.8	-0.8	7	8	-6.9
4	A	x,y,z	25	A	X-1/2,-Y+1/2,-Z+3/4	26	824.5	-2.4	5	8	-7.6
5	A	x,y,z	24	A	-X+1/2,Y-1/2,-Z+1/4	24	793.1	-9.3	2	2	-11.0
6	A	x,y,z	19	A	-Y+3/2,X-1/2,Z+1/4	21	442.4	-0.8	1	0	-1.2
7	A	x,y,z	13	A	-X+3/2,Y-1/2,-Z+1/4	11	378.9	0.8	3	2	-1.3
8	A	x,y,z	4	A	-X+1,-Y+1,Z-1/2	8	184.3	1.3	6	0	-1.3
9	A	x,y,z	5	A	X-1/2,-Y-1/2,-Z+3/4	2	57.4	-0.1	1	0	-0.5
10	A	x,y,z	5	A	Y,X,-Z	5	35.4	0.7	0	0	0.7

\* Predicted solvation free energy gain

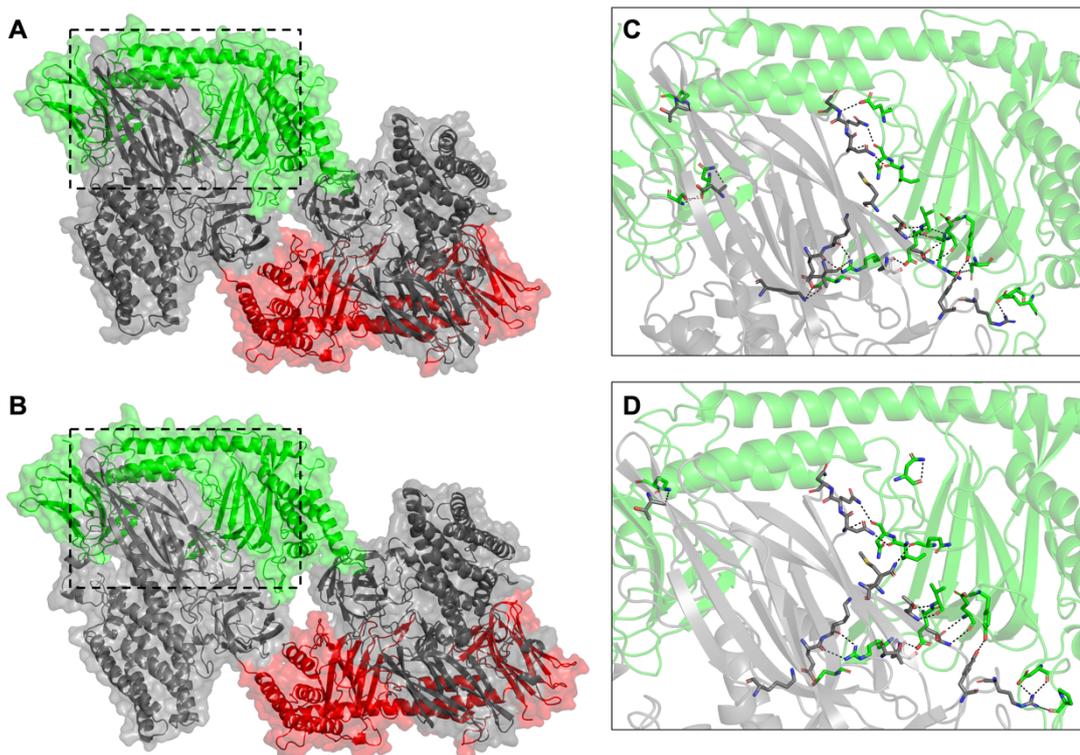
Table 3.3. Calculated structural properties of the molecular interfaces in the Cry8Ba2-L crystal.

Interface	Monomer 1			Monomer 2			Interface area (Å <sup>2</sup> )	Δ <sup>i</sup> G* (kcal mol <sup>-1</sup> )	H-bonds	Salt bridges	Binding energy (kcal mol <sup>-1</sup> )
	Chain	Symmetry	AA	Chain	Symmetry	AA					
1	A	x,y,z	100	A	-Y,-X,-Z+1/2	100	3407.4	-5.2	34	8	-23.3
2	A	x,y,z	25	A	-X+1/2,Y-1/2,-Z+1/4	26	829.1	-2.9	6	8	-8.5
3	A	x,y,z	20	A	X-1/2,-Y+1/2,-Z+3/4	29	803.2	-8.5	9	0	-12.5
4	A	x,y,z	21	A	-Y,-X+1,-Z+1/2	30	735.0	-0.6	5	4	-4.3
5	A	x,y,z	25	A	-Y+1/2,X-1/2,Z+1/4	24	692.9	-1.7	1	0	-2.1
6	A	x,y,z	25	A	-Y-1/2,X-1/2,Z+1/4	22	575.9	-1.3	1	0	-1.7
7	A	x,y,z	10	A	X-1/2,-Y-1/2,-Z+3/4	14	367.4	0.0	3	3	-2.5
8	A	x,y,z	8	A	-X,-Y,Z-1/2	4	188.0	1.2	5	0	-1.0

65

\* Predicted solvation free energy gain

The largest of these interfaces is composed of a head-to-tail dimer (**Fig 3.12**). In Cry8Ba2-L, the dimer interface area totals 3407.4 Å<sup>2</sup> and involves 34 hydrogen bonds as well as 8 salt bridges, with an estimated binding energy of -23.3 kcal mol<sup>-1</sup> (**Table 3.3**). In Cry8Ba2-S, the dimer interface area totals 3243.6 Å<sup>2</sup> and involves 44 hydrogen bonds as well as 8 salt bridges, with an estimated binding energy of -27.6 kcal mol<sup>-1</sup> (**Table 3.2**), representing the stronger interface of the two Cry8Ba2 dimers. In both crystals, the pro-domains of one monomer cup the toxin core of another monomer in a neighbouring strand (**Fig 3.12**), supporting the idea that the C-terminal region (pro-domains) of 130 kDa Cry proteins are involved in crystal formation and stability. A large majority of polar contacts formed (**Fig 3.12, Table 3.4, 3.5**), originate from DV of the C-terminal region and interact with DII/DIII of the other Cry8Ba2 monomer. The DV pro-domain residues involved in these interactions lie between three blocks of amino acids that are highly conserved (blocks 6 – 8) across the Cry class (**Fig 3.8**). Given their proximity to the Cry8Ba2 dimer interface, conservation of the amino acid sequences in block 6 – 8 may be required to maintain the correct secondary structure that allows the formation of Cry protein dimers.



**Figure 3.12. The structure of the Cry8Ba2 dimer interface.** (A, C) Cry8Ba2-S. (B, D) Cry8Ba2-L. In both crystals, a head-to-toe dimer forms, where the pro-domains of one monomer cup the toxin core of the other monomer. Toxin cores are coloured grey and pro-domains are coloured red and green. (C, D) Detailed view of the pro-domains of one monomer interacting with the toxin core. Polar contacts (black dashed lines) were identified in PyMOL using the default 3.6 Å cut-off. Residues identified by PISA to be involved in forming polar contacts (**Tables 3.4, 3.5**) are shown as sticks (carbon – green / red, oxygen – red, nitrogen – blue).

**Table 3.4. Cry8Ba2-S dimer interactions identified using PISA.**

Structure 1	Distance	Structure 2
<b>Hydrogen bonding</b>		
Tyr374 [OH]	2.4	Glu754 [OE2]
Thr382 [N]	3.2	Glu757 [O]
Thr384 [N]	2.8	Val755 [O]
Asn390 [ND2]	3.2	Ile741 [O]
Gln391 [NE2]	3.2	Asn742 [O]
Asn392 [N]	3.5	Glu744 [OE2]
Thr425 [N]	3.7	Gln1015 [OE1]
Tyr532 [OH]	3.4	Asp759 [OD2]
Tyr532 [OH]	2.7	Tyr766 [OH]
Arg535 [NH1]	3.5	Leu832 [O]
Arg535 [NH2]	3.8	Asp829 [O]
Asn742 [ND2]	3.4	Met386 [SD]
Val755 [N]	2.9	Thr384 [O]
Glu757 [N]	3.1	Thr382 [O]
Tyr766 [OH]	2.7	Tyr532 [OH]
Arg768 [NE]	3.5	Tyr532 [OH]
Arg768 [NH2]	3.7	Tyr532 [OH]
Arg773 [NH1]	3.0	Lys334 [O]
Arg773 [NH2]	3.3	Glu333 [O]
Arg773 [NH2]	2.8	Ile337 [O]
Gln1015 [NE2]	2.9	Thr425 [O]
Asn1026 [ND2]	3.3	Asp324 [O]
Ser1029 [OG]	3.2	Asp324 [OD1]
Glu754 [OE2]	2.4	Tyr374 [OH]
Glu757 [O]	3.2	Thr382 [N]
Val755 [O]	2.8	Thr384 [N]
Ile741 [O]	3.2	Asn390 [ND2]
Asn742 [O]	3.2	Gln391 [NE2]
Glu744 [OE2]	3.5	Asn392 [N]
Gln1015 [OE1]	3.7	Thr425 [N]
Asp759 [OD2]	3.4	Tyr532 [OH]
Leu832 [O]	3.5	Arg535 [NH1]
Asp829 [O]	3.8	Arg535 [NH2]
Met386 [SD]	3.4	Asn742 [ND2]
Thr384 [O]	2.9	Val755 [N]
Thr382 [O]	3.1	Glu757 [N]
Tyr532 [OH]	3.5	Arg768 [NE]

Tyr532 [OH]	3.7	Arg768 [NH2]
Lys334 [O]	3.0	Arg773 [NH1]
Glu333 [O]	3.3	Arg773 [NH2]
Ile337 [O]	2.8	Arg773 [NH2]
Thr425 [O]	2.9	Gln1015 [NE2]
Asp324 [O]	3.3	Asn1026 [ND2]
Asp324 [OD1]	3.2	Ser1029 [OG]
<b>Salt bridges</b>		
Lys258 [NZ]	2.8	Asp776 [OD1]
Lys258 [NZ]	2.9	Asp776 [OD2]
Arg535 [NH2]	3.9	Asp829 [OD1]
Arg535 [NH2]	2.8	Asp829 [OD2]
Asp776 [OD2]	2.9	Lys258 [NZ]
Asp776 [OD1]	2.8	Lys258 [NZ]
Asp829 [OD1]	3.9	Arg535 [NH2]
Asp829 [OD2]	2.8	Arg535 [NH2]

**Table 3.5. Cry8Ba2-L dimer interactions identified using PISA.**

Structure 1	Distance	Structure 2
<b>Hydrogen bonds</b>		
Gln371 [NE2]	3.4	Gln739 [O]
Tyr374 [OH]	2.5	Glu754 [OE2]
Thr382 [N]	3.4	Glu757 [O]
Thr384 [N]	2.8	Val755 [O]
Asn390 [ND2]	3.4	Ile741 [O]
Gln391 [NE2]	3.4	Asn742 [O]
Asn392 [N]	3.5	Glu744 [OE1]
Tyr532 [OH]	2.8	Tyr766 [OH]
Arg535 [NH1]	3.4	Pro833 [O]
Arg535 [NH2]	3.5	Asp829 [O]
Asn742 [ND2]	3.5	Met386 [SD]
Val755 [N]	2.9	Thr384 [O]
Glu757 [N]	3.2	Thr382 [O]
Tyr766 [OH]	2.8	Tyr532 [OH]
Arg773 [NH1]	3.0	Lys334 [O]
Arg773 [NH2]	2.6	Ile337 [O]
Arg773 [NH2]	3.3	Glu333 [O]
Gln1015 [NE2]	3.4	Thr425 [O]
Gln739 [O]	3.4	Gln371 [NE2]
Glu754 [OE2]	2.5	Tyr374 [OH]

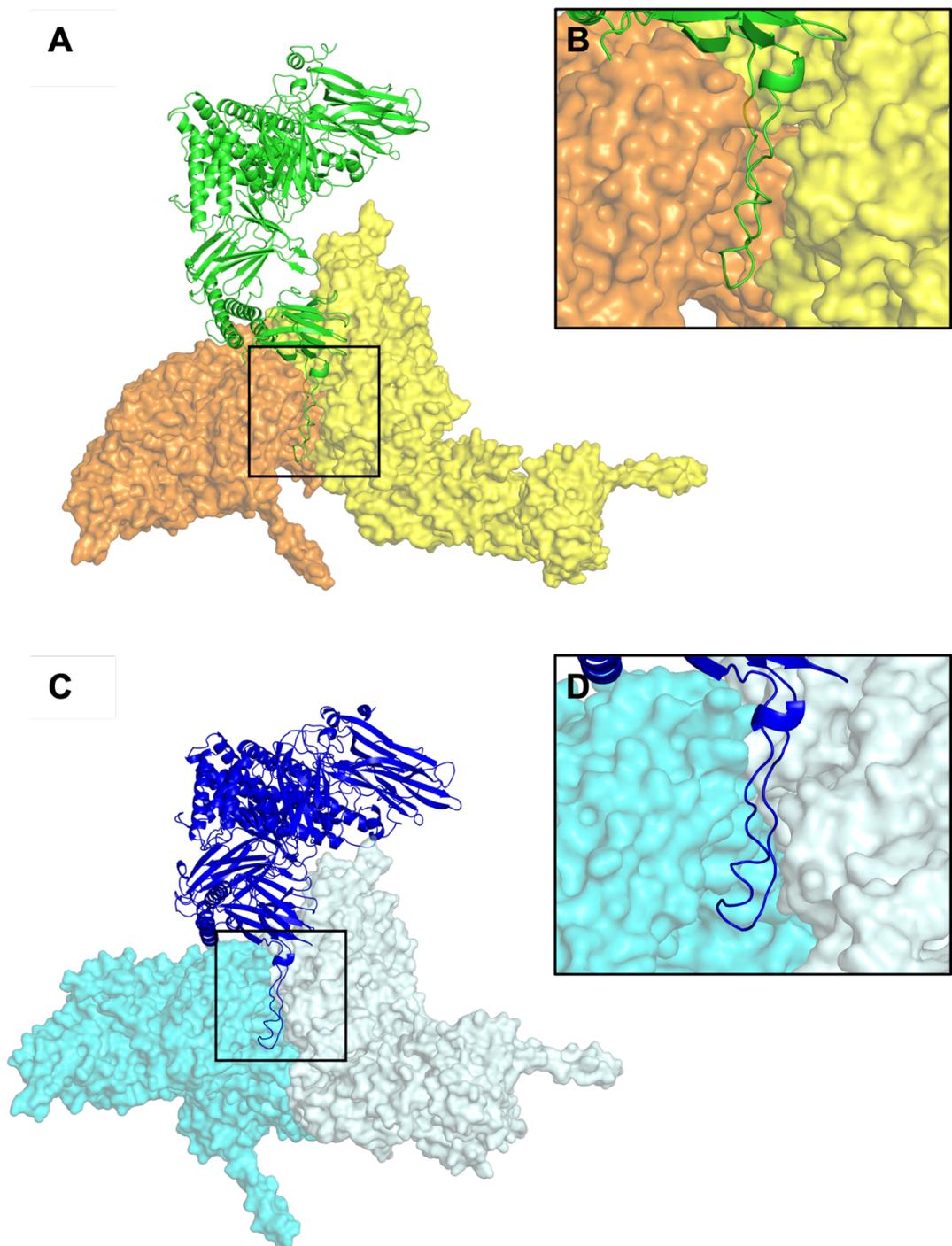
Glu757 [O]	3.4	Thr382 [N]
Val755 [O]	2.8	Thr384 [N]
Ile741 [O]	3.4	Asn390 [ND2]
Asn742 [O]	3.4	Gln391 [NE2]
Glu744 [OE1]	3.5	Asn392 [N]
Pro833 [O]	3.4	Arg535 [NH1]
Asp829 [O]	3.5	Arg535 [NH2]
Met386 [SD]	3.5	Asn742 [ND2]
Thr384 [O]	2.9	Val755 [N]
Thr382 [O]	3.2	Glu757 [N]
Lys334 [O]	3.0	Arg773 [NH1]
Glu333 [O]	3.3	Arg773 [NH2]
Ile337 [O]	2.6	Arg773 [NH2]
Thr425 [O]	3.4	Gln1015 [NE2]
<b>Salt bridges</b>		
Lys258 [NZ]	4.0	Asp776 [OD1]
Lys258 [NZ]	3.8	Asp776 [OD2]
Arg535 [NH2]	3.6	Asp829 [OD1]
Arg535 [NH2]	3.1	Asp829 [OD2]
Asp776 [OD1]	4.0	Lys258 [NZ]
Asp776 [OD2]	3.8	Lys258 [NZ]
Asp829 [OD1]	3.6	Arg535 [NH2]
Asp829 [OD2]	3.1	Arg535 [NH2]

Structural analysis of crystal packing also provides insight into the mechanisms of crystal dissolution upon exposure to non-neutral pH in the insect gut, which can range from slightly acidic to slightly basic depending on the target insect. Excluding one case of lepidopteran activity (Cry8Qa2 - Amadio et al. 2013), insect bioassays have demonstrated that the Cry8 subclass are active against the insect order Coleoptera (Asano et al. 2003; Shu et al. 2007; Yamaguchi et al. 2008; Guo et al. 2009; Zhang et al. 2013; Park et al. 2014; Shu et al. 2020). Coleopteran insect targets include the Chrysomelid (e.g., *Agelastica coerulea*) and Tenebrionoid (e.g., *Alphitobius diaperinus*) beetles, which exhibit neutral to slightly acidic gut pH, Scarabaeoid beetles (e.g., *Popillia japonica*), which exhibit alkaline gut pH, and Cerambycid beetles (e.g., *Anoplophora glabripennis*), for which the gut pH is unknown (Crowson 1981). Hence, both acidic pH (down to 3.6 in

coleopteran insects) and alkaline pH (up to 10.9 in coleopteran insects) (Biggs and McGregor 1996) appears to be physiologically relevant for crystal dissolution in the Cry8 subclass. Structural characterisation of the Cry8Ba2 crystals show that there are several intermolecular pH sensitive interactions, including salt bridges and hydrogen bonds, as well as one intramolecular disulphide bridge. Several of these interactions include titratable residues, such as histidine, lysine, aspartic acid, glutamic acid, and tyrosine (**Table 3.4, 3.5**). Upon exposure to low / high pH, the side chain groups can transition between protonated and deprotonated states, which can cause dissociation of hydrogen bonds and salt bridges at the site of crystal interfaces, ultimately leading to the dissolution of the crystal proteins (the first stage in the mechanism of action of these pesticidal proteins, following their ingestion by target insect larvae). All the identified salt bridges at the Cry8Ba2 dimer interfaces (**Table 3.4, 3.5**) involve aspartic acid (side-chain pKa = 3.65) bonded to either arginine (side-chain pKa = 12.48) or lysine (side-chain pKa = 10.53). Given that the side-chain pKa of arginine is too high to be of physiological relevance in the gut of known coleopteran insect targets, there may be a higher number of labile salt bridges at acidic pH in comparison to alkaline pH. Perturbations of side-chain pKa values from the normal, caused by the local environment in the tertiary structures (Aqvist et al. 1991), also cannot be ruled out .

In both Cry8Ba2 models, the electron density map showed continuous density for residues 35 – 1169 of the protein sequence, with a region of weak density existing between residues 1107 – 1129 and 1107 – 1130 for Cry8Ba2-S and Cry8Ba2-L, respectively. We attempted to manually build the missing residues by ensuring that there were no clashes with atoms of neighbouring molecules and that correct coordinate geometry existed, guided by electron density contoured at the lower 0.5 rms level. In both models, model building revealed an extended loop which breaks up the  $\beta$ -roll architecture of DVII. When viewed within the context of the crystal lattice, the loop appears to act as a peg, stabilising the sideways packing of neighbouring strings of molecules parallel to the  $4_1$  axis (**Fig 3.13**). To assess the nature of the loop interface, shape complementarity was

computed using the shape correlation (Sc) program (Lawrence and Colman 1993) available as a command-line executable (sc.exe) in the CCP4i2 package. Higher Sc values indicate tighter interfacial packing, with a value of 0 indicating perfect anti-complementarity and a value of 1 indicating perfect complementarity (Lawrence and Colman 1993). The Sc values of well-packed protein-protein interfaces have been found to range from approximately 0.55 – 0.75 (Lawrence and Colman 1993; Basu et al. 2021; Roy et al. 2022). Here, Sc values of 0.41 in the Cry8Ba2-S model and 0.49 in the Cry8Ba2-L model suggest that the loop interface exhibits a reasonably complementary shape. Taken together, these results suggest that a flexible interface between the loop and neighbouring molecules may exist, and this interface may aid the crystal packing of Cry8Ba2 monomers.



**Figure 3.13. DVII putative loop of Cry8Ba2 appears to stabilise crystal packing.** (A) The Cry8Ba2-S putative loop (green) interface within two neighbouring molecules (orange, yellow) (B) Close up Cry8Ba2-S putative loop interface (C) The Cry8Ba2-L putative loop (blue) interface within two neighbouring molecules (cyan, light cyan) (D) Close up Cry8Ba2-L putative loop interface.

### 3.2.2.5. Comparison with related protein structures

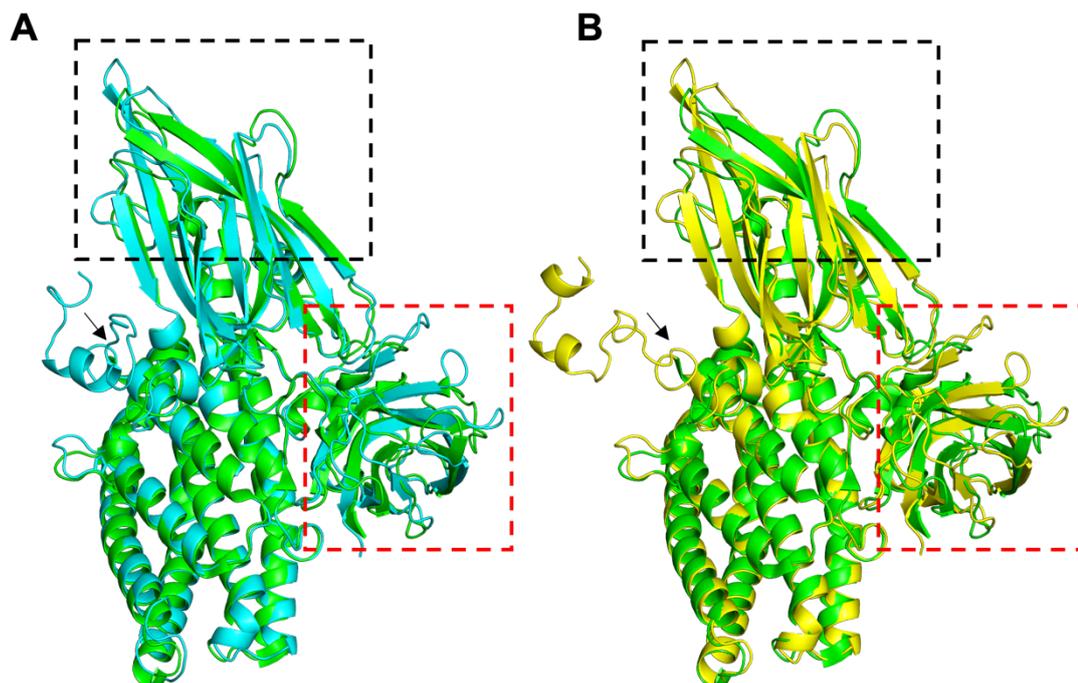
For comparison with related structures in the PDB, the DALI web server was used (Holm 2020). The best matches (**Table 3.6**) were identified as other pesticidal proteins belonging to the Cry family, including the activated forms of Cry8Ea1 (PDB 3EB7-B - Guo et al. 2009) and Cry3Bb1 (PDB 1JI6-A - Galitsky et al. 2001), as well as the mutated Cry1Ac- $\Delta$ 14C (PDB 4W8J-A - Evdokimov et al. 2014). Other structures of activated Cry proteins were identified, but only the most closely related structures have been compared here. Prb2Aa1 (formerly PirB, PDB 3X0U-A – Lee et al. 2015) was also identified. Prb2Aa1 is produced by the gram-negative bacterium *Vibrio parahaemolyticus* and functions as a binary protein alongside Pra2Aa1 (formerly PirA) to cause acute hepatopancreatic necrosis disease in shrimps (Lee et al. 2015). Despite low sequence identity (<10%), Pra/Prb has known structural similarity with DI and DII of Cry proteins and is thought to exert toxicity via similar mechanisms involving pore-formation. Of all the pesticidal proteins identified, the greatest structural similarity was identified between Cry8Ba2 and Cry8Ea1 (**Table 3.6**).

**Table 3.6. Structural similarity of Cry8Ba2 with other proteins.**

PDB ID_chain	Name (formerly)	Cry8Ba2-S		Cry8Ba2-L	
		Z-score*	RMSD (Å)	Z-score*	RMSD (Å)
3EB7_B	Cry8Ea1	36.5	1.4	53.4	1.4
4W8J_A	Cry1Ac1-Δ14C	33.4	3.3	41.2	2.9
1JI6_A	Cry3Bb1	32.1	1.8	40.7	1.7
3X0U_A	Prb2Aa1 (PirB)	13.8	3.5	24.1	3.5
2XON_L	β-1,4-galactanase	14.6	2.2	13.4	2.3
3ZM8_A	β-1,4-mannanase	13.3	10.9	13.9	10.6
5LF2_A	Laminin β2 LF	11.4	7.2	12.0	7.5
2GOM_A	C3-inhibitory domain, Efb-C	8.8	1.4	6.7	1.4
1CUN_A	Spectrin	5.9	9.4	7.0	13.3

\* DALI Z-scores larger than 2 are considered significant. Other activated Cry proteins were identified, but only the most closely related (Cry8Ea1, Cry1Ac1, Cry3Bb1) have been included here.

Cry8Ea1 is encoded by the *cry8Ea1* gene of the Chinese Bt185 isolate and has known activity against *H. parallela* (dark black chafer), a scarab beetle and pest of several crops in East Asia (Shu et al. 2009). The structure of the activated Cry8Ea1 toxin has been determined (PDB 3EB7, Guo et al. 2009). Similar to the Cry8Ba2 structure, Cry8Ea1 DI helix 2 is split into helix 2a and 2b, suggesting that this may be a recurring feature of the Cry8 subclass (Guo et al. 2009). Key differences between the Cry8Ba2 and Cry8Ea1 structures occur at the surface exposed loops of DII and DIII (**Fig 3.14**), corresponding with several insertions / deletions identified in the sequences. Given that these regions are hypothesized to be involved in receptor interaction, it is possible that Cry8Ba2 and Cry8Ea1 confer distinct target specificity. Cry8Ba2 DI also reveals the presence of an extended N-terminal region which does not appear in the chymotrypsin-activated Cry8Ea1 structure, likely due to it being cleaved during proteolytic processing (**Fig 3.14**).

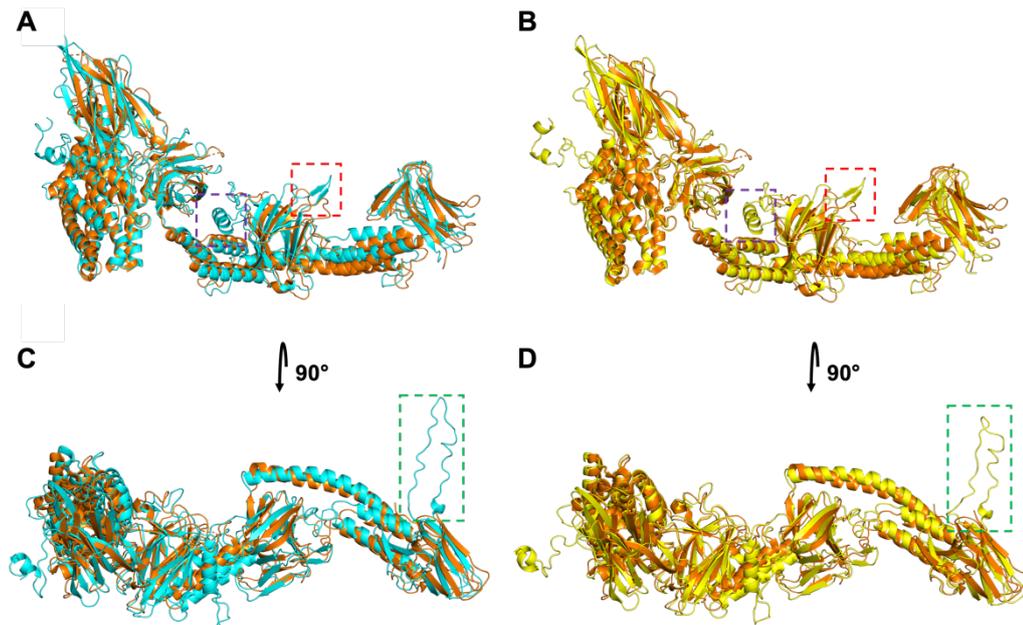


**Figure 3.14. Structural alignment of Cry8Ba2 with Cry8Ea1.** (A) Alignment of Cry8Ba2-S (cyan) and chymotrypsin-activated Cry8Ea1 (green). (B) Alignment of Cry8Ba2-L (yellow) and chymotrypsin-activated Cry8Ea1 (green). In both figures, DI – DIII (residues 35 – 658, inclusive of the N-terminal peptide) of the Cry8Ba2 structure are shown. Key differences occur at the exposed loop regions of DII (black box) and DIII (red box), corresponding with several insertions / deletions identified in a sequence alignment. The position of the Cry8Ba2 N-terminal peptide is also shown relative to the Cry8Ea1 chymotrypsin cleavage site (black arrow).

Prior to this work, one structure of a full-length 130 kDa Cry protein had been reported and that was of a highly-mutated form of Cry1Ac, Cry1Ac- $\Delta$ 14C (Evdokimov et al. 2014). To minimise protein aggregation prior to crystallisation, Evdokimov et al (2014) mutated 14 out of the 16 cysteine residues (by serine substitution), in addition to a F462V mutation (Evdokimov et al. 2014). Mutation of all 16 cysteines (where the first two are present within the N-terminal peptide, whilst the latter 14 – mutated in Cry1Ac- $\Delta$ 14C – are present within the extended C-terminus) produced unstable protein crystals, suggesting that the first two cysteine residues are important for

crystal stability. To solve the Cry1Ac- $\Delta$ 14C structure, protein crystals were expressed in *B. thuringiensis* strain EG10650 and solubilised at pH 10.5. The soluble protein was purified and subjected to crystal trials for conventional X-ray crystallography.

Within the protein core (DI – DIII), key differences between the Cry8Ba2 and Cry1Ac- $\Delta$ 14C structures occur at the surface exposed loops of DII and DIII, suggesting distinct target specificity. Again, within the C-terminal pro-domains (DIV – DVII), several amino acid insertions / deletions lead to differences in the placement of surface exposed loops. Despite this, the overall secondary structure of the pro-domains is widely similar. One key difference was identified in Cry8Ba2 DV, where residues 773-REIDTETY-780 form a short  $\beta$ -hairpin, not seen in the Cry1Ac- $\Delta$ 14C structure (**Fig 3.15A, 3.15B – red box**). Other notable differences occur at regions that were not seen in the electron density map of Cry1Ac- $\Delta$ 14C and hence, were hypothesized to be disordered (Cry1Ac1 residues 783 – 821 and 1063 – 1137). In the Cry8Ba2 structure, the first of these regions corresponds to a helix (residues 846 – 858) which interrupts the  $\beta$ -roll structure of DV and forms a disulphide bridge (Cys703-Cys847) with a helix of DIV (**Fig 3.7 and Fig 3.15A, 3.15B – purple box**). Cry8Ba2 Cys703 is not conserved in the Cry1Ac1 sequence. On the other hand, Cry8Ba2 Cys847 is conserved in the Cry1Ac1 sequence (Cys802 in Cry1Ac1) and has the potential to cross-link with Cys661, which is in close structural proximity to the unmodelled region in Cry1Ac- $\Delta$ 14C. The second of these regions corresponds to an extended loop of DVII (residues 1104 – 1131), which projects out perpendicular to the rest of the structure (**Fig 3.7 and Fig 3.15C, 3.15D – green box**) and appears to act as a peg, stabilising crystal packing in Cry8Ba2 (**Fig 3.13**). Given that the corresponding region was not seen in the Cry1Ac- $\Delta$ 14C electron density map (suggesting that it is disordered), it's possible that this structural feature also exists in Cry1Ac1.



**Figure 3.15. Structural alignment of Cry8Ba2 with Cry8Ea1.** (A, C) Alignment of Cry8Ba2-S (cyan) and Cry1Ac- $\Delta$ 14C (orange). (B, D) Alignment of Cry8Ba2-L (yellow) and Cry1Ac- $\Delta$ 14C (orange). Key differences include the presence of a short  $\beta$ -hairpin in DV (panels A, B – red box), as well as regions that were not seen in the electron density map of Cry1Ac- $\Delta$ 14C and hence, were hypothesized to be disordered (Cry1Ac1 residues 783 – 821 and 1063 – 1137). In Cry8Ba2, the first of these regions corresponds to a helix which interrupts the  $\beta$ -roll structure of DV and forms a disulphide bridge (Cys703-Cys847) with a helix of DIV (panels A, B – purple box). The second of these regions corresponds to an extended loop of DVII, which projects out perpendicular to the rest of the structure (panels C, D – green box).

PISA analysis of the Cry1Ac- $\Delta$ 14C crystal interfaces (**Table 3.7**) highlights the formation of nine crystal interfaces, including a dimer interface involving several hydrogen bonds and salt bridges, similar to those seen in Cry8Ba2 crystals (**Tables 3.2, 3.3**). In some Cry proteins (e.g., Cry1 subclass), the protoxin domain is cysteine rich. Previous studies have suggested that these cysteines are able to support crystal packing and stability by forming inter- and intramolecular disulphide cross-links which, in the reducing environment of the insect midgut, would dissociate, leading to

the release of protoxin monomers (Dastidar and Nickerson 1979; Couche et al. 1987; Bietlot et al. 1990; Evdokimov et al. 2014). However, Li et al (2019) demonstrated that cysteine substitution (to serine) in Cry1Ac1 had little effect on protein expression, crystal formation, or toxicity (Li et al. 2019). Since 14 of the 16 cysteine residues in Cry1Ac1 were mutated to elucidate the Cry1Ac- $\Delta$ 14C structure, intermolecular disulphide crosslinking is yet to be investigated structurally (Evdokimov et al. 2014). Like other Cry8 proteins, Cry8Ba2 does not contain a high number of cysteine residues, with only four cysteines present, two of which are in the pro-domains. Here, the Cry8Ba2 structure reveals one intramolecular disulphide bridge between pro-domain residues Cys703-Cys847. As previously discussed, this disulphide bridge appears to stabilise the intramolecular DIV-DV interface. Both DIV and DV are involved in the formation of the Cry8Ba2 dimer interface (**Fig 3.12**) and therefore, it is possible that Cys703-Cys847 plays a role in maintaining the structural fold required for formation of Cry8Ba2 dimers, and more widely, crystal contacts. Upon exposure to an alkaline, reducing environment, dissociation of Cys703-Cys847 (alongside loss of other interactions, such as pH labile hydrogen bonds and salt bridges, may destabilise Cry8Ba2 crystal packing, leading to crystal dissolution. Structural studies of Cry proteins with a high number of cysteines (e.g., non-mutated Cry1Ac1 and/or Cry1Ca1) are required to investigate the role of inter- and intramolecular disulphides further.

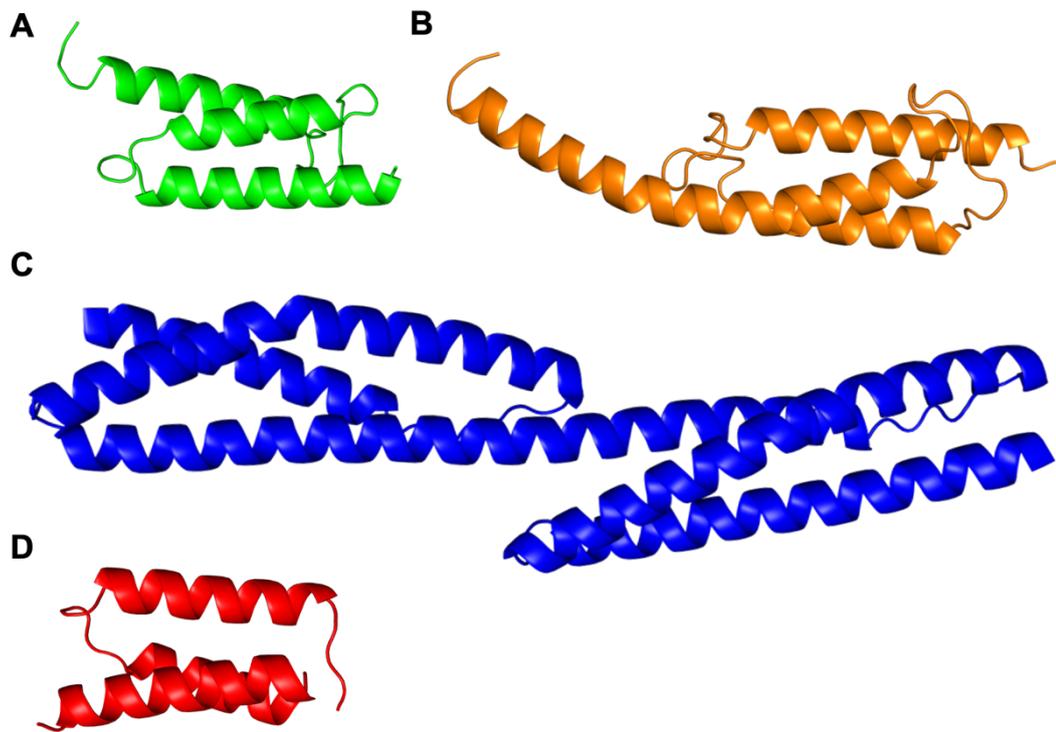
Table 3.7. Calculated structural properties of the molecular interfaces in Cry1Ac1- $\Delta$ 14C (PDB 4W8J).

Interface	Monomer 1			Monomer 2			Interface area ( $\text{\AA}^2$ )	$\Delta^iG^*$ (kcal mol $^{-1}$ )	H-bonds	Salt bridges	Binding energy (kcal mol $^{-1}$ )
	Chain	Symmetry	AA	Chain	Symmetry	AA					
1	A	x,y,z	95	A	-y,-x,-z+1/2	95	2817.6	-11.0	36	4	-28.5
2	A	x,y,z	37	A	-y-1/2,x+1/2,z+1/4	31	1068.0	-1.0	9	8	-7.9
3	A	x,y,z	24	A	x-1/2,-y-1/2,-z+3/4	29	884.2	-6.2	8	1	-10.1
4	A	x,y,z	26	A	-x-1/2,y-1/2,-z+1/4	31	877.3	-2.9	2	4	-5.2
5	A	x,y,z	26	A	-y-1,-x,-z+1/2	25	751.2	-6.8	1	4	-8.7
6	A	x,y,z	14	A	y,x,-z+1	14	408.7	-3.9	2	0	-4.8
7	A	x,y,z	12	A	-y+1/2,x+1/2,z+1/4	11	202.9	-1.6	0	0	-1.6
8	A	x,y,z	11	A	x-1/2,-y+1/2,-z+3/4	13	193.4	0.1	1	0	-0.4
9	A	x,y,z	5	A	-x,-y,z-1/2	8	183.4	-2.4	0	0	-2.4

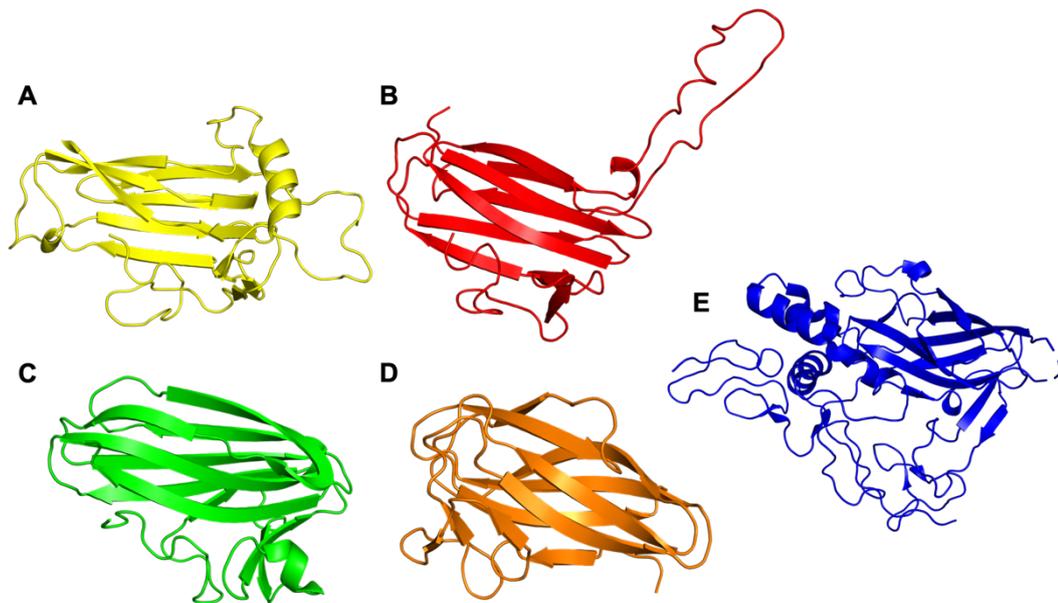
\*Predicted solvation free energy gain

Other proteins identified using DALI included Repeats 16 and 17 of *Gallus gallus*  $\alpha$ -Spectrin (PDB 1CUN-A, **Table 3.6, Fig 3.16** - Grum et al. 1999) and the complement C3-inhibitory domain of *Staphylococcus aureus* extracellular fibrinogen-binding protein, Efb-C (PDB 2GOM-A, **Table 3.6, Fig 3.16** - Hammel et al. 2007), both of which resemble Cry8Ba2 DIV and DVI helical bundles. *S. aureus* is a human pathogen that suppresses the complement system, a central component of human immunity, by producing Efb-C, which binds the helical C3d domain of the complement component 3 (C3) protein (Hammel et al. 2007). Spectrin is a highly conserved component of the cytoskeleton which functions to maintain the stability and structure of the cell wall and is conserved across all eukaryotes (Baines 2009). Spectrin tetramers associate with actin filaments, alongside other proteins which function as junctional complexes, to form a hexagonal scaffold-like structure on the intracellular membrane (Baines 2009). Structural similarity between Spectrin and DIV / DVI of Cry8Ba2 may suggest some capability of the pro-domains in assembling Cry protein at the membrane, increasing the concentrations for subsequent receptor interaction, oligomer assembly, and pore-formation.

Like Cry1Ac- $\Delta$ 14C, Cry8Ba2 DV and DVII were found to resemble carbohydrate binding modules of sugar hydrolases, including *Thermotoga maritima*  $\beta$ -1,4-galactanase (PDB 2XON-L, **Table 3.6, Fig 3.17** - Cid et al. 2010) and *Podospira anserina*  $\beta$ -1,4-mannanase (PDB 3ZM8-A, **Table 3.6, Fig 3.17** - Couturier et al. 2013). DALI also revealed structural similarity with the LF domain of the laminin, a basement membrane protein which plays roles in cell differentiation, adhesion, and migration (PDB 5LF2-A, **Table 3.6, Fig 3.17** - Pulido et al. 2017). The LF domain is a globular domain which interrupts the laminin-type epidermal growth factor-like (LE) domain of the laminin  $\beta$  chain and is also related to family 35 carbohydrate binding module (CBM35). There is ample precedent for the role of glycoconjugates as receptors for Cry proteins (Best et al. 2023), however, the specific role of carbohydrate binding in the C-terminal pro-domains is yet to be investigated.



**Figure 3.16. Structural similarity of Cry8Ba2 DIV and DVI with other proteins.** Structural similarity of Cry8Ba2-S (A) DIV (green) and (B) DVI (orange) with (C) Repeats 16 and 17 of *Gallus gallus*  $\alpha$ -Spectrin (blue, PDB 1CUN\_A) and the complement C3-inhibitory domain of *Staphylococcus aureus* extracellular fibrinogen-binding protein, Efb-C (red, PDB 2GOM\_A).



**Figure 3.17. Structural similarity of Cry8Ba2 DV and DVII with other proteins.** Structural similarity of Cry8Ba2-S **(A)** DV (yellow) and **(B)** DVII (red) with **(C)**  $\beta$ -1,4-galactanase (green, PDB 2XON\_L), **(D)**  $\beta$ -1,4-mannanase (orange, PDB 3ZM8\_A), and **(E)** Laminin (blue, PDB 5LF2\_A).

### 3.2.3. Insect feeding bioassays

Previous work has demonstrated that proteins in the Cry8 subclass are specifically active against members of the insect order Coleoptera (Asano et al. 2003; Shu et al. 2007; Yamaguchi et al. 2008; Guo et al. 2009; Zhang et al. 2013; Park et al. 2014; Shu et al. 2020). One case of low toxicity against *Anticarsia gemmatalis*, a member of the insect order Lepidoptera, has also been reported (Amadio et al. 2013). Despite this, the target insect range of the Cry8Ba2 variant is unknown. To investigate the target insect range of Cry8Ba2, insect feeding bioassays were performed against *T. molitor* and *Z. morio* larvae as described in **section 2.5.5**. Briefly, mealworm larvae were fed a diet of sweet potato / courgette that had been pre-soaked with high concentrations of Cry8Ba2 crystal samples. As negative controls, larvae were either fed a diet of sweet potato / courgette that had been pre-soaked with Cry48Aa1/Tpp49Aa1 crystal samples, a

mosquitocidal binary protein with no known coleopteran activity, or sweet potato / courgette that had not been soaked with any pesticidal protein samples. The diet was replenished every two days to ensure freshness and the survival rate of larvae was monitored for 14 days. After 14 days of exposure to high concentrations of Cry8Ba2, no significant difference in mortality was observed between larvae fed a diet with Cry8Ba2 or the negative controls, suggesting that Cry8Ba2 does not exert toxicity against *T. molitor* or *Z. morio*. Further work is required to identify the target insect range of Cry8Ba2.

### 3.3. Conclusions

In this chapter, the structures of two forms of Cry8Ba2 were elucidated from their natural crystals using SFX at an XFEL to a final resolution of 2.2 Å. The Cry8Ba2 structures represent the first structure of a full-length, non-mutated Cry protein in the long (~ 130 kDa) form and is consistent with that of a highly mutated structure of Cry1Ac, Cry1Ac-Δ14C (solved from reconstituted crystals), whereby the extended C-terminal pro-protein region is split into four domains (DIV, DV, DVI, DVII). The Cry8Ba2 structure highlights two features which were not seen in Cry1Ac-Δ14C due to lack of electron density. First, a helix which interrupts the β-roll structure of DIV and forms a disulphide bridge (Cys703-Cys847) with a helix of DIV. Secondly, an extended loop in DVII which projects out perpendicular to the structure and appears to act as a peg, stabilising the sideways packing of Cry8Ba2 monomers in the natural crystal. Both features are significant for maintaining the structural integrity of Cry8Ba2, supporting the hypothesis that the extended C-terminus is involved in maintaining crystal stability in the environment, prior to ingestion by the target insect.

Structural similarity of DIV and DVI with Spectrin (a cytoskeletal protein which forms scaffold-like structures of the intracellular membrane) led us to hypothesize a novel role for DIV and DVI, involving sequestering of Cry protein to the membrane for subsequent receptor interaction and pore-formation. Structural similarity of DV and DVII also suggests some sugar

binding capability of the pro-domains. Overall, the Cry8Ba2 structure sheds light on structural features which appear significant for crystal integrity, and points to additional functions which require further investigation.

Given that receptor interaction is essential for downstream insecticidal activity, the Cry8Ba2 structure is also significant for understanding the specificity of this pesticidal protein. Insect bioassays carried out in this work indicated that Cry8Ba2 does not display activity against two coleopteran insects, *T. molitor* and *Z. morio*. Additional bioassays are required to identify the target insect range which, once known, would enable downstream work aimed at studying the molecular mechanisms of Cry8Ba2, for which knowledge of the structure will be applicable.

#### 4. Structural analysis of the Cry48Aa1/Tpp49Aa1 pesticidal protein

Work in this chapter contributed to one published, peer-reviewed article where I am 1<sup>st</sup> author. With permission from the publisher, text and figures have been taken and adapted for my thesis.

**Williamson, L.J.**, Galchenkova, M., Best, H.L., Bean, R.J., Munke, A., Awel, S., Pena, G., Knoska, J., Schubert, R., Dörner, K., Park, H.W., Bideshi, D, K., Henkel, A., Kremling, V., Klopprogge, B., Lloyd-Evans, E., Young, M., Valerio, J., Kloos, M., Sikorski, M., Mills, G., Bielecki, J., Kirkwood, H., Kim, C., Wijn, R., Lorenzen, K., Lourdu, P. X., Rahmani Mashhour, A., Gelisio, L., Yefanov, O., Mancuso, A.P., Federici, B., Chapman, H.N., Crickmore, N., Rizkallah, P.J., Berry, C., Oberthür, D. 2023. Structure of the *Lysinibacillus sphaericus* Tpp49Aa1 pesticidal protein elucidated from natural crystals using MHz-SFX. *Proceedings of the National Academy of Sciences* 120(49), p. e2203241120. doi: [10.1073/pnas.2203241120](https://doi.org/10.1073/pnas.2203241120)

##### 4.1. Introduction

Mosquitoes act as vectors for several human diseases. To control mosquito populations, highly pathogenic strains of *L. sphaericus*, expressing the Tpp1Aa2/Tpp2Aa2 (BinA2/B2) pesticidal protein (Broadwell et al. 1990a; Charles et al. 1996), have been commercially applied worldwide. Despite its success, resistance to Tpp1Aa2/Tpp2Aa2 has been detected in both field treated populations and lab selected colonies of *Cx. pipiens* and *Cx. quinquefasciatus* mosquitoes (Silva-Filha et al. 1995; Yuan et al. 2000; Silva-Filha et al. 2021). In response to this, *L. sphaericus* isolates exhibiting toxicity against mosquito larvae resistant to Tpp1Aa2/Tpp2Aa2 were characterised, leading to the identification of the Cry48Aa1/Tpp49Aa1 (formerly Cry48Aa1/Cry49Aa1) pesticidal protein (Yuan et al. 2003; Jones et al. 2007; Jones et al. 2008). Given the significance of Cry48Aa1/Tpp49Aa1 in the field, additional work aimed at understanding the mechanism of action of this toxin pair was required.

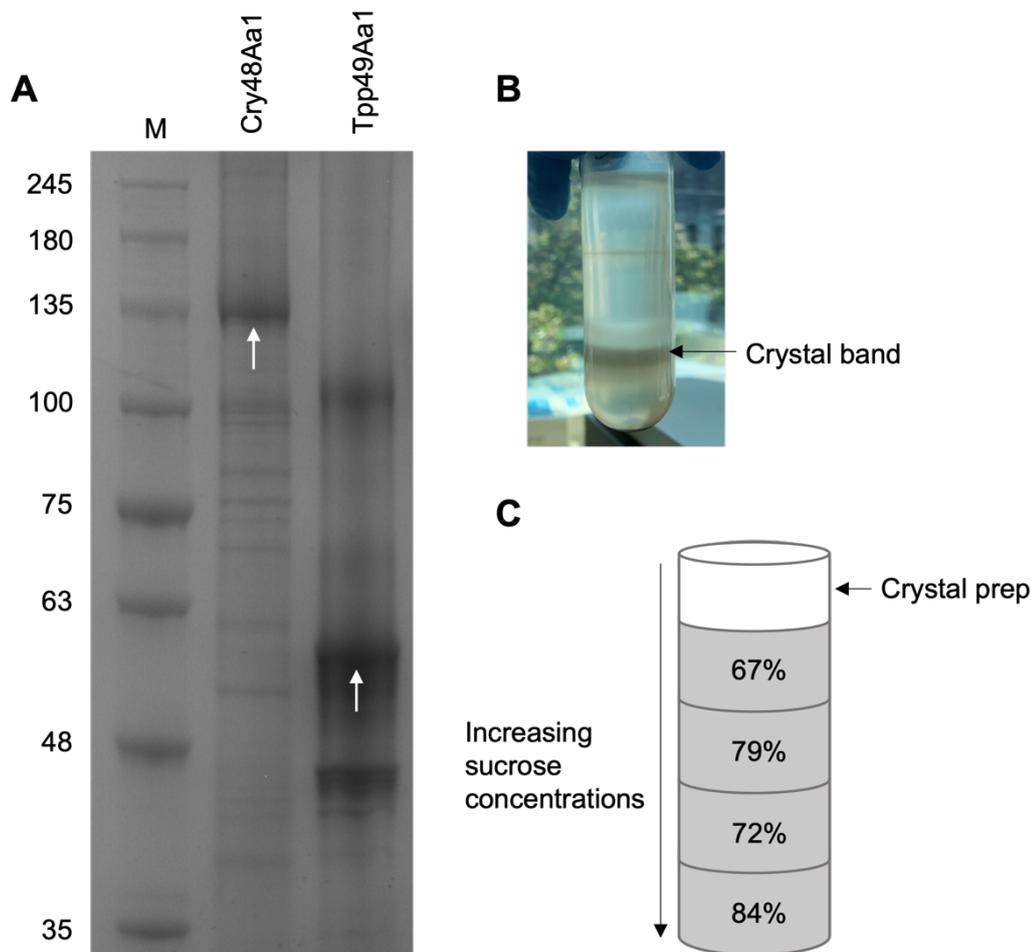
This chapter focuses on production of Cry48Aa1 and Tpp49Aa1 crystal proteins for characterisation of their activity and structure solution of Tpp49Aa1 using SFX at the European X-ray Free Electron Laser (EuXFEL, Hamburg, Germany). Complementary pH mixing studies were performed, enabling early structural changes in the crystal dissolution process to be investigated. In addition, the interaction of Tpp49Aa1 with its partner protein, Cry48Aa1, was predicted using computational modelling, enabling discussion surrounding models for mode of action. Finally, insect bioassays were performed, leading to the identification of new target mosquito species, thus expanding the potential use of Cry48Aa1-Tpp49Aa1 as a bioinsecticide. Throughout this chapter, procedures that were carried out by other researchers have been indented and italicised.

## **4.2. Results and Discussion**

### **4.2.1. Production of the Cry48Aa1 and Tpp49Aa1 crystal proteins**

#### **4.2.1.1. Crystal protein expression and isolation**

Recombinant *B. thuringiensis* subsp. *israelensis* 4Q7 bacterial strains encoding either the Cry48Aa1 or Tpp49Aa1 crystal proteins were obtained from Prof. Colin Berry (Cardiff University, UK). The *cry48Aa1* gene was expressed under the control of the *cyt1A* promoters in the pSTAB vector (Park et al. 1999) from the construct pSTABP135 (Jones et al. 2007), while the *tp49Aa1* gene was expressed under the control of its own promoter in the pHT304 vector from the construct pHTP49 (Jones et al. 2007). The Cry48Aa1 and Tpp49Aa1 protein sequences have been provided in **Appendices 2 and 3**, respectively. The strains were grown in Embrapa medium by incubating at 30 °C until sporulation reached 90% (typically between 48 - 72 h). Subsequently, the lysed sporulated culture was separated on a discontinuous sucrose gradient. Crystal enriched bands were extracted and analysed by SDS-PAGE (**Fig 4.1**). SDS-PAGE analysis revealed the enrichment of an ~135 kDa protein (Cry48Aa1) and an ~53 kDa protein (Tpp49Aa1) from the respective cultures.

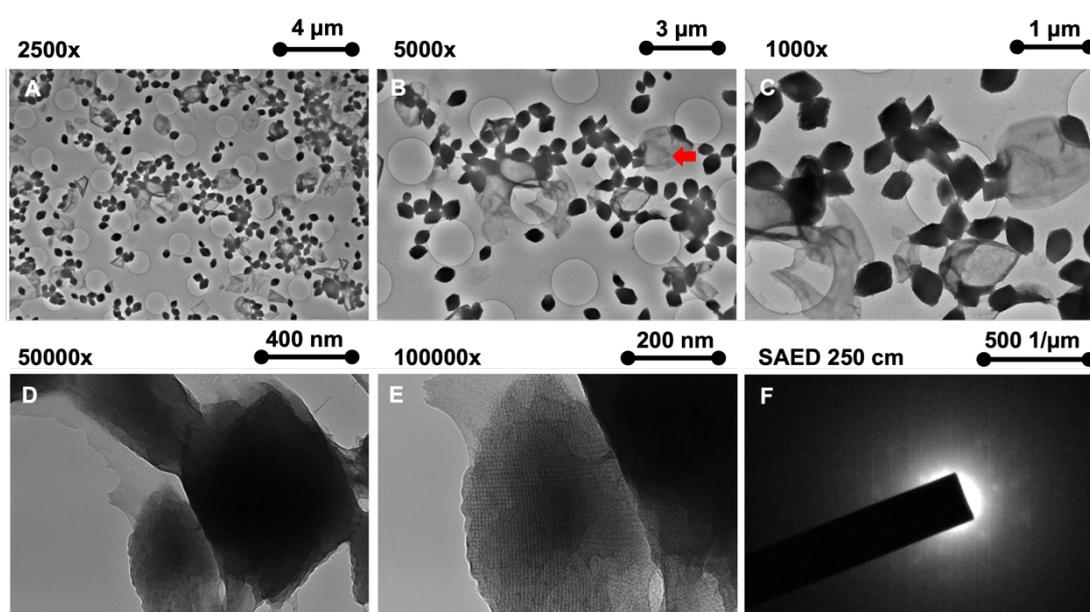


**Figure 4.1. Production of Cry48Aa1 and Tpp49Aa1 nanocrystals.** (A) Coomassie stained SDS-PAGE showing Cry48Aa1 (lane 2, ~135 kDa) and Tpp49Aa1 (lane 3, ~53 kDa) crystal samples. A band ~105 kDa is thought to correspond to the Tpp49Aa1 dimer. M (marker): BLUeye Pre-Stained Protein Ladder (kDa). (B) Sucrose density gradient of lysed sporulated culture showing example crystal band. (C) Diagram of sucrose gradient set up.

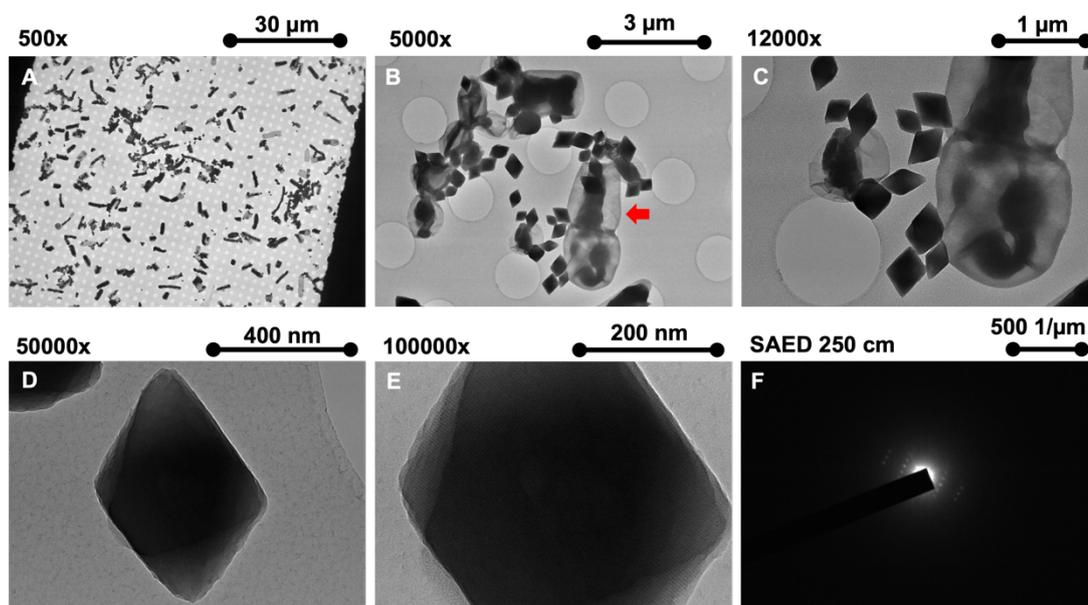
#### 4.2.1.2. Transmission electron microscopy (TEM) of nanocrystals

*TEM was employed to assess the crystal quality prior to data collection. TEM experiments were carried out by Dr. Robin Schubert in the XBI lab at the EuXFEL (Han et al. 2021), as previously described in **section 3.2.1.2**. For both proteins,*

TEM revealed tetragonal crystals with a diameter of approximately 500 nm (Fig 4.2, 4.3). The crystal borders of Cry48Aa1 (Fig 4.2D, 4.2E) appeared less defined in comparison to Tpp49Aa1 (Fig 4.3D, 4.3E). In SAED mode, electron diffraction was seen for both crystal proteins (Fig 4.2F, 4.3F). In both samples, remaining spore material (Fig 4.2B, 4.3B – red arrow) were identified. TEM analysis confirmed the presence of diffraction-quality crystals suitable for SFX studies.



**Figure 4.2. Transmission electron microscopy on negative stained Cry48Aa1 nanocrystals.** (A - E) Cry48Aa1 native nanocrystals with remaining spore material indicated by red arrow. (E) Crystal lattice of Cry48Aa1 nanocrystals. (F) Electron diffraction shown by selected area electron diffraction imaging (SAED). Images obtained from Dr. Robin Schubert (EuXFEL – Schenefeld, Germany).



**Figure 4.3. Transmission electron microscopy on negative stained Tpp49Aa1 nanocrystals.** (A – E) Tpp49Aa1 native nanocrystals with remaining spore material indicated by red arrow. (E) Crystal lattice of Tpp49Aa1 nanocrystals. (F) Electron diffraction shown by selected area electron diffraction imaging (SAED). Images obtained from Dr. Robin Schubert (EuXFEL – Schenefeld, Germany).

#### 4.2.2. Serial femtosecond crystallography (SFX) of Tpp49Aa1

To investigate the structure of Tpp49Aa1 in natural crystals, SFX at the EuXFEL was performed.

##### 4.2.2.1. Diffraction data collection and processing

The process of SFX diffraction data collection and processing has been previously described in relation to the Cry8Ba2 structure (**section 3.2.2.1**). Briefly, MHz SFX (Wiedorn et al. 2018a; Yefanov et al. 2019) diffraction data were collected for Tpp49Aa1 across two beamtimes (p2442 and p2819) on the SPB/SFX beamline at the EuXFEL using the novel nano-focus option to match the beam size to that of the crystals. Crystal samples

were prepared by filtration and centrifugation before being transferred to high-pressure reservoirs that were subsequently loaded onto the fast liquid-jet based injection system (Wiedorn et al. 2018b) with 3D-printed (Knoška et al. 2020) DFFN (Oberthuer et al. 2017). In beamtime p2442, the Tpp49Aa1 crystals diffracted to a resolution of 2.22 Å (PDB 7QA1). Subsequently, the resolution was significantly improved in beamtime p2819 where Tpp49Aa1 crystals diffracted to 1.62 Å (PDB 8BEY – diffraction pattern shown in **Fig 4.4**). In this chapter, the higher resolution structure will be presented.

*As previously described (**section 3.2.2.1**), the early stages of data processing were performed by Dr. Marina Galchenkova and Dr. Oleksandr Yefanov (DESY – Hamburg, Germany) following the method described by Wiedorn et al. (Wiedorn et al. 2018a). Here, the Cheetah program (Barty et al. 2014) was employed to perform hit finding, with a total of 2,458,059 diffraction patterns identified. From this, 157,582 of the strongest patterns were indexed and integrated using the XGandalf (Gevorkov et al. 2019) indexing method in CrystFEL (White et al. 2012; White et al. 2016) version 0.10.1. Diffraction patterns were indexed in the space group  $p2_12_12_1$  of the orthorhombic crystal system and the unit cell parameters were defined as:  $a = 79.65 \text{ \AA}$ ,  $b = 83.11 \text{ \AA}$ ,  $c = 156.91 \text{ \AA}$  and  $\alpha = \beta = \gamma = 90.0^\circ$  (**Table 4.1**). The indexed diffraction patterns were merged using the Partialator program in CrystFEL to produce a reflection (.hkl) file which was converted and exported as an MTZ file for data analysis and structure solution.*

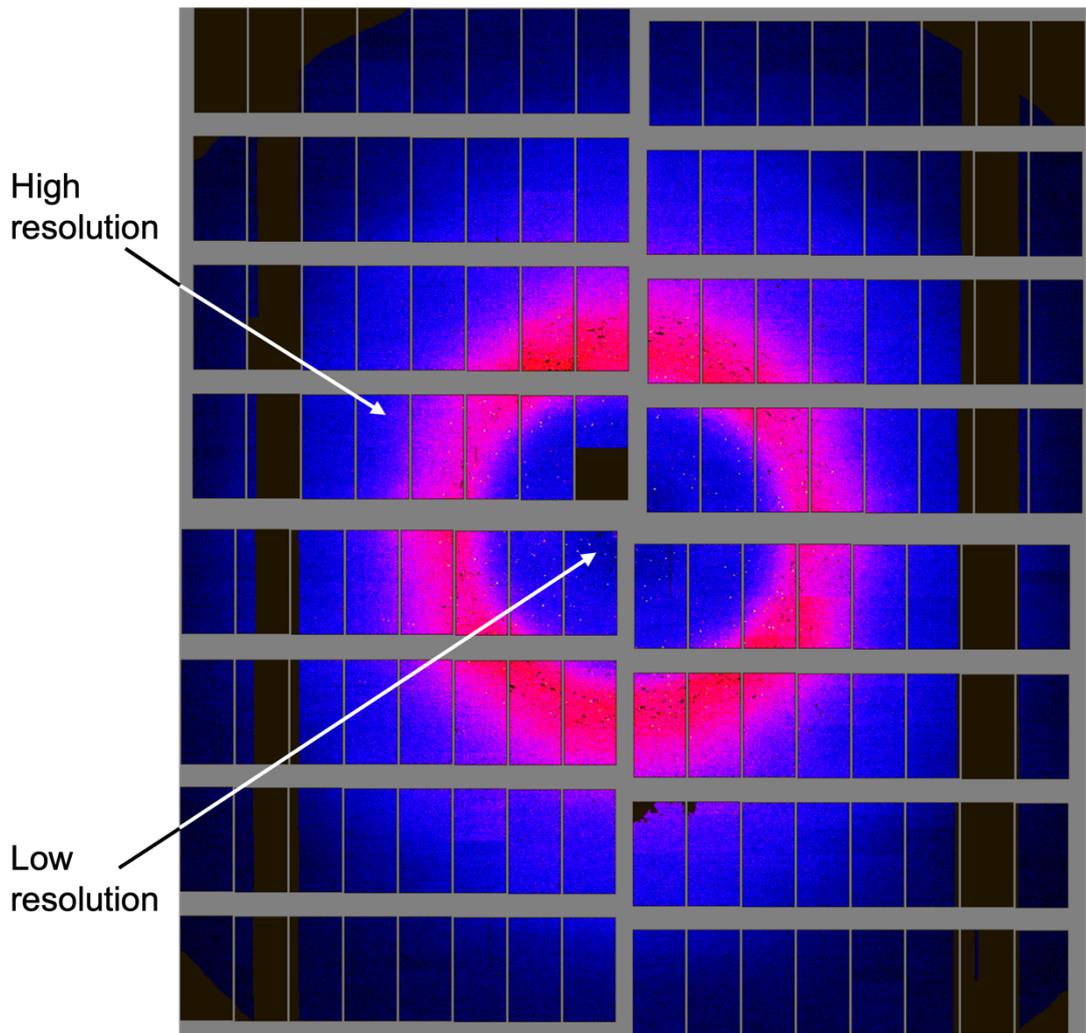


Figure 4.4. X-ray diffraction pattern (frame) of Tpp49Aa1.

**Table 4.1. Data collection and refinement statistics for Tpp49Aa1.**

<b>Data Collection</b>			
PDB ID	8BEX	8BEY	8BEZ
Beamline	SPB/SFX at European XFEL	SPB/SFX at European XFEL	SPB/SFX at European XFEL
X-ray Energy (keV)	9.3	9.3	9.3
Wavelength (Å)	1.33	1.33	1.33
<b>Crystal Data (figures in brackets refer to outer resolution shell)</b>			
pH	3	7	11
a,b,c (Å)	79.89, 82.84, 157.88	79.65, 83.11, 156.91	80.12, 83.22, 156.49
$\alpha,\beta,\gamma$ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Space group	P 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Resolution (Å)	1.78 – 36.01	1.62 - 24.98	1.75 – 36.12
Outer shell	1.78 – 1.80	1.62-1.64	1.75-1.77
R-split (%)	8.59 (231.27)	6.07 (299.67)	6.40 (177.89)
CC*	0.999 (0.492)	0.999 (0.484)	0.999 (0.632)
CC1/2	0.997 (0.138)	0.997 (0.133)	0.998 (0.249)
I / $\sigma$ (I)	10.50 (0.81)	10.61 (0.36)	14.81 (0.75)
Completeness (%)	99.51 (92.76)	100 (100)	99.82 (97.34)
Multiplicity	183.70 (6.3)	1028.30 (186.7)	323.70 (10.6)
Total Measurements	18469814 (38681)	136507469(163 5083)	34281129 (71761)
Unique Reflections	100541 (6177)	132750 (8756)	105904 (6782)
Wilson B-factor(Å <sup>2</sup> )	27.96	21.79	23.82
<b>Refinement Statistics</b>			
Refined atoms	7,234	7,357	7,324
Protein atoms	6,867	6,877	6,854
Non-protein atoms	0	0	0
Water molecules	367	480	470
R-work reflections	95,333	130,631	100,468
R-free reflections	5,033	6,921	5,313
R-work/R-free (%)	18.9 / 21.3	17.8 / 19.7	17.7 / 19.9
<b>rms deviations (target in brackets)</b>			
Bond lengths (Å)	0.011 (0.013)	0.013 (0.013)	0.012 (0.013)
Bond Angles (°)	1.472 (1.648)	1.446 (1.648)	1.443 (1.648)
<sup>1</sup> Coordinate error (Å)	0.105	0.077	0.088
Mean B value (Å <sup>2</sup> )	32.724	31.368	29.318
<b>Ramachandran Statistics (PDB Validation)</b>			
Favoured/allowed/Outliers	812 / 15 / 1	814 / 14 / 0	811 / 17 / 0
%	98.1 / 1.8 / 0.1	98.3 / 1.7 / 0	98 / 2 / 0

<sup>1</sup> Coordinate Estimated Standard Uncertainty in (Å), calculated based on maximum likelihood statistics. Table taken from Williamson et al., 2023.

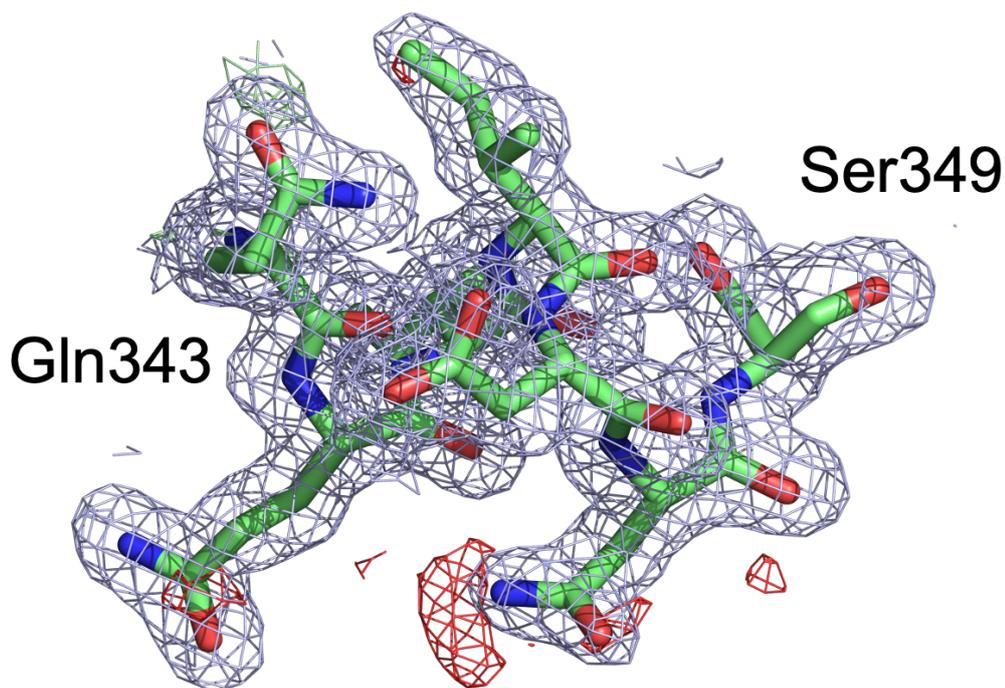
#### 4.2.2.2. Molecular replacement (MR) and structure solution

Several phasing methods exist and include MR, isomorphous replacement, and anomalous scattering. MR requires a structurally similar model usually with a sequence identity of >25%. Here, a blast search showed 28.3% sequence identity with the Tpp2Aa2 protein of the Tpp1Aa2/Tpp2Aa2 crystal structure at pH 7 (PDB 5FOY – chain B) and pH 5 (PDB 5G37 – chain B) (Colletier et al. 2016), as well as 25.4% sequence identity with the Tpp2Aa3 protein (PDB 3WA1) (Srisucharitpanit et al. 2014).

*The initial phasing of the lower resolution Tpp49Aa1 structure (PDB 7QA1) was performed by Dr. Dominik Oberthur (DESY – Hamburg, Germany). MR was performed using the phasing pipeline MRage in Phenix (McCoy et al. 2007; Adams et al. 2010) and L. sphaericus Tpp1Aa2/Tpp2Aa2 (PDB 5FOY and 5G37 – chain B) and L. sphaericus Tpp2Aa3 (PDB 3WA1) as templates, followed by an initial round of refinement using phenix.phase\_and\_build. Matthews' analysis using the MATTPROB web server was employed to predict the solvent content and number of molecules in the asymmetric unit. The asymmetric unit of the crystal contained two molecules which, according to PISA analysis performed later in this work, formed a stable dimer. As per this prediction, phasing was performed using a stoichiometry of two as input. The initial model was subjected to another round of Phaser, and phenix.autobuild (Terwilliger et al. 2007) was subsequently used for automated model building. The output model from these initial steps showed that two non-crystallographic symmetry (NCS)-related domains (A1 and A2 – the N-terminal domains) had been switched (e.g., A1-B2, A2-B1, where B1 and B2 represent the C-terminal domains), meaning that fragments from one monomer had been assigned as belonging to another monomer. To correct this error, Coot (Emsley et al. 2010) was used to switch the domains and the residues linking the A and*

*B* domains were built manually (A1-B1, A2-B2). Following this, the model was iteratively refined and built using *phenix.refine* (Afonine et al. 2012) and *Coot*.

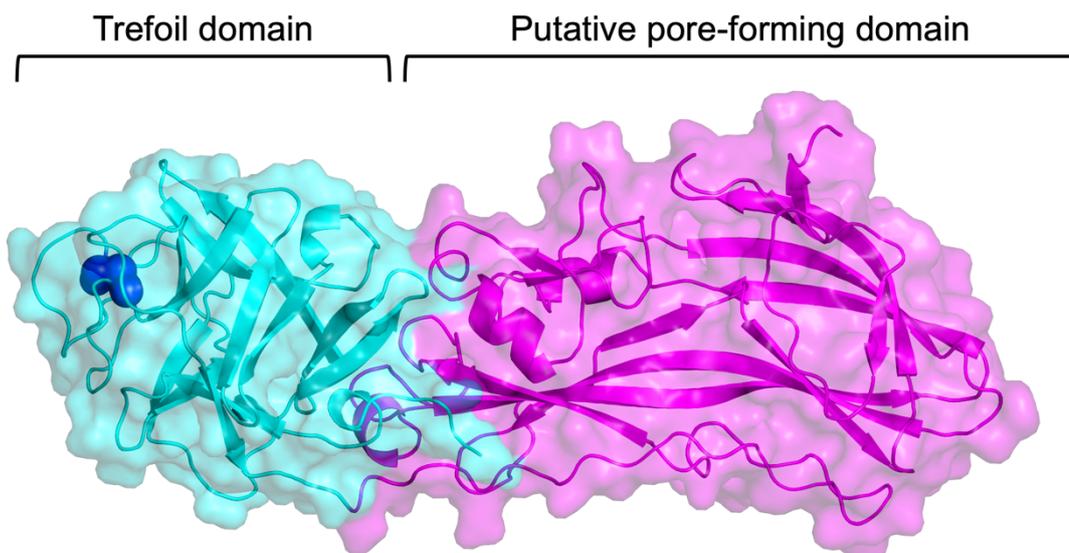
The higher resolution dataset collected in beamtime p2819 was subsequently phased using the model obtained at 2.2 Å (PDB 8BEY). MR was performed using the phasing pipeline Phaser in the CCP4i2 software suite (Potterton et al. 2018). This was followed by iterative rounds of refinement and manual model building cycles in Refmac5 (Murshudov et al. 2011) and *Coot*, respectively, until the R/Rfree values reached convergence. This led to a final model with R/Rfree 0.178/0.197 and a resolution of 1.62 Å (**Table 4.1**). The electron density map (**Fig 4.5**) showed continuous density for residues 49 – 464 of the protein sequence. Previous work has confirmed the presence of the first 48 residues in Tpp49Aa1 crystals produced by native *L. sphaericus* strains (Jones et al. 2007). To confirm whether these residues were also present in Tpp49Aa1 crystals recombinantly expressed in *B. thuringiensis*, we performed N-terminal sequencing as described in **section 2.5.9**. The sequence was identified as MXNQ(E), where X indicates a position at which the amino acid could not be identified, and the residue in brackets indicates a mismatch with the authentic Tpp49Aa1 sequence. The authentic N-terminal Tpp49Aa1 sequence, MENQI, otherwise corresponds with that identified by N-terminal sequencing (and no internal MXNQE sequence exists in the protein), suggesting that the N-terminal region is present in the crystals. Given that this region is not seen in the electron density map, it is, therefore, apparently disordered.



**Figure 4.5. Partial electron density map and model of the Tpp49Aa1 structures elucidated at pH 7.** 2Fo-Fc electron density map contoured at the 1.0  $\sigma$  level (coloured blue) and Fo-Fc electron density map contoured at the  $\pm 3.0$   $\sigma$  level (positive coloured green, negative coloured red). Protein backbone is coloured green, nitrogen atoms are coloured blue, oxygen atoms are coloured red. Residues Gln343 – Ser349 are shown.

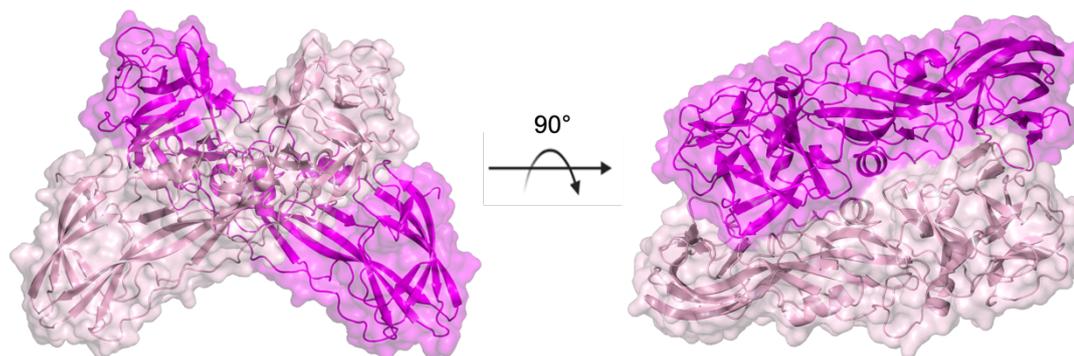
#### 4.2.2.3. General features of the Tpp49Aa1 structure

Structure solution revealed that the Tpp49Aa1 monomer (**Fig 4.6**) is composed of an N-terminal Ricin\_B-like\_Lectin (SSF50379) head domain comprising a  $\beta$ -trefoil fold (spanning residues 49 – 214) and a C-terminal Toxin\_10 (PF05431) tail domain with a  $\beta$ -sheet rich topology (spanning residues 215 – 464).



**Figure 4.6. Crystal structure of Tpp49Aa1.** The Tpp49Aa1 monomer is composed of two distinct domains. The N-terminal Ricin\_B-like\_lectin (SSF50379) head domain (cyan) comprises a  $\beta$ -trefoil fold and one disulphide bridge (dark blue spheres). The C-terminal Toxin\_10 (PF05431) tail domain (magenta) is  $\beta$ -sheet rich and exhibits structural homology with pore-forming Aerolysin-like proteins. Figure taken from Williamson et al., 2023.

Tpp49Aa1 crystals are formed by the packing of dimers that form an “X” structure with a large intermolecular interface (**Fig 4.7**). The two copies of Tpp49Aa1 are almost identical, as indicated by their all-atom RMSD of 0.681 Å.



**Figure 4.7. Tpp49Aa1 homodimers.** Tpp49Aa1 forms an ‘X’-shaped homodimer (magenta, light pink) with a large intermolecular interface. Figure adapted from Williamson et al., 2023.

For comparison with related structures in the PDB, the DALI web server was used (Holm 2020). The best matches were identified as other pesticidal proteins belonging to the Tpp family (PF05431), including Tpp1Aa2 (PDB 5FOY-A) from *L. sphaericus* (Colletier et al. 2016), Tpp2Aa2 (PDB 5FOY-B) from *L. sphaericus* (Colletier et al. 2016), Tpp2Aa3 (PDB 3WA1-A) from *L. sphaericus* (Srisucharitpanit et al. 2014), Tpp78Aa1 (PDB 7Y78-A) from *B. thuringiensis* (Cao et al. 2022), Tpp80Aa1 (PDB 8BAD-A) from *B. thuringiensis* (Best et al. 2022), and Tpp35Ab1 (PDB 4JP0-B) from *B. thuringiensis* (Kelker et al. 2014) (**Table 4.2**). Carbohydrate binding proteins, including the carbohydrate binding module (CBM)-13 of 1,3Gal43A (PDB 3VSF-A) from *Clostridium thermocellum* (Jiang et al. 2012) and the Hemagglutinin (HA1) subcomponent of the type-C 16S toxin (PDB 3AH1-B) from *Clostridium botulinum* (Nakamura et al. 2008), were also identified (**Table 4.2**). Overall, greatest structural similarity was identified between Tpp49Aa1 and Tpp2Aa2 (**Table 4.2**).

**Table 4.2. Structural similarity of Tpp49Aa1 with other proteins.**

<b>PDB ID_chain</b>	<b>Name (formally)</b>	<b>Z-score*</b>	<b>RMSD (Å)</b>
5FOY_B	Tpp2Aa2 (BinBA2)	41.4	2.1
3WA1_A	Tpp2Aa3 (BinBA3)	38.3	2.3
5FOY_A	Tpp1Aa2 (BinAA2)	37.9	2.5
7Y78_A	Tpp78Aa1 (Cry78Aa1)	34.1	2.7
8BAD_A	Tpp80Aa1 (Cry80Aa1)	33.4	4.2
4JP0_A	Tpp35Ab2 (Cry35Ab2)	29.3	3.5
3AH1_B	Hemagglutinin	18.6	2.2
3VSF_A	1,3Gal43A CBM13	17.3	2.0

\* DALI Z-scores larger than 2 are considered significant. Table taken from Williamson et al., 2023.

Two distinct domains were identified within the Tpp49Aa1 monomer. The N-terminal domain comprises a  $\beta$ -trefoil fold (**Fig 4.8A, 4.8B**), which itself is composed of six two-stranded  $\beta$ -hairpins arranged into three subdomains with pseudo three-fold symmetry, designated  $\alpha$  ( $\beta$ 1- $\beta$ 4),  $\beta$  ( $\beta$ 5- $\beta$ 8), and  $\gamma$  ( $\beta$ 9- $\beta$ 12), alongside one disulphide bridge, Cys91-Cys183 (**Fig 4.8B**). Viewed from the side, the  $\beta$ -trefoil fold can also be subdivided into a  $\beta$ -barrel with a triangular cap (**Fig 4.8A**).

$\beta$ -trefoil folds are characteristic of carbohydrate binding proteins, including ricin B-like lectins (Hazes 1996). They commonly consist of three homologous repeats, usually of approximately 40 – 45 residues, with a conserved QxW motif (where x denotes any residue) capable of binding carbohydrates. The repeating nature of the sequence gives rise to the three-fold symmetry. A role for carbohydrate moieties in eliciting pesticidal action of Cry proteins has been demonstrated in several studies (Knowles et al. 1984; Dennis et al. 1986; Sangadala et al. 2001; Jurat-Fuentes et al. 2002) and recently reviewed (Best et al. 2023). Carbohydrate-mediated receptor interaction has also been demonstrated in the Tpp family. Specifically, Best

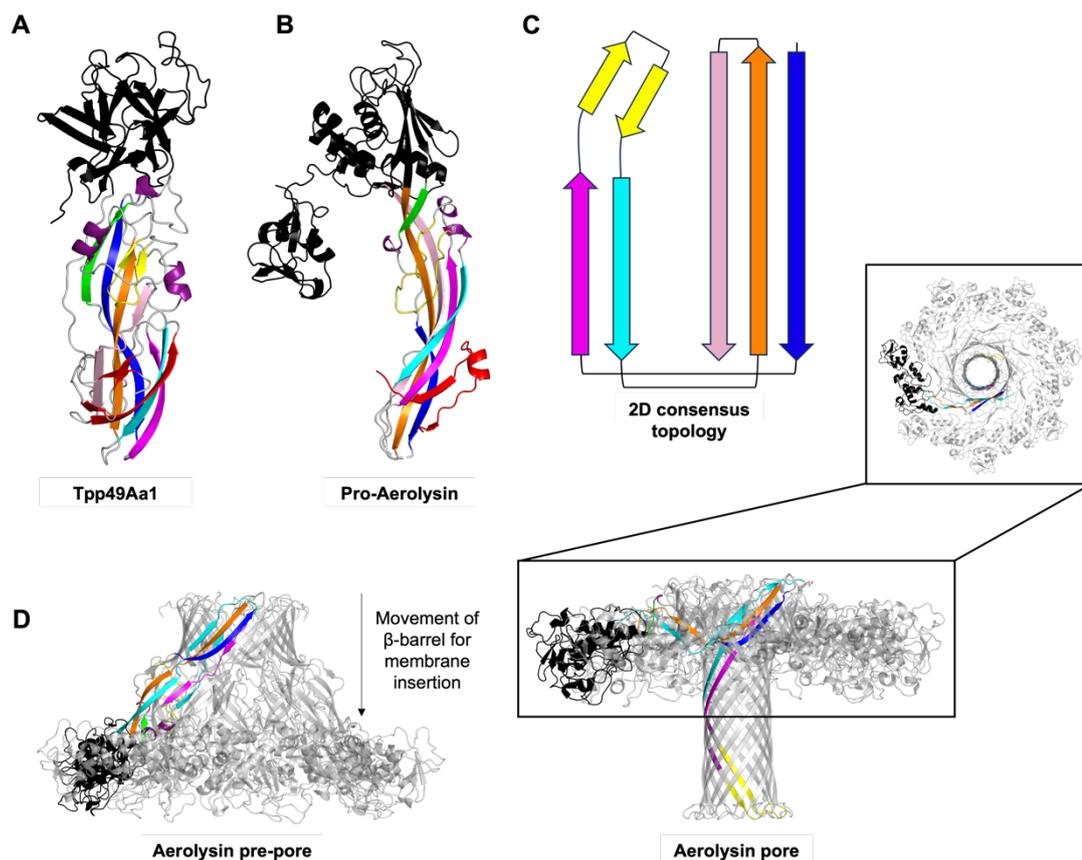
et al. (2022) showed that Tpp80Aa1 is able to interact with galactose-containing glycolipids. Furthermore, addition of galactose/GalNAc to toxin-treated *Cx. quinquefasciatus* MRA-918 cells had a protective effect against Tpp80Aa1, suggesting that binding of free galactose/GalNAc prevents Tpp80Aa1 from binding its putative receptor (Best et al. 2022). Hence, it is possible that carbohydrate binding may also be significant for pesticidal activity of the Cry48Aa1/Tpp49Aa1 toxin pair. In contrast, sequence analysis of Tpp49Aa1 indicates that several loop insertions distort the symmetry of the  $\beta$ -trefoil domain in comparison to other Tpp class members, including Tpp80Aa1, and other carbohydrate binding proteins, such as 1,3Gal43A CBM13, both of which display clear threefold-symmetry, as indicated by their protein structure and sequence alignment of the  $\alpha$ ,  $\beta$ ,  $\gamma$  subdomains (**Fig 4.8D, 4.8E**). Distortion of the trefoil domain is not uncommon within the Tpp class. Colletier et al. (2016) demonstrated that, relative to its partner protein, Tpp1Aa2, the  $\beta$ -trefoil fold of Tpp2Aa2 contains four loop insertions, including one disulphide-linked tether, Cys67-Cys161, which disrupts the trefoil's three-fold symmetry and, furthermore, appears to obstruct the  $\alpha$ -carbohydrate binding module. Given this, the authors hypothesized that carbohydrate binding may play less of a role in the mode of action of Tpp2Aa2 (Colletier et al. 2016). The Tpp2Aa2 Cys67-Cys161 disulphide bridge is conserved in the Tpp49Aa1 structure (**Fig 4.8B**), where it also appears to obstruct the  $\alpha$ -carbohydrate binding module. Further to this, the QxW motif is not present in the Tpp49Aa1  $\beta$ -trefoil sequence. Clearly, further studies aimed at deciphering the role glycoconjugates in Cry48Aa1/Tpp49Aa1 mode of action are required.



module. (C) Sequence alignment of the Tpp49Aa1  $\beta$ -trefoil subdomains (with the positions of  $\beta$ -strands highlighted cyan, magenta, and grey, respectively) shows several loop insertions which disrupt the threefold symmetry. (D, E) For comparison with Tpp49Aa1, top-down views of the N-terminal  $\beta$ -trefoil domains in (D) Tpp80Aa1 and (E) the carbohydrate binding module of 1,3Gal43A in complex with glycerol (purple sticks). Sequence alignment of the 3 subdomains in these proteins (below the structures in D and E), highlights the homologous repeats which give rise to the threefold symmetry in the  $\beta$ -trefoil fold in these proteins. The positions of  $\beta$ -strands within the 3 subdomains have been highlighted cyan, magenta, and grey, respectively. The “\*” indicates positions with a single, fully conserved residue; “:” indicates conservation between groups of strongly similar properties; “.” indicates conservation between groups of weakly similar properties. Sequence alignment was produced using Clustal Omega.

The C-terminal tail domain comprises a  $\beta$ -sheet rich topology characteristic of the Toxin\_10 family (PF05431), and wider Aerolysin (PF01117), ETX/MTX-2 (PF03318) superfamily of  $\beta$ -pore forming proteins (Parker et al. 1994; Szczesny et al. 2011). Across this superfamily, a core topology consisting of five antiparallel  $\beta$ -strands and a short  $\beta$ -hairpin, proposed to act as the transmembrane segment following pore formation (Iacovache et al. 2006), is conserved (**Fig 4.9A – C**). The length and twists of the  $\beta$ -strands vary across family members and further  $\beta$ -strands, in addition to the core five mentioned here, alongside  $\alpha$ -helices may also be present. The Tpp49Aa1 C-terminal domain comprises four additional  $\beta$ -strands (one preceding and three following the core topology), as well as three short  $\alpha$ -helices, two of which lie either side of the  $\beta$ -hairpin (**Fig 4.9A**). The pore formation process of Aerolysin (**Fig 4.9B**), which relies on the C-terminal domain, is well characterized, with cryoEM structures for both the pre-pore state (PDB 5JZH) and pore state (PDB 5JZT, embedded in lysomyristoylphosphatidylglycerol – LMPG – micelles) having been elucidated (**Fig 4.9D**) (Iacovache et al. 2016). The inactive pro-form of Aerolysin is secreted as a head-to-tail homodimer that is subsequently

activated by proteolytic cleavage to remove the C-terminal propeptide (Parker et al. 1994; Iacovache et al. 2016). In the inactive form, the C-terminal propeptide acts as a peg to prevent premature pore formation. Activation causes the homodimer to dissociate, allowing the N-terminal receptor binding domains (domains I and II) to interact with target N-glycosylated GPI-anchored receptors (Abrami et al. 2002). Domain I interacts with the N-linked sugar modifications, whilst domain II binds directly to the glycan core of the receptor (Abrami et al. 2002). Seven Aerolysin monomers oligomerize via the pore-forming C-terminal domain, leading to the formation of a pre-pore structure constituting an amphipathic  $\beta$ -barrel (Iacovache et al. 2016). Hinge-like movements cause the heptamer to collapse inwards, leading to insertion of the  $\beta$ -barrel into the cell membrane (Iacovache et al. 2016). Given the high structural homology between the C-terminal domain of Tpp49Aa1 and Aerolysin (**Fig 4.9A, 4.9B**), it is possible that Tpp49Aa1 forms pores by a similar mechanism.



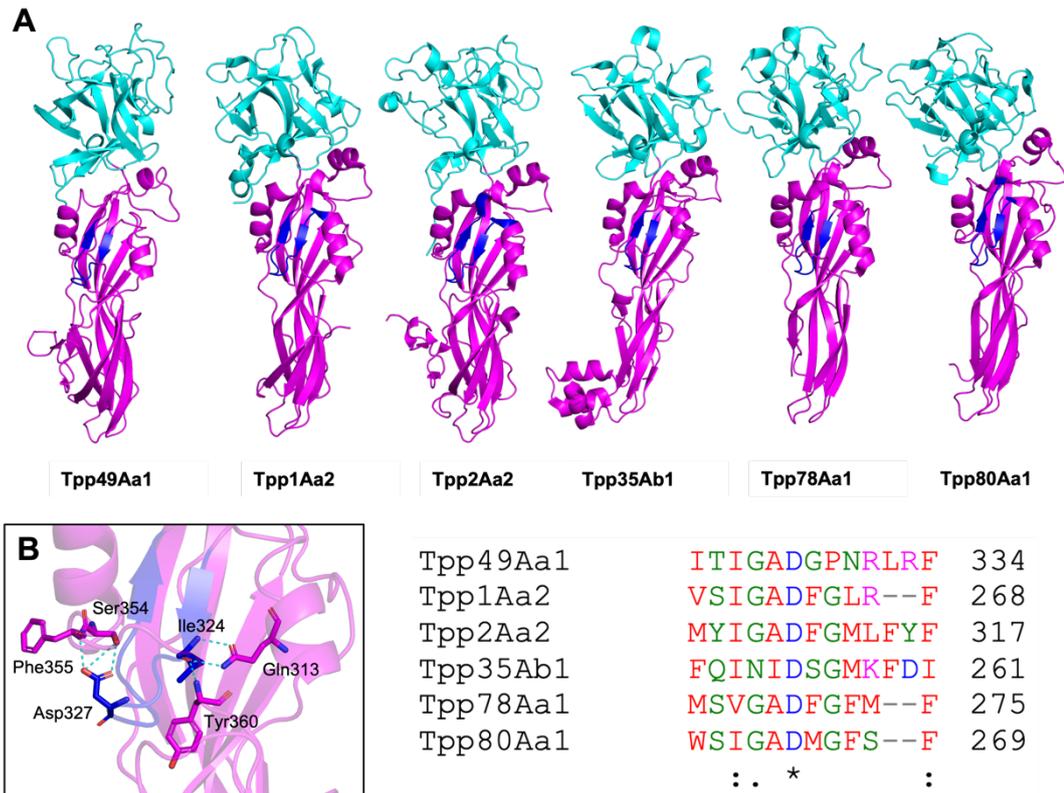
**Figure 4.9. Structural comparison of Tpp49Aa1 and Aerolysin C-terminal domains.** (A) Tpp49Aa1 C-terminal domain topology. Across the Aerolysin, ETX/MTX-2 superfamily of  $\beta$ -pore forming proteins, a core C-terminal domain topology consisting of five antiparallel  $\beta$ -strands ( $\beta$ 1 – blue,  $\beta$ 2 – magenta,  $\beta$ 3 – cyan,  $\beta$ 4 – orange,  $\beta$ 5 – light pink) and a short  $\beta$ -hairpin (yellow), proposed to act as the transmembrane segment during pore formation, is conserved. The length and twists of the  $\beta$ -strands vary across family members and additional  $\beta$ -strands, as well as short  $\alpha$ -helices, may also be present. The Tpp49Aa1 C-terminal domain comprises one additional  $\beta$ -strand (green) preceding the core topology and three additional  $\beta$ -strand (red) following the core topology, as well as three short  $\alpha$ -helices (purple), two of which lie either side of the  $\beta$ -hairpin. The receptor binding domain is coloured black. (B) Pro-Aerolysin (PDB 1PRE) C-terminal domain topology showing equivalent regions in the same colour scheme. Pro-Aerolysin contains a short C-terminal peptide (red) which is cleaved during activation. The two receptor binding domains are coloured black. (C) 2D schematic of the C-terminal domain core topology conserved across the Aerolysin, ETX/MTX-2 superfamily. (D)

Aerolysin pre-pore (PDB 5JZH) and quasi-pore (PDB 5JZT), showing equivalent regions of one monomer in the same colour scheme as the pro-form. The short  $\beta$ -hairpin seen in the pro-Aerolysin structure (yellow) forms the transmembrane  $\beta$ -barrel of the pore.

Related Tpp structures available in the PDB were identified using DALI (**Fig 4.10, Table 4.2**). All Tpp proteins identified display an overall architecture comprising an N-terminal  $\beta$ -trefoil fold and C-terminal putative pore-forming domain consistent with that of the Tpp49Aa1 structure. The greatest variability exists in the N-terminal domain, specifically within the exposed surface loops linking the trefoil subdomains ( $\alpha$ ,  $\beta$ ,  $\gamma$ ). Given that the N-terminal domain has been implicated in receptor binding (Guo et al. 2016), variations in the loop regions of the trefoil fold provide information about target specificity, and may be used to explain why Tpp49Aa1, which displays greatest structural similarity with Tpp2Aa2 (all-atom RMSD = 2.1 Å), is able to target Tpp1Aa2/Tpp2Aa2-resistant mosquitoes (Jones et al. 2007; Jones et al. 2008), possibly by interaction with distinct receptor classes.

All Tpp proteins identified from the DALI search contain a short  $\beta$ -hairpin with predominantly amphipathic structure (**Fig 4.10A**), which is conserved across the wider Aerolysin, ETX/MTX-2 superfamily, and known to unfold in Aerolysin to form the  $\beta$ -barrel pore in the target cell membrane (Iacovache et al. 2016). In the Tpp family, the  $\beta$ -hairpin is tucked under a loop, and held in place by several polar interactions between conserved residues, including aspartic acid (Asp327 in Tpp49Aa1), indicated in the sequence alignment (**Fig 4.10B**). These regions have been proposed as the transmembrane domains in Tpp49Aa1 (residues 322 - 334), Tpp1Aa2 (residues 256 – 268), Tpp2Aa2 (residues 302 – 317), Tpp78Aa1 (residues 265 – 275), Tpp80Aa1 (residues 259 – 269), and Tpp35Ab1 (residues 249 – 261) (Lacomel et al. 2021) and it has been hypothesized that the polar contacts between the  $\beta$ -hairpin and adjacent loop act to hold the putative transmembrane region in place, preventing premature pore formation. We speculate that this region may unfold to form the  $\beta$ -barrel pore in an

Aerolysin-like mechanism. However, given the necessity for Cry48Aa1 in the action of Tpp49Aa1, it also cannot be ruled out that pore formation occurs via unique mechanisms involving both proteins.

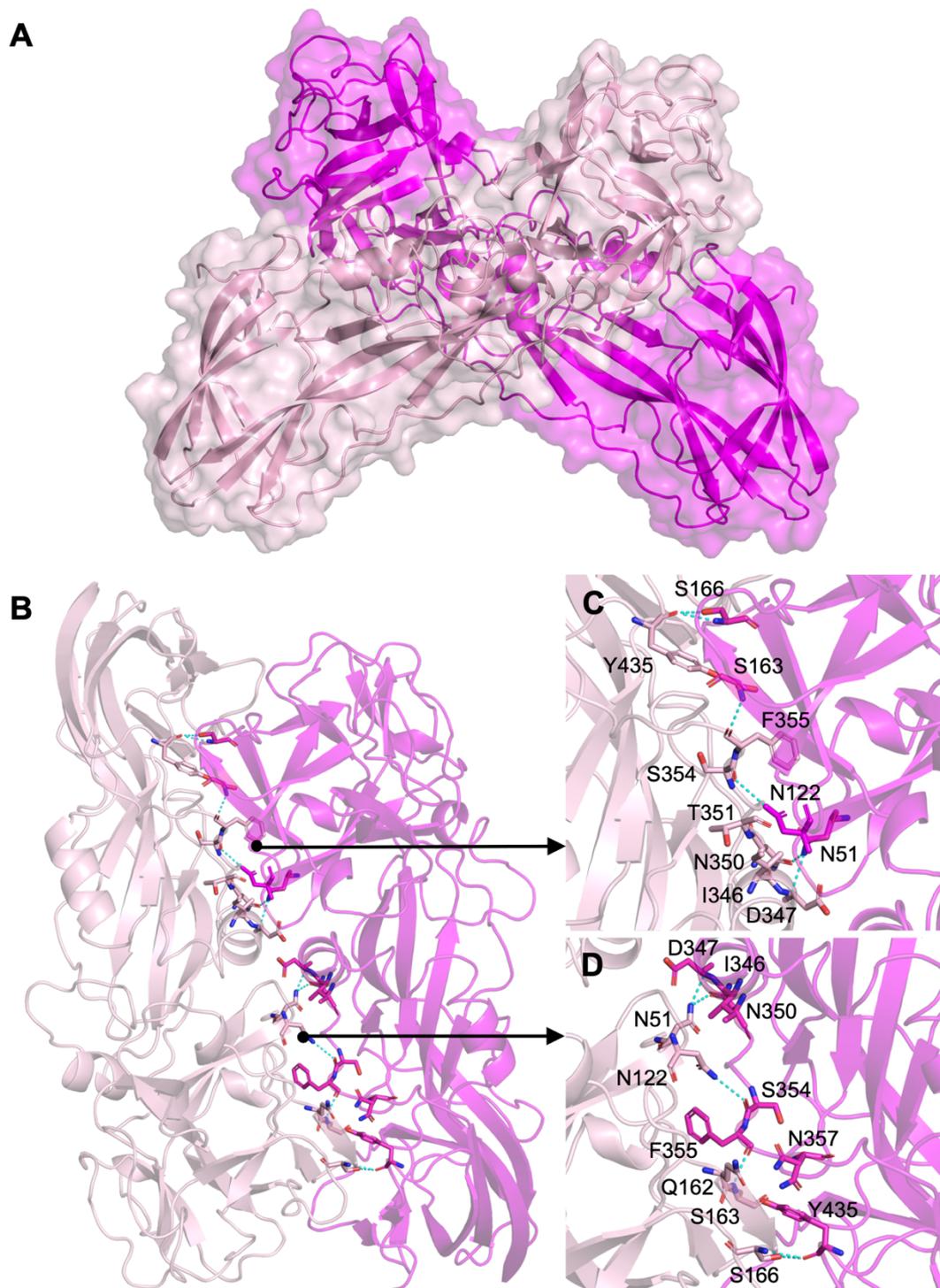


**Figure 4.10. Structural comparison of Tpp49Aa1 with other Tpp family members. (A)** The structures of Tpp family members: Tpp49Aa1 (PDB 8BEY-A), Tpp1Aa2 (PDB 5FOY-A) and Tpp2Aa2 (PDB 5FOY-B) from *L. sphaericus*, Tpp78Aa1 (PDB 7Y78-A), Tpp80Aa1 (PDB 8BAD-A) and Tpp35Ab1 (PDB 4JP0-B) from *B. thuringiensis*. The N-terminal Ricin\_B-like\_lectin head domain is shown in cyan and the C-terminal Toxin\_10 tail domain is shown in magenta. A conserved  $\beta$ -hairpin with predominantly amphipathic structure, tucked under a loop, is highlighted in blue. This region has been proposed as the transmembrane domains in Tpp49Aa1 (residues 322 - 334), Tpp1Aa2 (residues 256 - 268), Tpp2Aa2 (residues 302 - 317), Tpp78Aa1 (residues 265 - 275), Tpp80Aa1 (residues 259 - 269), and Tpp35Ab1 (residues 249 - 261). **(B)** The  $\beta$ -hairpin loop contains a conserved aspartic acid and isoleucine /

valine residue that forms polar contacts (cyan dashed lines) with neighbouring residues. Sequence alignment of the  $\beta$ -hairpin region. Residues are highlighted red – small + hydrophobic, blue – acidic, magenta – basic, green – hydroxyl + sulfhydryl + amine; “\*” indicates positions with a single, fully conserved residue; ‘:’ indicates conservation between groups of strongly similar properties; “.” indicates conservation between groups of weakly similar properties. Sequence alignment was produced using Clustal Omega. Figure adapted from Williamson et al., 2023.

Structure solution revealed that Tpp49Aa1 crystals are formed by the packing of dimers that form an “X” structure with a large intermolecular interface (**Fig 4.11**), also seen in the natural heterodimeric crystals of Tpp1Aa2/Tpp2Aa2 (Colletier et al. 2016). Analysis using PISA (Krissinel and Henrick 2007) indicates that the Tpp49Aa1 dimer interface has a total surface area of 1329.1 Å<sup>2</sup> and a predicted binding energy of -11.1 kcal mol<sup>-1</sup>. Monomers A and B contribute 41 and 42 residues respectively, forming 16 hydrogen bonds. No salt bridges or disulphide bonds were identified. In comparison, the Tpp1Aa2/Tpp2Aa2 dimer interface has a total surface area of 1833.1 Å<sup>2</sup> and a predicted binding energy of -22.5 kcal mol<sup>-1</sup>. Tpp1Aa2 contributes 49 residues to the interface, whilst Tpp2Aa2 contributes 63 residues, making 19 hydrogen bonds and 2 salt bridges. Taken together, this indicates a more stable complex for Tpp1Aa2/Tpp2Aa2 heterodimers than for Tpp49Aa1 homodimers. Interaction of the solubilised and activated forms of Tpp1Aa2 and Tpp2Aa2 is required for pesticidal activity of this toxin pair, possibly explaining the more stable complex identified in natural crystals (although it is also possible that Tpp1Aa2/Tpp2Aa2 dimers dissociate during crystal dissolution and a new interface, caused by interaction of the solubilised and activated protein, is formed). In contrast, interaction of Tpp49Aa1 may only occur in the crystal to support packing and stability prior to delivery to the target insect midgut. Indeed, an interesting feature of the Tpp family relates to differences in mode of action, in which some family members (Tpp36, Tpp78, Tpp80) act alone to exert pesticidal activity, whilst others require interaction with a second pesticidal protein. This may be

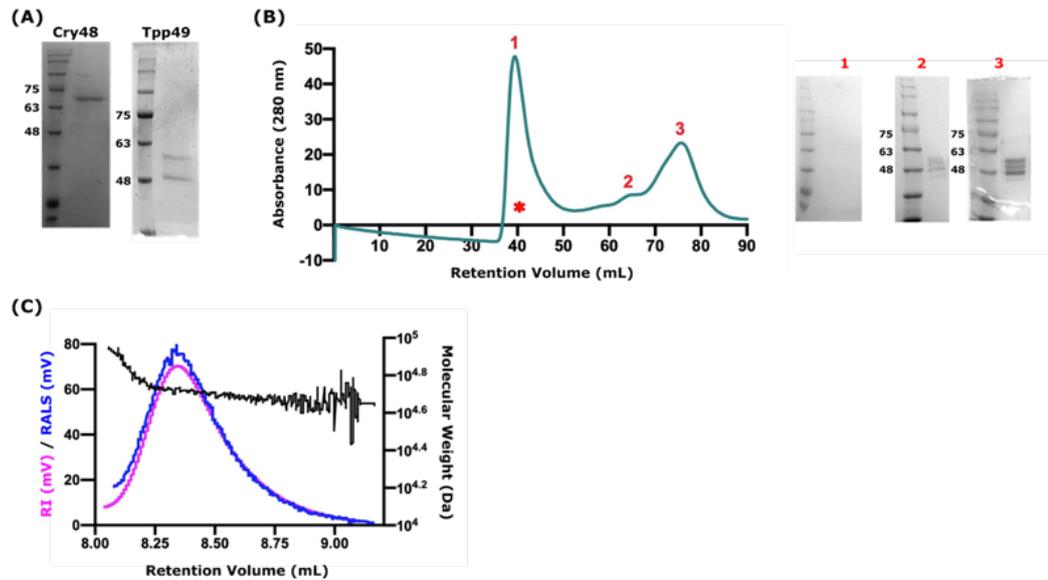
another Tpp family member (Tpp1/Tpp2) or a member of a different pesticidal protein family. For example, Tpp49Aa1 requires interaction with Cry48Aa1, whilst Tpp35 requires interaction with Gpp34.



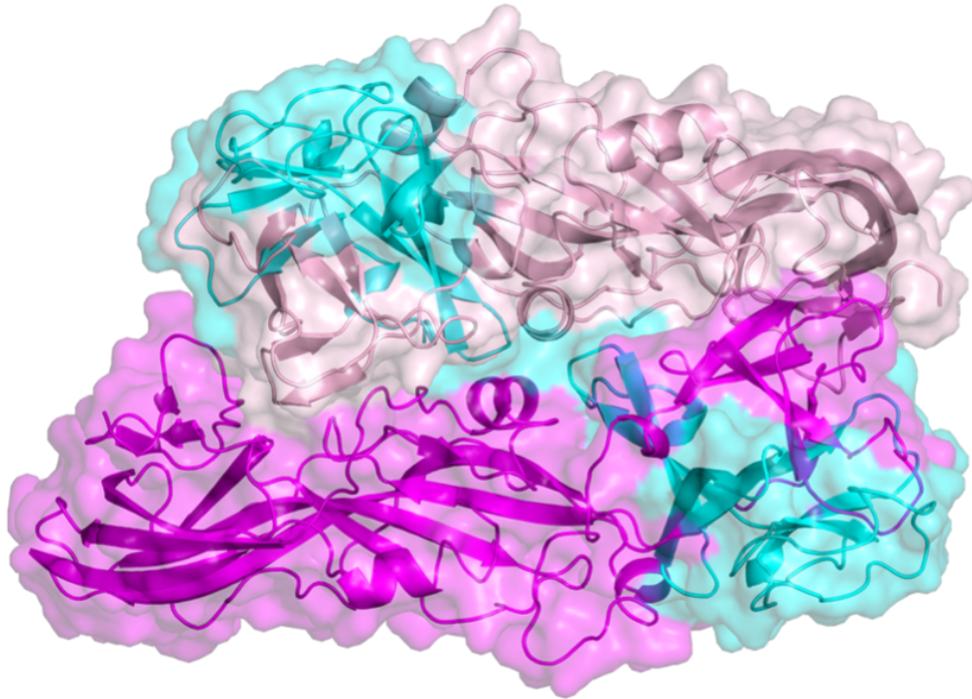
**Figure 4.11. Structure and interactions of the Tpp49Aa1 homodimer.** (A) Tpp49Aa1 crystals are assembled by the packing of dimers (monomer A – magenta, monomer B – light pink) that form an “X” structure with a large intermolecular interface. (B) Tpp49Aa1 dimer interface showing polar contacts (cyan dashed lines) identified in PyMOL. Residues involved are shown as sticks (carbon – magenta / light pink, oxygen – red, nitrogen – blue). (C, D)

Detailed view of Tpp49Aa1 dimer interactions. Figure taken from Williamson et al., 2023.

To investigate whether the dimer of Tpp49Aa1 is maintained in solution, the solubilised proteins were analysed by SEC, RALS and RI measurements in collaboration with Dr. Hannah Best (Cardiff University). In line with published work, solubilised Tpp49Aa1 migrates on SDS-PAGE as two bands, sized approximately 49 kDa and 55 kDa (**Fig 4.12**). The smaller molecular weight (MW) band may represent the activated Tpp49Aa1 species, cleaved by bacterial proteinases present in the sample preparation. SEC showed a main peak (~51 kDa) as well as a shoulder peak (~121 kDa), suggesting that Tpp49Aa1 is predominantly monomeric in solution, with some dimers persisting (**Fig 4.12**). Further analysis using RALS and RI, quantitative techniques that enable the MW of the protein to be calculated, indicated that, at pH 8.5, Tpp49Aa1 is present in solution at a single MW of 52.1 kDa (**Fig 4.12**), close to the expected MW of the monomer (53 kDa). Previous studies using dot blot assays suggest that the Tpp49Aa1 N-terminal region, Asn49 – Ser148, interacts with Cry48Aa1 (Guo et al. 2016). Given that this region is partially buried within the Tpp49Aa1 dimer interface (**Fig. 4.13**), we hypothesize that dissociation may be required to allow interaction with Cry48Aa1.



**Figure 4.12. Tpp49Aa1 is predominantly monomeric in solution.** (A) Cry48Aa1 and Tpp49Aa1 crystal proteins solubilised in 50 mM Na<sub>2</sub>CO<sub>3</sub> pH 10.5 + 0.05% β-mercaptoethanol. Cry48Aa1 is present at ~70 kDa and Tpp49Aa1 is present as two bands ~49 and ~55 kDa. (B) SEC profile of solubilised Tpp49Aa1 indicated three absorption peaks (280 nm). SDS-PAGE analysis of fractions collected, indicated that Tpp49Aa1 was present in peaks 2 and 3, corresponding to dimeric and monomeric protein, respectively. (C) Static light scattering (RALS – blue) and refractive index (RI – pink) analysis of protein-containing peaks 2 and 3 (combined) indicated a molecular weight (black) of 52.1 kDa when calibrated against bovine serum albumin (BSA), close to the expected molecular weight of 53 kDa. Figure taken from Williamson et al., 2023.



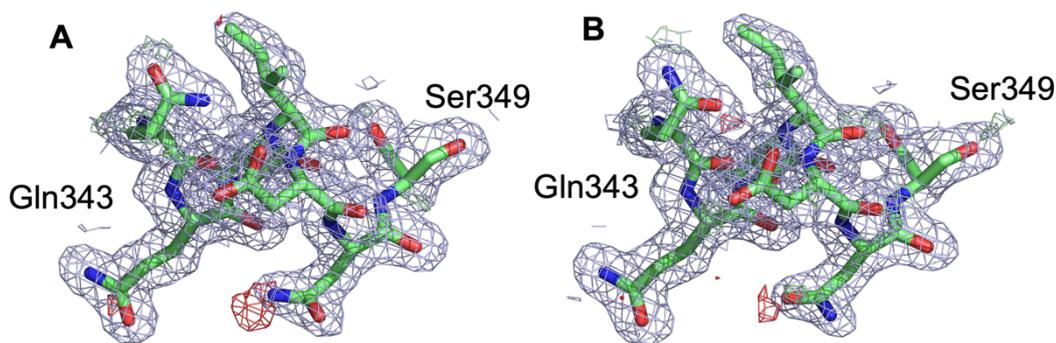
**Figure 4.13. Tpp49Aa1 regions shown to interact with Cry48Aa1 are partially buried in the dimer.** The Tpp49Aa1 dimer structure (magenta, light pink) demonstrates that regions shown to interact with Cry48Aa1 in dot blot assays (cyan) are partially buried. Figure taken from Williamson et al., 2023.

#### **4.2.2.4. pH mixing studies of the Tpp49Aa1 native crystals**

Since data are collected in a serial fashion, SFX can be applied to perform substrate mixing and time-resolved studies, yielding dynamic structural information relating to protein function. In the field of bacterial pesticidal proteins, pH mixing studies have been performed on native *B. thuringiensis* and *L. sphaericus* crystals to probe the early structural events leading to crystal dissolution. Specifically, in their work with Tpp1Aa2/Tpp2Aa2 native crystals Colletier et al. (2016) showed that exposure to elevated pH (pH 10) led to several early structural events, including the unravelling of an  $\alpha$ -helix present in the pro-peptide region of Tpp1Aa2, the loss of intermolecular hydrogen bonds that, at high pH, were

deprotonated leading to electrostatic repulsion, as well as the loss of pH-labile intermolecular salt bridges.

Given that Tpp49Aa1 crystals are readily solubilised in the alkaline environment of the mosquito larval midgut, crystals were mixed with pH 11 buffer using a T-junction mixer at the SPB/SFX instrument of the EuXFEL. Mixing occurred approximately 1 minute prior to the crystals entering the beam. A total of 707,992 diffraction patterns were collected from which 426,506 could be indexed in space group  $P2_12_12_1$ , ( $a = \sim 80.12 \text{ \AA}$ ;  $b = \sim 83.22 \text{ \AA}$ ;  $c = \sim 156.49 \text{ \AA}$ ;  $\alpha = \beta = \gamma = 90^\circ$ ) (**Table 4.1**). Changes in the dimensions of the unit cell in comparison to data collected at pH 7 ( $a = \sim 79.65 \text{ \AA}$ ;  $b = \sim 83.11 \text{ \AA}$ ;  $c = \sim 156.91 \text{ \AA}$ ;  $\alpha = \beta = \gamma = 90^\circ$ ) indicated possible movements in the protein structure (**Table 4.1**). MR was performed using the phasing pipeline Phaser (McCoy et al. 2007) in the CCP4i2 software suite (Potterton et al. 2018) and the native (pH 7) Tpp49Aa1 structure as the template. The output model was subjected to iterative rounds of manual model building using Coot (Emsley et al. 2010) and refinement using Refmac5 (Murshudov et al. 2011), leading to a final model with R/Rfree 0.177/0.199 and a resolution of 1.75  $\text{\AA}$  (**Table 4.1**). As in the pH 7 structure, the electron density maps (**Fig 4.14**) showed continuous density for residues 49 – 464 of the protein sequence.



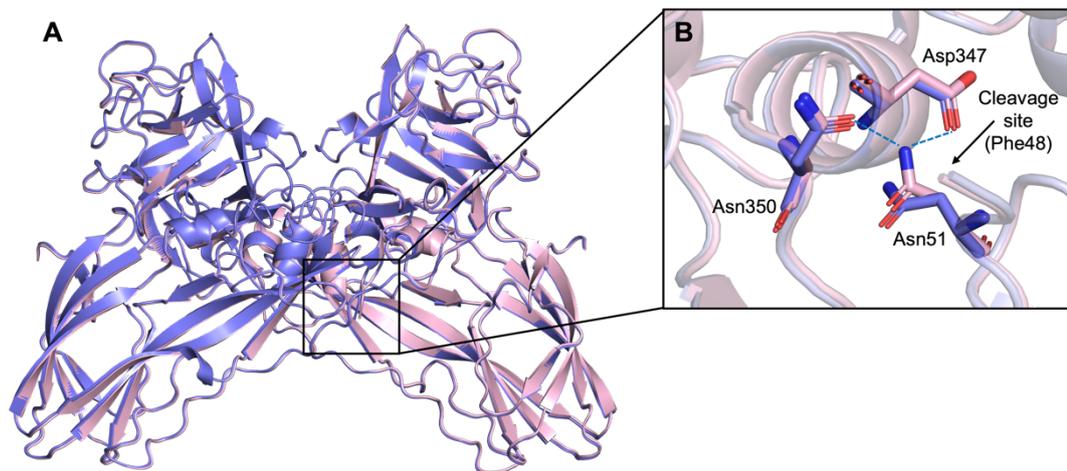
**Fig 4.14. Partial electron density map and model of the Tpp49Aa1 structures elucidated at (A) pH 3 (B) pH 11.** 2Fo-Fc electron density map contoured at the 1.0  $\sigma$  level (coloured blue) and Fo-Fc electron density map contoured at the  $\pm 3.0$   $\sigma$  level (positive coloured green, negative coloured red). Protein backbone is coloured green, nitrogen atoms are coloured blue, oxygen atoms are coloured red. Residues Gln343 – Ser349 are shown.

The pH 7 and pH 11 Tpp49Aa1 structures align closely, as indicated by their superposition (**Fig. 4.15**) and all-atom RMSD of 0.681 Å. The largest structural movements appeared to occur within the surface exposed loops of the Tpp49Aa1 monomer. Alkalinity also disrupted the eight interfaces produced by the packing of Tpp49Aa1 dimers into native crystals, identified by analysis using PISA (**Tables 4.3, 4.4, 4.5**). Specifically, an increase in pH caused the predicted overall binding energy of the eight crystal interfaces (excluding the dimer interface) to increase by 1.8 kcal mol<sup>-1</sup>, translating to a weaker affinity. In contrast, the predicted binding energy of the dimer decreased by 0.8 kcal mol<sup>-1</sup>, translating to a stronger affinity. Taken together, this may suggest that initial dissolution of the crystals releases dimers, prior to the dimers dissociating themselves.

Further to this, elevated pH induced a total of 18 changes in hydrogen bonds and salt bridges (**Table 4.6**) across both the crystal and dimer interfaces. Given the proximity to the pro-peptide cleavage site (Phe48), loss of hydrogen bonds Asn51[ND2]-Asp347[O] and Asn51[ND2]-Asn350[OD1], which are present at the dimer interface in both Tpp49Aa1 monomers,

constituted the most notable change (**Fig 4.15**). Specifically, we hypothesize that loss of these interactions would increase accessibility of the pro-peptide for subsequent proteolytic cleavage. Outside of the dimer interface, loss of a salt bridge, Asp429(OD1)-Arg267(NH1), supports the proposal that intermolecular salt bridges play a role in supporting crystal stability of *B. thuringiensis* and *L. sphaericus* pesticidal proteins.

To investigate the effect of low pH on Tpp49Aa1, crystals were also mixed with pH 3 buffer. A total of 466,741 diffraction patterns were collected from which 279,362 could be indexed in space group  $P2_12_12_1$ , ( $a = \sim 79.89 \text{ \AA}$ ;  $b = \sim 82.84 \text{ \AA}$ ;  $c = \sim 157.88 \text{ \AA}$ ;  $\alpha = \beta = \gamma = 90^\circ$ ), again indicating a change in the unit cell dimensions in comparison to the pH 7 dataset. MR was performed using the phasing pipeline Phaser (McCoy et al. 2007) in the CCP4i2 software suite (Potterton et al. 2018) and the native (pH 7) Tpp49Aa1 structure as the template. The output model was subjected to iterative rounds of manual model building using Coot (Emsley et al. 2010) and refinement using Refmac5 (Murshudov et al. 2011), leading to a final model with R/Rfree 0.189/0.213 and a resolution of 1.78  $\text{\AA}$  (**Table 4.1**). Although not physiologically relevant to the pH environment of mosquito larval guts, similar, but somewhat more extreme changes were seen in crystals exposed to low pH in comparison to elevated pH. Specifically, a decrease in pH caused the predicted overall binding energy of the eight crystal interfaces (excluding the dimer interface) to increase by 5.0 kcal mol<sup>-1</sup> (in comparison to an increase of 1.8 kcal mol<sup>-1</sup> in crystals exposed to high pH). Further to this, a total of 23 changes in hydrogen bonds and salt bridges (**Table 4.7**) were identified across both the crystal and dimer interfaces. Taken together, these results suggests that crystals may dissolve more readily at low pH.



**Figure 4.15. pH mixing studies of Tpp49Aa1.** (A) Superposition of Tpp49Aa1 structures obtained at pH 7 (pink) and pH 11 (purple). Pro-peptide cleavage site (Phe48) is boxed. (B) Loss of hydrogen bonds Asn51[ND2]-Asp347[O] and Asn51[ND2]-Asn350[OD1] at the pro-peptide cleavage site following mixing at pH 11.

**Table 4.3. Calculated structural properties of the molecular interfaces in the Tpp49Aa1 crystal at pH 7.**

Interface	Monomer 1			Monomer 2			Interface area (Å <sup>2</sup> )	$\Delta^iG^{**}$ (kcal mol <sup>-1</sup> )	H-bonds	Salt bridges	Binding energy (kcal mol <sup>-1</sup> )
	Chain	Symmetry	AA	Chain	Symmetry	AA					
1*	B	x,y,z	42	A	x,y,z	41	1329.1	-4.0	16	0	-11.1
2	A	x,y,z	13	B	x-1/2,-y-1/2,-z	13	470.2	-0.6	10	2	-5.8
3	B	x,y,z	12	A	-x,y-1/2,-z-1/2	13	445.1	-0.3	7	1	-3.8
4	B	x,y,z	12	A	-x+1/2,-y,z-1/2	10	383.4	-8.0	0	0	-8.0
5	B	x,y,z	20	B	x-1/2,-y-1/2,-z	15	361.3	-0.1	5	0	-2.3
6	A	x,y,z	12	A	-x,y-1/2,-z-1/2	8	147.9	0.9	1	0	0.5
7	A	x,y,z	6	B	x-1,y,z	6	169.1	-1.3	0	1	-1.7
8	A	x,y,z	3	B	-x-1/2,-y-1,z-1/2	3	91.6	0.5	1	3	-1.0
9	B	x,y,z	1	A	x,y-1,z	2	12.4	0.3	0	0	0.3
Total***	—	—	—	—	—	—	2081	-8.6	24	7	-21.8

\* Interface 1 refers to the Tpp49Aa1 dimer interface. \*\* Predicted solvation free energy gain. \*\*\* Total excludes dimer interface.

Table taken from Williamson et al., 2023.

**Table 4.4. Calculated structural properties of the molecular interfaces in the Tpp49Aa1 crystal at pH 11.**

Interface	Monomer 1			Monomer 2			Interface area (Å <sup>2</sup> )	$\Delta^i G^{**}$ (kcal mol <sup>-1</sup> )	H-bonds	Salt bridges	Binding energy (kcal mol <sup>-1</sup> )
	Chain	Symmetry	AA	Chain	Symmetry	AA					
1*	B	x,y,z	38	A	x,y,z	41	1312.8	-5.2	15	0	-11.9
2	A	x,y,z	13	B	x-1/2,-y-1/2,-z	13	477.8	0.2	9	1	-4.2
3	B	x,y,z	13	A	-x,y-1/2,-z-1/2	13	458.3	-1.0	7	1	-4.5
4	B	x,y,z	12	A	-x+1/2,-y,z-1/2	11	396.4	-7.0	0	0	-7.0
5	B	x,y,z	19	B	x-1/2,-y-1/2,-z	15	355.3	-0.5	4	0	-2.3
6	A	x,y,z	10	A	-x,y-1/2,-z-1/2	7	131.5	0.9	1	0	0.5
7	A	x,y,z	6	B	x-1,y,z	6	180.1	-1.4	0	1	-1.7
8	A	x,y,z	3	B	-x-1/2,-y-1,z-1/2	3	89.6	0.5	1	3	-1.1
9	B	x,y,z	2	A	x,y-1,z	3	12.8	0.3	0	0	0.3
Total***	—	—	—	—	—	—	2101.8	-8	22	6	-20

Annotations as in Table 4.3.

Table 4.5. Calculated structural properties of the molecular interfaces in the Tpp49Aa1 crystal at pH 3.

Interface	Monomer 1			Monomer 2			Interface area (Å <sup>2</sup> )	$\Delta^iG^{**}$ (kcal mol <sup>-1</sup> )	H-bonds	Salt bridges	Binding energy (kcal mol <sup>-1</sup> )
	Chain	Symmetry	AA	Chain	Symmetry	AA					
1*	B	x,y,z	40	A	x,y,z	40	1333.5	-5.0	16	0	-12.1
2	A	x,y,z	14	B	x-1/2,-y-1/2,-z	14	466.5	-1.1	6	0	-3.8
3	B	x,y,z	12	A	-x,y-1/2,-z-1/2	13	445.0	-0.4	9	0	-4.4
4	B	x,y,z	10	A	-x+1/2,-y,z-1/2	11	354.0	-7.7	0	0	-7.7
5	B	x,y,z	18	B	x-1/2,-y-1/2,-z	11	305.2	1.9	3	0	0.5
6	A	x,y,z	11	A	-x,y-1/2,-z-1/2	6	145.4	1.1	1	0	0.6
7	A	x,y,z	6	B	x-1,y,z	5	113.7	-1.0	0	1	-1.3
8	A	x,y,z	3	B	-x-1/2,-y-1,z-1/2	3	89.2	1.1	1	4	-0.9
9	B	x,y,z	1	A	x,y-1,z	2	7.1	0.2	0	0	0.2
Total***	–	–	–	–	–	–	1926.1	-5.9	20	5	-16.8

Annotations as in Table 4.3.

**Table 4.6. Tpp49Aa1 interactions perturbed by an increase in pH to 11.**

Interface	Monomer 1	Monomer 2	Interactions
1	x,y,z	x,y,z	<b>Loss of:</b> B:Asn51(ND2) – A:Asp347(O) B:Asn51(ND2) – A:Asn350(OD1) B:Gln162(OE1) – A:Asn357(ND2) B:Asn347(O) – A:Asn51(ND2) B:Asn350(OD1) – A:Asn51(ND2) B:Thr351(N) – A:Asn51(OD1) <b>Formation of:</b> B:Asn51(ND2) – A:Thr351(O) B:Asp120(O) – A:Gln343(NE2) B:Gln343(NE2) – A:Asp120(O) B:Thr351(O) – A:Asn51(ND2) B:Tyr435(OH) – A:Asn165(O)
2	x,y,z	x-1/2,-y-1/2,-z	<b>Loss of:</b> A:Asp429(OD1) – B:Arg267(NH1) * A:Tyr463(O) – B:Thr400(OG1)
3	x,y,z	-x,y-1/2,-z-1/2	<b>Loss of:</b> B:Tyr463(O) – A:Thr400(OG1) <b>Formation of:</b> B:Asp432(O) – A:Gln274(NE2)
4	x,y,z	-x+1/2,-y,z-1/2	
5	x,y,z	x-1/2,-y-1/2,-z	<b>Loss of:</b> B:Asp198(OD1) – B:Gln220(NE2)
6	x,y,z	-x,y-1/2,-z-1/2	
7	x,y,z	x-1,y,z	<b>Loss of:</b> A:His97(ND1) – B:Glu237(OE1)
8	x,y,z	-x-1/2,-y-1,z-1/2	<b>Loss of:</b> A:Asp88(N) – B:Asp88(OD1)
9	x,y,z	x,y-1,z	

\* Salt bridges. Table taken from Williamson et al., 2023.

**Table 4.7. Tpp49Aa1 interactions perturbed by a decrease in pH to 3.**

Interface	Monomer 1	Monomer 2	Interactions
1	x,y,z	x,y,z	<b>Loss of:</b> B:Gln162(OE1) – A:Asn357(ND2) <b>Formation of:</b> B:Asn51(OD1) – A:Thr351(N)
2	x,y,z	x-1/2,-y-1/2,-z	<b>Loss of:</b> A:Asp429(OD1) – B:Arg267(NH1) * A:Asp429(OD2) – B:Arg267(NH1) * A:Ser434(N) – B:Gln274(OE1) A:Asp429(OD2) – B:Arg267(NH1) A:Asp429(O) – B:Arg267(NH1) A:Asp432(O) – B:Gln274(NE2) A:Tyr463(O) – B:Thr400(OG1)
3	x,y,z	-x,y-1/2,-z-1/2	<b>Loss of:</b> B:Asp429(OD2) – A:Arg267(NH2) * B:Asp429(OD2) – A:Arg267(NH2) <b>Formation of:</b> B:Ser434(N) – A:Asp394(OD2) B:Asp432(O) – A:Gln274(NE2)
4	x,y,z	-x+1/2,-y,z-1/2	
5	x,y,z	x-1/2,-y-1/2,-z	<b>Loss of:</b> B:Arg196(NH2) – B:Gln220(OE1) B:Asn204(ND2) – B:Gln220(O) B:Asn206(ND2) – B:Asp394(O) <b>Formation of:</b> B:Asn204(ND2) – B:Gln220(OE1)
6	x,y,z	-x,y-1/2,-z-1/2	
7	x,y,z	x-1,y,z	<b>Loss of:</b> A:His97(ND1) – B:Glu237(OE1) * A:His96(ND1) – B:Glu237(OE1) <b>Formation of:</b> A:His97(NE2) – B:Glu237(OE2) *
8	x,y,z	-x-1/2,-y-1,z-1/2	<b>Loss of:</b> A:Asp88(N) – B:Asp88(OD1) <b>Formation of:</b> A:Asp88(OD2) – B:Arg92(NE) * A:Asp88(N) – B:Asp88(OD2)
9	x,y,z	x,y-1,z	

\* Salt bridges. Table taken from Williamson et al., 2023.

#### 4.2.3. Serial femtosecond crystallography (SFX) of Cry48Aa1

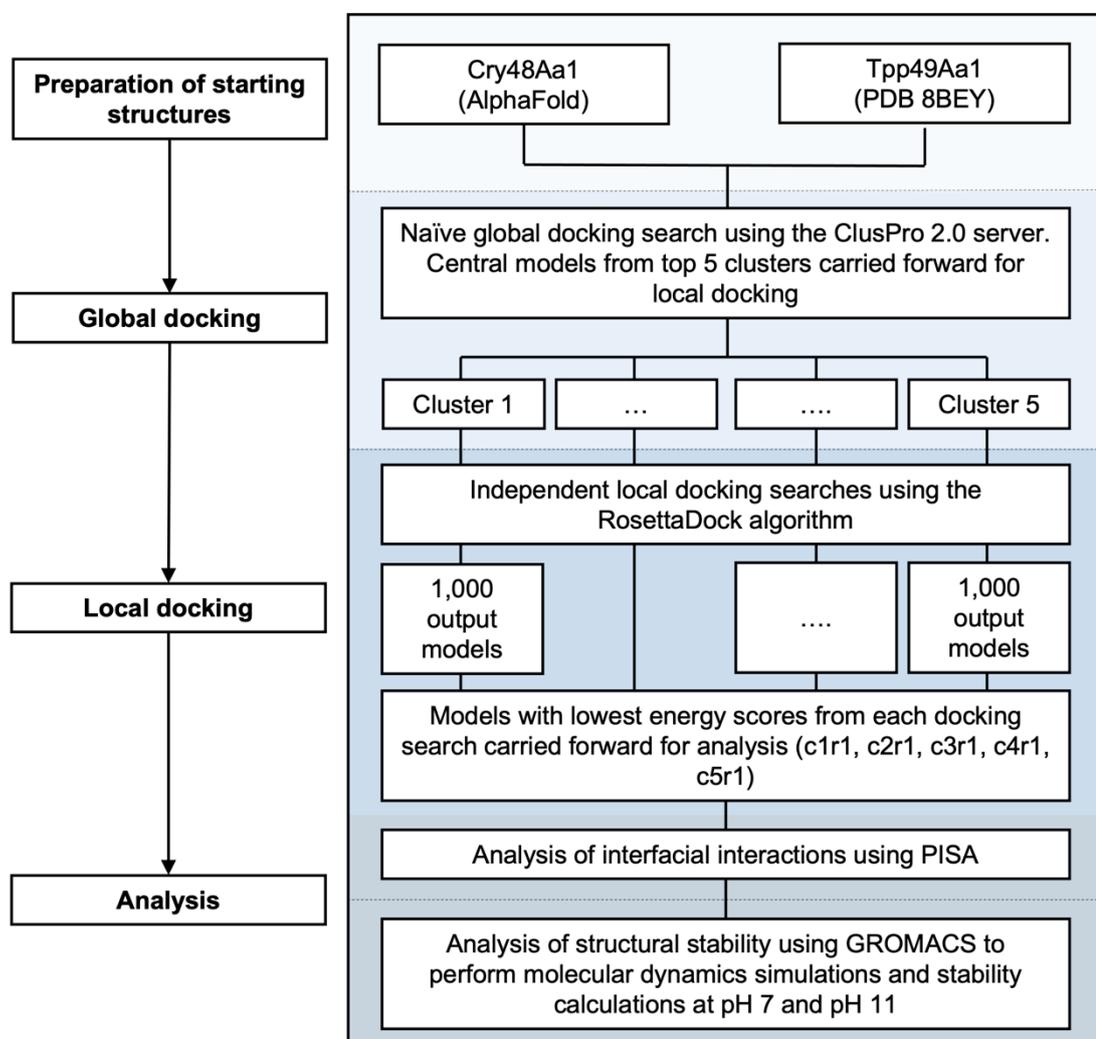
To investigate the structure of Cry48Aa1 (a seven domain Cry protein) in natural crystals, SFX studies were performed. Cry48Aa1 crystal proteins were prepared (**section 4.2.1.1**) and analysed by TEM (**section 4.2.1.2**) as previously shown. Diffraction data were collected across two beamtimes (p2442 and p2819) on the SPB/SFX beamline at the EuXFEL. However, across both beamtimes, Cry48Aa1 crystals diffracted poorly meaning that the structure could not be resolved.

#### 4.2.4. Prediction of the Cry48Aa1-Tpp49Aa1 interaction

Cell toxicity (De Melo et al. 2009) and mosquito larval bioassays have shown that Cry48Aa1 and Tpp49Aa1 are required in combination to elicit cell and/or insect death, with highest activity demonstrated at a 1:1 M ratios (Jones et al. 2008). In addition, dot-blot assays assessing deletion fragments have demonstrated an interaction between the two proteins, showing that the N-terminal region of Tpp49Aa1 (residues Asn49 – Ser148) is required for interaction with Cry48Aa1, whilst the C-terminal region (residues Ser349 – Asn464) is required for membrane interaction (Guo et al. 2016). The cooperation of this two-component system is not well understood and hence, to investigate further how this interaction may occur, computational modelling studies were performed.

An overview of the computational modelling procedure and rationale is provided here (**Fig 4.16**). The protein structures were first prepared for modelling, such that factors relating to their mode of action (including proteolytic processing and environment pH) had been considered. Following this, a molecular docking approach combining two programs (ClusPro and RosettaDock) was employed to perform global, unbiased protein-protein docking, followed by local refinement. Usually, the output structures are assessed according to the binding free energy, which itself is based on the thermodynamic hypothesis, stating that the native protein structure will exist at the lowest free energy (Anfinsen 1973). However, limitations exist and

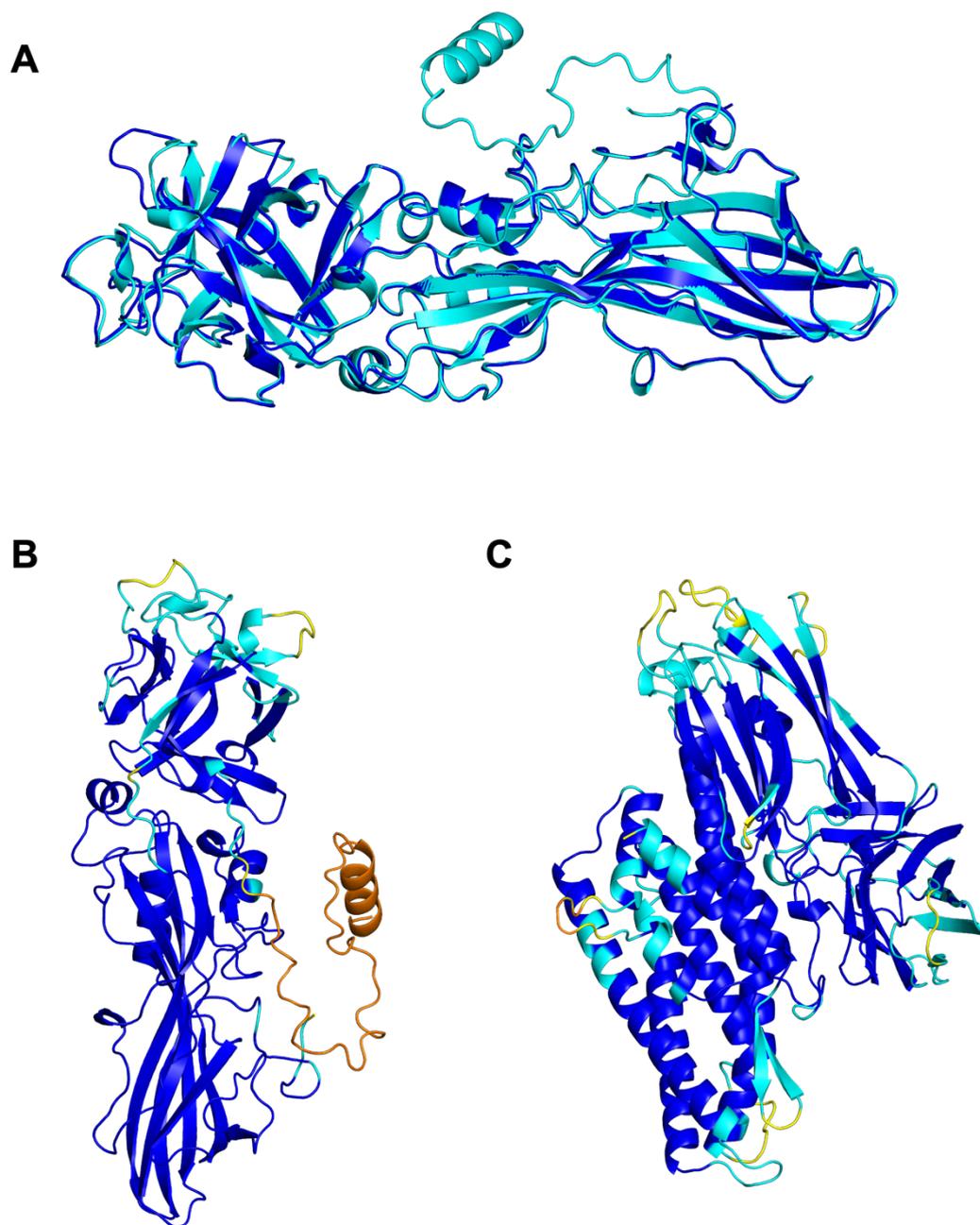
include the ability to identify the near-native structure from similarly scored, non-native structures, as well as the ability to model protein dynamics. In recent studies, MD has been employed as an additional method to evaluate docked protein-protein models, offering the ability to model the stability of protein-protein complexes which, in turn, may be used to assess the likelihood of a complex persisting in solution (Radom et al. 2018; Jandova et al. 2021; Karaca et al. 2022). Hence, the modelled Cry48Aa1-Tpp49Aa1 complexes were further scrutinised using MD simulations, alongside interface analysis and binding free energy calculations.



**Figure 4.16. Overview of Cry48Aa1-Tpp49Aa1 modelling.** Modified from Worthy et al., 2021.

#### 4.2.4.1. Preparation of Cry48Aa1 and Tpp49Aa1 starting structures

The structure of Tpp49Aa1 was obtained from the crystal structure elucidated as part of this work (PDB 8BEY). The Cry48Aa1 structure was not resolved and, therefore, the AF2 package (Jumper et al. 2021b) was used to predict the full-length structure. AF2 estimates a per-residue confidence metric known as the predicted local-distance difference test (pLDDT). The pLDDT score is embedded in the B-factor column of the PDB file and hence, can be visualised in molecular graphics programs, such as PyMOL. Regions of lower confidence are often associated with disordered (flexible) regions. Overall, the Cry48Aa1 structure was predicted with very high confidence, with some regions of medium confidence existing within the surface exposed loops of DI – DIII (**Fig 4.17**). In parallel, AF2 was used to predict the Tpp49Aa1 structure, thus enabling assessment of the suitability of AF2 to predict structures of pesticidal proteins by direct comparison between the predicted and experimentally resolved Tpp49Aa1 structure (**Fig 4.17**). The AF2 predicted Tpp49Aa1 model displayed an all-atom RMSD of 1.982 Å (calculated using PyMOL) when superposed with the Tpp49Aa1 structure (**Fig 4.17**). Taken together, this indicates the ability of AF2 to predict the structures of pesticidal proteins with high accuracy.



**Figure 4.17. Suitability of AlphaFold2 for prediction of pesticidal protein structures.** (A) Predicted Tpp49Aa1 structure (cyan) superposed with the crystal structure of Tpp49Aa1 elucidated at pH 7 (dark blue). The AF2 model of Tpp49Aa1 displayed an all-atom RMSD of 1.982 Å (calculated using PyMOL) with the crystal structure. (B) AF2 predictions of Tpp49Aa1 and (C) Cry48Aa1 coloured by the confidence metric (predicted local-distance difference test - pLDDT). pLDDT confidence scores are coloured dark blue (very high, pLDDT > 90), cyan (high, pLDDT > 70), yellow (low, pLDDT > 50),

orange (very low, pLDDT < 50). The N-terminal pro-peptide, not seen in the electron density map of Tpp49Aa1, is predicted with very low confidence.

Insect gut proteases (mainly trypsin-like and chymotrypsin-like) are involved in proteolytic cleavage of *B. thuringiensis* and *L. sphaericus* pesticidal proteins to produce the activated protein core (Rukmini et al. 2000). Jones et al. (2008) investigated the effect of *in vitro* incubation with *Cx. quinquefasciatus* gut extract, chymotrypsin, trypsin, or proteinase-K on the cleavage of Cry48Aa1 and Tpp49Aa1 proteins. Processing of Tpp49Aa1 produced similar sized products (~47 kDa) and Edman degradation identified an N-terminal cleavage site following Phe48 after treatment with *Cx. quinquefasciatus* gut extract (Jones et al. 2008). Hence, for computational modelling studies, residues Asn49 – Asn464 were extracted from the Tpp49Aa1 crystal structure (PDB 8BEY).

Processing of Cry48Aa1 produced two major products (~46 kDa and ~60 kDa) with N-terminal cleavage sites identified as Arg238 and Tyr52, respectively (Jones et al. 2008). The smaller of these products corresponds to a cleavage site at the loop between DI helix-5 and helix-6. Given that DI is thought to be involved in pore-formation, and hence required for activity, the larger of the two products was used for computational modelling studies. The precise location of the C-terminal cleavage site of Cry48Aa1 is unknown, but X-ray crystallography studies of activated Cry proteins indicate that the extended C-terminal region consisting of DIV – DVII is cleaved during activation. Hence, the Cry48Aa1 model was cleaved after Gln659, within a loop connecting DIII and DIV. Taken together, residues Asp53 – Gln659 were extracted from the Cry48Aa1 AF2 prediction for modelling.

Given that the environmental pH affects the protonation states of amino acids and, thus, their interactions, it was important to consider the pH at which computational modelling should be performed. Previously, Cry48Aa1/Tpp49Aa1 toxicity has been demonstrated against mosquito cell lines maintained at pH 7 – 7.4 (De Melo et al. 2009). In addition, interaction

of Cry48Aa1 and Tpp49Aa1 has been demonstrated in dot blot assays performed using PBS (pH 7.4) (Guo et al. 2016). Taken together, these studies suggest that Cry48Aa1/Tpp49Aa1 is functional at neutral pH. However, the luminal pH of the mosquito larval midgut ranges from ~8 in the gastric caecum (Boudko et al. 2001) to > 10 in the anterior gut and ~7.5 in the posterior gut (Dadd 1975). To reflect the rise of pH in the mosquito gut, and possibility that the Cry48Aa1-Tpp49Aa1 interaction occurs within the alkaline environment of the anterior midgut, we decided to perform modelling studies at both pH 7 and pH 11.

As default, the ClusPro and RosettaDock molecular docking programs (Chaudhury et al. 2011; Kozakov et al. 2017) assume a neutral pH, with the protonation states of titratable residues reflecting this. Hence, docking at pH 7 could be performed using default settings. In order to model the Cry48Aa1-Tpp49Aa1 interaction at pH 11, we initially attempted to protonate titratable residues using PROPKA in the PDB2PQR online server (Dolinsky et al. 2007), with the aim of carrying forward the protonated structures for molecular docking. However, the assigned protonation states of titratable residues could not be carried forward for protein-protein docking, as several programs investigated (ClusPro, RosettaDock, Haddock) were not able to maintain the input protonation state assigned by PROPKA, nor, in subsequent attempts, account for protonation states at extreme alkaline pH via their own internal algorithms. Specifically, the output structures produced by ClusPro and RosettaDock were reverted to protonation states reflecting a neutral pH environment. Given this limitation, molecular docking was performed at neutral pH. Subsequently, the modelled Cry48Aa1-Tpp49Aa1 complexes were protonated using PDB2PQR (Dolinsky et al. 2007), and the alternate protonation states of titratable residues at either pH 7 or pH 11 were modelled in subsequent MD simulations performed using the GROMACS software suite.

#### 4.2.4.2. Molecular docking studies

Molecular docking was performed using a naïve approach with no presumed Cry48Aa1-Tpp49Aa1 interface. First, a global search was carried out using the ClusPro 2.0 server (Kozakov et al. 2017). The Cry48Aa1 and Tpp49Aa1 starting structures were provided as the receptor and ligand, respectively. An FFT-based algorithm available in the PIPER program (Kozakov et al. 2006) is employed by ClusPro. Briefly, the ligand is placed on a moveable grid and the receptor on a fixed grid (Kozakov et al. 2017). For each grid point, the interaction energy is calculated as a correlation function, with billions of possible conformations sampled and scored. Subsequently, the 1,000 models with the lowest interaction energies are clustered according to RMSD, before undergoing energy minimisation (Kozakov et al. 2017). Since the largest (most populated) clusters are expected to represent the most likely models, clusters are ordered according to their size, from largest to smallest. The output consists of the model with the median interaction energy (Kozakov et al. 2017). Here, 30 clusters were identified. Of these clusters, the representative models from the 5 largest clusters (**Table 4.8**) were carried forward for refinement.

**Table 4.8. Output Cry48Aa1-Tpp49Aa1 clusters produced by ClusPro.**

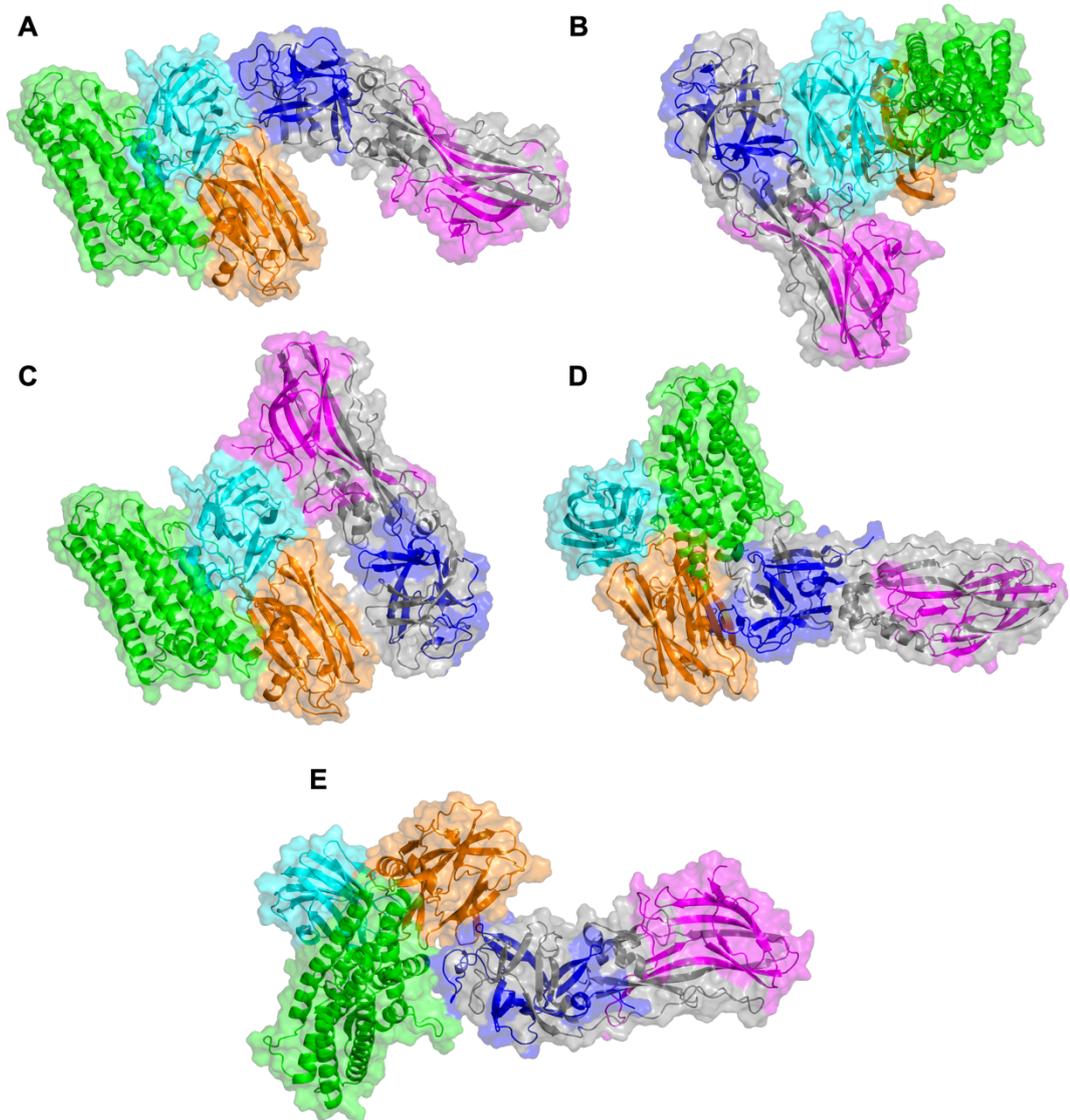
ClusPro		
Cluster	Cluster size	Total energy (kcal mol <sup>-1</sup> )
c1	53	-662.7
c2	42	-789.2
c3	37	-778.3
c4	36	-734.9
c5	33	-664.4

Following global docking, the RosettaDock source code, from the RosettaCommons software suite, was utilised to refine each of the models by performing independent local docking searches. Here, the *docking\_prepack\_protocol.macosclangrelease* executable was employed to prepack starting models output from ClusPro. Prepacking ensured that side chains were present in their lowest energy conformation. For each prepacking run, 25 models were produced and the model with the lowest energy score was carried forward for docking. To perform local docking searches, the *docking\_protocol.macosclangrelease* executable was employed. Unbound rotamer conformations were provided to improve the accuracy of docking (Wang et al. 2005) and 1,000 output models were produced and ranked according to their total energy score. From each docking search, the model with the lowest energy score (**Table 4.9, Fig. 4.18**) was identified and carried forward for further analysis. Of all 5 models identified, model c3r1 (originating from ClusPro cluster 3) exhibited the lowest total energy value of -1554.753 REU. From here on, models are referred to by their ClusPro cluster and Rosetta model rank (e.g., model c1r1 = ClusPro cluster 1, Rosetta rank 1).

**Table 4.9. Energy scores of models refined by RosettaDock.**

RosettaDock	
Model	Total energy (REU)
c1r1	-1532.5
c2r1	-1491.2
c3r1	-1554.8
c4r1	-1524.6
c5r1	-1489.2

REU = Rosetta energy units



**Figure 4.18. Modelled Cry48Aa1-Tpp49Aa1 complexes.** (A) c1r1 (B) c2r1 (C) c3r1 (D) c4r1 (E) c5r1. Cry48Aa1 DI is coloured green, DII is coloured orange, and DIII is coloured cyan. Tpp49Aa1 is coloured grey, with regions involved in Cry48Aa1 interaction coloured blue and regions involved in membrane interaction coloured magenta.

#### 4.2.4.3. Interface analysis

Protein interfaces are characterised by several features, including interface area, hydrophobic interactions, hydrogen bonding and salt bridge formation (Chothia and Janin 1975; Xu et al. 1997; Conte et al. 1999). Each feature plays an important role in interface stability and/or binding specificity and, therefore, was investigated here.

For interface analysis, PISA was utilised (Krissinel 2015), enabling prediction of the solvation free energy gain ( $\Delta^iG$ ), as well as calculation of interface area, the number of hydrogen bonds, salt bridges, and disulphide bridges (**Table 4.10**). Across all 5 models, interface area ranged from 408.0 to 981.1 Å<sup>2</sup> (**Table 4.10**). In addition, models exhibited  $\Delta^iG$  values ranging from +0.7 to -6.2 kcal mol<sup>-1</sup> (**Table 4.10**). A negative  $\Delta^iG$  is associated with a hydrophobic interface and positive protein affinity (Chothia and Janin 1975). Finally, all models were found to exhibit interfacial hydrogen bonding, whilst the majority (excluding c4r1) also exhibited salt bridge interactions (**Table 4.10**). These interactions contribute to interface stability, as well as binding specificity where side chains are involved (Xu et al. 1997). No models were found to exhibit interfacial disulphide bonding, suggesting that disulphides may not play significant roles in Cry48Aa1-Tpp49Aa1 binding. Previously, Guo et al. (2021) demonstrated that substitution of Tpp49Aa1 Cys91, Cys183 and Cys258 led to weaker Cry48Aa1 interaction in dot blots, suggesting that these residues either play a direct role in Cry48Aa1 interaction, or are required to maintain the correct fold of the Tpp49Aa1 binding site (Guo et al. 2021).

**Table 4.10. Interfacial interactions of Cry48Aa1-Tpp49Aa1 models.**

PISA					
Model	Interface Area (Å <sup>2</sup> )	$\Delta^i G$ (kcal mol <sup>-1</sup> )	No. H-bonds	No. salt bridges	Binding energy (kcal mol <sup>-1</sup> )
c1r1	689.0	-6.2	5	2	-9.1
c2r1	951.7	0.7	9	1	-3.6
c3r1	981.1	-1.8	7	1	-5.3
c4r1	817.4	-6.1	6	0	-8.8
c5r1	408.0	-2.1	2	1	-3.4

Interface analysis enabled us to discriminate somewhat between sampled models, however, limitations of computational docking protocols exist and include the inability to account for protein flexibility, as well as the fact that energy scores are often similar and hence, discriminating the near-native structure from non-native structures remains a challenge (Lensink et al. 2017; Radom et al. 2018). An emerging method employed to overcome these limitations is MD.

#### 4.2.4.4. Molecular dynamics (MD) simulations

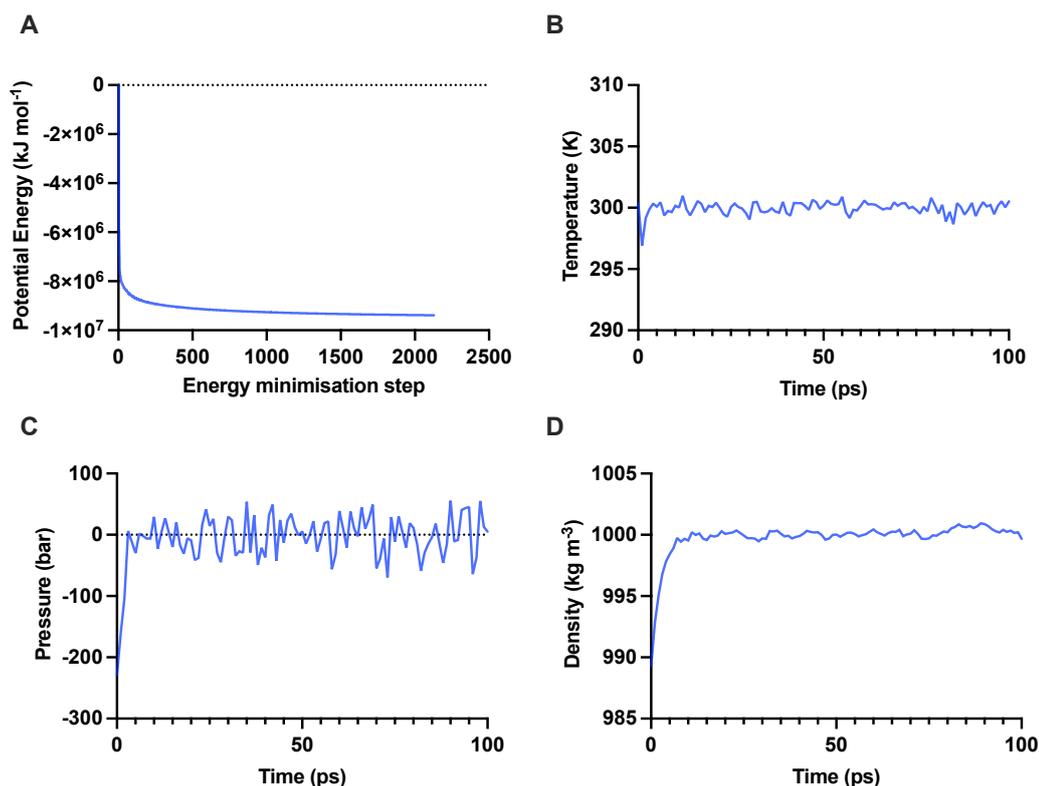
MD is an approach that utilises Newton's equations of motion to analyse the dynamics of atoms and molecules with respect to time (Hollingsworth and Dror 2018). With relation to protein-protein docking, several studies have utilised MD to evaluate docked models by performing RMSD and Rg calculations on output trajectories to analyse the structural stability of modelled interfaces (Radom et al. 2018; Jandova et al. 2021; Karaca et al. 2022). Using these calculations, authors were able to identify models consistent with experimental structures (Radom et al. 2018).

Here, all-atom MD simulations of the modelled Cry48Aa1-Tpp49Aa1 complexes were performed for 100 ns using GROMACS (Abraham et al. 2015). As previously discussed, molecular docking studies were limited by

the incapability to account for environment pH and the protonation states of titratable groups. Prior to MD simulations, the protonation states of titratable residues at pH 7 and pH 11 were estimated using PROPKA, available in the PDB2PQR online server (Dolinsky et al. 2007). To assign protonation states, the PARSE forcefield and CHARMM output naming scheme were used. Briefly, the pKa is the pH at which the acid is half associated and half dissociated. Hence, if  $\text{pH} = \text{pKa}$ , we would expect ~50% of all groups to be deprotonated, whilst if  $\text{pH} = \text{pKa} + 1$ , ~90% of all groups will be deprotonated, and if  $\text{pH} = \text{pKa} + 2$ , ~99% of all groups will be deprotonated. Since the protonation state of titratable groups is also affected by solvent environment and exposure, the protonation states of Cry48Aa1 and Tpp49Aa1 were calculated independently, and the protonated structures were then superposed with the modelled Cry48Aa1-Tpp49Aa1 complexes. The superposed, protonated structures were then carried forward for MD simulations using GROMACS.

Structure preparation in GROMACS has been described in depth in **section 2.6.4.3**. Briefly, the structures were first prepared using the pdb2gmx module and CHARMM 27 forcefield. The pdb2gmx module allowed interactive selection of the side-chain protonation states of lysine, arginine, aspartic acid, and glutamic acid, but not tyrosine. As default, pdb2gmx selects the protonated forms of lysine and arginine and the deprotonated forms of aspartic acid and glutamic acid, whilst optimising the protonation state of histidine based on optimal hydrogen bonding conformation, which itself is based upon maximum donor-acceptor distances and angles. These protonation states reflect the charges which occur at neutral pH. At pH 11, between 50 – 90% of the lysine side chains ( $\text{pKa}_3 = 10.53$ ) will be deprotonated, whilst arginine ( $\text{pKa}_3 = 12.48$ ), on the other hand, remains predominantly protonated. Hence, simulations modelled at pH 7 were run using the default residue charges, whilst in simulations modelled at pH 11, the -lys flag was used to select the protonation states of lysine residues interactively based on the PROPKA output. In addition, the -ss flag was used to incorporate the Cys91-Cys183 disulphide bridge.

Following this, the structures were relaxed by energy minimisation to ensure the correct geometry and removal of steric clashes within the system. Here, the steepest descent algorithm was applied using a maximum number of 50,000 steps in 0.01 step sizes, and a maximum force (Fmax) of 1000 kJ mol<sup>-1</sup> nm<sup>-1</sup>. To determine whether energy minimisation was successful, the potential energy (E<sub>pot</sub>) was plotted against the energy minimisation step (**Fig 4.19A**). A successful energy minimisation is identified by a negative E<sub>pot</sub> with an order of 10<sup>5</sup> / 10<sup>6</sup>. In addition, the maximum force (Fmax) should be no greater than the target value indicated above, indicating that these variables had been satisfied and a stable system had been generated.



**Figure 4.19. Energy minimisation and system equilibration of the Cry48Aa1-Tpp49Aa1 simulations.** An example of MD set-up has been illustrated for simulation of model c1r1 at pH 7. **(A)** Potential energy ( $E_{\text{pot}}$ ) plotted against the energy minimisation step. A successful energy minimisation is identified by a negative  $E_{\text{pot}}$  with an order of  $10^5 / 10^6$ . **(B)** Average temperature of the simulation. **(C)** Average pressure of the simulation. **(D)** Average density of the simulation.

Following energy minimisation, the temperature and pressure of the system were equilibrated. To stabilize the temperature of the system, an isothermal-isochoric ensemble was applied, without pressure-coupling. To ensure the system had reached the target temperature (300 K), the calculated temperature was plotted against time (**Fig 4.19B**). Here, the temperature is expected to reach the target value and remain stable across the duration of the simulation. To stabilize the pressure and density of the system, an isothermal-isobaric ensemble was applied. Pressure-coupling was performed using a Parrinello-Rahman barostat. Again, pressure and density were plotted against time (**Fig 4.19C and 4.19D**). Here, the pressure

of the system is expected to fluctuate, however, the density should replicate the expected experimental value for water ( $1000 \text{ kg m}^{-3}$ ) and remain stable over time. Both the isothermal-isochoric and isothermal-isobaric equilibration stages were performed for 100 ps.

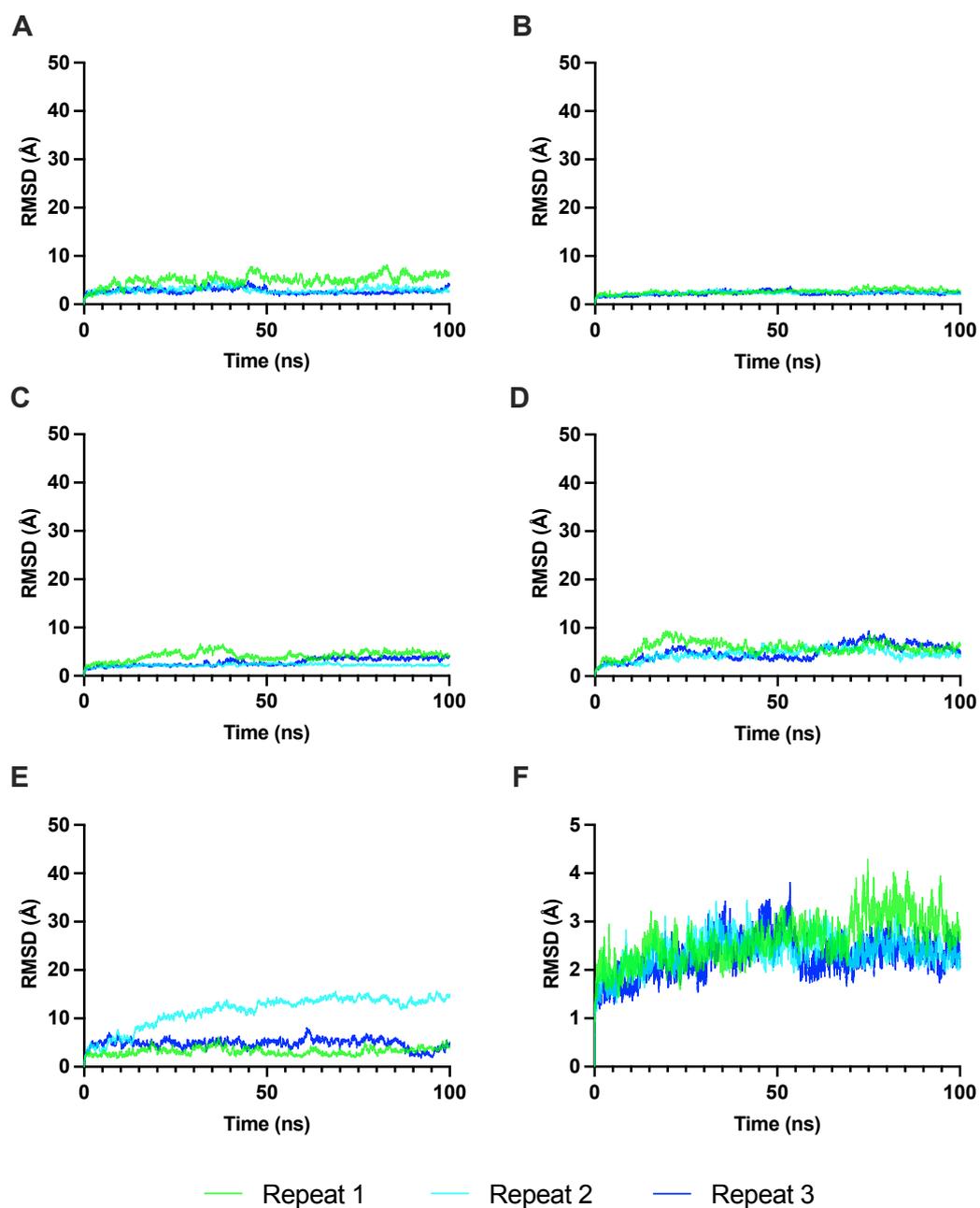
Following energy minimisation and system equilibration, 100 ns production MD simulations were performed as described in **section 2.6.4.6**. The output trajectories were extracted and converted to account for any jumps caused by the periodic boundary conditions. The resulting trajectory files were carried forward for RMSD and Rg analysis.

#### **4.2.4.5. Root-mean-square deviation (RMSD) trajectory analysis**

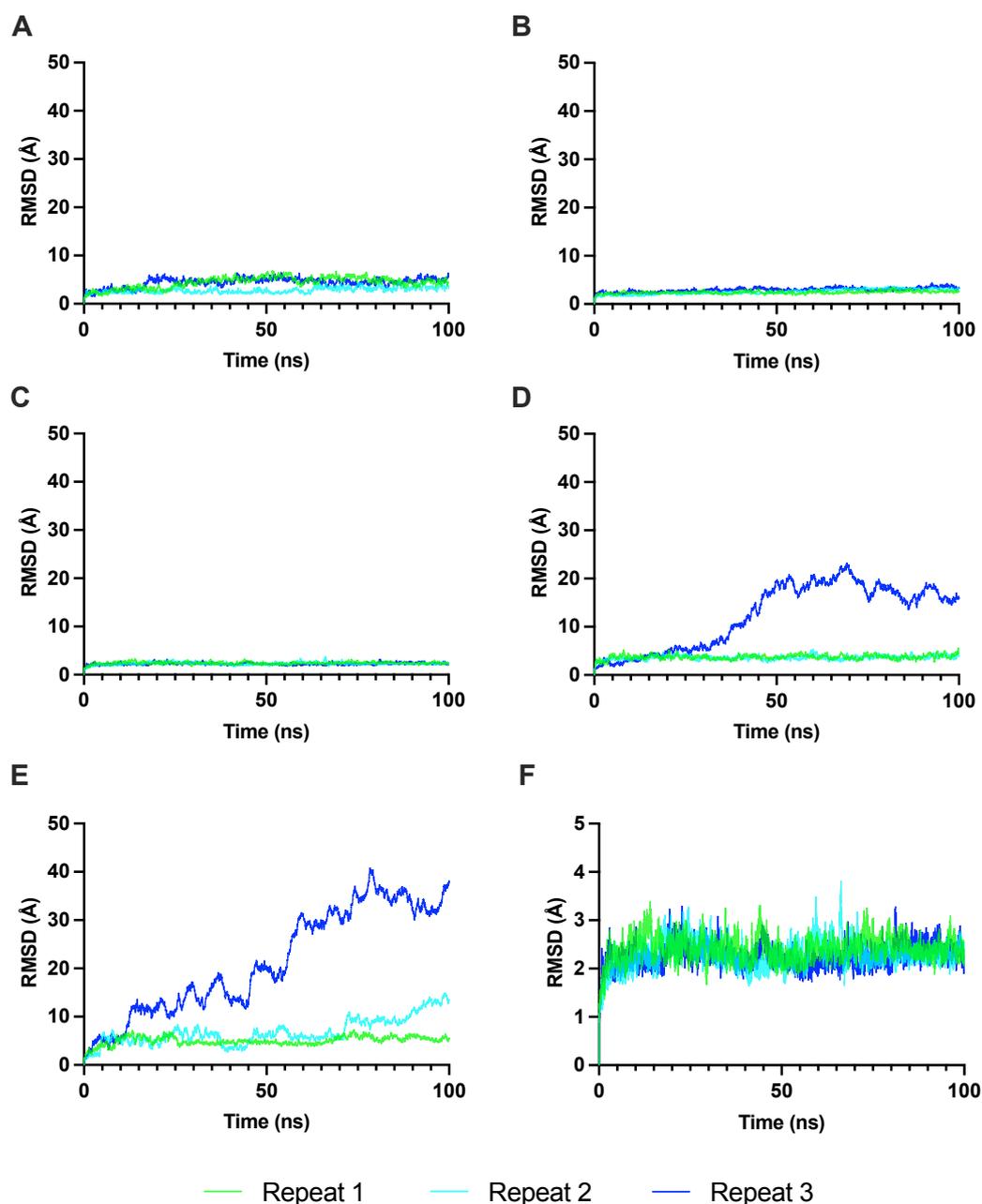
To assess the structural stability of modelled Cry48Aa1-Tpp49Aa1 complexes at both pH 7 (**Fig 4.20**) and pH 11 (**Fig 4.21**), the RMSD of the position of C $\alpha$  backbone atoms was calculated as a function of time using the GROMACS command line `gmx_rms` module. RMSD calculates the deviation of a selection of atoms in reference to their initial starting position. High RMSD values, therefore, correlate with large changes in the structure, indicating significant instability. On the other hand, low RMSD values correlate with small changes in the structure, indicating stability. In a previous study, RMSD analysis was utilised to assess the stability of modelled complexes and identify models consistent with experimental structures (Radom et al. 2018).

At pH 7, the lowest RMSD values were obtained for model c2r1 (**Fig 4.20B**), ranging from 1.19 – 4.29 Å, and model c3r1 (**Fig 4.20C**), ranging from 1.12 – 6.77 Å. At pH 11, the lowest RMSD values were also obtained for model c3r1 (**Fig 4.21C**), ranging from 1.28 – 3.82 Å, and model c2r1 (**Fig 4.21B**), ranging from 1.32 – 4.45 Å. RMSD values obtained for model c3r1 were lower at pH 11 (in comparison to pH 7) indicating increased stability at an elevated pH. Taken together, this analysis suggests that models c2r1 and c3r1 exhibit greater structural stability compared to the other models

analysed and, thus, are more likely to persist in solution. Ranges calculated exclude RMSD values obtained during the first 1 ns of the MD simulations, corresponding to the 'warm-up' phase.



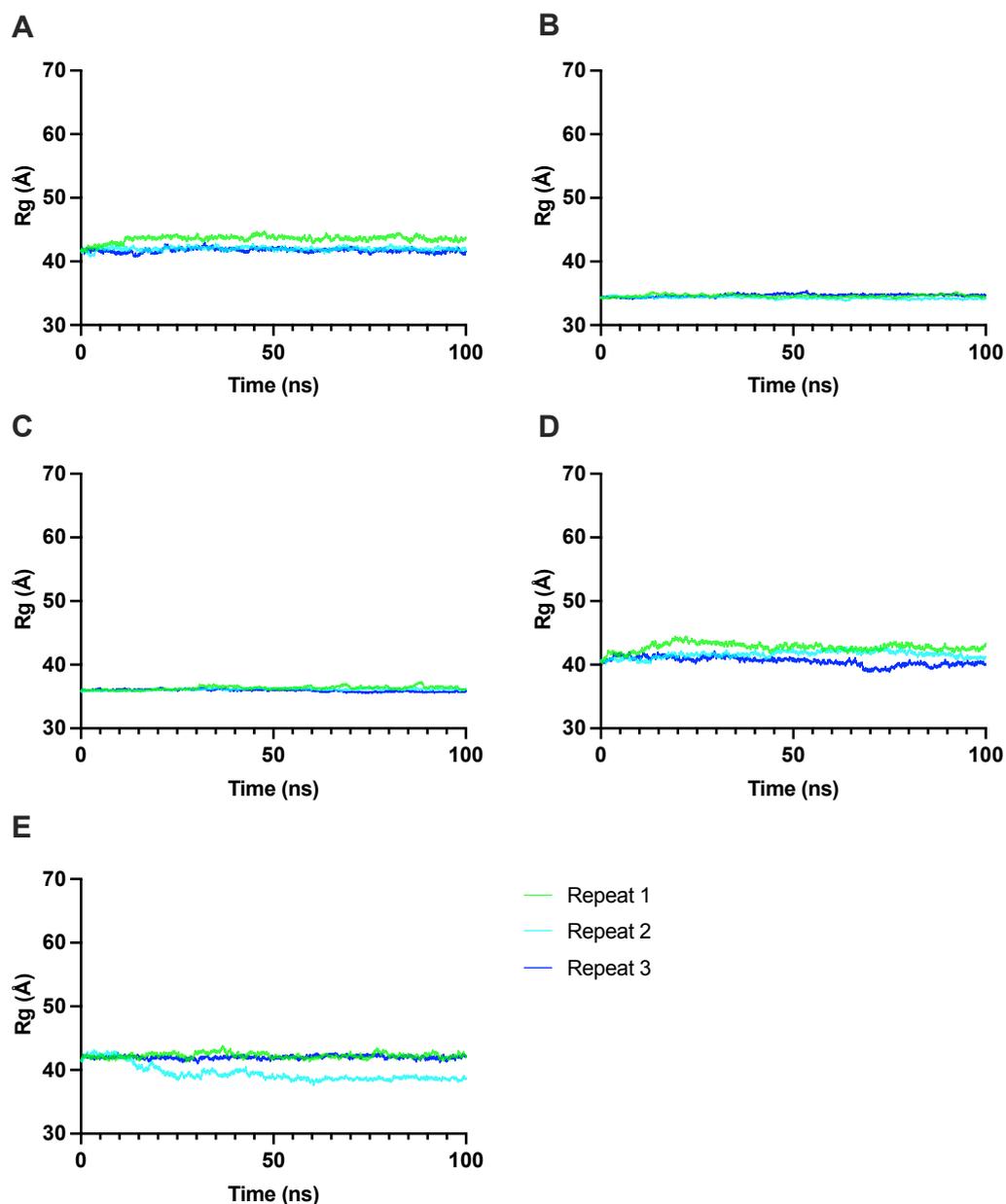
**Figure 4.20. RMSD of Cry48Aa1-Tpp49Aa1 models throughout 100 ns simulations performed at pH 7.** High RMSD values indicate significant structural instability, whilst low RMSD values indicate structural stability. (A) c1r1 (B) c2r1 (C) c3r1 (D) c4r1 (E) c5r1 (F) Zoom of the most stable structure (c2r1) indicated by RMSD analysis. MD simulations were repeated three times (green, blue, and cyan).



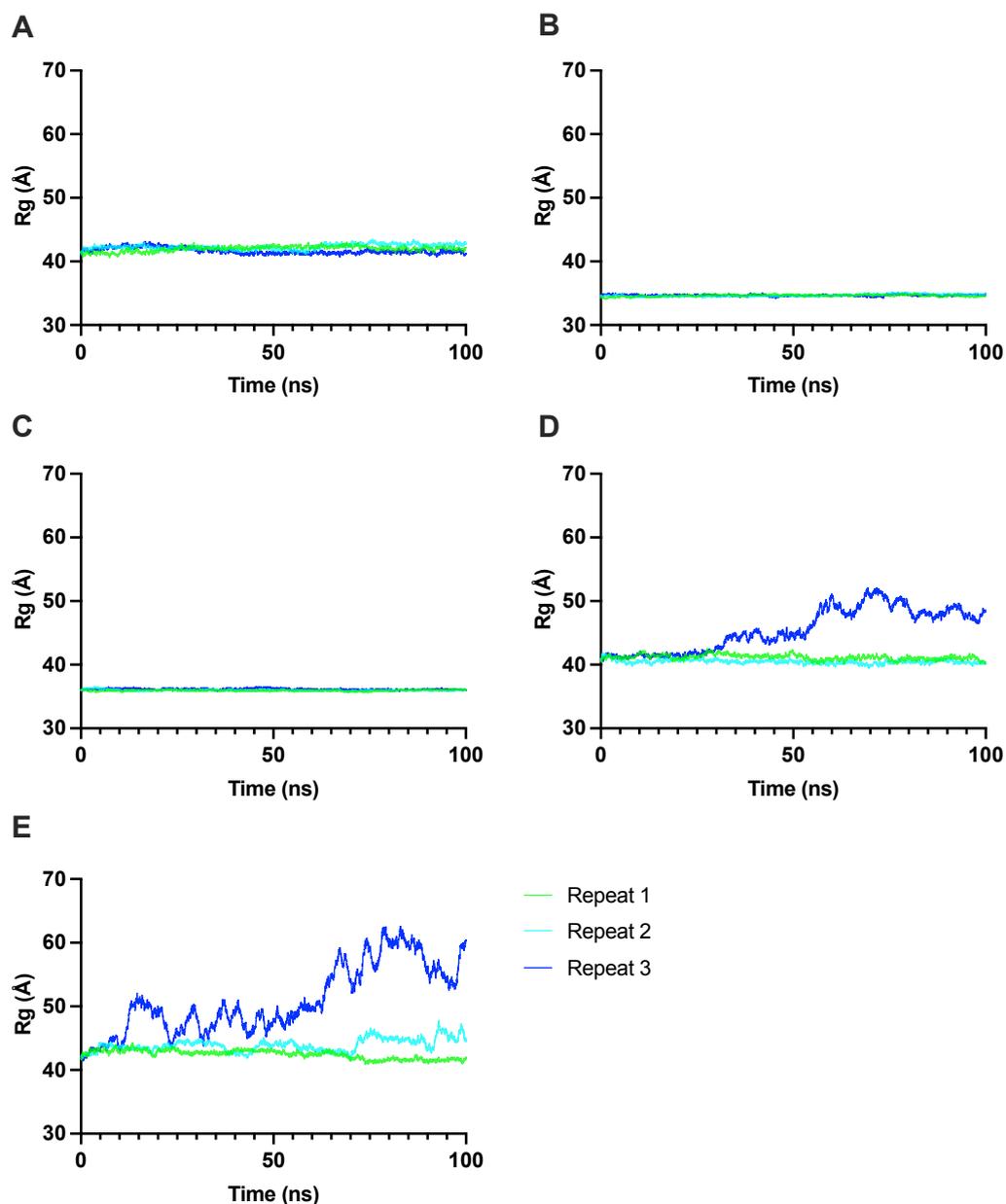
**Figure 4.21. RMSD of Cry48Aa1-Tpp49Aa1 models throughout 100 ns simulations performed at pH 11.** High RMSD values indicate significant structural instability, whilst low RMSD values indicate structural stability. (A) c1r1 (B) c2r1 (C) c3r1 (D) c4r1 (E) c5r1 (F) Zoom of the most stable structure (c3r1) indicated by RMSD analysis. MD simulations were repeated three times (green, blue, and cyan).

#### 4.2.4.6. Radius of gyration (Rg) trajectory analysis

Further investigation of the structural stability of modelled Cry48Aa1-Tpp49Aa1 complexes at both pH 7 (**Fig 4.22**) and pH 11 (**Fig 4.23**) was performed by calculating the Rg as a function of time using the GROMACS command line `gmx_gyrate` module. Rg analyses the mass and position of atoms in relation to the centre of mass of the molecule and, thus, indicates the compactness of the overall structure. In this case, a stably folded structure is expected to maintain a steady Rg value following equilibration in MD simulations. Consistent with RMSD analysis, models c2r1 and c3r1 were found to exhibit stable Rg values for all three simulation repeats at both pH 7 (**Fig 4.22B, 4.22C**) and pH 11 (**Fig 4.23B, 4.23C**).



**Figure 4.22. Rg of Cry48Aa1-Tpp49Aa1 models across 100 ns simulations performed at pH 7.** Radius of gyration (Rg) indicates the compactness of the overall structure. Stably folded structures are expected to maintain a steady Rg value. (A) c1r1 (B) c2r1 (C) c3r1 (D) c4r1 (E) c5r1. MD simulations were repeated three times (green, blue, and cyan).



**Figure 4.23. Rg of Cry48Aa1-Tpp49Aa1 models across 100 ns simulations performed at pH 11.** Radius of gyration (Rg) indicates the compactness of the overall structure. Stably folded structures are expected to maintain a steady Rg value. (A) c1r1 (B) c2r1 (C) c3r1 (D) c4r1 (E) c5r1. MD simulations were repeated three times (green, blue, and cyan).

#### 4.2.4.7. Selection of most likely Cry48Aa1-Tpp49Aa1 model

As described above, we aimed to establish a model of the Cry48Aa1-Tpp49Aa1 complex by utilising a combined global and local docking approach, whilst considering the luminal pH of the mosquito larval midgut, which ranges from ~8 in the gastric caecum (Boudko et al. 2001) to > 10 in the anterior gut and ~7.5 in the posterior gut (Dadd 1975). ClusPro was employed to perform global docking, whilst RosettaDock was employed to perform local docking refinement. ClusPro ranks output models by performing a cluster analysis, whereby the native structure is expected to exist within the largest cluster, which, in this work, corresponds to model c1 (**Table 4.8**). On the other hand, RosettaDock ranks output models according to their total energy score, which itself is based upon the thermodynamic hypothesis. Hence, models produced by ClusPro were locally refined and re-ranked according to their final total energy score. At this stage, model c3r1 represented the most likely complex, with the lowest total energy score of -1554.753 REU (**Table 4.9**).

The interfaces of all five models were found to contain a network of hydrogen bonds, with all (except model c4r1) also containing salt bridges (**Table 4.10**). MD simulation analysis provided further insight into the stability of modelled Cry48Aa1-Tpp49Aa1 complexes at varied environmental pH. Specifically, RMSD and Rg analysis demonstrated the most stable complex to be model c2r1 at pH 7 (**Fig 4.20**) and model c3r1 at pH 11 (**Fig 4.21**). In model c2r1, Cry48Aa1 interacts via DII and DIII with both the N- and C-terminal domains of Tpp49Aa1, whilst Cry48Aa1 DI is not involved (**Fig 4.18B**). Model c2r1 is consistent with previous work showing that the Tpp49Aa1 N-terminal region, Asn49 – Ser148, interacts with Cry48Aa1 (Guo et al. 2016). In addition, the Tpp49Aa1 C-terminal region, which is required for interaction with the membrane (Guo et al. 2016), as well as the apical loops of Cry48Aa1 domain II, known to be involved in receptor binding in this class of pesticidal proteins, and Cry48Aa1 domain I, the Cry toxin pore-forming domain, are not blocked by the proposed interaction (**Fig 4.18B**). In contrast, a large proportion of the Tpp49Aa1 C-terminal region

required for interaction with the membrane is blocked by the proposed interaction in model c3r1 (**Fig 4.18C**). The stability of model c2r1 was comparable at both pH 7 and pH 11, suggesting that this interaction may be able to occur throughout all regions (gastric caecum, anterior gut, posterior gut) of the mosquito larval midgut, whilst model c3r1 exhibited higher structural stability at pH 11. Taken together, the proposed models identify several putative interfacial residues which are targets for experimental validation using protein mutagenesis techniques, work that is clearly required to probe the Cry48Aa1-Tpp49Aa1 interaction further. In addition, further work is required to investigate whether Cry48Aa1/Tpp49Aa1 exhibits regional binding at a specific luminal pH.

#### **4.2.5. Insect bioassays**

Previous work demonstrated that Cry48Aa1/Tpp49Aa1 was active against only one (*Cx. quinquefasciatus*) of eight insect species tested (*Anthonomus grandis*, *A. gemmatalis*, *S. frugiperda*, *Plutella xylostella*, *Chironomus riparius*, *Ae. aegypti*, *An. gambiae*), suggesting a narrow target insect range (Jones et al. 2008). To validate the known target insect range against *Cx. quinquefasciatus*, mosquito larvae bioassays were performed as described in **section 2.5.4**. After 24 hours exposure to a high dose of the Cry48Aa1/Tpp49Aa1 pair, mortality was observed in *Cx. quinquefasciatus*, but not *An. gambiae* or *Ae. aegypti* larvae. In addition, no mortality was seen from a high dose of the individual toxin components. These results are in line with those published in the literature, confirming the known target range, and viability of the Cry48Aa1 and Tpp49Aa1 proteins produced for this study.

*In other work, cellular toxicity of Cry48Aa1/Tpp49Aa1 was demonstrated against Cx. quinquefasciatus cell lines (MRA-918 and Hsu), as well as a cell line derived from the Culex tarsalis mosquito species (Ct) (Williamson et al. 2023). Having demonstrated toxicity against a Cx. tarsalis derived-cell line, we next sought to investigate toxicity against other*

mosquito species, including *Cx. tarsalis*, *Aedes albopictus*, and *Anopheles stephensi*. Since we were unable to access these mosquito colonies in the UK, insect bioassays were performed in collaboration with Prof. Brian Federici (University of California, Riverside), Dr. Hyun-Woo Park and Dr. Dennis K. Bideshi (California Baptist University). Mortality (**Table 4.11**) was observed against all three species assayed 48 hours following administration of a 1:1 ratio (w/w) Cry48Aa1/Tpp49Aa1, with LC<sub>50</sub> values of 91 ng/mL (*Cx. tarsalis*), 111 ng/mL (*Ae. albopictus*), and 173 ng/mL (*An. stephensi*). Control experiments (data not shown in **Table 4.11**) confirmed toxicity against *Cx. quinquefasciatus*, but not *Ae. aegypti* or *An. gambiae*. Given their role as vectors of human disease, the identification of *Cx. tarsalis*, *Ae. albopictus*, and *An. stephensi* as new targets significantly enhances the impact of Cry48Aa1/Tpp49Aa1 as a bioinsecticide.

**Table 4.11. Cry48Aa1/Tpp49Aa1 mosquito larval bioassays.**

Mosquito Species *	Exposure (hours)	LC <sub>50</sub> (ng/mL) (95% fiducial limits)	LC <sub>95</sub> (ng/mL) (95% fiducial limits)
<i>Aedes albopictus</i>	24	607 (332 – 1,110)	4,800 (2,630 – 8,800)
	48	111 (68 – 183)	1,360 (823 – 2,230)
<i>Anopheles stephensi</i>	24	818 (450 – 1,490)	3,830 (2,100 – 6,960)
	48	173 (118 – 253)	712 (450 – 965)
<i>Culex tarsalis</i>	24	1,890 (578 – 6,160)	301,000 (92,100 – 982,000)
	48	91 (61 – 134)	436 (295 – 646)

\* LC<sub>50</sub> and LC<sub>95</sub> values are not shown for control experiments performed against *Cx. quinquefasciatus*, *Ae. aegypti* and *An. gambiae*.

### 4.3. Conclusions

In this chapter, the structure of Tpp49Aa1 was elucidated from its natural crystals using SFX at an XFEL to a final resolution of 1.62 Å. The Tpp49Aa1 structure provides insight into structural features that may be

significant for function, including a conserved  $\beta$ -hairpin in the C-terminal domain, thought to act as the transmembrane region during pore-formation, alongside several loop insertions and a disulphide bridge in the N-terminal domain that appear to distort the three-fold symmetry of the  $\beta$ -trefoil fold which, itself, is implicated in carbohydrate binding in other protein families. When compared with other pesticidal proteins, Tpp49Aa1 displayed highest structural similarity with Tpp2Aa2, with the largest differences occurring in the exposed loops of the N-terminal head domain that is implicated in receptor binding.

The Tpp49Aa1 structure also provided insight into natural crystal packing, showing that Tpp49Aa1 forms a homodimer with a large intermolecular interface and these dimers pack together to form the native crystal. Subsequently, RALS and RI measurements illustrated that, upon crystal dissolution, the Tpp49Aa1 homodimer dissociates into its monomeric form. Given that regions shown to interact with Cry48Aa1 are partially buried within the Tpp49Aa1 dimer, we hypothesized that dissociation is required for downstream interaction with Cry48Aa1. pH mixing in the XFEL beam also enabled the early structural events leading up to crystal dissolution to be investigated, where the loss of hydrogen bonds near to the N-terminal pro-peptide cleavage site was hypothesized to increase accessibility to gut proteases.

Using the Tpp49Aa1 structure and an AF2 model of Cry48Aa1, a model of the Cry48Aa1-Tpp49Aa1 interaction was predicted and can be used to direct mutagenesis studies aimed at investigating this interaction and the synergistic activity of Cry48Aa1/Tpp49Aa1 further. Finally, mosquito larval bioassays were performed, leading to the identification of three new target species, substantially increasing the significance and impact of Cry48Aa1/Tpp49Aa1 as a bioinsecticide.

## 5. Characterisation of the Tpp1Aa2/Tpp2Aa2 pesticidal protein and its interaction with the receptor protein, Cqm1

### 5.1. Introduction

The Tpp1/Tpp2 binary pesticidal protein of the Toxin-10 Pesticidal Protein class is produced by highly mosquitocidal *L. sphaericus* strains (Priest et al. 1997; Berry and Crickmore 2017) and is composed of two structurally similar proteins: Tpp1 (~42 kDa) and Tpp2 (~51 kDa) (Baumann et al. 1988; Oei et al. 1992; Charles et al. 1997). Importantly, Tpp1/Tpp2 is active against the larvae of *Culex* and *Anopheles* mosquitoes involved in the transmission of malaria and West Nile virus (Berry et al. 1993), and hence, has been commercially applied worldwide with much success. Structures of both Tpp2Aa3 (PDB 3WA1, solved from reconstituted crystals - Srisucharitpanit et al. 2014) and the Tpp1Aa2/Tpp2Aa2 complex (PDB 5FOY - pH 7, PDB 5FOZ - pH 10, solved from natural co-crystals - Colletier et al. 2016) have been elucidated. In both Tpp1/Tpp2, distinct N-terminal  $\beta$ -trefoil and C-terminal  $\beta$ -sheet rich domains exist, with the latter sharing structural homology with the Aerolysin family of pore-forming proteins (Colletier et al. 2016; Lacomel et al. 2021). In Tpp1/Tpp2 susceptible species, receptor binding proteins have been identified as *Culex quinquefasciatus* maltase 1 (Cqm1 - Sharma et al. 2018), for which a structure has been solved (Sharma and Kumar 2019), *Culex pipiens* maltase 1 (Cpm1 - Silva-Filha et al. 1999), and *Anopheles gambiae* maltase 3 (Agm3 - Opota et al. 2008). An orthologous non-binding protein, *Aedes aegypti* maltase 1 (Aam1), has been identified in the refractory mosquito species, *Ae. aegypti* (Nielsen-Leroux and Charles 1992; Ferreira et al. 2010). Despite knowledge of the receptor binding partner, the molecular details of the toxin-receptor interactions have remained largely unknown. Since toxin specificity is determined by the recognition of target receptors, insight into the molecular details of toxin-receptor interactions will support future work aimed at developing new bioinsecticides with enhanced potency, stability, and target insect range.

This chapter focuses on production of Tpp1Aa2, Tpp2Aa2, and Cqm1 recombinant proteins using *E. coli* as an expression host followed by purification methods under native conditions. Mosquitocidal activity of Tpp1Aa2/Tpp2Aa2 was assessed using mosquito larval bioassays, whilst  $\alpha$ -glucosidase activity of Cqm1 was confirmed using enzyme assays. In addition, the interaction of the Tpp2 protein with its receptor, Cqm1, was predicted using computational modelling, and compared with that of the Tpp2Aa2-Cqm1 structure subsequently elucidated using single-particle cryoEM in collaboration with researchers from Monash University. Directed by the Tpp2Aa2-Cqm1 structure, mutagenesis studies were performed with the aim of engineering the Tpp1Aa2/Tpp2Aa2 pesticidal protein to broaden the target insect range. Throughout this chapter, work carried out by other researchers has been indented and italicised.

## **5.2. Results and Discussion**

### **5.2.1. Expression of Tpp1Aa2 and Tpp2Aa2 in *E. coli***

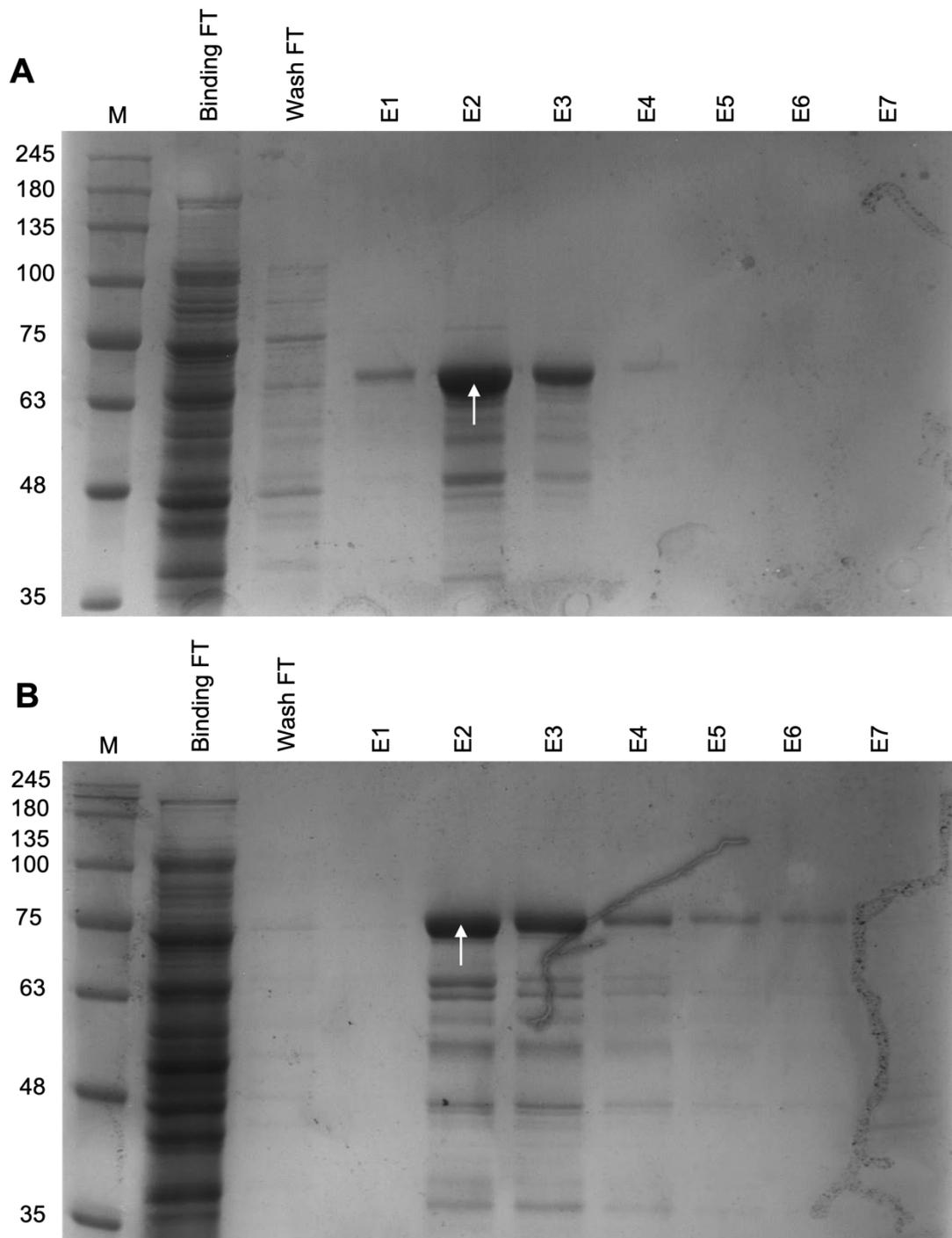
Clones encoding the Tpp1Aa2 and Tpp2Aa2 proteins were obtained from Prof. Colin Berry (Cardiff University, UK). The pGEX-based expression vectors comprised the *tpp1Aa2* gene excluding the first 8 residues or the full-length *tpp2Aa2* gene with encoded N-terminal GST tags. The *tpp1Aa2* and *tpp2Aa2* genes were expressed under the control of the *tac* promoter, which is induced by the lactose analog, IPTG. The pGEX expression vector also contained an internal *lacIq* gene, the product of which is a repressor protein that binds to the *tac* promoter operator region, preventing expression prior to induction with IPTG. Inclusion of a thrombin cleavage site (Leu-Val-Pro-Arg-||-Gly-Ser) following the signal sequence enabled site-specific cleavage of the N-terminal GST tag for single-particle cryoEM studies. Inclusion of an ampicillin resistance gene enabled selection of clones containing the expression construct. The gene sequences of the pGEX-Tpp1Aa2 and pGEX-Tpp2Aa2 clones were validated by Sanger sequencing using pGEX-5

and pGEX-3 primers (**Table 2.7**). The Tpp1Aa2 and Tpp2Aa2 protein sequences have been provided in **Appendices 4 and 5**, respectively.

The constructs were expressed using protocols previously optimised for soluble protein expression in the Berry laboratory. Briefly, constructs were expressed in *E. coli* BL21 (DE3) by incubating at 37°C until an OD<sub>600</sub> of ~0.6 was obtained before induction with 1 mM IPTG at 37°C for 4 hours. GST-tagged Tpp1Aa2 and Tpp2Aa2 proteins were extracted from the cells as described in **section 2.3.3** and purified as shown below.

### **5.2.2. Purification of Tpp1Aa2 and Tpp2Aa2**

To obtain the GST-Tpp1Aa2 and GST-Tpp2Aa2 fusion proteins, GST-affinity chromatography was performed on whole cell lysates as described in **section 2.4.2**. Briefly, cell lysates were applied to equilibrated columns containing glutathione sepharose beads and the flow-through was collected. Columns were incubated for 1 hour to allow binding of the GST fusion protein. The binding flow-through was collected (**Fig 5.1A, 5.1B – lane 2**) and the column was washed with 30 mL PBS (**Fig 5.1A, 5.1B – lane 3**). GST fusion protein was eluted into 7 x 1 mL fractions of 50 mM TrisHCl, 10 mM reduced glutathione, pH 8.0. Elution fractions obtained were analysed by SDS-PAGE (**Fig 5.1A, 5.1B – lanes 4 – 10**), showing that GST-Tpp1Aa2 and GST-Tpp2Aa2 eluted into fractions 1 – 4 and 2 – 6, respectively. Fractions containing the GST-Tpp1Aa2 (~68 kDa) and GST-Tpp2Aa2 (~77 kDa) proteins were separately pooled, concentrated, and stored at -20°C until downstream applications.



**Figure 5.1. GST affinity purification of GST-Tpp1Aa2 and GST-Tpp2Aa2.** GST-Tpp1Aa2 and GST-Tpp2Aa2 were expressed using previously optimised conditions (1mM IPTG, 4 hours, 37°C) and purified via GST-tagged affinity purification. Binding / wash flow-through (FT) and elution fractions (E1 – E7) obtained were analysed by SDS-PAGE. (A) Coomassie stained SDS-PAGE showing partial purification of GST-Tpp1Aa2 (white arrow, ~68 kDa) (B)

Coomassie stained SDS-PAGE showing partial purification of GST-Tpp2Aa2 (white arrow, ~77 kDa). M (marker): BLUeye Pre-Stained Protein Ladder.

### 5.2.3. Protein activation using trypsin treatment

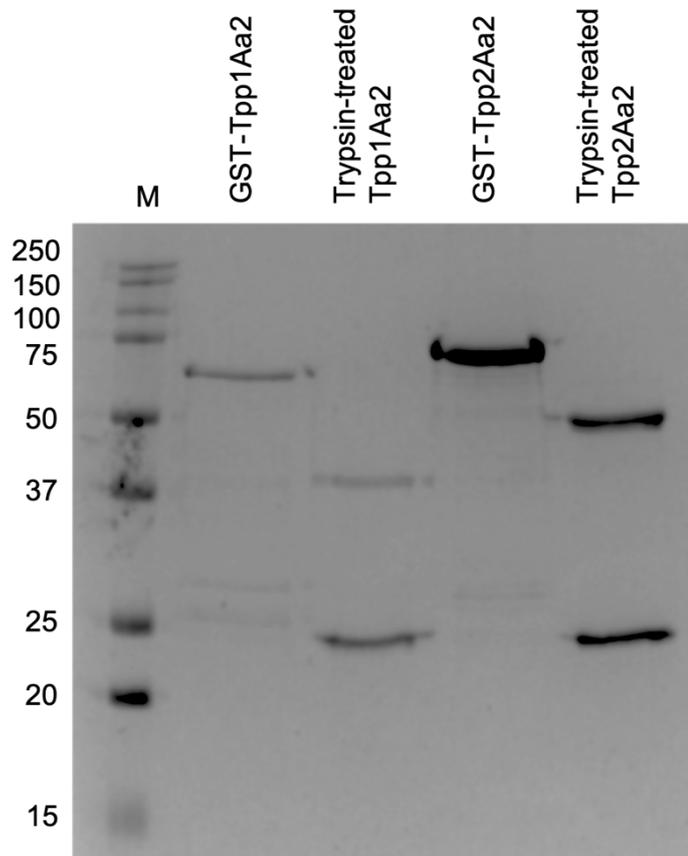
Once ingested by mosquito larvae, the protoxin forms of Tpp1 (~42 kDa) and Tpp2 (~51 kDa) are proteolytically cleaved by midgut proteases, resulting in the production of activated Tpp1 (~39 kDa) and Tpp2 (~43 kDa) (Baumann et al. 1988; Oei et al. 1992; Charles et al. 1997). Previous studies have shown that deletion of 10 residues from the N-terminus and 17 residues from the C-terminus of the Tpp1Aa2 protein (both correlating to predicted chymotrypsin sites) produces a protein which displays toxicity against *Cx. quinquefasciatus* cells and exhibits similar electrophoretic properties as Tpp1Aa2 activated by mosquito larval midgut juice (Broadwell et al. 1990b). In contrast, trypsin-treatment of Tpp1Aa2 is thought to cleave 16 or 19 residues from the N-terminus and 7 residues from the C-terminus (Broadwell et al. 1990b).

With relation to Tpp2Aa2, deletion of 21 amino acids from the N-terminus and 53 amino acids from the C-terminus, both corresponding to predicted chymotrypsin sites, resulted in the production of a protein which displayed identical electrophoretic properties as Tpp2Aa2 activated by midgut juice and retained its activity against *Cx. pipiens* larvae (Clark and Baumann 1990). Removal of 41 residues at the N-terminus and 61 residues at the C-terminus (corresponding to the next predicted chymotrypsin sites) led to a loss of toxicity against *Cx. pipiens* larvae, suggesting that these regions of Tpp2Aa2 are required for protein activity (Clark and Baumann 1990).

To investigate proteolytic processing of the recombinant Tpp1Aa2 and Tpp2Aa2 proteins produced as part of this work, trypsin cleavage was performed as described in **section 2.5.3**. Trypsin preferentially cleaves peptide bonds at the C-terminal side of Lysine and Arginine residues. Since

trypsin cleavage occurred within the protein, removal of the N-terminal GST tag was achieved simultaneously. Trypsin treatment was assessed by SDS-PAGE and showed complete trypsin cleavage occurred following incubation at room temperature for 18 hours (**Fig 5.2**).

To confirm the site of trypsin cleavage in Tpp2Aa2, N-terminal sequencing was performed for 7 cycles. The resulting sequence FTNYPLN is identical to the authentic Tpp2Aa2 sequence. Based on these results, the N-terminal trypsin cleavage site was shown to be between Lys17 and Phe18. This region is not seen in either of the published Tpp2Aa2 crystal structures (PDB 5FOY, PDB 3WA1), suggesting that it is disordered.



**Figure 5.2. Trypsin activation of GST-Tpp1Aa2 and GST-Tpp2Aa2.** Coomassie stained SDS-PAGE showing complete trypsin activation of GST-Tpp1Aa2 (~68 kDa) and GST-Tpp2Aa2 (~77 kDa) to produce ~39 kDa and ~43 kDa proteins, respectively. Remaining trypsin (~23 kDa) can be seen in lanes 3 and 5. M (marker): Precision Plus Protein Standards Dual Color.

#### 5.2.4. Mosquito larval bioassays of Tpp1Aa2/Tpp2Aa2

Previous work has shown that Tpp1Aa2/Tpp2Aa2 is active against *Cx. quinquefasciatus* and *An. gambiae* mosquito larvae (Berry et al. 1993). To validate the viability of Tpp1Aa2 and Tpp2Aa2 recombinant protein produced here, mosquito larval bioassays were performed as described in **section 2.5.4**. After 24 hours exposure to a high dose of a 1:1 w/w mix of the Tpp1Aa2/Tpp2Aa2 pair (200  $\mu$ L 500  $\mu$ g/mL total – estimated by BCA assay),

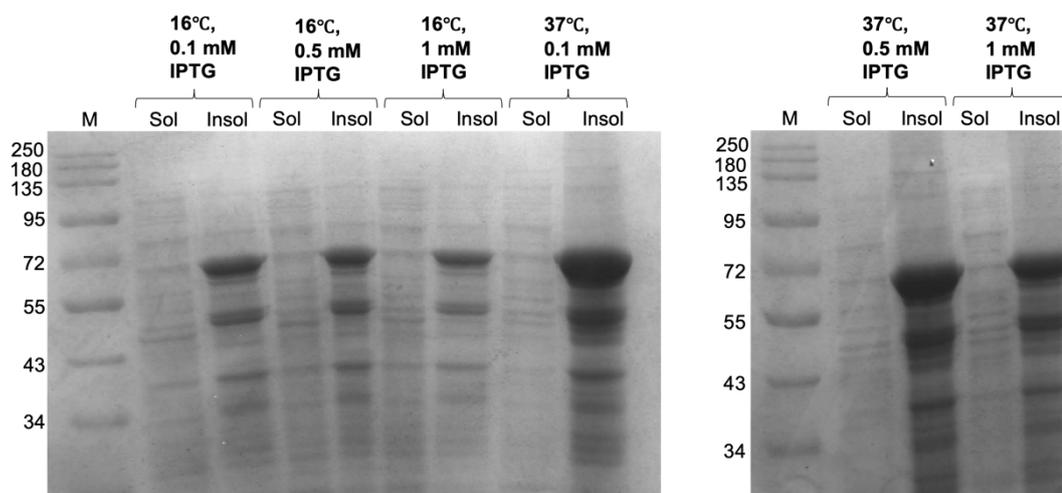
mortality was observed in *Cx. quinquefasciatus* and *An. gambiae* larvae, but not *Ae. aegypti*. In addition, no mortality was seen from a high dose of the individual toxin components (200  $\mu$ L 500  $\mu$ g/mL) or buffer alone controls against any of the mosquito targets. These results are in line with those published in the literature (Berry et al. 1993), confirming the known target range and viability of the Tpp1Aa2 and Tpp2Aa2 proteins produced for this study.

#### 5.2.5. Expression of Cqm1 in *E. coli*

In *Cx. quinquefasciatus*, the Tpp1/Tpp2 receptor has been identified as Cqm1 (Sharma et al. 2018). To investigate the interaction of the Tpp2Aa2 protein with Cqm1, Cqm1 was recombinantly expressed. The clone encoding the Cqm1 protein was obtained from A/Prof. Michelle Dunstone (Monash University, Australia). The pET28a expression vector comprised residues 20 – 557 of the *cqm1* gene, lacking the GPI anchor attachment region and signal peptide, with an N-terminal 6 x histidine tag. Inclusion of a thrombin cleavage site (Leu-Val-Pro-Arg-||-Gly-Ser) enabled efficient cleavage of the histidine tag after protein expression. Inclusion of a kanamycin resistance gene enabled selection of clones containing the expression construct. The gene sequence of the pET28a-Cqm1 expression vector was validated by Sanger sequencing using T7 and T7 terminator primers (**Table 2.7**). Throughout this work, the Cqm1 sequence is numbered according to the PDB entry of the Cqm1 crystal structure (PDB 6K5P), which, itself, was solved from an expression vector encoding residues 23 – 600 of the insect *cqm1* gene. Hence, our Cqm1 sequence is numbered -3 – 537. Where comparisons to Aam1 and Agm3 are made, the Aam1 and Agm3 sequences are numbered as they exist in the insect, inclusive of the GPI anchor and signal peptide. The Cqm1 protein sequence and the relevant numbering scheme used have been provided in **Appendix 6**.

Conditions for Cqm1 expression were optimised by testing a range of IPTG concentrations (0.1 mM, 0.5 mM, and 1 mM) at induction temperatures of 16°C and 37°C. Soluble and insoluble samples were obtained as

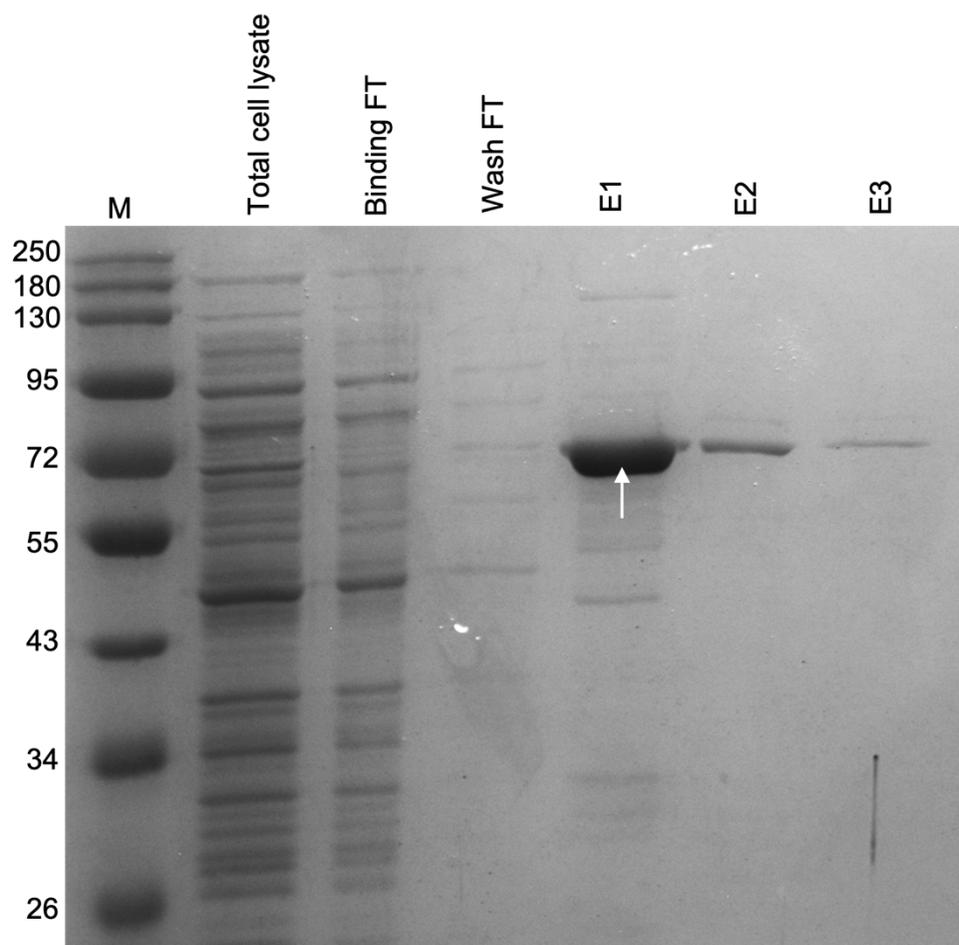
described in **section 2.3.1** and subsequently analysed using SDS-PAGE, showing that insoluble protein expression was reduced at lower temperatures, whilst marginal differences were seen in soluble protein expression (**Fig 5.3, ~64 kDa**). Given that there may have been increased production of soluble protein at lowered temperatures, the pET28a-Cqm1 construct was induced with 1 mM IPTG and expressed at 16°C overnight for all further protein preparations. Following expression, the 6xHis-Cqm1 fusion protein was extracted from cells as described in **section 2.3.2** and purified as shown below.



**Figure 5.3. Cqm1 protein expression at varied temperature and IPTG concentration.** SDS-PAGE analysis of Cqm1 protein expression (~64 kDa) at 16°C or 37°C using 0.1, 0.5, or 1 mM IPTG concentrations. Sol and Insol correspond to the soluble and insoluble cell proteins of induced cells, respectively. M (marker): Color Prestained Protein Standard, Broad Range.

### 5.2.6. Purification of Cqm1

To obtain the 6xHis-Cqm1 fusion protein (~64 kDa), nickel affinity chromatography was performed on whole cell lysates as described in **section 2.4.1**. Briefly, cell lysates were applied to equilibrated Protino Ni-TED 2000 packed columns and the flow-through was collected. The columns were washed with 8 mL LEW buffer and the 6xHis-Cqm1 fusion protein was eluted into 3 x 3 mL fractions of LEW-imidazole elution buffer, showing that Cqm1 eluted into fractions 1 – 3, with most of the protein eluting in the first fraction (**Fig 5.4**). Fractions containing the 6xHis-Cqm1 protein were pooled, concentrated, and stored at -20°C until downstream applications.

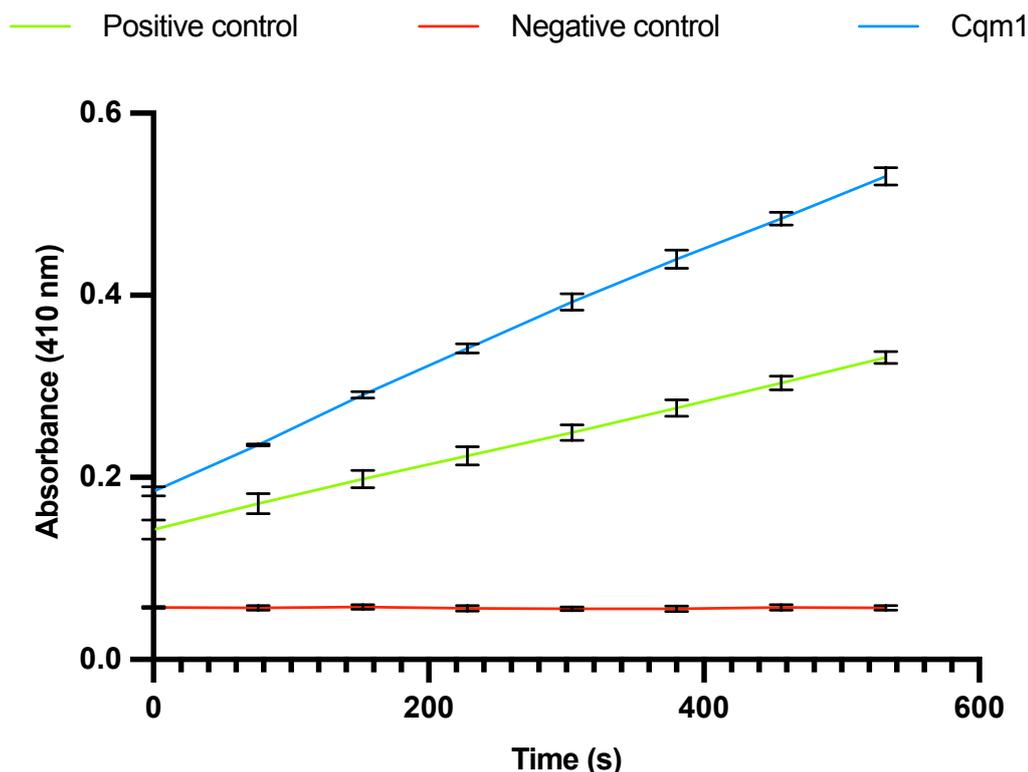


**Figure 5.4. Nickel affinity purification of 6xHis-Cqm1.** 6xHis-Cqm1 was expressed and purified via nickel affinity purification. Elution fractions obtained were analysed by coomassie stained SDS-PAGE showing partial purification of 6xHis-Cqm1 (white arrow, ~64 kDa). M (marker): Color Prestained Protein Standard, Broad Range.

### 5.2.7. Glucosidase assays

Previous work has demonstrated that Cqm1 exhibits  $\alpha$ -glucosidase activity (Sharma et al. 2018). The  $\alpha$ -glucosidases hydrolyse  $\alpha$ -1,4-linked polysaccharides to release glucose. To validate the viability of recombinant Cqm1 protein produced here,  $\alpha$ -glucosidase activity was assayed using the Abcam Alpha-Glucosidase Activity Assay Kit, as described in **section 2.5.5**.

Briefly, Cqm1 protein was mixed with the substrate (*p*-nitrophenol- $\alpha$ -D-glucopyranoside - colourless) to release *p*-nitrophenol, which is yellow in colour and can be measured colorimetrically (at 410 nm). Samples (10  $\mu$ L of  $\sim$ 100  $\mu$ g/mL Cqm1 – estimated using a BCA assay) were plated in triplicate into a 96-well microplate and made up to 50  $\mu$ L with  $\alpha$ -glucosidase assay buffer. A positive control (included in the kit) and negative control (buffer alone) were also plated in triplicate. Following addition of the  $\alpha$ -glucosidase substrate mix, the release of *p*-nitrophenol was monitored by measuring absorbance at 410 nm for 60 minutes every 76 seconds. For each time point, the average absorbance reading across the three repeats was calculated, and the absorbance against time was plotted for the linear portion of the curve (the first 10 minutes) (**Fig 5.5**). An increase in *p*-nitrophenol was detected for the positive control (**Fig 5.5 – green line**), whilst no increase in *p*-nitrophenol was detected for the negative control (**Fig 5.5 – red line**), indicating that the kit was functioning as expected. For Cqm1 (**Fig 5.5 – blue line**), an increase in *p*-nitrophenol confirmed  $\alpha$ -glucosidase activity and the validity of recombinant protein produced.



**Figure 5.5. Recombinant Cqm1 displays  $\alpha$ -glucosidase activity.**  $\alpha$ -glucosidase activity of Cqm1 was confirmed by an increase in p-nitrophenol, measured colorimetrically ( $A_{410 \text{ nm}}$ ). Positive (green) and negative (red) controls are shown. Error bars represent the standard error (n = 3).

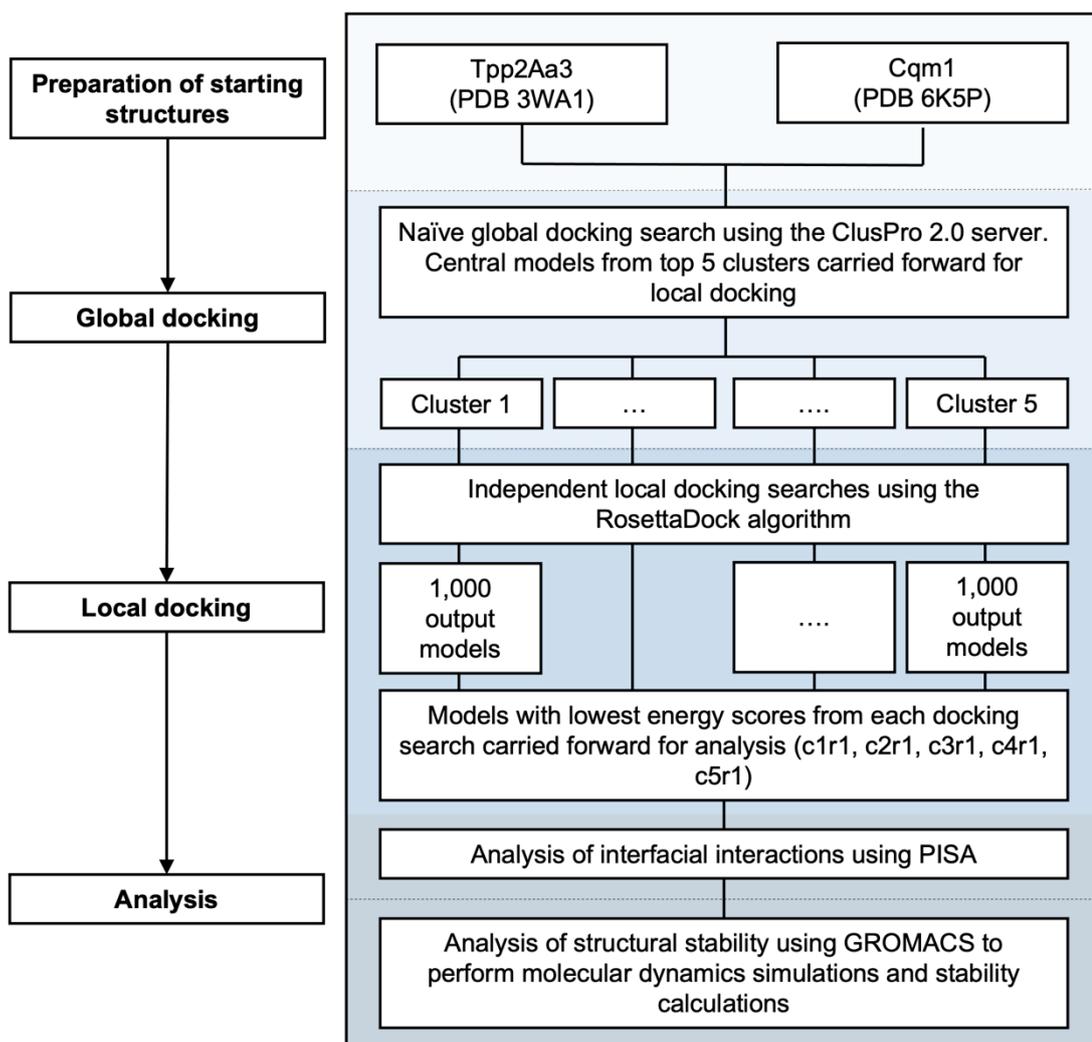
### 5.2.8. Prediction of the Tpp2Aa3-Cqm1 interaction

The specificity of *B. thuringiensis* and *L. sphaericus* pesticidal proteins is mediated by binding to target receptors present on midgut epithelial cells. Therefore, structural analysis of pesticidal proteins in complex with their target receptors will increase knowledge of specificity, which itself may be applied to downstream applications including protein engineering to produce pesticidal proteins with enhanced potency, stability and/or broadened target insect range. Despite this, no structures of toxin-receptor complexes have been elucidated for any *B. thuringiensis* or *L. sphaericus* pesticidal proteins.

In *Cx. quinquefasciatus*, the Tpp1/Tpp2 receptor binding protein has been identified as Cqm1, a GPI-anchored  $\alpha$ -glucosidase which normally

functions in carbohydrate digestion (Sharma et al. 2018). Orthologous receptor binding proteins have also been identified in Tpp1/Tpp2 susceptible species, *Cx. pipiens* (Cpm1 - Silva-Filha et al. 1999) and *An. gambiae* (Agm3 - Opota et al. 2008). An orthologous non-binding protein has been identified in the refractory (non-susceptible) species, *Ae. aegypti* (Aam1 - Ferreira et al. 2010). In *Culex* mosquitoes, receptor binding is thought to be mediated by Tpp2 which, once bound to the receptor, binds its partner protein, Tpp1 (Oei et al. 1992; Charles et al. 1997). Receptor binding has been investigated using several techniques, including surface plasmon resonance studies which demonstrated that Tpp2 binds Cqm1 with high affinity ( $K_d$  9.8 nM), whereas the interaction of Tpp1 with Cqm1 is ~ 1000 fold weaker ( $K_d$  2.2  $\mu$ M), suggesting that Tpp1 interaction is transient and non-specific (Sharma et al. 2018). This is in line with the current working model that suggests that Tpp2 acts as the receptor binding component, whilst Tpp1 acts as the pore-forming component. In addition, small-angle neutron scattering (SANS) studies have produced a low resolution, *ab initio* model of the Tpp2-Cqm1 complex (Sharma et al. 2020). Further studies are required to elucidate the full molecular details of complex structure and formation.

Here, we aimed to establish a model of the Tpp2-Cqm1 complex using a similar approach to that applied in the prediction of the Cry48Aa1-Tpp49Aa1 complex (**section 4.2.4**). Briefly, the protein structures were first prepared for modelling such that factors relating to their mode of action (including proteolytic processing and environment pH) had been considered. Following this, we utilised ClusPro (Kozakov et al. 2017) and RosettaDock (Chaudhury et al. 2011) to perform global and local docking, respectively. Following docking, the interfacial interactions of candidate models were assessed using PISA (Krissinel 2015) and, in addition, GROMACS (Abraham et al. 2015) was applied to perform MD simulations enabling the structural stability of proposed models to be evaluated. An overview of the computational modelling approach is provided in **Fig 5.6**.



**Figure 5.6. Overview of Tpp2-Cqm1 modelling.** Modified from Worthy et al., 2021.

#### 5.2.8.1. Preparation of Tpp2 and Cqm1 starting structures

Crystal structures for Tpp2 variants, namely Tpp2Aa2 in the Tpp1Aa2/Tpp2Aa2 natural co-crystal (PDB 5FOY) and trypsin-activated Tpp2Aa3 (PDB 3WA1), have been elucidated. Tpp2 variants (Aa1, Aa2, Aa3, Aa4) display high sequence identity, differing by a maximum of 6 amino acids between any two variants. Tpp2Aa2 and Tpp2Aa3 differ by 3 amino acids present in the C-terminal pore-forming domain.

Given that it is widely accepted that *B. thuringiensis* / *L. sphaericus* crystal proteins undergo solubilisation and proteolytic activation prior to receptor interaction, the structure of trypsin-activated Tpp2Aa3 was obtained for docking (Srisucharitpanit et al. 2014). Insect gut proteases (mainly trypsin-like and chymotrypsin-like) are involved in proteolytic cleavage of pest proteins to produce the activated protein core (Rukmini et al. 2000). At the time of these computational modelling studies, N-terminal sequencing of Tpp2Aa2 (**section 5.2.3**) had not been performed (due to Covid lockdown access restrictions), and therefore, the proteolytic activation sites of Tpp2 were unknown. However, previous studies have demonstrated that the N-terminal region of Tpp2Aa2, Asn33 – Leu158, is involved in receptor binding (Romão et al. 2011). In addition, removal of unstructured terminal regions is recommended for docking in ClusPro (Kozakov et al. 2017). Hence, for docking studies performed here, the unstructured N-terminal region of the Tpp2Aa3 protein was removed, resulting in the truncation of residues 19-33, leaving residues 34 – 407 remaining. The structure of the monomeric Cqm1 receptor was obtained by extracting one subunit (residues 6-536) from the tetrameric crystal structure (PDB 6K5P - Sharma and Kumar 2019). Starting structures were prepared using the PyMOL Molecular Graphics System (version 2.4.0).

The luminal pH of the mosquito larval midgut ranges from ~8 in the gastric caecum (Boudko et al. 2001) to > 10 in the anterior gut and ~7.5 in the posterior gut (Dadd 1975). Previously, regional binding of Tpp2Aa2 has been demonstrated in the gastric caecum and posterior midgut of *Cx. quinquefasciatus* larvae (Oei et al. 1992). In addition, Tpp1Aa2/Tpp2Aa2 toxicity has been demonstrated against mosquito cell lines maintained at pH 7 – 7.4 (Davidson and Titus 1987), and pull-down assays have demonstrated an interaction of Tpp2Aa2 and Cqm1 at pH 7.4 (Romão et al. 2011; Ferreira et al. 2014). Taken together, these studies demonstrate that Tpp2Aa2-Cqm1 interaction occurs at approximately neutral pH and hence, modelling studies were performed at pH 7, with the protonation states of titratable residues reflecting this.

### 5.2.8.2. Molecular docking studies

Molecular docking was performed using a naïve approach with no presumed Tpp2Aa3-Cqm1 interface. First, a global search was carried out using the ClusPro 2.0 server, the algorithm of which has been described in detail in **section 2.6.2.1** (Kozakov et al. 2017). The Tpp2Aa3 and Cqm1 starting structures were provided as the ligand and receptor, respectively. Briefly, ClusPro clusters output models according to RMSD. Clusters are ordered according to their size, with the largest (most populated) clusters expected to represent the most likely models. Of the 30 clusters identified, the representative models from the 5 largest clusters (**Table 5.1**) were carried forward for refinement.

**Table 5.1. Output Tpp2Aa3-Cqm1 clusters produced by ClusPro.**

ClusPro		
Cluster	Cluster size	Total energy (kcal mol <sup>-1</sup> )
c1	55	-668.0
c2	51	-628.4
c3	50	-620.7
c4	36	-666.6
c5	35	-618.9

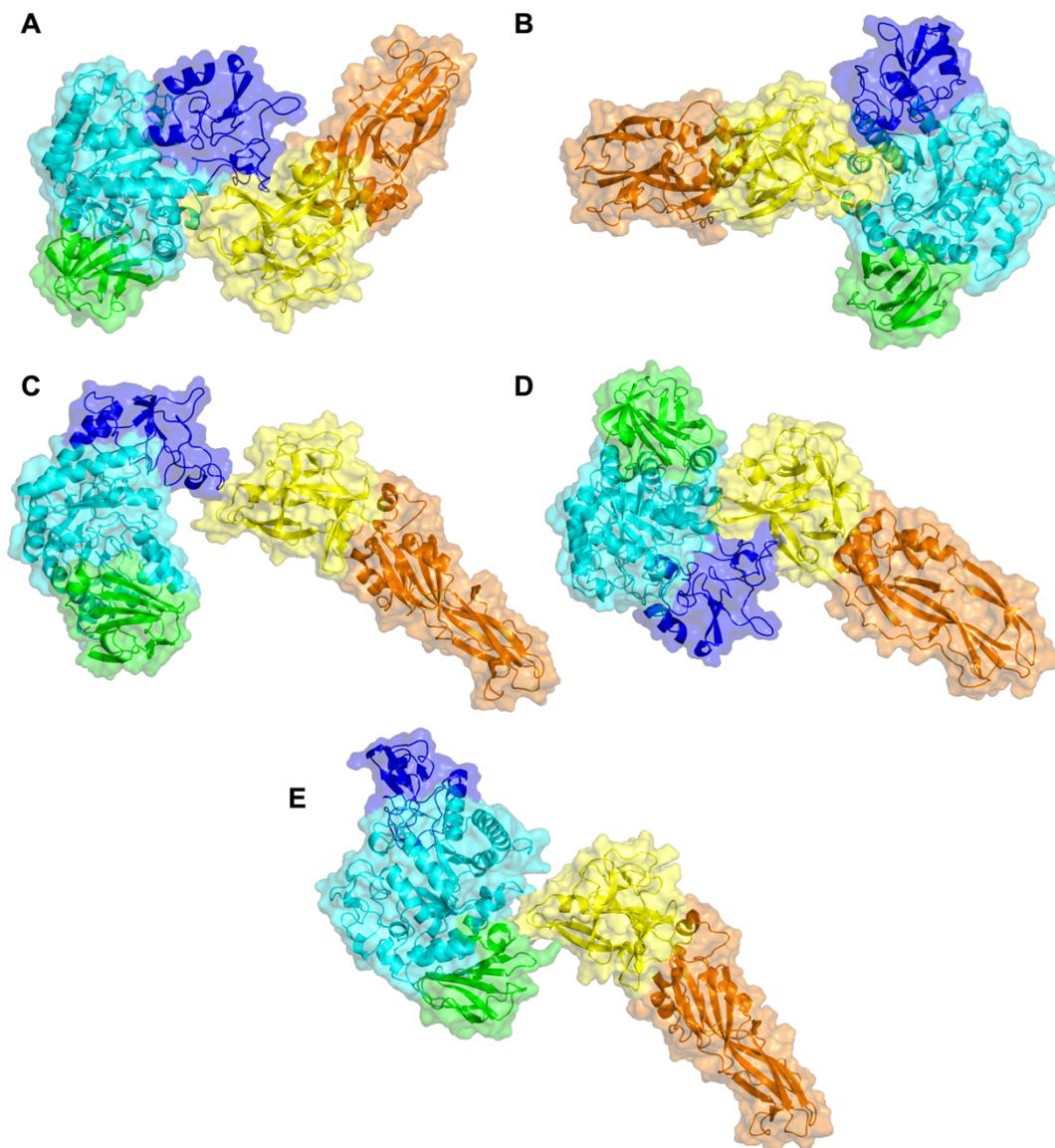
Following global docking, the RosettaDock source code, from the RosettaCommons software suite (described in detail in **section 2.6.2.2**) (Chaudhury et al. 2011), was utilised to refine each of the models by performing independent local docking searches. Briefly, the *docking\_prepack\_protocol.macosclangrelease* executable was employed to prepack starting models, ensuring that side chains were present in their lowest energy conformation. For each prepacking run, 25 models were produced and the model with the lowest energy score was carried forward for docking. Local docking searches were performed using the

*docking\_protocol.macosclangrelease* executable. Unbound rotamer conformations were provided to improve the accuracy of docking (Wang et al. 2005) and 1,000 output models were produced and ranked according to their total energy score. From each docking search, the 5 models with the lowest energy scores (**Table 5.2**) were identified and carried forward for further analysis. Of all 5 models identified, model c3r1 exhibited the lowest total energy value of -1421.7 REU, followed by model c4r1 at -1410.6 REU. In all models, the N-terminal region of Tpp2Aa3 was predicted to interact with Cqm1 (**Fig 5.7**), in line with previous studies showing that residues Asn33 – Leu158 are involved in receptor binding. From here on, models are referred to by their ClusPro cluster and Rosetta model rank (e.g., model c1r1 = ClusPro cluster 1, Rosetta rank 1).

**Table 5.2. Energy scores of models refined by RosettaDock.**

RosettaDock	
Model	Total Energy (REU)
c1r1	-1390.6
c2r1	-1388.7
c3r1	-1421.7
c4r1	-1410.6
c5r1	-1379.9

REU = Rosetta Energy Units



**Figure 5.7. Modelled Tpp2Aa3-Cqm1 complexes.** (A) c1r1 (B) c2r1 (C) c3r1 (D) c4r1 (E) c5r1. Tpp2Aa3 (N-terminal domain – yellow, C-terminal domain – orange) and Cqm1 (domain A – dark blue, domain B – cyan, domain C – green). All models are consistent with previous studies suggesting that receptor binding occurs via the Tpp2 N-terminal domain.

### 5.2.8.3. Interface analysis

As previously described (**section 4.2.4.3**), protein interfaces are characterised by several features, including interface area, hydrophobic interactions, hydrogen bonding, and salt bridge formation (Chothia and Janin 1975; Xu et al. 1997; Conte et al. 1999). To analyse these interfacial features, PISA (Krissinel 2015) was utilised. Across all 5 models, interface area ranged from 201.7 to 945.2 Å<sup>2</sup> (**Table 5.3**). In addition, models exhibited  $\Delta^iG$  values ranging from -2.3 to -9.1 kcal mol<sup>-1</sup> (**Table 5.3**). A negative  $\Delta^iG$  is associated with a hydrophobic interface and positive protein affinity (Chothia and Janin 1975). Of all the models analysed, model c4r1 exhibited the lowest (most negative)  $\Delta^iG$  values. Finally, all models were found to exhibit interfacial hydrogen bonding, whilst model c4r1 also exhibited salt bridge interactions (**Table 5.3**). These interactions contribute to interface stability, as well as binding specificity where side chains are involved (Xu et al. 1997). No models were found to exhibit interfacial disulphide bridges, suggesting that disulphides do not play significant roles in the Tpp2Aa3-Cqm1 interaction. To investigate the dynamics and stability of modelled Tpp2Aa3-Cqm1 complexes further, MD simulations were performed.

**Table 5.3. Interfacial interactions of Tpp2Aa3-Cqm1 models.**

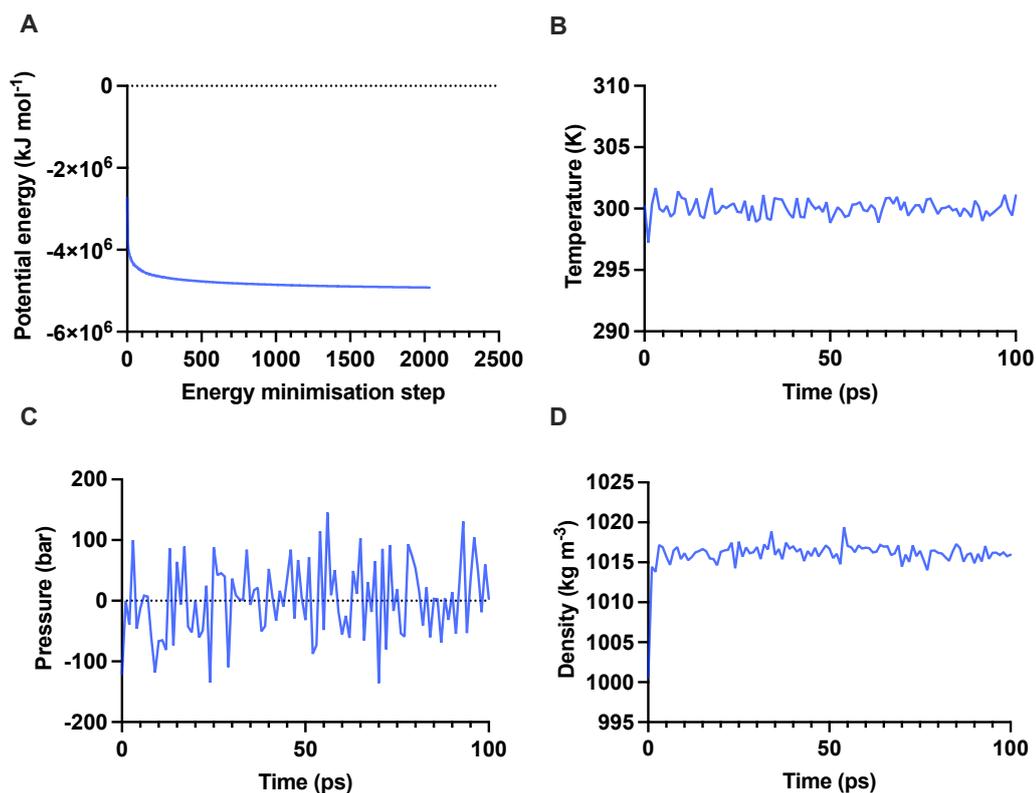
PISA					
Model	Interface Area (Å <sup>2</sup> )	$\Delta^iG$ (kcal mol <sup>-1</sup> )	No. H-bonds	No. salt bridges	Binding energy (kcal mol <sup>-1</sup> )
c1r1	718.8	-6.0	3	0	-7.3
c2r1	732.8	-8.7	3	0	-10.0
c3r1	201.7	-2.8	2	0	-3.7
c4r1	945.2	-9.1	3	2	-11.1
c5r1	310.6	-2.3	2	0	-3.2

#### 5.2.8.4. Molecular dynamics (MD) simulations

MD uses Newton's equations of motion to model the dynamics of atoms and molecules with respect to time (Hollingsworth and Dror 2018) and, in the past, has been utilised to evaluate the structural stability of modelled interfaces throughout output trajectories (Radom et al. 2018).

Here, all-atom MD simulations of the modelled Tpp2Aa3-Cqm1 complexes were performed for 100 ns using GROMACS (Abraham et al. 2015) and a similar approach described for prediction of the Cry48Aa1-Tpp49Aa1 interaction in **section 4.2.4.4**. Briefly, the structures were first prepared using the pdb2gmx module and AMBER99SB forcefield (Hornak et al. 2006; Lindorff-Larsen et al. 2010). As default, the pdb2gmx module assigns protonation states of titratable residues according to a neutral pH environment. The protonation state of histidine was chosen automatically according to optimal hydrogen bonding conformation.

To ensure correct structural geometry and remove steric clashes within the system, energy minimisation was performed using a maximum number of 50,000 steps in 0.01 step sizes, and a maximum force ( $F_{max}$ ) of  $1000 \text{ kJ mol}^{-1} \text{ nm}^{-1}$  (**Fig 5.8A**). Following energy minimisation, the temperature, pressure, and density of the system was equilibrated (**Fig 5.8B, 5.8C, 5.8D**). Finally, 100 ns MD simulations were performed as described in **section 2.6.4.6**.



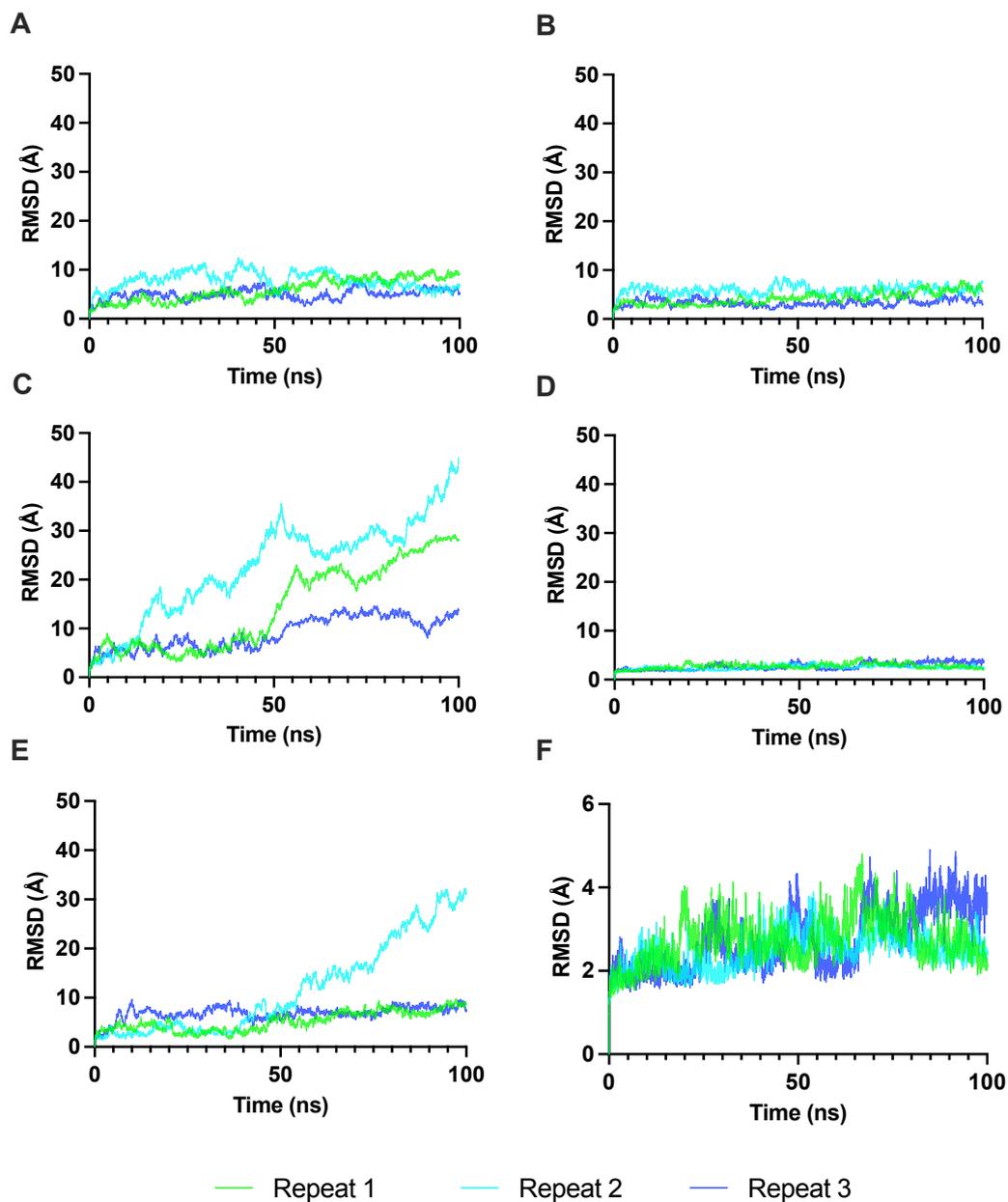
**Figure 5.8. Energy minimisation and system equilibration of the Tpp2Aa3-Cqm1 simulations.** An example of MD set-up has been illustrated for simulation of model c1r1. (A) Potential energy ( $E_{\text{pot}}$ ) plotted against the energy minimisation. A successful energy minimisation is identified by a negative  $E_{\text{pot}}$  with an order of  $10^5 / 10^6$ . (B) Average temperature of the simulation. (C) Average pressure of the simulation. (D) Average density of the simulation.

#### 5.2.8.5. Root-mean-square deviation (RMSD) trajectory analysis

To assess the structural stability of modelled Tpp2Aa3-Cqm1 complexes, the RMSD of the position of backbone atoms was calculated as a function of time using the GROMACS command line `gmx_rms` module (**Fig 5.9**). As previously described, RMSD calculates the deviation of a selection of atoms in reference to their initial starting position. High RMSD values, therefore, correlate with large changes in the structure, indicating significant

instability, whilst low RMSD values correlate with small changes in the structure, indicating stability.

Within the presented MD simulations, the lowest RMSD values were obtained for model c4r1 (**Fig 5.9D**), ranging from 1.42 – 4.91 Å. In contrast, the RMSD values of other models were found to fluctuate. For example, the calculated RMSDs of model c3r1 (**Fig 5.9C**) continued to increase throughout the duration of the simulations, indicating a highly unstable structure. Taken together, this analysis suggests that model c4r1 represents a more stable structure than other models analysed and thus, is more likely to persist in solution. Ranges calculated exclude RMSD values obtained during the first 1 ns of the MD simulations, corresponding to the ‘warm-up’ phase.

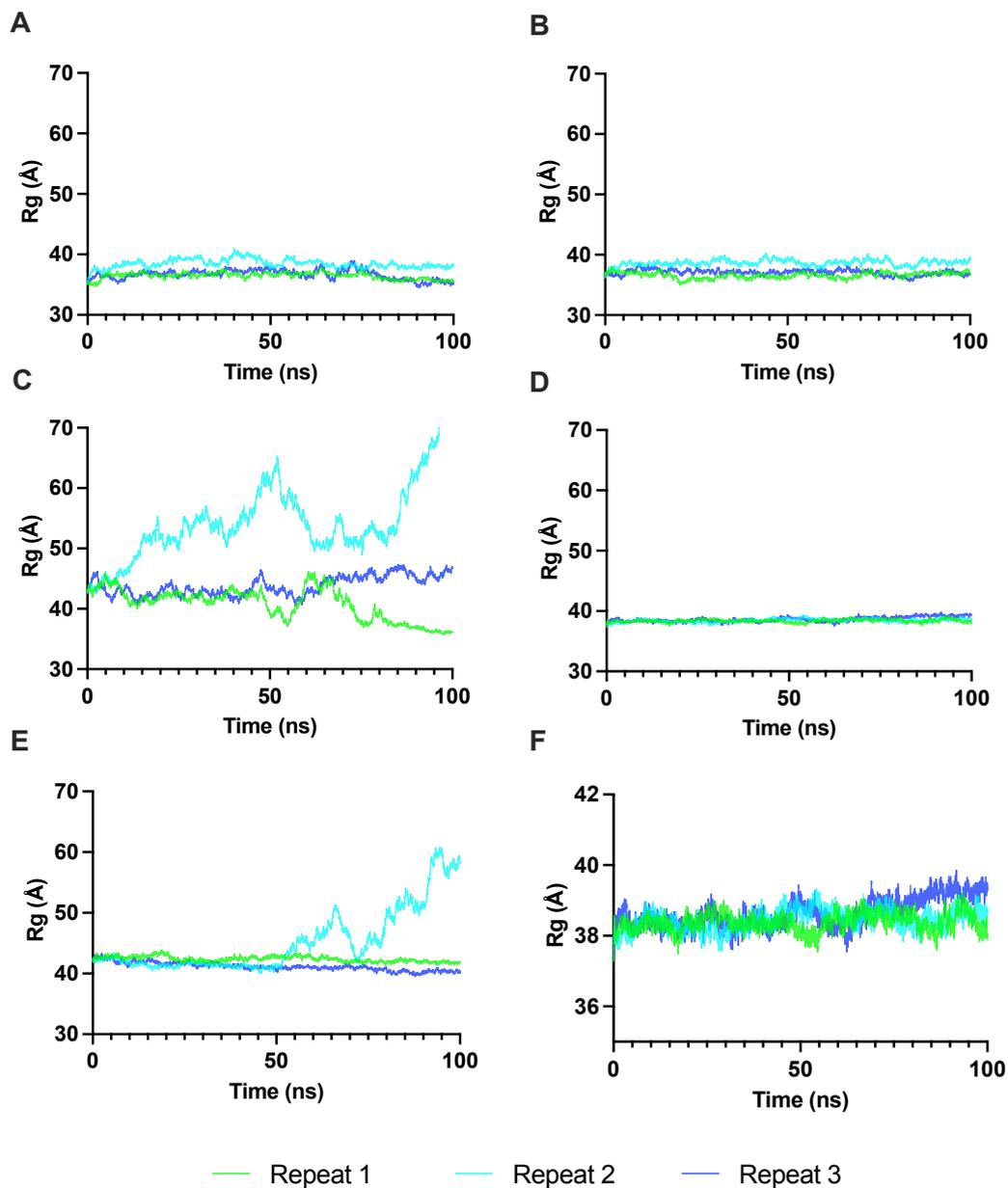


**Figure 5.9. RMSD of Tpp2Aa3-Cqm1 models throughout 100 ns simulations.** High RMSD values indicate significant structural instability, whilst low RMSD values indicate structural stability. (A) c1r1 (B) c2r1 (C) c3r1 (D) c4r1 (E) c5r1 (F) Zoom of the most stable structure (c4r1) indicated by RMSD analysis. MD simulations were repeated three times (green, blue, and cyan).

#### 5.2.8.6. Radius of gyration (Rg) trajectory analysis

Further investigation of the structural stability of the modelled Tpp2Aa3-Cqm1 complex was performed by calculating the Rg as a function of time (**Fig 5.10**). As previously described, Rg analyses the mass and position of atoms in relation to the centre of mass of the molecule and thus, indicates the compactness of the overall structure. In this case, a stably folded structure is expected to maintain a steady Rg value.

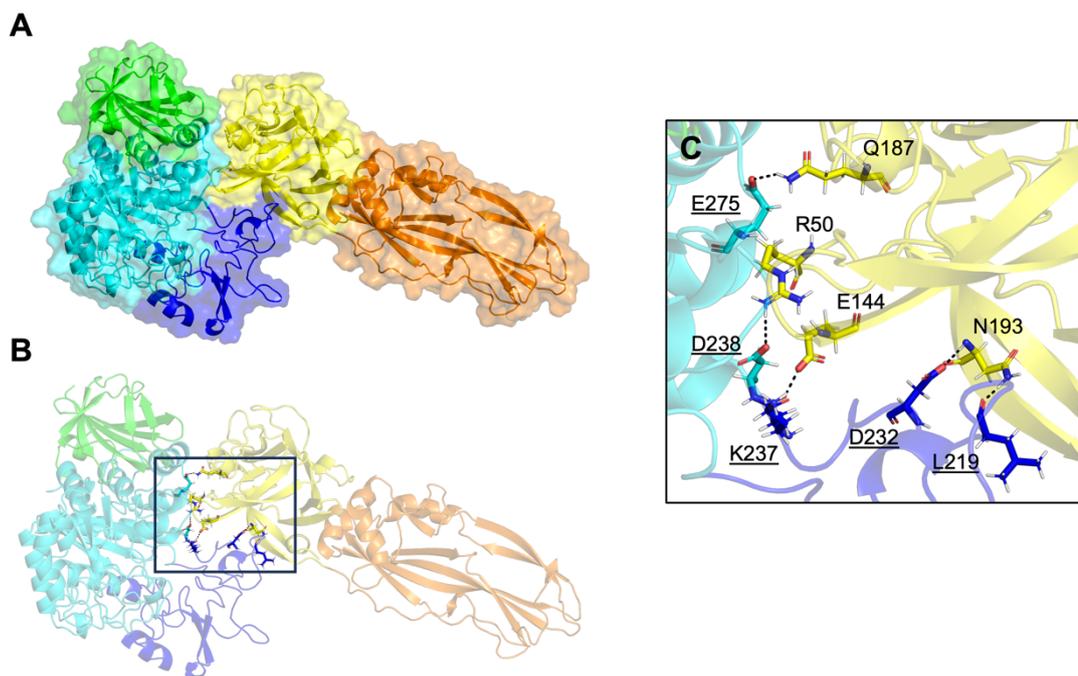
Within the presented MD simulations, models c1r1, c2r1, and c4r1 (**Fig 5.10**) were found to exhibit stable Rg values for all simulation repeats. Consistent with RMSD analysis, the Rg values of model c3r1 largely fluctuated over time, indicating that model c3r1 is unstable, whilst model c4r1 exhibited the most consistent Rg values. Model c3r1, repeat 1, exhibited Rg values which decreased over time, suggesting that the complex became more compact. However, visualisation of the MD trajectory confirmed that the structure moved out of its original binding pose, in line with its instability.



**Figure 5.10. Rg of Tpp2Aa3-Cqm1 models across 100 ns simulations.** Radius of gyration (Rg) indicates the compactness of the overall structure. Stably folded structures are expected to maintain a steady Rg value. (A) c1r1 (B) c2r1 (C) c3r1 (D) c4r1 (E) c5r1 (F) Zoom of the most stable structure (c4r1) indicated by RMSD analysis. MD simulations were repeated three times (green, blue, and cyan).

#### 5.2.8.7. Selection of most likely Tpp2Aa3-Cqm1 model

As described above, we aimed to establish a model of the Tpp2Aa3-Cqm1 complex by utilising a combined global and local docking approach. A total of 5 models were evaluated using PISA (Krissinel 2015) to perform interface analysis and GROMACS (Abraham et al. 2015) to perform MD simulations. Of all 5 models investigated, model c3r1 exhibited the lowest total energy value of -1421.7 REU. The thermodynamic hypothesis states that the native protein structure will exist at the lowest free energy (Anfinsen 1973) and hence, c3r1 was a possible candidate structure of the Tpp2Aa3-Cqm1 complex. Despite this, model c3r1 was found to be highly unstable in 100 ns MD simulations, indicated by RMSD and Rg analysis, suggesting that this structure may not persist in solution. In contrast, model c4r1, which exhibited the next best energy score of -1410.6 REU, remained stable throughout 100 ns MD simulations. Taken together, we hypothesized that model c4r1 (**Fig 5.11**) may represent a similar structure to the native Tpp2Aa3-Cqm1 structure.



**Figure 5.11. Proposed Tpp2Aa3-Cqm1 model c4r1.** (A) Surface and cartoon representation of the modelled Tpp2Aa3-Cqm1 complex. Tpp2Aa3 N-terminal domain – yellow, C-terminal domain – orange. Cqm1 domain A – dark blue, domain B – cyan, domain C – green. (B, C) Several polar contacts (black dashed lines), including hydrogen bonds and salt bridges, were identified in the Tpp2Aa3-Cqm1 model using PISA. Residues involved in polar contacts are shown as sticks (carbon – cyan / blue / yellow, oxygen – red, nitrogen – blue). To distinguish Tpp2Aa3 / Cqm1 residues, Cqm1 residue labels have been underlined.

**Table 5.4. Tpp2Aa3-Cqm1 model c4r1 interactions identified by PISA.**

<b>Interfacial residues*</b>	
<b>Tpp2Aa3</b>	<b>Cqm1</b>
Glu36	Arg215
Ile37	Asp216
Lys40	Glu217
Tyr42	Pro218
Lys45	Leu219 – H
Lys47	Ser220
Tyr48	Gly221
Arg50 – S	Trp222
Gln134	Gly223
Gly136	Pro225
Asp139	Asp229
Tyr140	Asp232 – H
Ile141	His233
Thr142	Ile234
Gly143	Tyr235
Glu144 – S	Lys237 – S
Gln145	Asp238 – S
Phe146	Pro240
Phe147	Tyr243
Leu164	Tyr271
Tyr180	Ser272
Pro181	Ser273
Pro182	Ile274
Ala183	Glu275 – H
Ser184	Gly276
Gln187 – H	Leu279
Val192	Gln298
Asn193 – H	Tyr301
Ser194	Asp302
Ser195	Gln307
Phe196	Gly311
	Ser314
	Ser315
	Trp318
	Asn322

\*Table rows do not correspond to pairs of Tpp2-Cqm1 interacting residues.

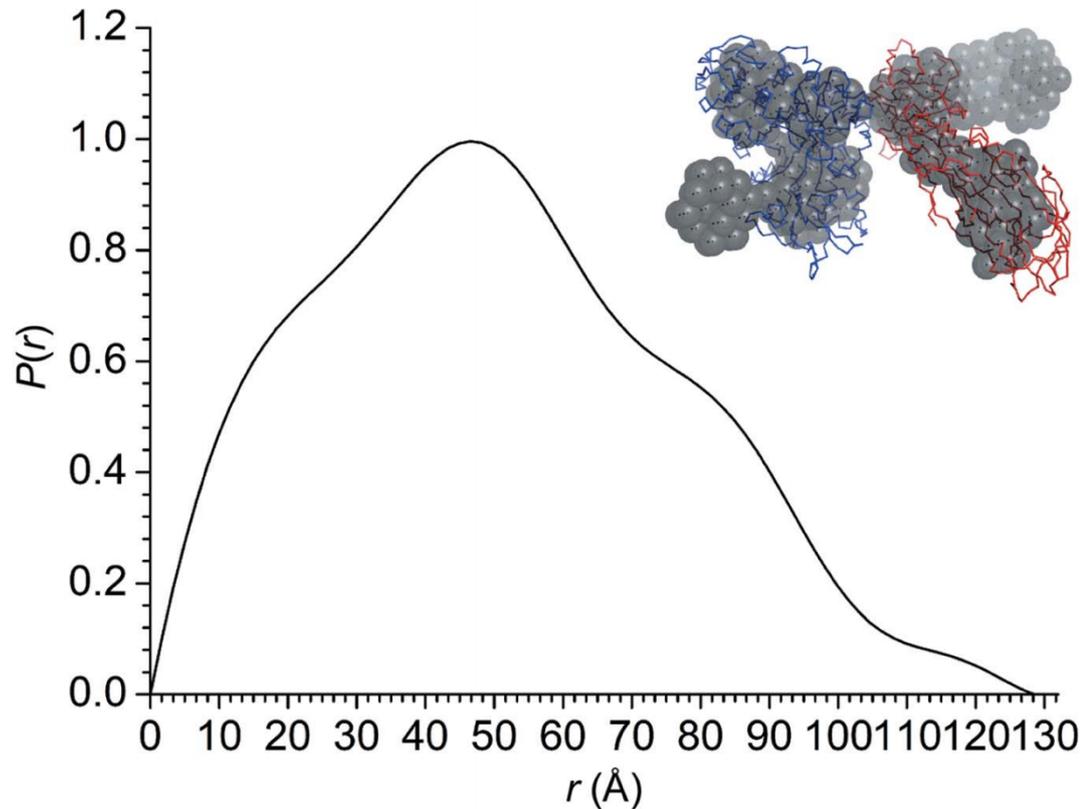
'H' indicates a residue involved in hydrogen bonding and 'S' indicates a

residue involved in salt bridge formation. Residues implicated in Tpp2-Cqm1 interaction in published mutagenesis work have been highlighted in grey.

In order to validate models produced by computational docking procedures, experimental data are required, for which several published studies exist. In Tpp2, N- and C-terminal truncations have identified the N-terminal region, Asn33-Leu158, as sufficient for receptor binding (Romão et al. 2011). Additionally, mutagenesis studies targeting residues Tyr42, 85-Ile-Arg-Phe-87, and 147-Phe-Gln-Phe-Tyr-150 have shown either a lack of, or reduced Cqm1 binding (Singkhamanan et al. 2010; Romão et al. 2011; Singkhamanan et al. 2013). Similarly, in Cqm1, mutagenesis studies have identified the region Ser129 – Ala12 (Ser108-Ala291 in the clone used in this work) as significant for Tpp2 binding (Ferreira et al. 2014). Moreover, mutagenesis studies targeting 159Gly-Gly160 (137Gly-Gly138 in the clone used in this work) have shown a lack of Tpp2 binding (Ferreira et al. 2014). These results suggest a role for these motifs in Tpp2-Cqm1 complex formation. Despite this, we cannot be sure whether targeted residues are directly or indirectly involved in binding. For example, residues that are located at the interface surface may be directly involved in the formation of interfacial contacts, whereas residues buried in the protein structure may be significant for maintaining the general fold of the binding site and thus, mutations within these regions may lead to large structural movements which disrupt the interfacial interactions and, hence, complex formation.

Several interfacial residues were identified in the Tpp2Aa3-Cqm1 model (c4r1) using PISA (**Table 5.4**). Of those interfacial residues, Tyr42 and Phe147 have been implicated in receptor binding in the above mutagenesis studies (Romão et al. 2011; Singkhamanan et al. 2013, ). In addition, in model c4r1, the Tpp2Aa3 interface is entirely composed of the N-terminal region (**Fig 5.11 – coloured in yellow**), consistent with deletion studies showing that the Tpp2 N-terminal region Asn33-Leu158 is involved in receptor binding (Romão et al. 2011). However, in contrast to the published mutagenesis studies, model c4r1 does not correspond with the *ab initio*

model (**Fig 5.12**) produced using data collected from SANS (Sharma et al. 2020). Specifically, the SANS-based model suggests that Tpp2 interacts with Cqm1 on the opposite side of Cqm1 domain B to that predicted in model c4r1 produced here (**Fig. 5.11**). This may be due to the inherent inaccuracies of SANS, which include the fact that multiple *ab initio* models with an equally good fit may be reconstructed from SANS scattering data (Jacques and Trehella 2010; Tuukkanen et al. 2016). Furthermore, enantiomorphic models produce exactly the same scattering pattern and hence, may not be differentiated (Tuukkanen et al. 2016).



**Figure 5.12. *Ab initio* SANS-based shape model of Tpp2-Cqm1.** Contrast-matched small angle neutron scattering (SANS) data suggests that the Cqm1 dimer dissociates to a monomer upon Tpp2 interaction. The atomic structures of Tpp2 (red) and Cqm1 (blue) were fitted into the *ab initio* shape model (grey spheres) to model the molecular details of the interaction. The low-resolution model suggests that Tpp2 interacts with Cqm1 domain B, but at the opposite side to that suggested in our model, c4r1. Figure taken from Sharma et al., 2020.

In collaboration with researchers at Monash University, the Tpp2Aa2-Cqm1 structure was subsequently elucidated using single-particle cryoEM, allowing direct comparison between the experimental structure and model produced here.

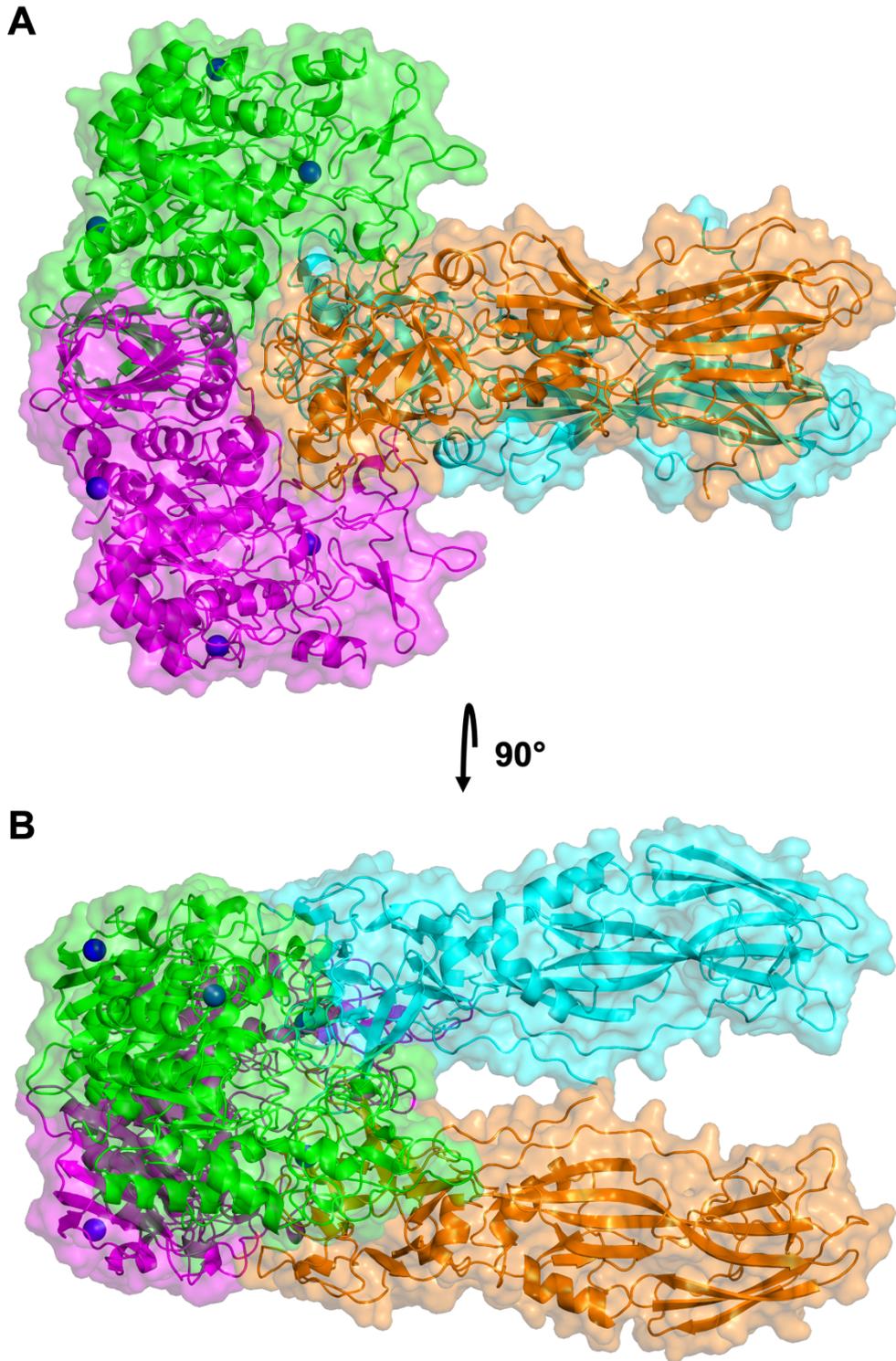
## 5.2.9. Structural analysis of the Tpp2Aa2-Cqm1 complex using single-particle cryogenic electron-microscopy (cryoEM)

### 5.2.9.1. Data collection and processing

*The Tpp2Aa2-Cqm1 structure was elucidated using single-particle cryoEM by Dr. Bradley Spicer, Hari Venugopal, Dr. Christopher Lupton, and A/Prof. Michelle Dunstone (Monash University – Melbourne, Australia). To elucidate the structure, purified Tpp2Aa2-Cqm1 complexes (obtained by incubating at 1.5:1 monomeric molar ratio of pro-form of Tpp2Aa2 and Cqm1) in 50 mM Tris, 150 mM NaCl, pH 8.0 were concentrated and prepared in cryoEM grids for screening by Dr. Bradley Spicer. Data collection for single-particle analysis was performed by Dr. Christopher Lupton. A total of 5040 movies were collected from which particles were selected for 2D classification. 2D classification revealed several classes (**Appendix 7**) which were used for 3D reconstruction (performed by Dr. Bradley Spicer) using C2 symmetry, leading to a final map with a global resolution of 2.42 Å (gold-standard Fourier shell correlation graph has been presented in **Appendix 7**). Local refinement of loops within the Tpp2Aa2 C-terminal domain, which displayed lower local resolution (possibly due to increased flexibility) was required. In addition, Tpp2Aa2 residues 137 – 140, present at the Tpp2Aa2-Cqm1 interface, sterically clashed with Cqm1 in the initial model, and thus were manually rebuilt, suggesting that receptor-binding induced movements occur within this region.*

### 5.2.9.2. Comparison of the predicted model with the Tpp2Aa2-Cqm1 structure

Structure solution of the Tpp2Aa2-Cqm1 complex revealed a tetramer consisting of the Cqm1 dimer and two Tpp2Aa2 monomers (**Fig 5.13**).



**Figure 5.13. Tpp2Aa2-Cqm1 structure elucidated using single-particle cryo-electron microscopy.** (A) Front and (B) side view of Tpp2Aa2-Cqm1 tetramer consisting of the Cqm1 dimer (green, magenta) and two monomers of Tpp2Aa2 (cyan, orange). Each Cqm1 monomer was found to coordinate three calcium ions (dark blue spheres).

Prior to elucidating the Tpp2Aa2-Cqm1 structure using cryoEM techniques, we predicted the Tpp2Aa3-Cqm1 interaction using a combined ClusPro-Rosetta-GROMACS modelling approach (**section 5.2.8**). Molecular docking techniques are increasingly used to complement wet-lab techniques and thus, it is of interest to compare the experimental structure with that of the predicted model to assess the accuracy of the approach used.

Structural alignment of the predicted model (**Fig 5.11 – model c4r1**) with the Tpp2Aa2-Cqm1 structure (**Fig 5.14**) indicates an all-atom RMSD (including outliers) of 7.537 Å. The global RMSD score, derived from the distance between superimposed atomic coordinates, is the most widely applied quantitative measure of structural similarity. However, this method is often compromised by outliers, such as long disordered loops and relative domain movements, which can compromise the accuracy of superposition (Kufareva and Abagyan 2012). Consequently, structures with minor discrepancies, such as flexible terminal regions, may exhibit considerable differences in backbone RMSD. Weighted RMSD calculations allow atomic subsets to be selected, thus mitigating the impact of unstructured regions. However, RMSD values are also impacted by protein size, which becomes particularly relevant when comparing models of varying size.

Given the above limitations, we sought to compare our models to that of the experimentally resolved Tpp2-Cqm1 structure using other structural comparison tools, in addition to the RMSD assessment. To do so, QS-score (Bertoni et al. 2017) and DockQ-wave (Basu and Wallner 2016) assessments, available in the Compare to Reference analysis of the SWISS-MODEL Structure Assessment tool (<https://swissmodel.expasy.org/assess>) were applied. Given that the Tpp2-Cqm1 dimer was predicted, only chains B and C of the experimentally resolved Tpp2Aa2-Cqm1 structure were compared to the models. QS-score and DockQ-wave assessments are inherently superposition free and thus, overcome limitations associated with comparisons based upon superposition. QS-score measures the distance differences of inter-chain contacts and, thus, quantifies the similarity between interfaces. A four-level classification is used to discriminate the QS-score

output: incorrect (QS-score < 0.1), acceptable (0.1 ≤ QS-score < 0.3), medium (0.3 ≤ QS-score < 0.7), high (QS-score ≥ 0.7). DockQ-wave is a modified version of the DockQ scoring assessment used in CAPRI. DockQ is derived by combining three measurements, namely Fnat (the fraction of conserved native contacts, where a contact is defined as an inter-chain residue pair with at least one heavy atom within 5 Å of each other), LRMS (RMSD of ligand positions following superposition of the receptor atoms), and iRMS (RMSD between interface positions). Again, a four-level classification is used to discriminate the DockQ-wave output: incorrect (DockQ < 0.23), acceptable (0.23 ≤ DockQ < 0.49), medium (0.49 ≤ DockQ < 0.8), high (DockQ ≥ 0.8).

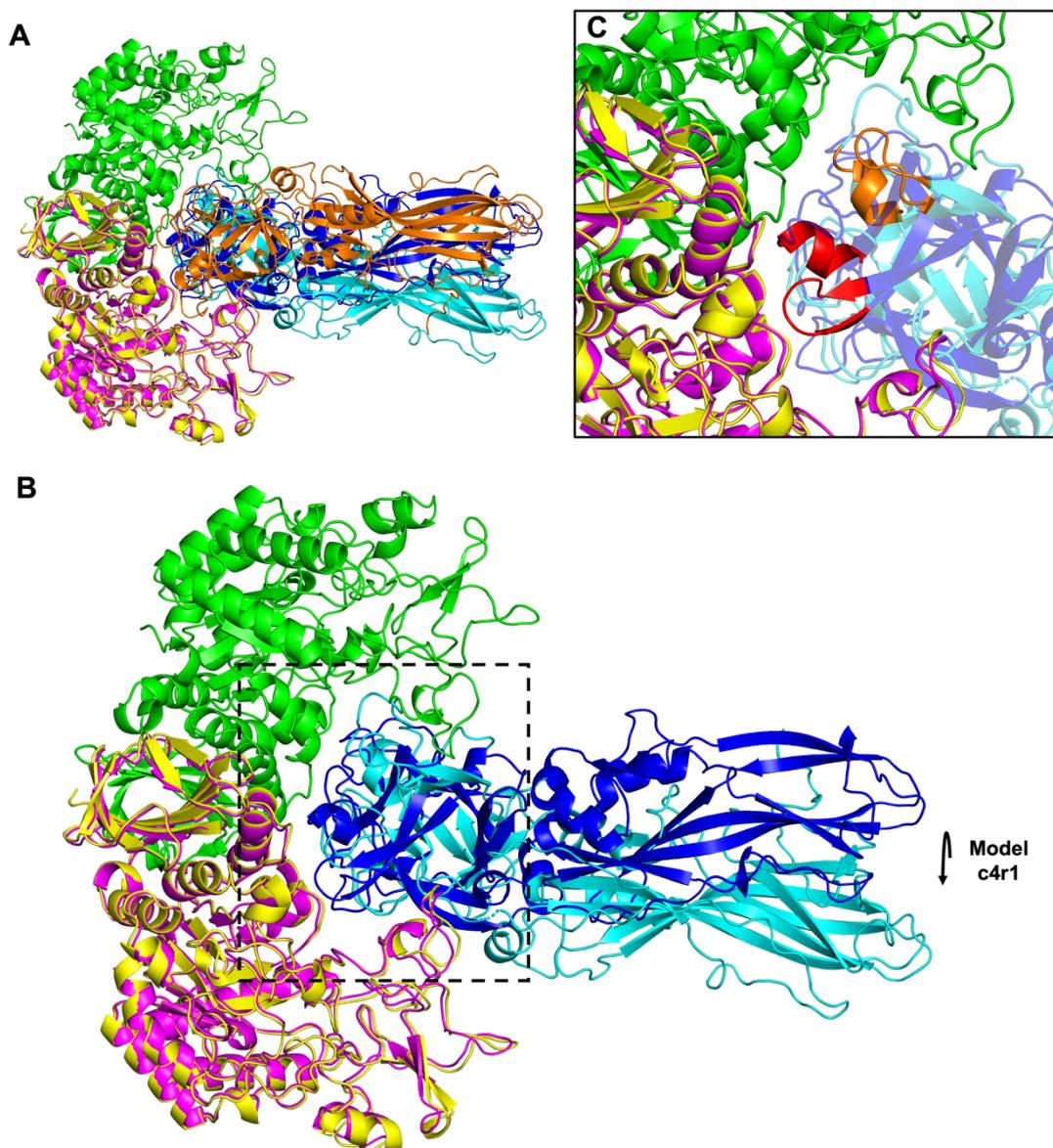
**Table 5.5. Calculated scores of the predicted Tpp2-Cqm1 models when compared with the experimentally resolved Tpp2Aa2-Cqm1 structure (chains B and C).**

Model	RMSD (Å)	QS-score	DockQ
c1r1	14.428	0.00	0.04
c2r1	25.952	0.00	0.01
c3r1	19.996	0.00	0.01
c4r1	7.537	0.21	0.11
c5r1	25.768	0.00	0.00

The predicted model (model c4r1) exhibited a QS-score of 0.21 (classified as an acceptable prediction) and a DockQ score of 0.11 (classified as an incorrect prediction), as indicated in **Table 5.5**. Compared with the experimental structure, Tpp2 is rotated 90° in model c4r1 (**Fig 5.14 – arrow**). In model c4r1, 3 hydrogen bonds (Tpp2-Cqm1 residues: Gln187-Glu275, Asn193-Asp232, and Asn193-Leu219) and 2 salt bridges (Tpp2-Cqm1 residues: Arg50-Asp238 and Glu144-Lys237) were predicted. Analysis of the cryoEM structure (presented later in **section 5.2.9.4**) indicated that several of these residues form hydrogen bonds in the cryoEM structure (Tpp2 residues Arg50, Gln187 and Cqm1 residues Leu219, Asp232 and Glu275). However, in model c4r1, the partner residues have been predicted

incorrectly. Superposition of the Cqm1 monomers of model c4r1 and the cryoEM structure (**Fig 5.14**) enabled the relative position of the Tpp2 monomers to be visualised and demonstrated that, despite residing at the interface in model c4r1, Tpp2 residues Arg50 and Gln187 are positioned 12.0 Å and 10.9 Å, respectively, from their actual positions in the cryoEM structure (calculated using the measure distance tool in PyMOL). Despite this, out of all 5 models produced by the combined ClusPro-Rosetta approach (models c1r1 – c5r1), model c4r1 displays the highest degree of similarity with the cryoEM structure as per all-atom RMSD, QS-score and DockQ score calculations (**Table 5.5**). This demonstrates that use of MD simulations to model the stability of predicted ClusPro-Rosetta complexes as an additional step to evaluate the models was advantageous as, had we relied on predicted energy scores alone, model c3r1 would have been identified as the most-likely model.

Taken together, these data show that the combined ClusPro-RosettaDock-GROMACS approach can be applied to produce modelled protein-protein complexes that can identify some interfacial residues correctly. Such models are good starting points for downstream wet-lab techniques, such as mutagenesis experiments probing interfacial interactions. In contrast to a low-resolution SANS model (**Fig 5.12**), which predicted dissociation of the Cqm1 dimer upon Tpp2 interaction to produce a 1:1 heterodimer (Sharma et al. 2020), structure solution of the Tpp2Aa2-Cqm1 complex by cryoEM revealed a tetramer consisting of the Cqm1 dimer and two Tpp2Aa2 monomers (**Fig 5.13**). Since the SANS-based model was published prior to this work, modelling studies performed here assumed a heterodimer consisting of Cqm1 and Tpp2Aa2 monomers. Given that the cryoEM structure demonstrates that each Tpp2Aa2 monomer interacts with both Cqm1 monomers, the accuracy of our modelling may have been improved if the model of the tetramer was predicted, instead of the dimer.



**Figure 5.14. Accuracy of ClusPro-Rosetta-GROMACS modelling approach.** (A) Structural alignment of the Cqm1 monomer in the predicted Tpp2-Cqm1 model (model c4r1 – Cqm1 - yellow, Tpp2Aa3 - dark blue) with the cryoEM structure of Tpp2Aa2-Cqm1 (Cqm1 monomers - green, magenta, Tpp2Aa2 monomers - cyan, orange) indicates an all-atom RMSD (including outliers) of 7.537 Å and shows the relative binding position of Tpp2. (B) The same structural alignment with the second Tpp2Aa2 monomer in the cryoEM structure (orange in panel A) removed to enable better visualisation of the predicted vs experimentally resolved Tpp2 binding site (dashed black box). Compared with the cryoEM structure, the Tpp2 monomer in the predicted model is rotated 90° around the x-axis (indicated by the arrow). (C) Zoom of

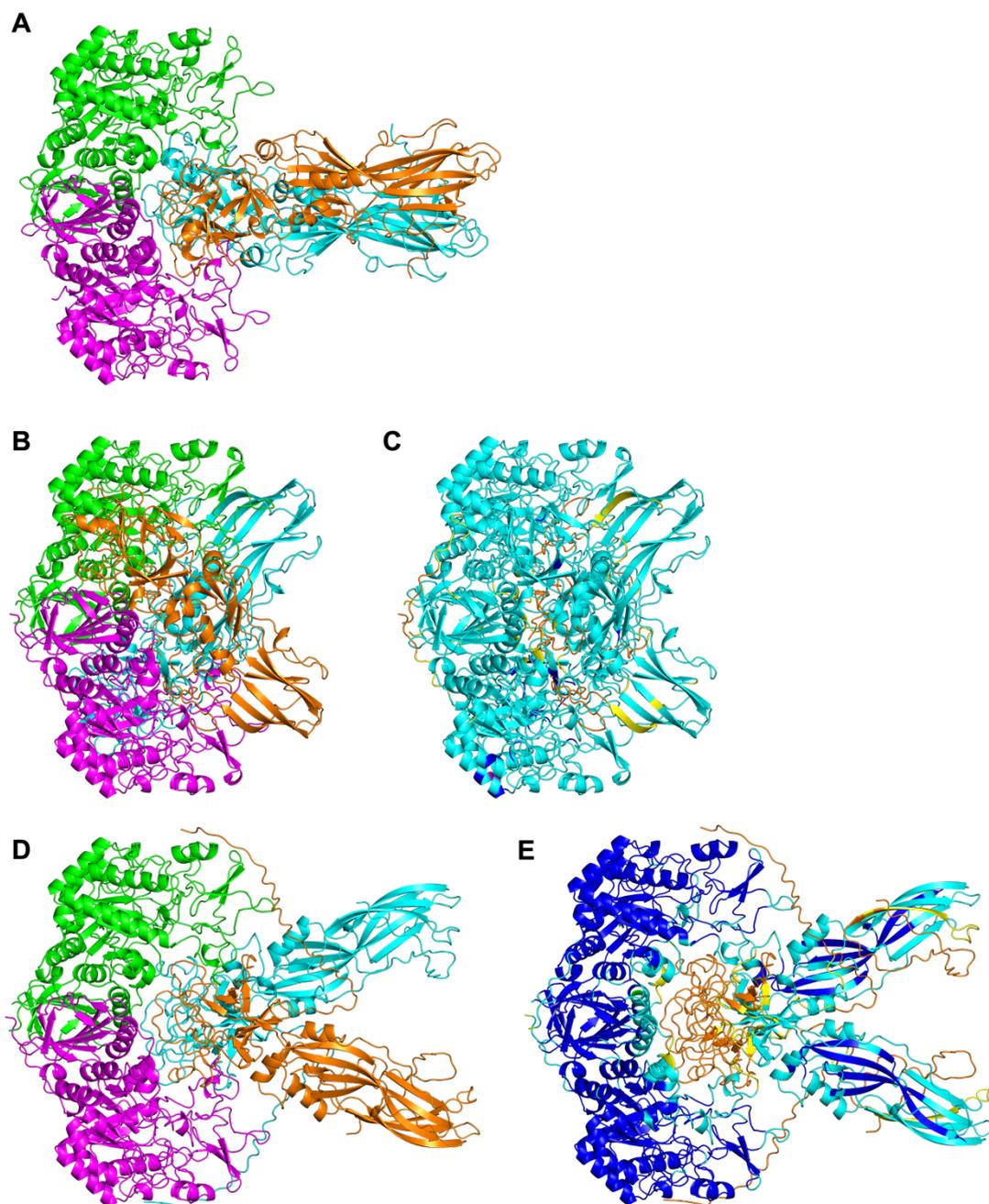
the predicted and experimentally resolved interface. The relative positions of a  $\beta$ -hairpin loop and short helix (model c4r1 – red, cryoEM structure – orange) have been indicated.

### 5.2.9.3. Modelling Tpp2-Cqm1 using Alphafold-Multimer and ColabFold

Following completion of initial modelling studies of Tpp2Aa3-Cqm1, AF2 (Jumper et al. 2021b), a deep-learning based prediction tool which uses a convolutional neural network trained on structures available in the PDB, was released. This led to spin-off algorithms such as AlphaFold-Multimer (Evans et al. 2022) and ColabFold (Mirdita et al. 2022) for prediction of protein-protein interactions. To investigate the accuracy of AlphaFold-Multimer and ColabFold in comparison to the modelling approach used here, the Tpp2Aa2-Cqm1 structure was predicted. To allow direct comparison with the cryoEM structure, the Tpp2Aa2-Cqm1 structure was predicted by providing the Tpp2Aa2 and Cqm1 sequences corresponding to the pGEX-Tpp2Aa2 and pET28a-Cqm1 clones used in this work. A stoichiometry of 2:2 was provided as input to reflect the tetrameric complex.

The models predicted by AlphaFold-Multimer and ColabFold have been presented in **Fig 5.15**. In both cases, models were evaluated by their confidence (pLDDT) scores, which are ranked as follows – very high (pLDDT > 90), high (pLDDT > 70), low (pLDDT > 50), and very low (pLDDT < 50). Confidence scores can be used to evaluate the accuracy of the model on a per-residue basis. Regions with very high confidence are suitable for applications which require high accuracy, including but not limited to, the characterisation of binding interfaces (Evans et al. 2022). On the other hand, regions with very low confidence, often predicted with a ribbon representation, should not be interpreted (Evans et al. 2022). Very low confidence scores have also been shown to act as a predictive measure of intrinsically disordered regions (Wilson et al. 2022).

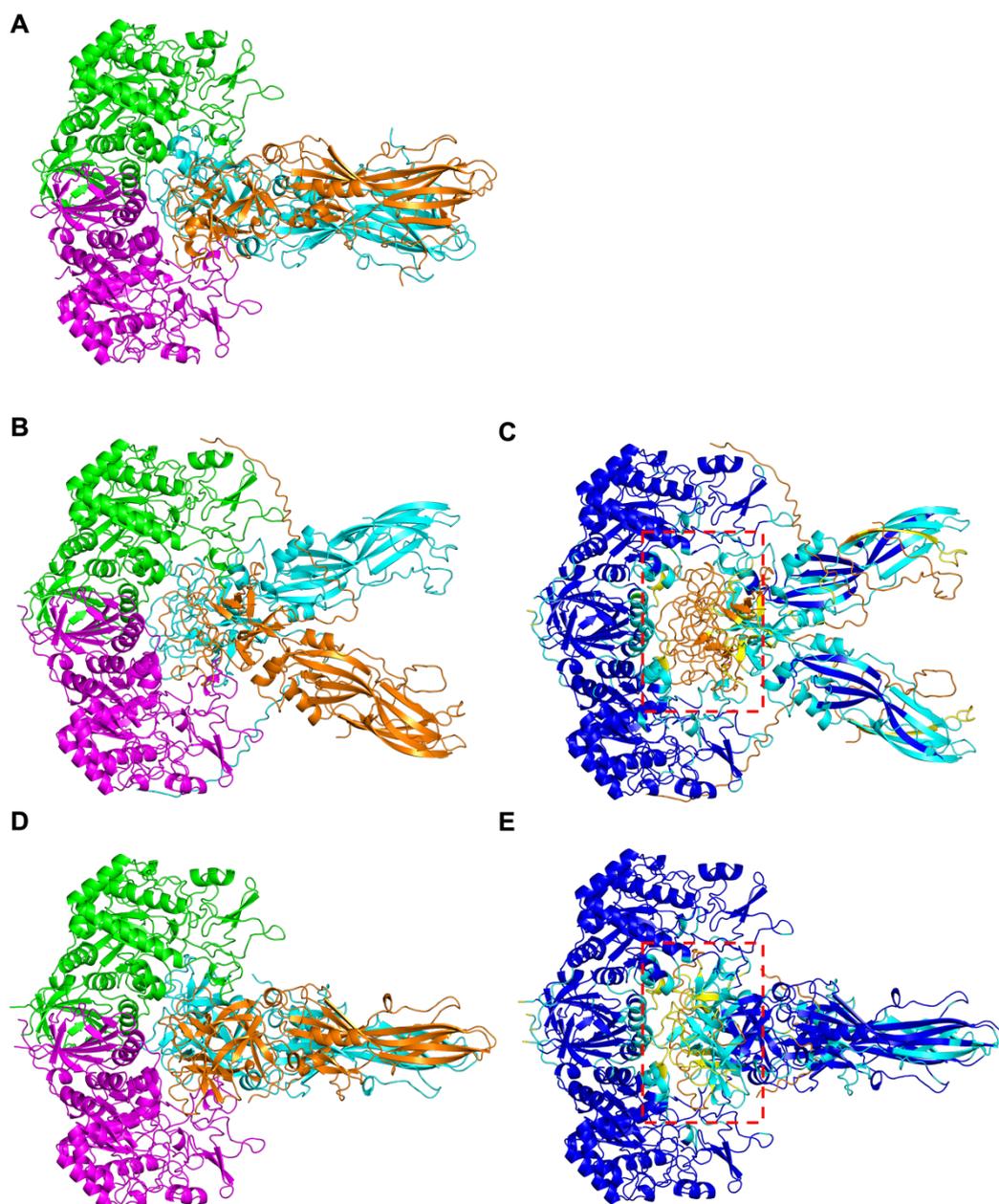
Despite exhibiting high confidence scores across the majority of the model, AlphaFold-Multimer was not able to predict the Tpp2Aa2-Cqm1 structure accurately (**Fig 5.15B, 5.15C, Table 5.6** – all-atom RMSD of 21.926 Å, QS-score of 0.03 and DockQ score of 0.36), with large areas of disordered structure and regions of physically impossible steric clashes predicted, including intertwining links in which independent chains were looped around one another. Such topological links have been observed in nature but always involve covalent modifications and/or disulphide bridges and occur in distinct groups, such as intrinsically disordered proteins (Sugase et al. 2007). In the case of folded proteins, complex formation may lead to conformational changes at the interface site, however, based on physical principles, we would not expect the protein to unfold. The topologically intertwining loops have been noted in other AlphaFold-Multimer targets and are likely artifacts that should be disregarded (Hou et al. 2023). In contrast, ColabFold predicted the Tpp2Aa2-Cqm1 structures with better accuracy (**Fig 5.15D, 5.15E, Table 5.6** – all-atom RMSD of 15.220 Å, QS-score of 0.25 and DockQ score of 0.34). Despite this, large regions of structure were predicted with very low confidence, including regions at the N- and C-terminal ends and the Tpp2Aa2 interface, suggesting that the model was not applicable for studying binding interactions. Indeed, interface analysis using PISA indicated that ColabFold had not predicted any of the hydrogen bonds or salt bridges seen in the cryoEM structure.



**Figure 5.15. Accuracy of modelling by AlphaFold-Multimer and ColabFold.** (A) Tpp2Aa2-Cqm1 structure elucidated by cryoEM (Cqm1 - green, magenta, Tpp2Aa2 - cyan, orange) (B) Tpp2Aa2-Cqm1 model predicted by AlphaFold-Multimer in the same colour scheme and (C) coloured according to the AlphaFold confidence (pLDDT) scores. pLDDT confidence scores are coloured dark blue (very high, pLDDT > 90), cyan (high, pLDDT > 70), yellow (low, pLDDT > 50), orange (very low, pLDDT < 50). Alignment with the cryoEM structure indicates an all-atom RMSD of 21.926 Å (including outliers). The majority of the structure is predicted with high confidence.

However, the presence of disordered loops and physically impossible clashes between loops of different chains suggest the model should be disregarded. (D) Tpp2Aa2-Cqm1 model predicted by ColabFold in the same colour scheme as the cryoEM structure and (E) coloured according to the AlphaFold confidence (pLDDT) scores. Alignment with the cryoEM structure indicates an all-atom RMSD of 15.220 Å (including outliers). ColabFold predicts the Tpp2Aa2-Cqm1 structure with better accuracy than AlphaFold-Multimer, however, large regions of structure are predicted with very low confidence, including regions at the N- and C-terminal ends, and the Tpp2Aa2 interface, suggesting that this model is not applicable for studying binding interactions.

To investigate whether the model could be improved by trimming away the disordered regions predicted with low confidence at both the N- and C-termini, we next attempted to model the Tpp2Aa2-Cqm1 interaction by removing 21 residues from the N-terminus and 53 residues from the C-terminus of the Tpp2Aa2 sequence. In addition to their low confidence scores, these residues were chosen as they correspond to predicted chymotrypsin cleavage sites. In a previous study, cleavage at these sites produced activated Tpp2Aa2 protein with similar electrophoretic properties to Tpp2Aa2 that had been activated by larval midgut juice and, further to this, retained its activity (when used in combination with Tpp1Aa2) against *Cx. pipiens* mosquito larvae (Clark and Baumann 1990). Removal of these residues substantially improved the ColabFold model (**Fig 5.16D, 5.16E**), which displayed an all-atom RMSD of 8.795 Å, QS-score of 0.69 and DockQ score of 0.59 when compared with the experimentally resolved structure (**Table 5.6**). The confidence scores of interfacial Tpp2Aa2 residues were also improved (**Fig 5.16E**). Interface analysis using PISA indicated that ColabFold was able to predict several hydrogen bonds (involving both partner residues) seen in the cryoEM structure (Tpp2-Cqm1 residues: Tyr197-Asp216, Phe196-Leu219, Ser195-Asp232). Taken together, this indicates that ColabFold was able to predict the Tpp2Aa2-Cqm1 structure more accurately following cleavage of disordered N- and C-terminal residues. In Tpp2Aa2, these regions may correspond to N- and C-terminal peptides cleaved in the midgut of *Cx. pipiens* mosquitoes (Clark and Baumann 1990).



**Figure 5.16. Improving the ColabFold model.** (A) Tpp2Aa2-Cqm1 structure elucidated by cryoEM (Cqm1 - green, magenta, Tpp2Aa2 - cyan, orange) (B) Full-length Tpp2Aa2-Cqm1 model predicted by ColabFold in the same colour scheme and (C) coloured according to the AlphaFold confidence (pLDDT) scores. pLDDT confidence scores are coloured dark blue (very high, pLDDT > 90), cyan (high, pLDDT > 70), yellow (low, pLDDT > 50), orange (very low, pLDDT < 50). (D) Tpp2Aa2-Cqm1 model predicted by ColabFold after removal of 21 N-terminal and 53 C-terminal residues in the same colour scheme as the cryoEM structure and (E) coloured according to the AlphaFold confidence (pLDDT) scores. Alignment with the cryoEM structure indicates

an improved all-atom RMSD of 8.795 Å (including outliers). The confidence scores of the Tpp2Aa2 interface (red box in panels C and E) were also improved.

**Table 5.6. Calculated scores of the predicted Tpp2-Cqm1 AlphaFold / ColabFold models when compared with the experimentally resolved Tpp2Aa2-Cqm1 structure (chains A, B, C, D).**

Model	RMSD (Å)	QS-score	DockQ
AlphaFold-Multimer	21.926	0.03	0.36
ColabFold (full-length)	15.220	0.25	0.34
ColabFold (N/C-term cleaved)	8.795	0.69	0.59

#### 5.2.9.4. General features of the Tpp2Aa2-Cqm1 complex

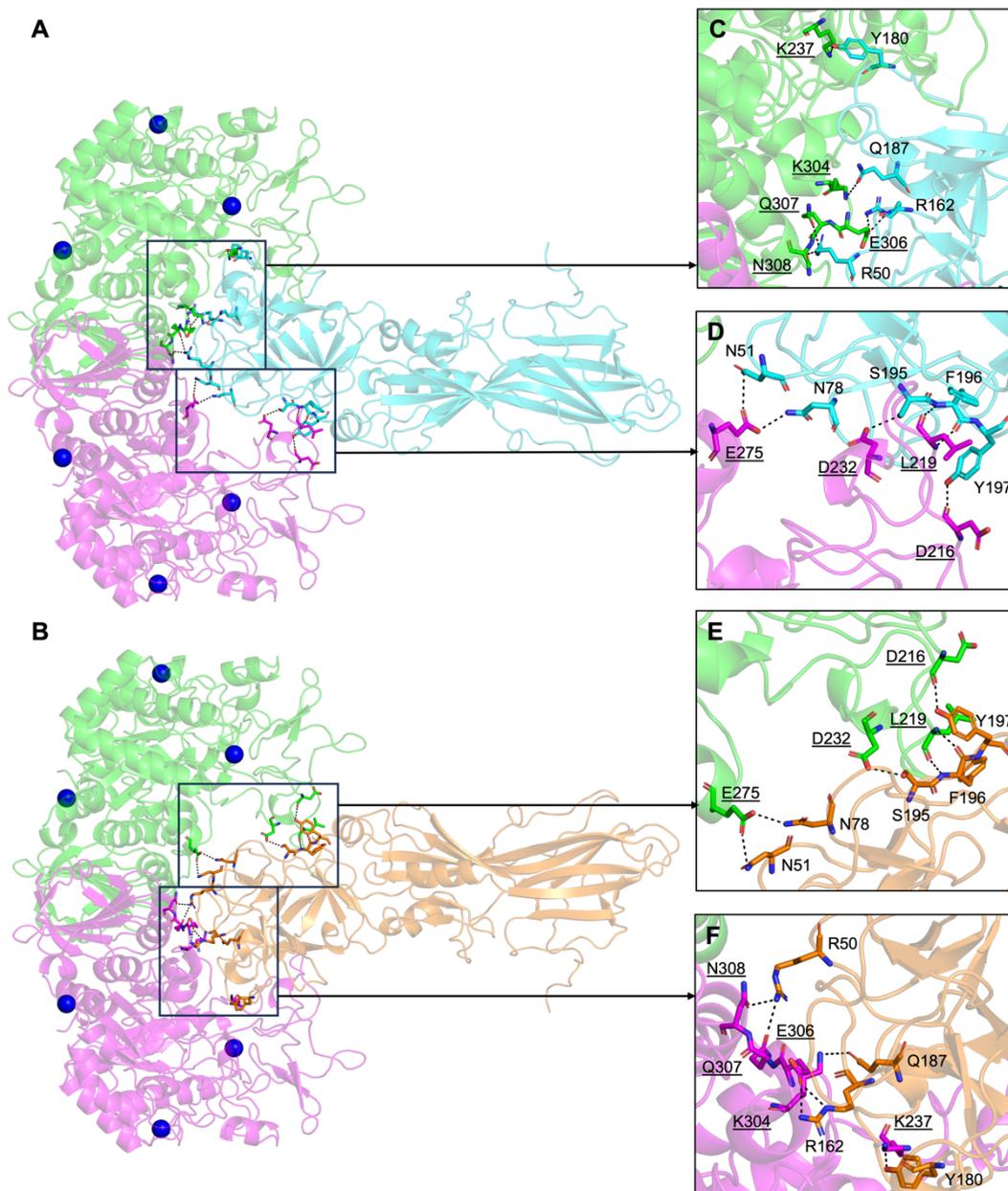
As described above, structure solution of the Tpp2Aa2-Cqm1 complex revealed a tetramer consisting of the Cqm1 dimer and two Tpp2Aa2 monomers (**Fig 5.13**). Further analysis using PISA (**Table 5.7**) demonstrated that each Tpp2Aa2 monomer forms a discontinuous interface, consisting of two distinct contact regions, one with each of the Cqm1 monomers, involving a network of hydrogen bonds and salt bridges (**Fig 5.17**). Tpp2Aa2 (chain B) interacts with the first Cqm1 monomer (chain A) via an interface sized 958.2 Å<sup>2</sup>, involving 7 hydrogen bonds and 2 salt bridges, and the second Cqm1 monomer (chain C) via an interface sized 430.6 Å<sup>2</sup>, involving 6 hydrogen bonds (**Table 5.7**). Similarly, the second Tpp2Aa2 monomer (chain D) interacts with the first Cqm1 monomer (chain C) via an interface sized 952.6 Å<sup>2</sup>, involving 5 hydrogen bonds and 2 salt bridges, and the second Cqm1 monomer (chain A) via an interface sized 439.0 Å<sup>2</sup>, involving 6 hydrogen bonds (**Table 5.7**). PISA identified a small interface (123.9 Å<sup>2</sup>) between the two Tpp2Aa2 monomers, involving 14 residues from chain B and 12 residues from chain D. Interaction of the two

Cqm1 monomers represents the largest interface within the complex at 1448.4 Å<sup>2</sup> and involves 13 hydrogen bonds and 20 salt bridges (**Table 5.7**). Each Cqm1 monomer was found to coordinate three calcium ions. In line with previous work (Romão et al. 2011), PISA analysis demonstrated that the Tpp2Aa2 monomer interacts with Cqm1 predominantly via its N-terminal domain (excluding one interfacial residue originating from the Tpp2Aa2 C-terminal domain – Lys256). A list of the hydrogen bonds and salt bridges formed is provided in **Table 5.8**.

**Table 5.7. Calculated structural properties of the molecular interfaces in Tpp2Aa2-Cqm1 identified using PISA.**

Interface	Structure 1			Structure 2			Interface area (Å <sup>2</sup> )	$\Delta^iG^*$ (kcal mol <sup>-1</sup> )	H-bonds	Salt bridges	Binding energy (kcal mol <sup>-1</sup> )
	Chain	Protein	AA	Chain	Protein	AA					
1	A	Cqm1	40	C	Cqm1	40	1448.4	6.5	13	20	-6.7
2	B	Tpp2Aa2	31	A	Cqm1	40	958.2	-5.7	7	2	-9.5
3	B	Tpp2Aa2	20	C	Cqm1	18	430.6	-2.8	6	0	-5.5
4	D	Tpp2Aa2	31	C	Cqm1	40	952.6	-7.3	5	2	-10.3
5	D	Tpp2Aa2	19	A	Cqm1	18	439.0	-3.0	6	0	-5.7
6	B	Tpp2Aa2	14	D	Tpp2Aa2	12	123.9	1.3	0	0	1.3

\*  $\Delta^iG$  = solvation free energy gain

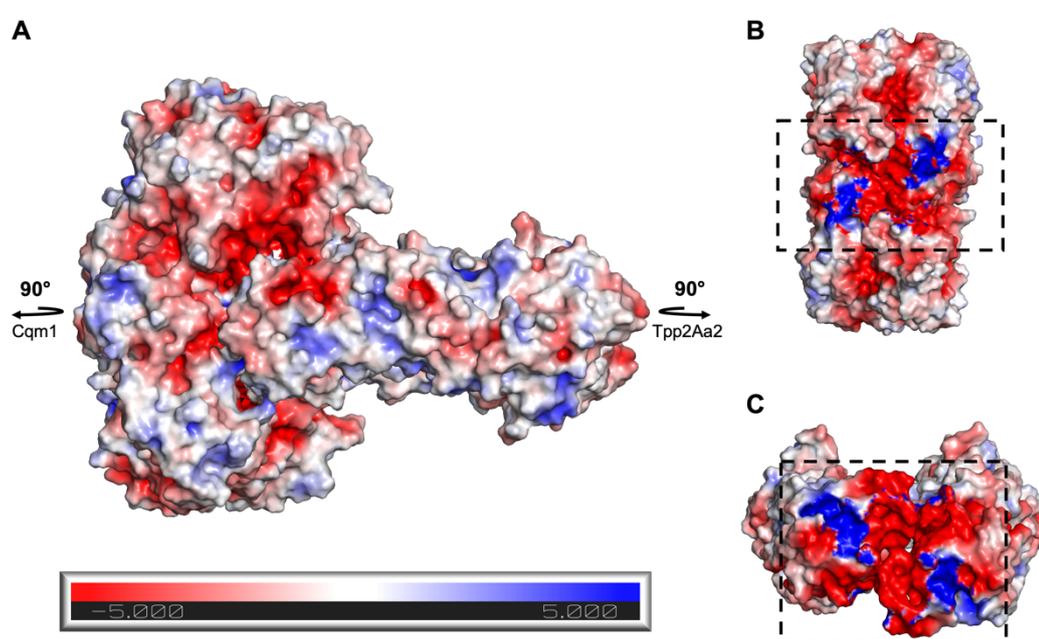


**Figure 5.17. Tpp2Aa2 forms a discontinuous interface with each Cqm1 monomer.** (A, B) Each Tpp2Aa2 monomer (cyan, orange) forms a discontinuous interface, consisting of two distinct interfaces, each interacting with one of the Cqm1 monomers (green, magenta). Calcium ions are shown as dark blue spheres. (C – F) Several polar contacts (black dashed lines), including hydrogen bonds and salt bridges, are formed between Tpp2Aa2 and Cqm1. Residues involved in polar contacts are shown as sticks (carbon – cyan / orange / green / magenta, oxygen – red, nitrogen – blue). To distinguish Tpp2Aa2 / Cqm1 residues, Cqm1 residue labels have been underlined.

**Table 5.8. Polar interactions within the Tpp2Aa2-Cqm1 interfaces.**

Interface	Chains	Hydrogen bonds		Salt bridges	
		Tpp2Aa2	Cqm1	Tpp2Aa2	Cqm1
1	B – A	Arg50 [NH1]	Gln307 [OE1]	Arg162 [NE]	Glu306 [OE1]
		Arg50 [NH2]	Gln307 [OE1]	Arg162 [NH2]	Glu306 [OE1]
		Arg50 [NH2]	Asn308 [OD1]		
		Arg162 [O]	Glu306 [N]		
		Tyr180 [OH]	His207 [NE2]		
		Tyr180 [OH]	Lys237 [NZ]		
		Gln187 [OE1]	Lys304 [NZ]		
2	B – C	Tyr197 [OH]	Asp216 [O]		
		Phe196 [N]	Leu219 [O]		
		Ser195 [OG]	Asp232 [OD2]		
		Asn51 [ND2]	Glu275 [OE1]		
		Asn78 [ND2]	Glu275 [OE2]		
		Phe196 [O]	Leu219 [N]		
3	D – C	Arg50 [NH2]	Asn308 [OD1]	Arg162 [NE]	Glu306 [OE1]
		Arg50 [NH2]	Gln307 [OE1]	Arg162 [NH2]	Glu306 [OE1]
		Arg162 [O]	Glu306 [N]		
		Tyr180 [OH]	Lys237 [NZ]		
		Gln187 [OE1]	Lys304 [NZ]		
4	D – A	Tyr197 [OH]	Asp216 [O]		
		Phe196 [N]	Leu219 [O]		
		Ser195 [OG]	Asp232 [OD2]		
		Asn51 [ND2]	Glu275 [OE1]		
		Asn78 [ND2]	Glu275 [OE2]		
		Phe196 [O]	Leu219 [N]		

To investigate the electrostatic compatibility of the Tpp2Aa2-Cqm1 interface, PROPKA (available in the PDB2PQR web server - Dolinsky et al. 2007) was used to assign the protonation state of titratable residues according to an environment pH 8. The PyMOL APBS electrostatics plugin (Jurrus et al. 2018) was then employed to visualise the surface charge of the complex. This analysis indicated that, at pH 8, the binding pocket of Cqm1 consists of regions of highly positive and highly negative charge that are electrostatically compatible with opposing regions of highly negative and highly positive charge at the Tpp2Aa2 interface (**Fig 5.18**).



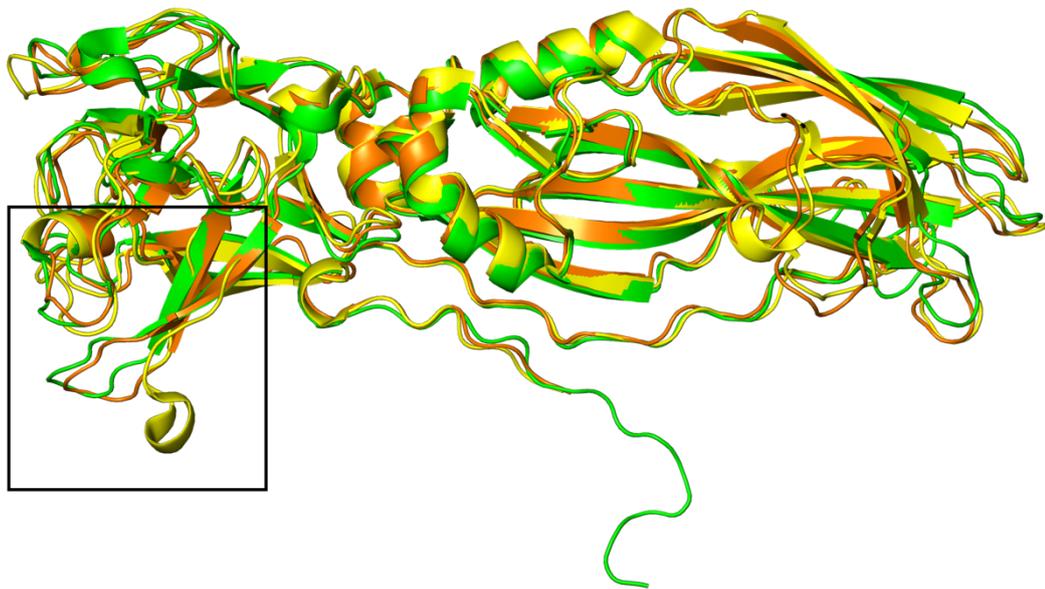
**Figure 5.18. Electrostatic compatibility of the Tpp2Aa2-Cqm1 interface.** (A) Tpp2Aa2-Cqm1 complex coloured by surface charge. Arrows depict the separation and 90° rotation of the Cqm1 dimer and Tpp2Aa2 monomers to obtain the orientations seen in panels B and C. (B) Cqm1 binding pocket coloured by surface charge. (C) Tpp2Aa2 interface coloured by surface charge. At pH 8, the binding pocket of Cqm1 consists of regions of highly positive and highly negative charge (dashed box) which appear to be electrostatically compatible with opposing regions of highly negative and highly positive charge (dashed box) at the Tpp2Aa2 interface.

It is generally accepted that, following the dissolution of crystals in the alkaline environment of the midgut, the Tpp1Aa2 and Tpp2Aa2 proteins are proteolytically cleaved, leading to receptor binding and toxin internalisation into the midgut cells. *In vitro* proteolytic activation of the Tpp1Aa2 and Tpp2Aa2 proteins following treatment with mosquito larval midgut juice and trypsin has been demonstrated (Baumann et al. 1988; Oei et al. 1992; Charles et al. 1997). In our cryoEM experiments, however, we utilised the pro-form of Tpp2Aa2 and show that it can interact with the target receptor, Cqm1. An alternative mechanism could be that proteolytic activation occurs following toxin internalisation into the midgut cells. Indeed, mosquito larval “midgut juice” is produced by grinding up larval midguts and so we cannot be sure whether proteases obtained in the extraction originate from the midgut lumen or from intracellular compartments. In larger insect larvae, it is possible to extract the midgut juice leaving the midgut tissue behind, but this is impossible for mosquito larvae given their small size range. Despite this, it may still be the case that proteolytic activation occurs prior to receptor interaction *in vivo* but, *in vitro*, the N-terminal peptide of pro-Tpp2Aa2 does not obstruct the Tpp2Aa2-Cqm1 interface and thus, the complex is still able to form. The first 27 residues of Tpp2Aa2 were not seen in the final cryoEM map and, hence, were not modelled into the structure, thus, we cannot be sure where these regions lie in relation to the Tpp2Aa2-Cqm1 interface. However, it is possible that the N-terminal peptide resides away from the binding interface, given that the first residue seen in the Tpp2Aa2 structure (Thr28) is directed away from the interface and towards the C-terminal domain. Further binding assays are required to investigate and compare the binding kinetics of pro-Tpp2Aa2 and activated Tpp2Aa2 with Cqm1.

#### 5.2.9.5. Comparison with related structures

Comparisons of the structure of Tpp2Aa2 in the Tpp2Aa2-Cqm1 cryoEM complex were made with non-bound Tpp2 structures. Structures for Tpp2 variants in both their trypsinized state (PDB 3WA1) and natural Tpp1/Tpp2 co-crystal (PDB 5FOY) have been elucidated (Srisucharitpanit et al. 2014; Colletier et al. 2016). To investigate differences in the secondary structure, structural alignments were performed (**Fig 5.19**). The receptor-bound Tpp2Aa2 elucidated here (chain B) displayed an all-atom RMSD of 1.754 Å when aligned with the Tpp2Aa2 monomer extracted from the Tpp1Aa2/Tpp2Aa2 natural co-crystals (PDB 5FOY, chain B). In comparison, receptor-bound Tpp2Aa2 displayed an all-atom RMSD of 1.540 Å when aligned with trypsin-activated Tpp2Aa3 (PDB 3WA1), indicating highest structural similarity with the trypsin-activated Tpp2Aa3.

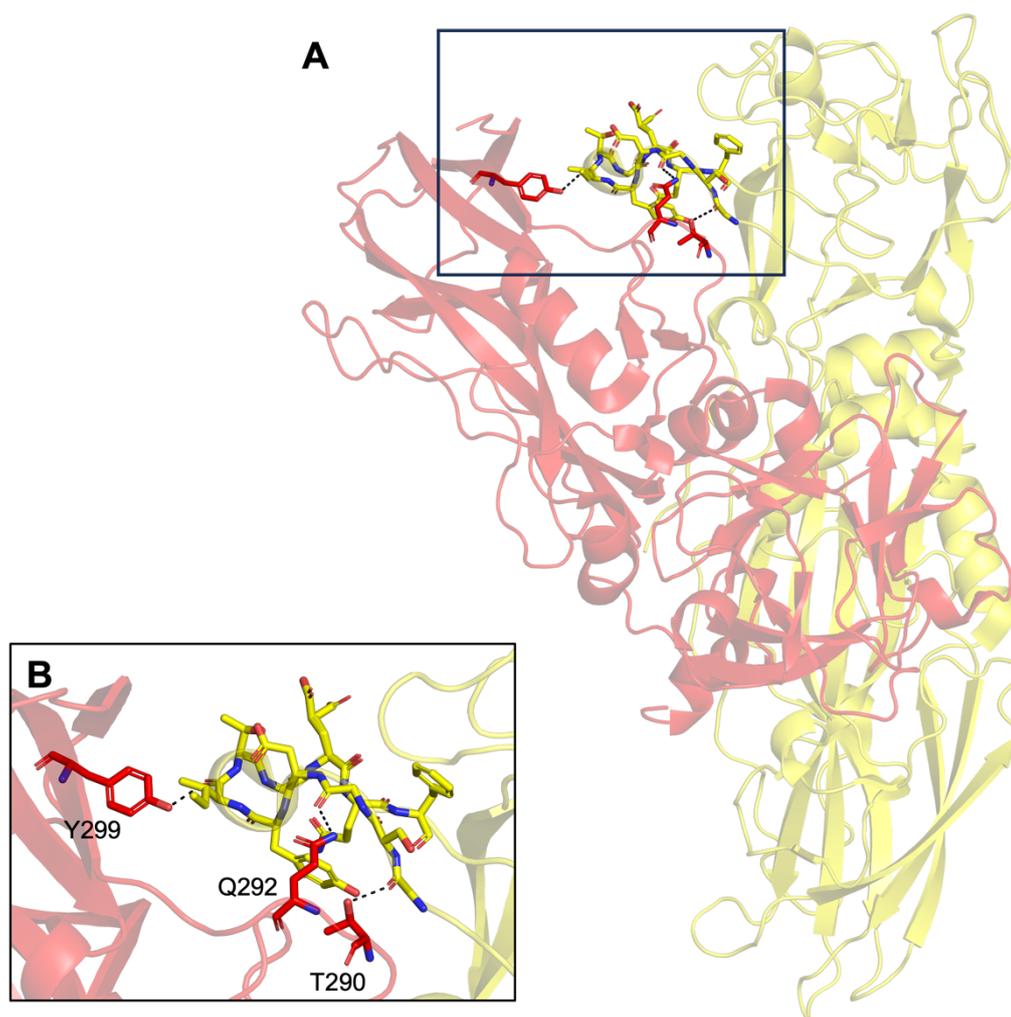
One notable difference was identified between Tpp2Aa2 residues 137 – 145, which, in the Tpp1Aa2/Tpp2Aa2 co-crystal, form a loop containing a small helix (**Fig 5.19**). This is in contrast to trypsin-activated Tpp2Aa3 and receptor-bound Tpp2Aa2, where residues 137 – 145 form an extended loop (**Fig 5.19**). PISA analysis of the Tpp1Aa2/Tpp2Aa2 natural co-crystal (**Table 5.9**) indicates that this region forms part of the dimer interface (**Fig 5.20**), with Ile141 (which forms a hydrogen bond with Tyr299 of Tpp1Aa2) indicated by the per-residue binding energy prediction of PISA, to have the most stabilizing effect in comparison to all other Tpp2Aa2 interfacial residues. This region also appears to play a role in the Tpp2Aa2-Cqm1 interface elucidated here. In the initial cryoEM model, Tpp2Aa2 residues 137 – 140 were clashing with Cqm1 and thus, this region was manually rebuilt. This steric clash indicates structural movement of Tpp2Aa2 residues 137 – 140 upon receptor interaction. Given that these residues form part of the Tpp2Aa2-Cqm1 interface, the unravelling of the short helix may be a significant step in the sequential mode of action of this protein, from crystal dissolution through to receptor binding.



**Figure 5.19. Comparison of receptor bound Tpp2Aa2 with other Tpp2 variants.** (A) Structural superposition of receptor-bound Tpp2Aa2 (orange) with trypsin-activated Tpp2Aa3 (PDB 3WA1 – green) and the Tpp2Aa2 monomer extracted from the Tpp1Aa2/Tpp2Aa2 natural co-crystal (PDB 5FOY – yellow). One notable difference was identified between the Tpp2 structures. In the Tpp1Aa2/Tpp2Aa2 co-crystal structure, residues 136 – 146 form a loop containing a small helix, in contrast to an extended loop in both the trypsin-activated Tpp2Aa3 and receptor-bound Tpp2Aa2 elucidated here (box).

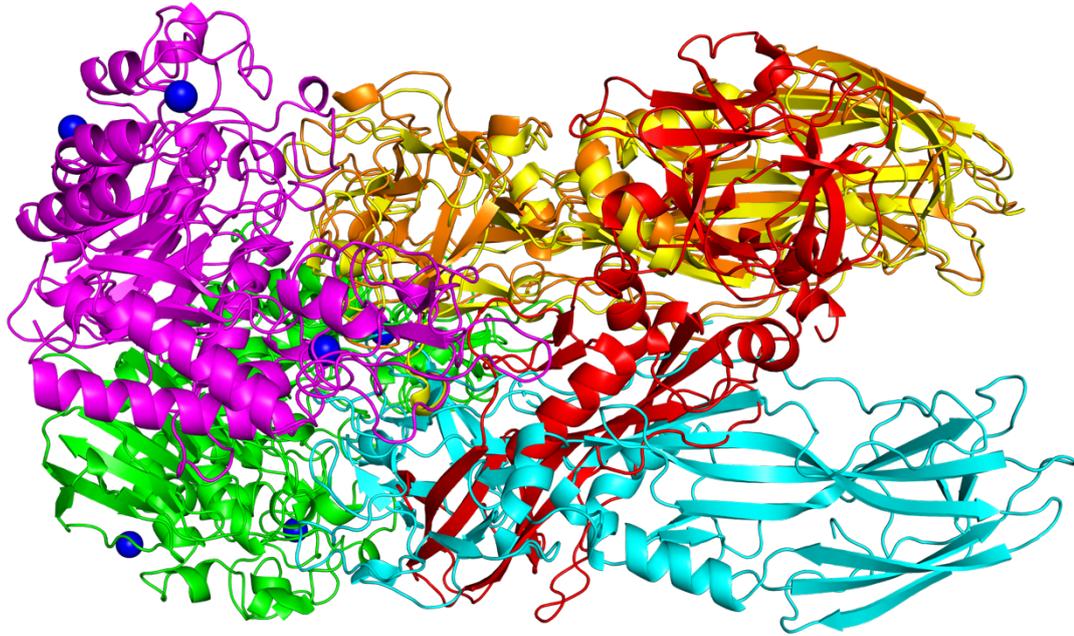
**Table 5.9. Intermolecular interactions in the Tpp1Aa2/Tpp2Aa2 co-crystal (PDB 5FOY).**

<b>Tpp2Aa2</b>	<b>Distance</b>	<b>Tpp1Aa2</b>
<b>Hydrogen bonds</b>		
Gln336 [NE2]	3.7	Glu98 [O]
Thr337 [OG1]	2.9	Lys100 [O]
Gln448 [N]	3.6	Ser148 [OG]
Gln448 [N]	3.9	Asn149 [O]
Asn33 [ND2]	3.3	Thr280 [O]
Asn33 [ND2]	3.0	Thr285 [O]
Gly136 [N]	2.8	Asn289 [O]
Leu210 [O]	3.5	Met1 [N]
Gln336 [O]	2.9	Lys100 [N]
Tyr418 [O]	2.8	Met103 [N]
Gln448 [OXT]	3.2	Ser148 [OG]
Gln448 [O]	3.3	Lys150 [NZ]
Arg328 [O]	3.8	Tyr281 [OH]
Asn31 [O]	2.7	Lys284 [NZ]
Asn33 [O]	3.1	Gln286 [NE2]
Gln134 [O]	2.8	Asn289 [ND2]
Gly136 [O]	2.8	Thr290 [OG1]
Gly138 [O]	3.2	Gln292 [NE2]
Ile141 [O]	2.7	Tyr299 [OH]
<b>Salt bridges</b>		
Lys234 [NZ]	3.5	Asp8 [OD1]
Arg328 [NH1]	4.0	Asp61 [OD1]



**Figure 5.20. Tpp1Aa2/Tpp2Aa2 co-crystal.** (A) Tpp1Aa2/Tpp2Aa2 co-crystal structure with Tpp1 shown in red and Tpp2 shown in yellow. The position of Tpp2Aa2 residues 136 - 146 relative to the dimer interface has been shown. (B) Polar contacts (black dashed lines) formed by Tpp2Aa2 residues 136 - 146 in the co-crystal. The corresponding Tpp1Aa2 residues have been labelled.

Following dissolution of the Tpp1Aa2/Tpp2Aa2 natural crystals, it remains unclear whether the proteins dissociate into their individual components, or if the heterodimeric complex persists. Here, alignment of the Tpp2Aa2-Cqm1 complex with the Tpp1Aa2/Tpp2Aa2 co-crystal (**Fig 5.21**) reveals that the large intermolecular interface of the latter is blocked by the second Tpp2Aa2 monomer binding to the receptor protein, Cqm1. This suggests that dissociation of the Tpp1Aa2/Tpp2Aa2 complex is required prior to receptor interaction. Indeed, sequential binding of Tpp2Aa2 to Cqm1, followed by Tpp1Aa2 to Tpp2Aa2 has been hypothesized. Structural rearrangement of Tpp2Aa2 residues 137 – 145 (that play significant interfacial roles in the respective complexes) from a short helix in the Tpp1Aa2/Tpp2Aa2 co-crystal to an extended loop in the Tpp2Aa2-Cqm1 complex, further supports dissociation of the Tpp1Aa2/Tpp2Aa2 complex prior to receptor interaction. Taken together, we hypothesize that dissociation of the Tpp1/Tpp2 complex must occur prior to receptor interaction.



**Figure 5.21. Superposition of the Tpp2Aa2-Cqm1 complex with the Tpp1Aa2/Tpp2Aa2 natural co-crystal.** Alignment of the Tpp2Aa2-Cqm1 complex (Cqm1 monomers – green, magenta, Tpp2Aa2 monomers – orange, cyan) with the Tpp1Aa2/Tpp2Aa2 co-crystal (Tpp1 – red, Tpp2 – yellow) reveals that the large intermolecular interface between Tpp2 and the Tpp1 molecule in their co-crystal complex is blocked in the Tpp2Aa2-Cqm1 structure by the second Tpp2Aa2 (cyan). This suggests that dissociation of the Tpp1Aa2/Tpp2Aa2 complex is required prior to receptor interaction.

Cqm1 is the functional receptor protein of Tpp1/Tpp2 in *Cx. quinquefasciatus* (Sharma et al. 2018; Sharma and Kumar 2019). Across other Tpp1/Tpp2 susceptible species, orthologous receptor binding proteins have been identified as Cpm1 (*Cx. pipiens*) and Agm3 (*An. gambiae*) (Silva-Filha et al. 1999; Opota et al. 2008). In addition, an orthologous non-binding protein, Aam1, has been identified in the refractory (non-susceptible) mosquito species, *Ae. aegypti* (Nielsen-Leroux and Charles 1992; Ferreira et al. 2010). As is the case for Cqm1, Cpm1, and Agm3, Aam1 is expressed in the midgut epithelial cells as a GPI-anchored membrane protein (Ferreira et al. 2010). Furthermore, Aam1 displays higher sequence identity than that of Agm3 with Cqm1 (Aam1/Cqm1 sequence identity = 74%, Agm3/Cqm1

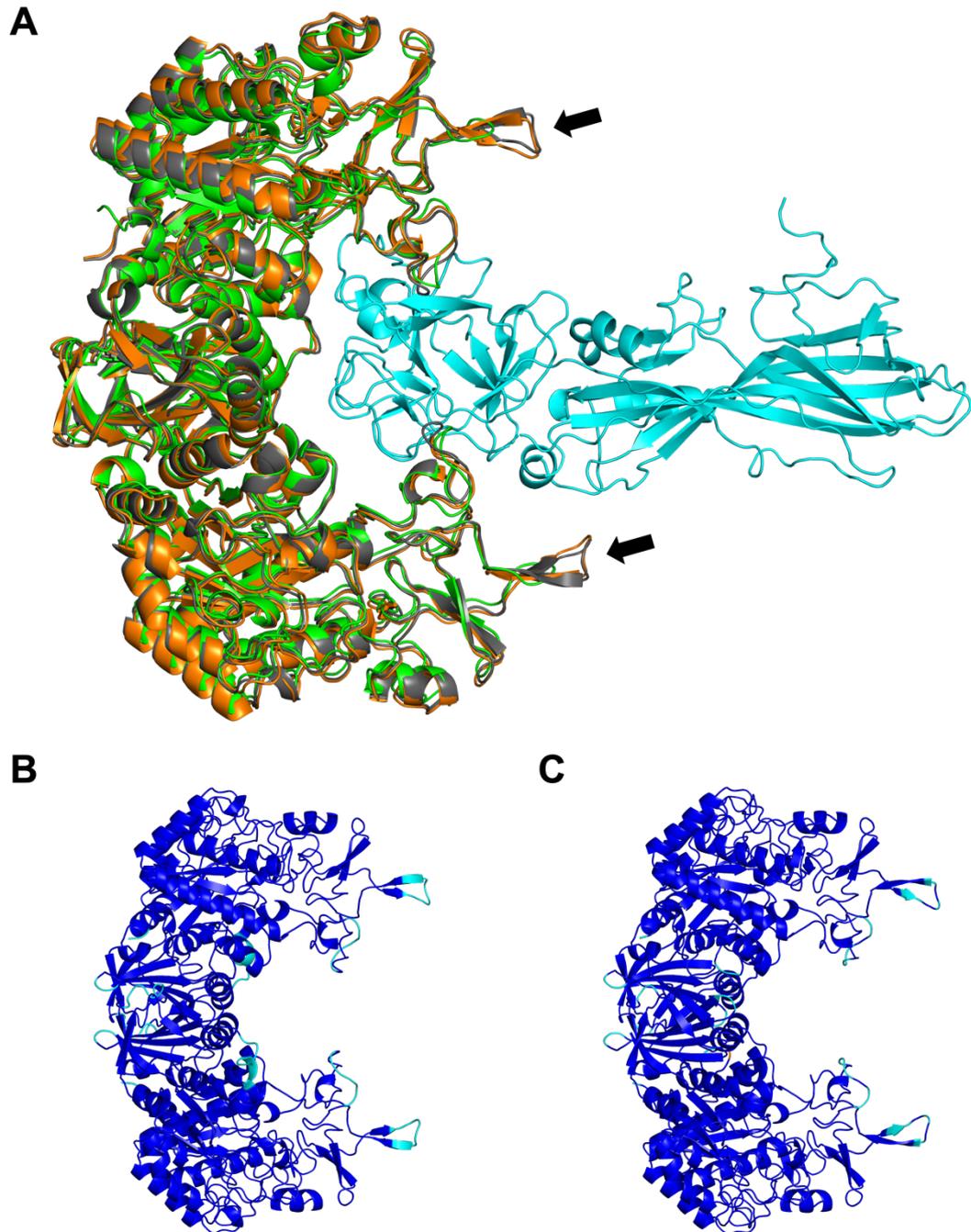
sequence identity = 67%) (Ferreira et al. 2010). Despite this, Aam1 is unable to bind Tpp1/Tpp2 and this lack of binding has been implicated in the refractory nature of *Ae. aegypti* (Ferreira et al. 2010).

Given that *Ae. aegypti* represents a significant mosquito pest involved in the transmission of several human diseases, including yellow fever, dengue and Zika virus, efforts have been made to investigate the lack of Tpp1/Tpp2 susceptibility further. In comparison to Cqm1, Aam1 is heavily glycosylated, with four confirmed N-glycosylation sites localised externally on the modelled structure (do Nascimento et al. 2017). Despite this, Ferreira et al (2010) demonstrated that deglycosylation of *Cx. quinquefasciatus* or *Ae. aegypti* CHAPS extracts (solubilized protein samples from brush border membrane fractions) had no effect on Tpp2 binding in subsequent pull-down assays. In addition, mutant studies targeting the four Aam1 N-glycosylation sites demonstrated that abolition of these sites in Aam1, as well as insertion in Cqm1, had no effect on Tpp2 binding (do Nascimento et al. 2017). Differential N-glycosylation did, however, influence the catalytic activity of Aam1 (do Nascimento et al. 2017). In line with these studies, Cqm1 N-glycosylation sites predicted by do Nascimento et al. (2017) reside away from the Tpp2 binding site in the Tpp2Aa2-Cqm1 cryoEM structure elucidated here. Furthermore, Aam1 N-glycosylation sites also reside away from the Tpp2 binding site when Aam1 is aligned with the Tpp2Aa2-Cqm1 structure, suggesting that differential N-glycosylation does not affect Tpp2 interaction.

In other work, a mutagenesis study targeting Cqm1 residues 159Gly-Gly160 (137Gly-Gly138 in the Cqm1 clone used as part of this work) demonstrated that these residues are required for Tpp2 interaction (Ferreira et al. 2014). Given that these residues are not conserved in the Aam1 protein, this study provided a possible explanation for lack of Tpp2 binding. However, the Tpp2Aa2-Cqm1 structure elucidated here shows that Cqm1 residues 137Gly-Gly138 form a short, exposed loop in domain A, within a region distinct to the Tpp2Aa2 binding pocket (**Fig 5.22 – arrow**). In contrast, the AF2 model of Aam1 predicts a longer  $\beta$ -hairpin within the

corresponding region, and this structural feature is conserved in the Agm3 prediction, of the Tpp1/Tpp2 susceptible species, *An. gambiae* (**Fig 5.22**). Taken together, this region is unlikely to be the specific difference causing lack of binding and, therefore, it is possible that other molecular features underly the lack of Tpp2 interaction with Aam1.

To investigate the lack of Tpp2Aa2 binding to Aam1, the AF2 predictions of Aam1 (available on the AF2 database – AF-Q16SN4) and Agm3 (predicted using AF2 installed on the Supercomputing Wales high-performance cluster, Hawk since only the monomeric prediction was available on the AF2 database) were aligned with Cqm1 in the Tpp2Aa2-Cqm1 structure, indicating respective all-atom RMSDs of 1.395 Å and 1.608 Å (**Fig 5.22**). In parallel, a sequence alignment was performed, demonstrating several substitutions in the Aam1 sequence relative to Agm3 and Cqm1 (**Fig 5.23**).



**Figure 5.22. Comparison of Aam1 and Agm3 with the Tpp2Aa2-Cqm1 complex.** (A) Structural alignment of the AlphaFold2 predictions of Aam1 (grey, available in the AF2 database under entry AF-Q16SN4-F1) and Agm3 (orange, predicted using AF2 installed on Hawk) with the Tpp2Aa2-Cqm1 structure (showing the Cqm1 dimer - green - and one Tpp2Aa2 monomer - cyan) indicated respective all-atom RMSDs of 1.395 Å and 1.608 Å. The position of Cqm1 residues 137Gly-Gly138 has been indicated by an arrow. (B) AlphaFold2 prediction of Aam1 and (C) Agm3 coloured by confidence (pLDDT)

score. In both structures, the GPI anchor has been removed for clarity. pLDDT confidence scores are coloured dark blue (very high, pLDDT > 90), cyan (high, pLDDT > 70), yellow (low, pLDDT > 50), orange (very low, pLDDT < 50).

Although the majority of Cqm1 interfacial residues identified in the Tpp2Aa2-Cqm1 structure are conserved in both Agm3 and Aam1, there are four cases where Cqm1 interfacial residues are conserved in Agm3 but substituted in Aam1 (**Fig 5.23 – highlighted in grey - Aam1 residues Ser248, Glu265, Glu364, Gly410**). These substitutions may disrupt binding to Tpp2 and hence, are possible targets for mutagenesis studies.

In addition, several Aam1 residues were found to clash sterically (donor-donor / acceptor-acceptor, calculated using the show\_contacts PyMOL plugin available at: [https://pymolwiki.org/index.php/Show\\_contacts](https://pymolwiki.org/index.php/Show_contacts)) with Tpp2Aa2, following structural alignment of the Aam1 AF2 prediction and Tpp2Aa2-Cqm1 structure (**Table 5.10**). Some of these residues were also found to clash sterically when Agm3 was aligned with the Tpp2Aa2-Cqm1 structure. Given that Agm3 can bind Tpp2Aa2, steric clashes identified in both Agm3 and Aam1 may not be significant for the difference in binding (**Table 5.10**). One Aam1 residue (Asp508) that is substituted in comparison to the Cqm1 sequence also appears to clash with Tpp2Aa2 residue Glu65 (**Fig 5.23, Table 5.10**). Given that Agm3 does not appear to clash with Tpp2Aa2 at this site, this Aam1 residue, and the Tpp2Aa2 residue that it clashes with, represent further targets for mutagenesis studies.

These structural analyses provide insight into the specificity of Tpp2Aa2 against susceptible and refractory species. In addition, these analyses provide testable hypotheses, including target residues for protein engineering (mutagenesis) studies aimed at broadening the target insect range of Tpp1/Tpp2.



fully conserved residue; ‘:’ indicates conservation between groups of strongly similar properties; “.” indicates conservation between groups of weakly similar properties. Sequence alignment was produced using Clustal Omega. For consistency with the Tpp2Aa2-Cqm1 structure elucidated as part of this work, the Cqm1 sequence does not include the GPI anchor and signal peptide and is numbered -3 – 537. The Aam1 and Agm3 sequences include the GPI anchor and signal peptide.

**Table 5.10. Aam1 / Agm3 residues that sterically clash with Tpp2Aa2.**

Aam1		Agm3	
Aam1	Tpp2Aa2	Agm3	Tpp2Aa2
Pro244	*Phe196	Pro250	*Phe196
Gly246	*Gln145	Trp253	*Gln145
Trp247	Gly136, Ser137, Glu144		
Glu300	*Asn51	Glu305	*Asn51
		His335	Asn66, Arg162
		Asp336	Arg162
Lys329	Ser184, Glu185		
Gly336	*Arg50	Asn338	*Arg50
Asp508	Glu65		

\*Tpp2Aa2 residues that clash with both Aam1 and Agm3. Table aligned on the basis of Tpp2 residues.

## 5.2.10. Engineering Tpp1Aa2/Tpp2Aa2 to expand the target insect range

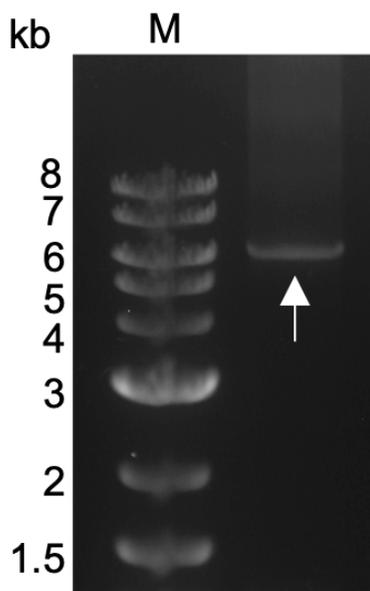
### 5.2.10.1. Rationale for Tpp2Aa2 mutation

To investigate whether the mosquitocidal activity of Tpp1Aa2/Tpp2Aa2 could be expanded to include *Ae. aegypti*, target mutations based on the Tpp2Aa2-Cqm1 structure were identified. As described above (**section 5.2.9.5**), sequence alignment of the Aam1 and Cqm1 orthologous proteins indicated several Cqm1 interfacial residues (relative to the Tpp2Aa2-Cqm1 structure) that were substituted in the Aam1 sequence. In addition, structural alignment of the AF2 prediction of Aam1 with the Tpp2Aa2-Cqm1 structure indicated several steric clashes. One Aam1 residue – Asp508 – was predicted to sterically clash with Glu65 in Tpp2Aa2. Hence, we hypothesized that mutation of Tpp2Aa2 residue 65 to alanine (E65A) would remove the steric clash, increasing the likelihood that Tpp2Aa2 would be able to interact with Aam1, thus broadening the target insect range to include *Ae. aegypti*. To investigate this further, we sought to produce a pGEX-Tpp2Aa2 clone containing the E65A mutation.

### 5.2.10.2. Site-directed mutagenesis (SDM) of Tpp2Aa2

Whole plasmid site-directed mutagenesis (SDM) was used to introduce the E65A mutation into the *tpp2Aa2* gene as described in **section 2.2.5** and **Tables 2.5, 2.6**. To do so, the plasmid vector containing the *tpp2Aa2* gene (pGEX-Tpp2Aa2) was amplified by PCR using a pair of phosphorylated oligonucleotide primers containing the GCT mutation encoding alanine at residue position 65 (forward primer containing **GCT** mutation: 5' - CTT CAA GTA TCG **CTA** ATT GCC CAT CTA AC - 3', reverse non-mutagenic primer: 5' - GAA ATT CGG TTT TTG ATA AAC CAT AAC C - 3'). The resulting DNA product was analysed by gel electrophoresis (**Fig 5.24**), showing a band of ~6.3 kb, corresponding to the size of the pGEX-Tpp2Aa2 construct. The

linear DNA product was ligated to recircularised the product as described in **section 2.2.5**. Subsequently, the recircularised plasmid was transformed into 5-alpha *E. coli* cells and plated on LB supplemented with ampicillin antibiotic for selection. Colonies were picked for sequencing, which confirmed successful incorporation of the GCT codon.

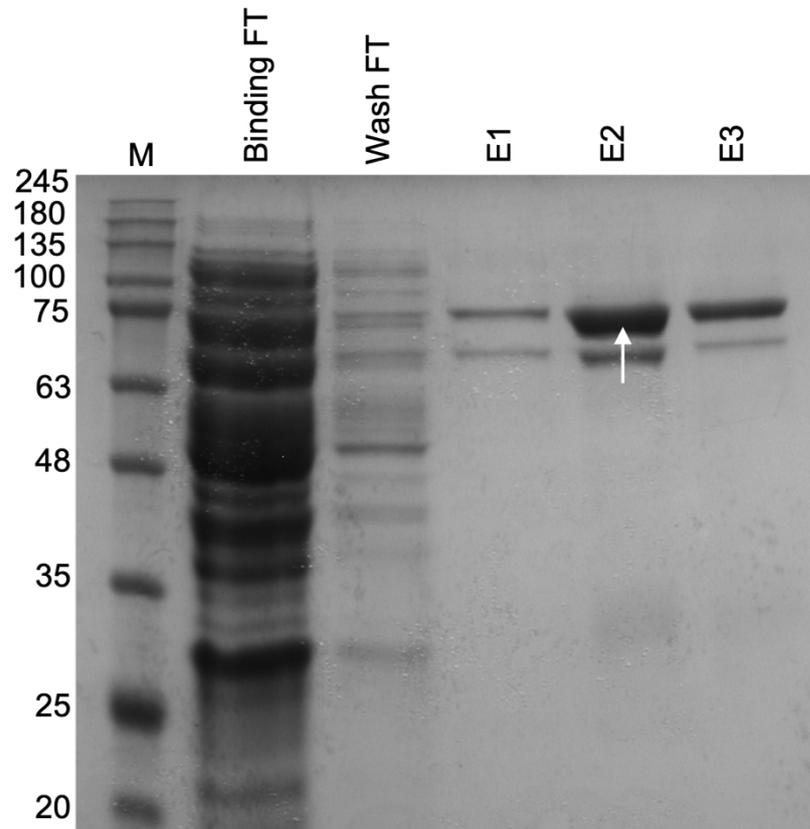


**Figure 5.24. Site-directed mutagenesis of pGEX-Tpp2Aa2 to introduce the E65A mutation.** Gel electrophoresis of the resulting DNA product showed a band ~ 6.3 kb, corresponding to the size of the pGEX-Tpp2Aa2 construct.

### 5.2.10.3. Expression of Tpp2Aa2-E65A mutant

To study the effect of the E65A mutation on Tpp2Aa2, protein samples were prepared as previously presented (**sections 5.2.1 and 5.2.2**). Briefly, the pGEX-tpp2aa2-E65A construct was transformed into *E. coli* BL21 (DE3). Cells were incubated at 37°C until an OD<sub>600</sub> of 0.6 was obtained, before induction with 1 mM IPTG at 37°C for 4 hours. The GST-tagged Tpp2Aa2-E65A protein was extracted from the cells and purified as shown below.

To obtain the GST-Tpp2Aa2-E65A fusion protein, GST-affinity chromatography was performed on whole cell lysates as described in **section 2.4.2**. Briefly, cell lysates were applied to equilibrated columns containing glutathione sepharose beads and the flow-through was collected. Columns were incubated for 1 hour to allow binding of the GST fusion protein. The binding flow-through was collected (**Fig 5.25 – lane 2**) and the column was washed with 30 mL PBS (**Fig 5.25 – lane 3**). GST fusion protein was eluted into 3 x 3 mL fractions of 50 mM TrisHCl, 10 mM reduced glutathione, pH 8.0. Elution fractions obtained were analysed by SDS-PAGE (**Fig 5.25 – lanes 3 – 5**), showing that GST-Tpp2Aa2-E65A eluted into fractions 1 – 3. Fractions containing the GST-Tpp2Aa2-E65A protein were pooled, concentrated, and stored at -20°C until downstream applications.



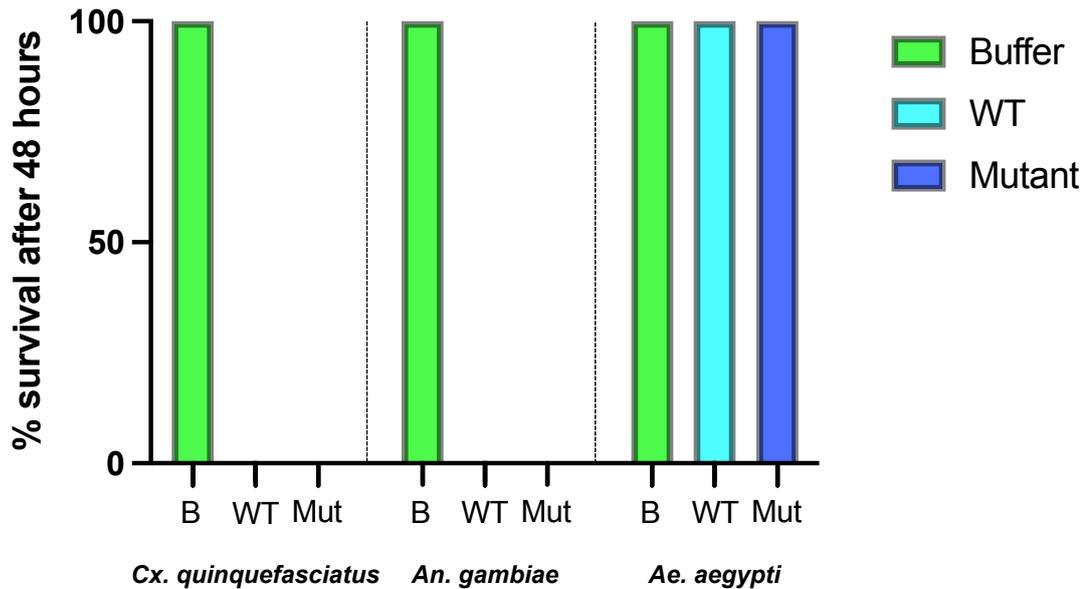
**Figure 5.25. GST affinity purification of GST-Tpp2Aa2-E65A.** Coomassie stained SDS-PAGE showing purification of GST-Tpp2Aa2-E65A (white arrow, ~ 77 kDa). GST-Tpp2Aa2-E65A was expressed using previously optimised conditions (1mM IPTG, 4 hours, 37°C) and purified via GST-tagged affinity purification. Binding / wash flow-through (FT) and elution fractions obtained were analysed by SDS-PAGE. M (marker): BLUeye Pre-Stained Protein Ladder.

#### 5.2.10.4. Mosquito larval bioassays

Structural alignment of the AF2 prediction of Aam1 with the Tpp2Aa2-Cqm1 structure indicated that Aam1 residue Asp508 may sterically clash with Tpp2Aa2 residue Glu65. Thus, we hypothesized that the E65A mutation of Tpp2Aa2 would remove this steric clash, increasing the likelihood that Tpp2Aa2 would be able to interact with Aam1, thus broadening the target insect range to include *Ae. aegypti*.

To investigate the activity of the mutant Tpp2Aa2-E65A protein, single dose mosquito bioassays were performed against larvae (n = 5) of *Cx. quinquefasciatus*, *An. gambiae*, and *Ae. aegypti*. A 1:1 w/w mixture of either WT Tpp1Aa2 and WT Tpp2Aa2, or WT Tpp1Aa2 and mutant Tpp2Aa2-E65A (200  $\mu$ L 500  $\mu$ g/mL total – estimated by BCA assay) were administered to second / third instar larvae and their survival was monitored across a 48-hour period. Buffer alone was included as a negative control.

In line with previous work, percentage survival of *Cx. quinquefasciatus* and *An. gambiae* larvae following administration of WT Tpp1Aa2/Tpp2Aa2 was 0% after a 48-hour period, whereas 100% survival rate was identified for *Ae. aegypti*, confirming that the WT proteins were functioning as expected (**Fig 5.26 – cyan**). Similarly, a 0% survival rate was identified for *Cx. quinquefasciatus* and *An. gambiae* following administration of WT Tpp1Aa2/mutant Tpp2Aa2-E65A, whilst a 100% survival rate was identified for *Ae. aegypti*, indicating that the Tpp2Aa2 E65A mutation had not interfered with normal activity against *Cx. quinquefasciatus* and *An. gambiae*, nor had it widened the activity to include the otherwise refractory *Ae. aegypti* species (**Fig 5.26 – dark blue**). A 100% survival rate was identified for all three mosquito species following administration of a buffer alone control, confirming the validity of these experiments (**Fig 5.26 – green**).



**Figure 5.26. Larval bioassays investigating activity of Tpp1Aa2/Tpp2Aa2-E65A.** The percentage survival of three mosquito species following administration of wildtype (WT) Tpp1Aa2/Tpp2Aa2 (cyan – positive control), wildtype Tpp1Aa2 and mutant Tpp2Aa2-E65A (dark blue), or buffer alone (green – negative control) has been plotted. Larval bioassays demonstrate that the Tpp2Aa2 E65A mutation does not expand the target insect range of Tpp1/Tpp2 to include the refractory species, *Ae. aegypti*. As expected, a 0% survival rate of susceptible species, *Cx. quinquefasciatus* and *An. gambiae*, was seen following administration of WT Tpp1Aa2/Tpp2Aa2. B = buffer alone, WT = wildtype, Mut = mutant (n = 5).

### 5.3. Conclusions

In this chapter, the interaction of the Tpp2 protein with its mosquitocidal receptor, Cqm1, was investigated. In the first instance (and due to Covid lockdown access restrictions), computational modelling was performed to predict the interaction of Tpp2-Cqm1. Global docking, followed by local refinement, was performed using two programs, ClusPro and RosettaDock. MD simulations were used to simulate the dynamics of modelled complexes and subsequently, model stability was evaluated by calculating the RMSD and Rg of output trajectories. Model c4r1 was

selected as the most-likely structure due to its stability throughout MD simulations.

Subsequently, single-particle cryoEM was performed in collaboration with researchers at Monash University. Recombinantly produced Cqm1 and Tpp2Aa2 (in combination with its partner protein, Tpp1Aa2) were validated using  $\alpha$ -glucosidase assays and mosquito larvae bioassays, respectively. Single-particle cryoEM revealed a tetrameric complex consisting of the Cqm1 dimer and two Tpp2Aa2 monomers. These interactions involved several hydrogen bonds and salt bridges. The success of the combined ClusPro-RosettaDock modelling approach ultimately came down to its ability to identify the interfacial interactions seen in the Tpp2Aa2-Cqm1 structure. Specifically, the ClusPro-RosettaDock approach was able to predict several interfacial residues but, importantly, was not able to correctly predict their partner residues. Despite this, it was clear that the use of MD simulations to evaluate the stability of modelled complexes was advantageous, as it enabled us to correctly identify the most-likely structure out of all 5 models investigated. Following the release of AlphaFold-Multimer and the spin-off program, ColabFold, we built on this modelling work by predicting the tetrameric complex of Tpp2Aa2-Cqm1, however, neither approach was able to yield an accurate model. Interestingly, when we repeated the ColabFold modelling by removing 21 N-terminal and 53 C-terminal Tpp2Aa2 residues (predicted with low-confidence in the full-length model), we obtained a model with remarkable similarity to that of the Tpp2Aa2-Cqm1 structure, with several polar interactions predicted correctly. Taken together, this work highlights the significance of downstream experimental work to validate modelled protein-protein complexes. Despite this, molecular docking represents an effective approach to explore protein-protein interactions and can be used to direct and focus downstream experiments, including mutagenesis studies probing interfacial interactions.

Structure solution of the Tpp2Aa2-Cqm1 complex represents the first toxin-receptor complex to be solved within the group of *B. thuringiensis* and *L. sphaericus* pesticidal proteins. The Tpp1Aa2/Tpp2Aa2 binary toxin is one

of the major bioactive components of bioinsecticides in use for the control of *Culex* mosquito species. Thus, the Tpp2Aa2-Cqm1 structure is significant for the continued success of this product and in turn, the control of devastating mosquito-borne diseases. The structure provides insight into several aspects of Tpp1Aa2/Tpp2Aa2 mode of action, including a requirement for dissociation of the Tpp1Aa2/Tpp2Aa2 complex (which exists in the native co-crystal) for downstream Tpp2Aa2 receptor interaction. Further to this, the ability of pro-Tpp2Aa2 to interact with the receptor is unexpected and generates additional questions surrounding the sequence of events (crystal dissolution – proteolytic activation – receptor interaction) which occur following toxin ingestion.

In addition, the Tpp2Aa2-Cqm1 structure is significant for understanding toxin specificity which, in turn, will be important for enhancing toxin potency, stability and target insect range. In this work, the Tpp2Aa2-Cqm1 structure was aligned with the AlphaFold prediction of Aam1, an orthologous non-binding protein found in *Ae. aegypti* mosquitoes. *Ae. aegypti* are involved in the transmission of yellow fever, dengue, and Zika virus, and hence, the ability to target this species with Tpp1Aa2/Tpp2Aa2 would be advantageous. Based on structural and sequence alignments, a mutation (E65A) was made in Tpp2Aa2 with the goal of relieving predicted steric clashes with Aam1. Mosquito larvae bioassays indicated that this attempt was unsuccessful but, despite this, the workflow highlights how the Tpp2Aa2-Cqm1 structure can be applied in future protein engineering studies. Several other target residues were identified, and a combination of mutations may be required to elicit binding.

## 6. Discussion

### 6.1. Overview

In this work, a combination of SFX, cryoEM, and computational modelling techniques were successfully applied to study the structures of pesticidal proteins and their interactions with target receptors or partner proteins. The essential themes and novel findings pertaining to each toxin studied (Cry8Ba2, Cry48Aa1/Tpp49Aa1, Tpp1Aa2/Tpp2Aa2) have already been addressed. To be considered further in this discussion are (i) the wider implications for the application of pesticidal proteins as bioinsecticides, and (ii) the importance of molecular structure in protein studies and engineering, including the continued application and development of structural biology techniques. Finally, the limitations of the work and future research directions will be discussed.

### 6.2. Application of pesticidal proteins as bioinsecticides and beyond

Globally, agricultural yields face a significant threat from insect pests, leading to food insecurity and substantial economic losses (IPPC Secretariat 2021). Furthermore, some insect pests, such as mosquitoes, are involved in transmitting devastating human diseases (Chala and Hamde 2021), highlighting the need for effective methods to control their populations. In the past, chemical pesticides have been successfully applied to control insect pest populations, however, their use is associated with detrimental effects to human health, food safety, beneficial insects, and the environment (Pathak et al. 2022). Biological pesticides have emerged as a highly specific and biocompatible alternative (Ayilara et al. 2023), and microbials, including *B. thuringiensis* (Palma et al. 2014) and *L. sphaericus* (Silva-Filha et al. 2014) bacterial strains and the pesticidal proteins that they produce, represent the most widely applied forms. Given their significance, it is critical that we understand the mechanisms by which *B. thuringiensis* and *L. sphaericus* exert their pesticidal activity, both to counteract emerging field resistance and

to develop new bioinsecticides with enhanced potency, stability, and target insect range.

This work aimed to analyse and characterise the structures of pesticidal proteins and their interactions with target receptors to support their continued application as bioinsecticides. The main portion of this work focused on two binary pesticidal proteins with mosquitocidal activity, namely Cry48Aa1/Tpp49Aa1 (chapter 4) and Tpp1Aa2/Tpp2Aa2 (chapter 5). At present, Tpp1Aa2/Tpp2Aa2 is one of the major bioactive components of bioinsecticides in use for the control of *Culex* mosquito species (Regis et al. 2001; Lacey 2007), to prevent both the transmission of human diseases, such as West Nile virus, and to prevent nuisance biting in, for example, Southern France (Karch et al. 1990). Despite its successful application, several cases of field resistance have been identified (Silva-Filha et al. 2021). In most cases, resistance is caused by the selection for mutations in the Cpm1/Cqm1 receptor, which lead to the expression of truncated proteins without the GPI anchor, such that the receptor is no longer presented on the epithelial cell surface (Nielsen-Leroux et al. 1995; Oliveira et al. 2004). Hence, the Tpp2Aa2-Cqm1 structure does not have implications for overcoming the major mechanism of insect resistance. Despite this, knowledge of the Tpp2Aa2-Cqm1 interaction might have relevance for other forms of resistance where Tpp2 binding is retained (Nielsen-Leroux et al. 1997). In addition, the Tpp2Aa2-Cqm1 structure provides insight into the specificity of Tpp1Aa2/Tpp2Aa2 and hence, is expected to support protein engineering studies aimed at enhancing potency (through increased binding affinity) or broadening the target insect range (by targeting orthologous non-binding proteins, such as Aam1 in the refractory species, *Ae. aegypti*). The Tpp2Aa2-Cqm1 structure may also act as a model to study the specificity of other pesticidal proteins from the same class. On the other hand, our work on Cry48Aa1/Tpp49Aa1 is expected to benefit studies aimed at overcoming insect resistance given that this protein is specifically able to target *Culex* populations which are resistant to Tpp1Aa2/Tpp2Aa2 (Yuan et al. 2003). Moreover, our work demonstrating an expanded target insect range against *An. stephensi*, *Ae. albopictus*, and *Cx. tarsalis* highlights the potential for

Cry48Aa1/Tpp49Aa1 to be used as a multi-target bioinsecticide. The three new mosquito targets identified are significant vectors of human disease. *Cx. tarsalis*, for example, is identified as the primary vector for West Nile virus in the upper midwest region of the USA (Dunphy et al. 2019). *Ae. albopictus*, on the other hand, is the predominant vector for emerging arboviruses, such as dengue (Lwande et al. 2020). Additionally, *An. stephensi* has recently been identified as a malaria vector in Africa (Sinka et al. 2020). Notably, *An. stephensi* is among the few anopheline species found in central urban areas in Africa and poses a risk to 126 million people. In response, the World Health Organisation issued a vector alert, advocating for targeted control measures and prioritized surveillance (Sinka et al. 2020).

In addition to the above, the structure of the first full-length, non-mutated Cry protein in the long (~130 kDa) form, Cry8Ba2, was also resolved. The target range of Cry8Ba2 is unknown, thus limiting structural inferences relating to target insect specificity. Despite this, the Cry8Ba2 structure provided insight into the role of the extended pro-domain region and packing of Cry proteins into native *B. thuringiensis* crystals. The evolution and natural ability of *B. thuringiensis* and *L. sphaericus* to produce crystals of hundreds of proteins originating from several distinct structural classes (Cry, Tpp, App, Mpp, Gpp, Cyt, Xpp) is fascinating and has been discussed previously (Tetreau 2021b). In biological terms, the packing of pesticidal proteins into highly stable crystals ensures their long-term storage and functionality until they are ingested by the target insect (along with the spore). However, this natural crystallisation process also has several possible biotechnological applications. The first relates to the possibility of utilising the *B. thuringiensis* / *L. sphaericus* crystallisation machinery to crystallise other proteins of interest (Tetreau 2021a). This would be beneficial for X-ray crystallography studies, given that *in vitro* protein crystallisation represents the rate-limiting step of structure solution. The second, and one that is being explored by other groups, relates to use of *B. thuringiensis* / *L. sphaericus* crystals to package, stabilize, and protect molecules of interest for applications in biotechnology or medicine. One published and patented example exists and that is of Cry3Aa, an example of

a short (~65kDa) Cry protein (Chan et al. 2010; Chan et al. 2019; Yang et al. 2019). Briefly, the large solvent channels present in native Cry3Aa crystals (Sawaya et al. 2014; Tetreau 2021b) have been utilised to package proteins of interest (including biopharmaceutical molecules) by either fusing the protein of interest to the toxin or by co-expressing them. This is an ongoing field of research and one that might be supported by the structural analysis of native crystals and identification of interactions / features significant for maintaining crystal integrity.

Like any product, challenges relating to commercialisation exist. In relation to bioinsecticides, registration is widely perceived to be the most significant barrier. Indeed, although the demand and global market share for bioinsecticides continues to increase, the European Union (and United Kingdom) significantly fall behind the United States, China, India, and Brazil in the number of bioinsecticides registered (Frederiks and Wesseler 2019). It has been suggested that this is due to a lengthy and complex regulatory procedures governing biopesticide registration that, in the past, have been based on the framework for chemical pesticides (Chandler et al. 2011). This highlights the need for improved frameworks that focus on increased efficiency, without comprising safety standards (Frederiks and Wesseler 2019). In this context, knowledge of a protein's 3D structure will be beneficial by offering the potential to predict off-target activity and ensure the safety of new products (Moar et al. 2017).

Additional challenges surround the uptake of products by farmers and/or governing bodies. One example relates to the cost-benefit comparison of niche market products. Specifically, in comparison to chemical pesticides, bioinsecticides are highly selective, which, itself, offers the advantage of reduced off-target effects to beneficial species. However, this means that, where a product targets only one species of insect, multiple pest management strategies must be implemented. This can be overcome by producing bioinsecticide products which combine multiple pesticidal proteins, or by rational protein engineering to broaden the target insect range, for which knowledge of the protein structure is required.

### 6.3. Importance of molecular structure in protein studies

As previously alluded to, the ability to elucidate a protein's 3D structure is essential for understanding its function and holds significant importance in the context of rational protein engineering. In the case of *B. thuringiensis* and *L. sphaericus* pesticidal proteins, such insight is imperative for understanding mechanism of action, including toxin specificity, and has implications for downstream protein engineering to overcome insect resistance, as well as to improve potency, stability, and target insect range. In addition, knowledge of the structure has relevance for regulatory procedures and the registration of new bioinsecticides.

Given the above, it is essential that the native structures of pesticidal proteins and their complexes (e.g., pro-proteins, activated proteins alone and in complex with target receptors, and pore structures) can be elucidated. In the past, such structural analyses have been limited. In the case of the native crystals, direct structural analysis has been limited by crystal size. This has meant that, up until recently, the native crystals were solubilised and reconstituted in laboratory crystal trials for structural analysis by conventional crystallography. Often, solubilisation leads to auto-activation of the pro-protein and thus, information surrounding the pro-domains is lost. In this context, our work emphasizes the utility of SFX at an XFEL to generate high-resolution structures from naturally occurring *B. thuringiensis* and *L. sphaericus* crystals. The ability to study the native crystal was of significant value in the case of Cry8Ba2 presented in chapter 3, where approximately half of the nascent pro-protein is cleaved during proteolytic activation. Furthermore, direct structural analysis of the native crystal provided insight into the interactions that are significant for crystal integrity, as in the Tpp49Aa1 structure presented in chapter 4, which demonstrated the packing of dimers with a large intermolecular interface into native crystals. Relating to this, pH mixing in the XFEL beam enabled the early structural events leading up to crystal dissolution to be investigated, as demonstrated for Tpp49Aa1 crystals, where loss of hydrogen bonds near to the N-terminal

pro-peptide cleavage site was hypothesized to increase accessibility to gut proteases. In addition, insight into mechanisms of natural crystallisation may be applied to develop new pesticidal proteins with increased crystalline stability in the open environment (Tetreau 2021a). It is important to note, however, that an increase in crystalline stability may limit the bioavailability of the soluble, active form. Notably, the solubility of crystals has been identified as a factor influencing toxicity. In a study by Peng et al. (2011), native Cry7Ba1 crystals did not exhibit toxicity against *P. xylostella*. However, when the Cry7Ba1 crystals were solubilized *in vitro* at pH 12.5, subsequent bioassays using the solubilised protein demonstrated toxicity. Furthermore, a single mutation conferred solubility of the crystals at pH 9.5 and, in addition to this, bioassays using the mutant crystal protein demonstrated toxicity against *P. xylostella*, confirming that solubility was a factor influencing Cry7Ba1 toxicity (Peng et al. 2011).

Further to the above, by elucidating the structures of two forms of Cry8Ba2 from diffraction data collected in one session, we showed for the first time that mixed populations of crystals can be applied simultaneously to the XFEL beam and that, subsequently, the data can be separated, allowing the structures to be solved individually. This has wider implications for the structural analysis of mixed crystal protein strains, such as *B. thuringiensis* strain YBT-1518 that produces App6Aa2, Xpp55Aa1, and Cry5Ba2. In more recent work not presented in this thesis, we have also demonstrated the ability of SFX to elucidate the structures of native *B. thuringiensis* crystals in complex with sugars by pre-soaking the crystals prior to diffraction data collection. This technique is commonly used in conventional crystallography using crystals grown *in vitro* but, to our knowledge, is yet to be attempted with native crystals at an XFEL. With this in mind, there is huge potential for SFX to study the sugar binding capabilities of various pesticidal protein classes, an area of research that remains under-investigated (Best et al. 2023).

Structural analyses of protein complexes, including toxin-receptor complexes and pore structures, are also required to understand the

mechanism of action of *B. thuringiensis* and *L. sphaericus* pesticidal proteins more thoroughly. In this context, our work on the Tpp2Aa2-Cqm1 structure in chapter 5 emphasizes the application of single-particle cryoEM for the study of large complexes. The “resolution revolution” of cryoEM enabled the Tpp2Aa2-Cqm1 structure to be solved to an atomic resolution of 2.42 Å, thus enabling details of the intermolecular interactions to be identified. This enabled hypotheses to be generated surrounding target specificity that, in turn, directed protein engineering studies aimed at enhancing the target insect range of Tpp1Aa2/Tpp2Aa2. Overall, single-particle cryoEM is applicable to the study of large protein complexes, as demonstrated for the Tpp2Aa2-Cqm1 complex, and has great potential for structural analysis of even larger complexes, such as *B. thuringiensis* / *L. sphaericus* pore structures, as has been demonstrated for *B. thuringiensis* Vip3Aa16 (Núñez-Ramírez et al. 2020) and Vip3Bc1 (Byrne et al. 2021).

In this work, we also made use of computational modelling techniques to support the structural analysis of pesticidal proteins. One of the most impactful tools available throughout this research was AF2 (Jumper et al. 2021b). The ability to predict a protein’s 3D structure quickly and reasonably accurately was unrivalled and enabled us to make informed decisions surrounding experimental design (e.g., beamline set up for SFX studies at the XFEL presented in chapters 3 and 4) and perform downstream docking studies where the structures of proteins of interest could not be elucidated (e.g., Cry48Aa1 in the molecular docking study of Cry48Aa1-Tpp49Aa1 in chapter 4). In the elucidation of other protein structures not presented in this thesis (Mpp60, MD32), the models predicted by AF2 were successfully used as templates for molecular replacement where, in previous attempts, the use of homology models had not yielded a successful output. AF2 is also likely to aid the development of computational pipelines for the discovery and characterisation of novel pesticidal proteins in nature (Berry and Crickmore 2017). This will be particularly advantageous in the Agrotech industry, as such a computational pipeline will enable newly discovered genes, identified at an increasing rate due to improvements in next generation sequencing

techniques, to be narrowed down, prior to laboratory testing and development.

In addition to protein structure prediction techniques, we also made use of MD to assess the structural stability of docked complexes (produced using the combined ClusPro-RosettaDock-GROMACS approach) and identify the most-likely structure. We were able to assess the success of this approach in chapter 5 whereby, downstream of docking predictions, we elucidated the structure of the Tpp2Aa2-Cqm1 complex. The predicted Tpp2-Cqm1 model identified some interfacial residues correctly but did not successfully predict the partner residues of these contacts. Despite this, the use of MD was advantageous, as out of all the models assessed, the model that was most structurally stable in MD simulations also displayed the highest structural similarity to the experimentally resolved structure, supporting the use of MD to evaluate protein-protein docking results.

#### **6.4. Future directions and limitations**

The research presented in this thesis has revealed several areas with potential for further development.

In collaboration with researchers at DESY and the EuXFEL, chapter 3 focused on the structural analysis of Cry8Ba2 in its natural crystal form. Cry8Ba2 DV and DVII display structural similarity with carbohydrate binding modules and hence, to build on this work, the novel role for sugar binding within the pro-domain region would be investigated. To do so, glycan microarrays would be used to analyse glycan-protein interactions of pro-domain fragments. Subsequently, sugar binding sites could be identified by soaking the natural crystals prior to SFX studies. The Cry8Ba2 structure also revealed interesting features surrounding the packing of molecules into natural crystals, including the identification of an extended loop in DVII that appears to act as a peg, stabilising the packing of Cry8Ba2 molecules. To investigate the role of this loop further, site-directed mutagenesis would be applied to delete the loop and investigate the effect on crystal formation and

stability. In addition, two forms of Cry8Ba2 crystals were identified, the biological relevance of which is unknown. It is possible that the two crystals exhibit different solubility characteristics, and it would be interesting to investigate this further. However, such work may be limited by the ability to produce and separate the two crystal populations for subsequent experiments. Another limitation of this work relates to the unknown target insect range of Cry8Ba2. Hence, a final area of future work would include additional insect bioassays to investigate the target insect range that, once known, would enable downstream work aimed at studying the molecular mechanisms of Cry8Ba2, such as receptor binding proteins, for which knowledge of the Cry8Ba2 structure will be applicable.

In collaboration with researchers at DESY and the EuXFEL, chapter 4 focused on the structural analysis of Tpp49Aa1 in its natural crystal form to gain insight into the Cry48Aa1/Tpp49Aa1 binary toxin. Despite collecting diffraction data across three beamtimes, the Cry48Aa1 structure could not be resolved. In future work, other structural analysis techniques, such as microED, would be attempted. Nonetheless, the Tpp49Aa1 structure was used to model the interaction with the partner protein, Cry48Aa1 (predicted using AF2), enabling models of activity to be discussed. To validate this model, mutagenesis studies targeting predicted interfacial interactions are required. Finally, Cry48Aa1/Tpp49Aa1 is active against *Culex* mosquitoes that are resistant to Tpp1Aa2/Tpp2Aa2 (Yuan et al. 2003) and hence, this toxin has implications for overcoming insect resistance. However, in natural strains, Cry48Aa1 is produced at much lower concentrations than Tpp49Aa1. This is interesting given that optimal activity is achieved when both proteins are present at a 1:1 M ratios (Jones et al. 2007) and highlights the need for optimised production of Cry48Aa1 if the toxin is to be successfully applied in the field.

In collaboration with researchers at Monash University, chapter 5 focused on the structural analysis of the Tpp2Aa2-Cqm1 complex and provided insight into target specificity. Downstream of this work, we applied the Tpp2Aa2-Cqm1 structure to generate hypotheses surrounding the

inability of Aam1 to bind Tpp2Aa2. Aam1 is a protein found in refractory *Ae. aegypti* that is orthologous to the Cqm1 receptor but is unable to bind Tpp2Aa2 (Ferreira et al. 2010). Given that this lack of binding is associated with the refractory nature of *Ae. aegypti*, such work has implications for broadening the target insect range of Tpp1Aa2/Tpp2Aa2. Specifically, we performed a structural alignment of the AF2-predicted Aam1 model with the Tpp2Aa2-Cqm1 structure and identified steric clashes between Aam1 and Tpp2Aa2 in order to identify target residues for mutagenesis and protein engineering. Our initial mutant (Tpp2Aa2-E65A) did not retarget Tpp1/Tpp2, however, several other mutations could be made. In addition, a favoured upgrade to the workflow would be to use computational tools to design a Tpp2Aa2 mutant with the desired ability to bind Aam1. To do so, the Aam1 model would be aligned with the Tpp2Aa2-Cqm1 structure to identify an initial binding pose. Short MD simulations would enable steric clashes to be removed. Downstream of this, protein design programs, such as Protein MPNN (Dauparas et al. 2022) or RosettaDiffusion (Fudge 2023; Watson et al. 2023), would be applied to design a Tpp2Aa2 mutant with the ability to bind Aam1 at the chosen binding site. The Tpp2Aa2 mutant would be expressed, purified, and assessed in binding assays. Aside from this, the Tpp2Aa2-Cqm1 structure also generated questions surrounding the mode of action of Tpp1Aa2/Tpp2Aa2, as it is generally accepted that, following the dissolution of crystals in the alkaline environment of the midgut, the Tpp1Aa2 and Tpp2Aa2 proteins are proteolytically cleaved, subsequently leading to receptor binding. In our cryoEM experiments, however, we utilised the pro-form of Tpp2Aa2 and show that it can interact with the target receptor, Cqm1. To investigate this further, cryoEM experiments of the activated Tpp2Aa2 in complex with Cqm1, alongside surface plasmon resonance experiments evaluating the kinetics of the interactions, would be performed. Given that Tpp1Aa2/Tpp2Aa2 is a binary toxin, a final area of future work would be aimed at investigating the involvement of Tpp1Aa2, with the ultimate aim being structural elucidation of the Tpp1Aa2-Tpp2Aa1-Cqm1 complex.

In chapters 4 and 5, computational modelling tools were applied to predict the structures of the Cry48Aa1-Tpp49Aa1 and Tpp2Aa2-Cqm1 complexes prior to experimental elucidation. A few areas for improvement were noted. Specifically, the combined Rosetta and ClusPro docking protocol did not account for backbone flexibility and, therefore, where binding-induced conformational backbone movements occur, structure prediction was limited. To model backbone flexibility more effectively, we could use the Rosetta Backrub protocol to produce conformational ensembles of the unbound Tpp2 and Cqm1 proteins prior to docking (Smith and Kortemme 2008). In the past, this method has improved the prediction accuracy of flexible targets (Marze et al. 2018). Here, we sought to account for protein flexibility by incorporating MD into our evaluation approach. Indeed, MD and subsequent backbone RMSD analysis has been utilised in the past to discriminate native structures, from similarly scored, non-native structures (Radom et al. 2018). However, this approach is also limited by the possibility of models becoming kinetically trapped in the free energy landscape (Radom et al. 2018). If additional time was available, this issue could be addressed by performing MD simulations at increasing temperatures (Radom et al. 2018). Further limitations include the length of MD simulations used. Protein-protein docking events occur in the  $\mu$ s to ms timescale and therefore, the 100 ns simulations performed here may not capture binding/unbinding events (Pan et al. 2019). However, due to the necessary computing power required to produce  $\mu$ s to ms simulations, such studies were deemed beyond the scope of this work. A final limitation related to the ability to predict the Cry48Aa1-Tpp49Aa1 interaction at both pH 7 and 11. Specifically, we were unable to identify molecular docking programs that could predict protein-protein interactions at pH 11 (representative of the mosquito larval anterior gut with a pH > 10) by retaining precalculated input protonation states. Hence, molecular docking was performed without considering pH and, downstream of this, the protonation states of titratable residues were incorporated into the models for subsequent MD simulations. However, even then, GROMACS (Abraham et al. 2015) was only able to consider the protonation states of lysine, aspartic acid, glutamine, cystine, and histidine, and thus excluded tyrosine ( $pK_{a3} =$

10.07). In addition, the use of constant protonation states in MD simulations has its own limitations as changes in conformational state are often coupled with changes in protonation patterns. In the future, the use of constant state pH simulations that sample protonation states on the fly could be considered (Buslaev et al. 2022).

## **6.5. Summary**

To summarise this work, a combination of SFX, cryoEM, and computational modelling were successfully applied to study the structures of pesticidal proteins and their interactions with target receptors or partner proteins. Taken together, the work highlights the importance of an interdisciplinary approach to capture the structural journey of *B. thuringiensis* and *L. sphaericus* pesticidal proteins and, in turn, understand their mechanisms of action, with the aim of developing new and improved bioinsecticides.

## 7. References

Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B. and Lindah, E. 2015. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1–2, pp. 19–25. doi: 10.1016/j.softx.2015.06.001.

Abrami, L. et al. 2002. The glycan core of GPI-anchored proteins modulates aerolysin binding but is not sufficient: The polypeptide moiety is required for the toxin-receptor interaction. *FEBS Letters* 512(1–3), pp. 249–254. doi: 10.1016/S0014-5793(02)02274-3.

Adams, P.D. et al. 2010. PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography* 66(2), pp. 213–221. doi: 10.1107/S0907444909052925.

Afonine, P. V. et al. 2012. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallographica Section D: Biological Crystallography* 68(4), pp. 352–367. doi: 10.1107/S0907444912001308.

Agirre, J. et al. 2023. The CCP4 suite: integrative software for macromolecular crystallography. *Acta crystallographica. Section D, Structural biology* 79(6), pp. 449–461. doi: 10.1107/S2059798323003595.

Alder, B.J. and Wainwright, T.E. 1957. Phase transition for a hard sphere system. *The Journal of Chemical Physics* 27(5), pp. 1208–1209. doi: 10.1063/1.1743957.

Alder, B.J. and Wainwright, T.E. 1959. Studies in Molecular Dynamics. *The Journal of Chemical Physics* 31(2), pp. 459–466. doi: 10.1063/1.1730376.

Alford, R.F. et al. 2017. The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *Journal of Chemical Theory and Computation* 13(6), pp. 3031–3048. doi: 10.1021/acs.jctc.7b00125.

Álvarez, F., Martín Camargo, A., Messéan, A., Lenzi, P. and Streissl, F. 2022. Assessment of the 2020 post-market environmental monitoring report on the cultivation of genetically modified maize MON 810 in the EU. *EFSA Journal* 20(7). doi: 10.2903/j.efsa.2022.7406.

Amadio, A.F., Navas, L.E., Sauka, D.H., Berretta, M.F., Benintende, G.B. and Zandomeni, R.O. 2013. Identification, cloning and expression of an insecticide cry8 gene from *Bacillus thuringiensis* INTA Fr7-4. *Journal of molecular microbiology and biotechnology* 23(6), pp. 401–409. doi: 10.1159/000353206.

Anfinsen, C.B. 1973. Principles that govern the folding of protein chains. *Science* 181(4096), pp. 223–230. doi: 10.1126/science.181.4096.223.

- Aquila, A. et al. 2012. Time-resolved protein nanocrystallography using an X-ray free-electron laser. *Optics Express* 20(3), p. 2706. doi: 10.1364/oe.20.002706.
- Aqvist, J., Luecke, H., Quijcho, F.A. and Warshel, A. 1991. Dipoles localized at helix termini of proteins stabilize charges. *Proceedings of the National Academy of Sciences* 88(5), pp. 2026–2030. doi: 10.1073/pnas.88.5.2026.
- Arenas, I., Bravo, A., Soberón, M. and Gómez, I. 2010. Role of alkaline phosphatase from *Manduca sexta* in the mechanism of action of *Bacillus thuringiensis* Cry1Ab toxin. *Journal of Biological Chemistry* 285(17), pp. 12497–12503. doi: 10.1074/jbc.M109.085266.
- Asano, S.I., Yamashita, C., Iizuka, T., Takeuchi, K., Yamanaka, S., Cerf, D. and Yamamoto, T. 2003. A strain of *Bacillus thuringiensis* subsp. *galleriae* containing a novel cry8 gene highly toxic to *Anomala cuprea* (Coleoptera: Scarabaeidae). *Biological Control* 28(2), pp. 191–196. doi: 10.1016/S1049-9644(03)00060-4.
- Atsumi, S., Inoue, Y., Ishizaka, T., Mizuno, E., Yoshizawa, Y., Kitami, M. and Sato, R. 2008. Location of the *Bombyx mori* 175 kDa cadherin-like protein-binding site on *Bacillus thuringiensis* Cry1Aa toxin. *FEBS Journal* 275(19), pp. 4913–4926. doi: 10.1111/j.1742-4658.2008.06634.x.
- Ayilara, M.S. et al. 2023. Biopesticides as a promising alternative to synthetic pesticides: A case for microbial pesticides, phytopesticides, and nanobiopesticides. *Frontiers in Microbiology* 14, p. 1040901. doi: 10.3389/fmicb.2023.1040901.
- Bae, B., Ohene-Adjei, S., Kocherginskaya, S., Mackie, R.I., Spies, M.A., Cann, I.K.O. and Nair, S.K. 2008. Molecular basis for the selectivity and specificity of ligand recognition by the family 16 carbohydrate-binding modules from *Thermoanaerobacterium polysaccharolyticum* ManA. *Journal of Biological Chemistry* 283(18), pp. 12415–12425. doi: 10.1074/jbc.M706513200.
- Baines, A.J. 2009. Evolution of spectrin function in cytoskeletal and membrane networks. *Biochemical Society Transactions* 37(4), pp. 796–803. doi: 10.1042/BST0370796.
- Barty, A., Kirian, R.A., Maia, F.R.N.C., Hantke, M., Yoon, C.H., White, T.A. and Chapman, H. 2014. Cheetah: Software for high-throughput reduction and analysis of serial femtosecond X-ray diffraction data. *Journal of Applied Crystallography* 47(3), pp. 1118–1131. doi: 10.1107/S1600576714007626.
- Basu, S., Chakravarty, D., Bhattacharyya, D., Saha, P. and Patra, H.K. 2021. Plausible blockers of Spike RBD in SARS-CoV2—molecular design and underlying interaction dynamics from high-level structural descriptors. *Journal of Molecular Modeling* 27(6), pp. 1–32. doi: 10.1007/s00894-021-04779-0.

- Basu, S. and Wallner, B. 2016. DockQ: A quality measure for protein-protein docking models. *PLoS ONE* 11(8), p. e0161879. doi: 10.1371/journal.pone.0161879.
- Baumann, L., Broadwell, A.H. and Baumann, P. 1988. Sequence Analysis of the Mosquitocidal Toxin Genes Encoding 51.4-and 41.9-Kilodalton Proteins from *Bacillus sphaericus* 2362 and 2297. *J Bacteriol.* 170(5), pp. 2045–2050.
- Bean, R.J., Aquila, A., Samoylova, L. and Mancuso, A.P. 2016. Design of the mirror optical systems for coherent diffractive imaging at the SPB/SFX instrument of the European XFEL. *Journal of Optics* 18(7), p. 074011. doi: 10.1088/2040-8978/18/7/074011.
- Bernardi, O. et al. 2012. Assessment of the high-dose concept and level of control provided by MON 87701 × MON 89788 soybean against *Anticarsia gemmatilis* and *Pseudoplusia includens* (Lepidoptera: Noctuidae) in Brazil. *Pest Management Science* 68(7), pp. 1083–1091. doi: 10.1002/ps.3271.
- Berry, C. and Crickmore, N. 2017. Structural classification of insecticidal proteins – Towards an in silico characterisation of novel toxins. *Journal of Invertebrate Pathology* 142, pp. 16–22. doi: 10.1016/j.jip.2016.07.015.
- Berry, C., Hindley, J., Ehrhardt, A.F., Grounds, T., De Souza, I. and Davidson, E.W. 1993. Genetic determinants of host ranges of *Bacillus sphaericus* mosquito larvicidal toxins. *Journal of Bacteriology* 175(2), pp. 510–518. doi: 10.1128/jb.175.2.510-518.1993.
- Bertoline, L.M.F., Lima, A.N., Krieger, J.E. and Teixeira, S.K. 2023. Before and after AlphaFold2: An overview of protein structure prediction. *Frontiers in Bioinformatics* 3, p. 1120370. doi: 10.3389/FBINF.2023.1120370.
- Bertoni, M., Kiefer, F., Biasini, M., Bordoli, L. and Schwede, T. 2017. Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. *Scientific Reports* 7(1). doi: 10.1038/s41598-017-09654-8.
- Best, H.L., Williamson, L.J., Heath, E.A., Waller-Evans, H., Lloyd-Evans, E. and Berry, C. 2023. The Role of Glycoconjugates as Receptors for Insecticidal Proteins. *FEMS Microbiology Reviews* 47(4), p. fuad026. doi: 10.1093/femsre/fuad026.
- Best, H.L., Williamson, L.J., Lipka-Lloyd, M., Waller-Evans, H., Lloyd-Evans, E., Rizkallah, P.J. and Berry, C. 2022. The Crystal Structure of *Bacillus thuringiensis* Tpp80Aa1 and Its Interaction with Galactose-Containing Glycolipids. *Toxins* 14(12), p. 863. doi: 10.3390/toxins14120863.
- Bietlot, H.P.L., Vishnubhatla, I., Carey, P.R., Pozsgay, M. and Kaplan, H. 1990. Characterization of the cysteine residues and disulphide linkages in the protein crystal of *Bacillus thuringiensis*. *Biochemical Journal* 267(2), pp. 309–315. doi: 10.1042/bj2670309.

- Biggs, D.R. and McGregor, P.G. 1996. Gut pH and amylase and protease activity in larvae of the New Zealand grass grub (*Costelytra zealandica*; Coleoptera: Scarabaeidae) as a basis for selecting inhibitors. *Insect Biochemistry and Molecular Biology* 26(1), pp. 69–75. doi: 10.1016/0965-1748(95)00064-X.
- Boonserm, P., Davis, P., Ellar, D.J. and Li, J. 2005. Crystal structure of the mosquito-larvicidal toxin Cry4Ba and its biological implications. *Journal of molecular biology* 348(2), pp. 363–382. doi: 10.1016/J.JMB.2005.02.013.
- Boonserm, P., Mo, M., Angsuthanasombat, C. and Lescar, J. 2006. Structure of the functional form of the mosquito larvicidal Cry4Aa toxin from *Bacillus thuringiensis* at a 2.8-angstrom resolution. *Journal of Bacteriology* 188(9), pp. 3391–3401. doi: 10.1128/JB.188.9.3391-3401.2006.
- Boudko, D.Y., Moroz, L.L., Linser, P.J., Trimarchi, J.R., Smith, P.J.S. and Harvey, W.R. 2001. In situ analysis of pH gradients in mosquito larvae using non-invasive, self-referencing, pH-sensitive microelectrodes. *Journal of Experimental Biology* 204(4), pp. 691–699. doi: 10.1242/jeb.204.4.691.
- Boutet, S. et al. 2012. High-Resolution Protein Structure Determination by Serial Femtosecond Crystallography. *Science* 337(6092), pp. 362–364. doi: 10.1126/science.1217737.
- Bravo, A. et al. 2004. Oligomerization triggers binding of a *Bacillus thuringiensis* Cry1Ab pore-forming toxin to aminopeptidase N receptor leading to insertion into membrane microdomains. *Biochimica et Biophysica Acta - Biomembranes* 1667(1), pp. 38–46. doi: 10.1016/j.bbamem.2004.08.013.
- Broadwell, A.H., Baumann, L. and Baumann, P. 1990a. The 42- and 51-kilodalton mosquitocidal proteins of *Bacillus sphaericus* 2362: Construction of recombinants with enhanced expression and in vivo studies of processing and toxicity. *Journal of Bacteriology* 172(5), pp. 2217–2223. doi: 10.1128/jb.172.5.2217-2223.1990.
- Broadwell, A.H., Clark, M.A., Baumann, L. and Baumann, P. 1990b. Construction by site-directed mutagenesis of a 39-kilodalton mosquitocidal protein similar to the larva-processed toxin of *Bacillus sphaericus* 2362. *Journal of Bacteriology* 172(7), pp. 4032–4036. doi: 10.1128/jb.172.7.4032-4036.1990.
- Burton, S.L., Ellar, D.J., Li, J. and Derbyshire, D.J. 1999. N-acetylgalactosamine on the putative insect receptor aminopeptidase N is recognised by a site on the domain III lectin-like fold of a *Bacillus thuringiensis* insecticidal toxin. *Journal of Molecular Biology* 287(5), pp. 1011–1022. doi: 10.1006/jmbi.1999.2649.

- Buslaev, P., Aho, N., Jansen, A., Bauer, P., Hess, B. and Groenhof, G. 2022. Best Practices in Constant pH MD Simulations: Accuracy and Sampling. *Journal of Chemical Theory and Computation* 18(10), pp. 6134–6147. doi: 10.1021/acs.jctc.2c00517.
- Byrne, M.J. et al. 2021. Cryo-EM structures of an insecticidal Bt toxin reveal its mechanism of action on the membrane. *Nature Communications* 12(1), pp. 1–9. doi: 10.1038/s41467-021-23146-4.
- Cao, B. et al. 2022. The crystal structure of Cry78Aa from *Bacillus thuringiensis* provides insights into its insecticidal activity. *Communications Biology* 5(1), pp. 1–14. doi: 10.1038/s42003-022-03754-6.
- Carroll, J. and Ellar, D.J. 1993. An analysis of *Bacillus thuringiensis*  $\delta$ -endotoxin action on insect-midgut-membrane permeability using a light-scattering assay. *European Journal of Biochemistry* 214(3), pp. 771–778. doi: 10.1111/j.1432-1033.1993.tb17979.x.
- Carroll, J. and Ellar, D.J. 1997. Analysis of the large aqueous pores produced by a *Bacillus thuringiensis* protein insecticide in *manduca sexta* midgut-brush-border-membrane vesicles. *European Journal of Biochemistry* 245(3), pp. 797–804. doi: 10.1111/j.1432-1033.1997.00797.x.
- Carugo, O. and Djinović-Carugo, K. 2023. Structural biology: A golden era. *PLoS Biology* 21(6), p. e3002187. doi: 10.1371/journal.pbio.3002187.
- Chala, B. and Hamde, F. 2021. Emerging and Re-emerging Vector-Borne Infectious Diseases and the Challenges for Control: A Review. *Frontiers in Public Health* 9, p. 715759. doi: 10.3389/fpubh.2021.715759.
- Chan, M.K., Lee, M.M. and Heater, B. 2019. *Fusion Proteins with Improved Properties* US Patent 2019/0185522 [Patent].
- Chan, M.K., Nair, M.S. and Lee, M.M. 2010. *Biomaterials, Compositions, and Methods* US Patent 2010/0322977 A1 [Patent].
- Chandler, D., Bailey, A.S., Mark Tatchell, G., Davidson, G., Greaves, J. and Grant, W.P. 2011. The development, regulation and use of biopesticides for integrated pest management. *Philosophical Transactions of the Royal Society B: Biological Sciences* 366(1573), pp. 1987–1998. doi: 10.1098/rstb.2010.0390.
- Chapman, H.N. et al. 2007. Femtosecond time-delay X-ray holography. *Nature* 448(7154), pp. 676–679. doi: 10.1038/nature06049.
- Chapman, H.N. et al. 2011. Femtosecond X-ray protein nanocrystallography. *Nature* 470(7332), pp. 73–78. doi: 10.1038/nature09750.
- Chari, A. and Stark, H. 2023. Prospects and Limitations of High-Resolution Single-Particle Cryo-Electron Microscopy. *Annual Review of Biophysics* 52, pp. 391–411. doi: 10.1146/annurev-biophys-111622-091300.

- Charles, J.F. 1987. Ultrastructural midgut events in culicidae larvae fed with *Bacillus sphaericus* 2297 spore/crystal complex. *Annales de l'Institut Pasteur. Microbiologie* 138(4), pp. 471–484. doi: 10.1016/0769-2609(87)90064-0.
- Charles, J.F., Nielsen-LeRoux, C. and Delécluse, A. 1996. *Bacillus sphaericus* toxins: Molecular biology and mode of action. *Annual Review of Entomology* 41(1), pp. 451–472. doi: 10.1146/annurev.en.41.010196.002315.
- Charles, J.F., Silva-Filha, M.H., Nielsen-LeRoux, C., Humphreys, M.J. and Berry, C. 1997. Binding of the 51- and 42-kDa individual components from the *Bacillus sphaericus* crystal toxin to mosquito larval midgut membranes from *Culex* and *Anopheles* sp. (Diptera: Culicidae). *FEMS Microbiology Letters* 156(1), pp. 153–159. doi: 10.1111/j.1574-6968.1997.tb12721.x.
- Chaudhury, S., Berrondo, M., Weitzner, B.D., Muthu, P., Bergman, H. and Gray, J.J. 2011. Benchmarking and analysis of protein docking performance in Rosetta v3.2. *PLoS ONE* 6(8), p. e22477. doi: 10.1371/journal.pone.0022477.
- Cheng, X., Sardana, R., Kaplan, H. and Altosaar, I. 1998. Agrobacterium-transformed rice plants expressing synthetic cryIA(b) and cryIA(c) genes are highly toxic to striped stem borer and yellow stem borer. *Proceedings of the National Academy of Sciences* 95(6), pp. 2767–2772. doi: 10.1073/pnas.95.6.2767.
- Chothia, C. and Janin, J. 1975. Principles of protein-protein recognition. *Nature* 256(5520), pp. 705–708. doi: 10.1038/256705a0.
- Chua, E.Y.D. et al. 2022. Better, Faster, Cheaper: Recent Advances in Cryo-Electron Microscopy. *Annual Review of Biochemistry* 91, pp. 1–32. doi: 10.1146/annurev-biochem-032620.
- Cid, M., Pedersen, H.L., Kaneko, S., Coutinho, P.M., Henrissat, B., Willats, W.G.T. and Boraston, A.B. 2010. Recognition of the helical structure of  $\beta$ -1,4-galactan by a new family of carbohydrate-binding modules. *Journal of Biological Chemistry* 285(46), pp. 35999–36009. doi: 10.1074/jbc.M110.166330.
- Clark, M.A. and Baumann, P. 1990. Deletion analysis of the 51-kilodalton protein of the *Bacillus sphaericus* 2362 binary mosquitocidal toxin: Construction of derivatives equivalent to the larva-processed toxin. *Journal of Bacteriology* 172(12), pp. 6759–6763. doi: 10.1128/jb.172.12.6759-6763.1990.
- Cokmus, C., Davidson, E.W. and Cooper, K. 1997. Electrophysiological Effects of *Bacillus sphaericus* Binary Toxin on Cultured Mosquito Cells. *Journal of Invertebrate Pathology* 69(3), pp. 197–204. doi: 10.1006/jipa.1997.4660.

Colletier, J.P. et al. 2016. De novo phasing with X-ray laser reveals mosquito larvicide BinAB structure. *Nature* 539(7627), pp. 43–47. doi: 10.1038/nature19825.

Conte, L. Lo, Chothia, C. and Janin, J. 1999. The atomic structure of protein-protein recognition sites. *Journal of Molecular Biology* 285(5), pp. 2177–2198. doi: 10.1006/jmbi.1998.2439.

Couche, G.A., Pfannenstiel, M.A. and Nickerson, K.W. 1987. Structural disulfide bonds in the *Bacillus thuringiensis* subsp. *israelensis* protein crystal. *Journal of Bacteriology* 169(7), pp. 3281–3288.

Couturier, M., Roussel, A., Rosengren, A., Leone, P., Stålbbrand, H. and Berrin, J.G. 2013. Structural and biochemical analyses of glycoside hydrolase families 5 and 26  $\beta$ -(1,4)-mannanases from *Podospora anserina* reveal differences upon manno-oligosaccharide catalysis. *Journal of Biological Chemistry* 288(20), pp. 14624–14635. doi: 10.1074/jbc.M113.459438.

Cowtan, K. 2006. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallographica Section D: Biological Crystallography* 62(9), pp. 1002–1011. doi: 10.1107/S0907444906022116.

Crickmore, N. et al. 1998. Revision of the Nomenclature for the *Bacillus thuringiensis* Pesticidal Crystal Proteins. *Microbiology and Molecular Biology Reviews* 62(3), p. 807. doi: 10.1128/MMBR.62.3.807-813.1998.

Crickmore, N., Berry, C., Panneerselvam, S., Mishra, R., Connor, T.R. and Bonning, B.C. 2020. A structure-based nomenclature for *Bacillus thuringiensis* and other bacteria-derived pesticidal proteins. *Journal of Invertebrate Pathology* 186, p. 107438. doi: 10.1016/j.jip.2020.107438.

Croll, T.I. 2018. ISOLDE: A physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallographica Section D: Structural Biology* 74(6), pp. 519–530. doi: 10.1107/S2059798318002425.

Crowson, R.A. 1981. *The Biology of the Coleoptera*. Academic Press, New York. Available at: [https://books.google.co.uk/books?hl=en&lr=&id=WMXYBAAAQBAJ&oi=fnd&pg=PP1&dq=Crowson,+R.+A.,+1981.+The+biology+of+the+Coleoptera.+Academic+Press,+London+%3B+New+York.&ots=MsW-bylgNX&sig=nCoWo8ByOjxHe\\_miLv2ctqHIMxE#v=onepage&q=Crowson%2C+R.+A.%2C+1981.+The](https://books.google.co.uk/books?hl=en&lr=&id=WMXYBAAAQBAJ&oi=fnd&pg=PP1&dq=Crowson,+R.+A.,+1981.+The+biology+of+the+Coleoptera.+Academic+Press,+London+%3B+New+York.&ots=MsW-bylgNX&sig=nCoWo8ByOjxHe_miLv2ctqHIMxE#v=onepage&q=Crowson%2C+R.+A.%2C+1981.+The) [Accessed: 16 September 2023].

Dadd, R.H. 1975. Alkalinity within the midgut of mosquito larvae with alkaline-active digestive enzymes. *Journal of Insect Physiology* 21(11), pp. 1847–1853. doi: 10.1016/0022-1910(75)90252-8.

- Darboux, I., Nielsen-LeRoux, C., Charles, J.F. and Pauron, D. 2001. The receptor of *Bacillus sphaericus* binary toxin in *Culex pipiens* (Diptera: Culicidae) midgut: Molecular cloning and expression. *Insect Biochemistry and Molecular Biology* 31(10), pp. 981–990. doi: 10.1016/S0965-1748(01)00046-7.
- Dastidar, P.G. and Nickerson, K.W. 1979. Interchain crosslinks in the entomocidal bacillus thuringiensis protein crystal. *FEBS Letters* 108(2), pp. 411–414. doi: 10.1016/0014-5793(79)80575-X.
- Dauparas, J. et al. 2022. Robust deep learning–based protein sequence design using ProteinMPNN. *Science* 378(6615), pp. 49–56. doi: 10.1126/science.add2187.
- Davidson, E.W. and Titus, M. 1987. Ultrastructural effects of the *Bacillus sphaericus* mosquito larvicidal toxin on cultured mosquito cells. *Journal of Invertebrate Pathology* 50(3), pp. 213–220. doi: 10.1016/0022-2011(87)90085-1.
- Dennis, R.D., Wiegandt, H., Haustein, D., Knowles, B.H. and Ellar, D.J. 1986. 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. *Biomedical Chromatography* 1(1), pp. 31–37. doi: 10.1002/bmc.1130010108.
- Derbyshire, D.J., Ellar, D.J. and Li, J. 2001. Crystallization of the *Bacillus thuringiensis* toxin Cry1Ac and its complex with the receptor ligand N-acetyl-D-galactosamine. *Acta Crystallographica Section D: Biological Crystallography* 57(12), pp. 1938–1944. doi: 10.1107/S090744490101040X.
- Dolinsky, T.J., Czodrowski, P., Li, H., Nielsen, J.E., Jensen, J.H., Klebe, G. and Baker, N.A. 2007. PDB2PQR: Expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Research* 35(Web Server issue), pp. W522–W525. doi: 10.1093/nar/gkm276.
- Donovan, W.P. et al. 2006. Discovery and characterization of Sip1A: A novel secreted protein from *Bacillus thuringiensis* with activity against coleopteran larvae. *Applied Microbiology and Biotechnology* 72(4), pp. 713–719. doi: 10.1007/s00253-006-0332-7.
- Dunphy, B.M., Kovach, K.B., Gehrke, E.J., Field, E.N., Rowley, W.A., Bartholomay, L.C. and Smith, R.C. 2019. Long-term surveillance defines spatial and temporal patterns implicating *Culex tarsalis* as the primary vector of West Nile virus. *Scientific Reports* 9(1), p. 6637. doi: 10.1038/s41598-019-43246-y.
- Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. 2010. Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography* 66(4), pp. 486–501. doi: 10.1107/S0907444910007493.

Estruch, J.J., Warren, G.W., Mullins, M.A., Nye, G.J., Craig, J.A. and Koziel, M.G. 1996. Vip3A, a novel *Bacillus thuringiensis* vegetative insecticidal protein with a wide spectrum of activities against lepidopteran insects. *Proceedings of the National Academy of Sciences* 93(11), pp. 5389–5394. doi: 10.1073/pnas.93.11.5389.

Evans, R. et al. 2022. Protein complex prediction with AlphaFold-Multimer. *bioRxiv*, p. 2021.10.04.463034. doi: 10.1101/2021.10.04.463034.

Evdokimov, A.G., Moshiri, F., Sturman, E.J., Rydel, T.J., Zheng, M., Seale, J.W. and Franklin, S. 2014. Structure of the full-length insecticidal protein Cry1Ac reveals intriguing details of toxin packaging into in vivo formed crystals. *Protein Science* 23(11), pp. 1491–1497. doi: 10.1002/pro.2536.

Ferreira, L.M., Romão, T.P., De-Melo-Neto, O.P. and Silva-Filha, M.H.N.L. 2010. The orthologue to the Cpm1/Cqm1 receptor in *Aedes aegypti* is expressed as a midgut GPI-anchored  $\alpha$ -glucosidase, which does not bind to the insecticidal binary toxin. *Insect Biochemistry and Molecular Biology* 40(8), pp. 604–610. doi: 10.1016/j.ibmb.2010.05.007.

Ferreira, L.M., Romão, T.P., Nascimento, N.A., Costa, M.C.M.F., Rezende, A.Ô.M., De-Melo-Neto, O.P. and Silva-Filha, M.H.N.L. 2014. Non conserved residues between Cqm1 and Aam1 mosquito  $\alpha$ -glucosidases are critical for the capacity of Cqm1 to bind the Binary toxin from *Lysinibacillus sphaericus*. *Insect Biochemistry and Molecular Biology* 50(1), pp. 34–42. doi: 10.1016/j.ibmb.2014.04.004.

Foloppe, N. and MacKerell, A.D. 2000. All-Atom Empirical Force Field for Nucleic Acids: I. Parameter Optimization Based on Small Molecule and Condensed Phase Macromolecular Target Data. *Journal of Computational Chemistry* 21(2), pp. 86–104. doi: 10.1002/(SICI)1096-987X(20000130)21:2<86::AID-JCC2>3.0.CO;2-G.

Frederiks, C. and Wesseler, J.H.H. 2019. A comparison of the EU and US regulatory frameworks for the active substance registration of microbial biological control agents. *Pest Management Science* 75(1), pp. 87–103. doi: 10.1002/ps.5133.

Fudge, J.B. 2023. Diffusion model expands RoseTTAFold's power. *Nature Biotechnology* 41(8), p. 1072. doi: 10.1038/s41587-023-01919-0.

Gahan, L.J., Gould, F. and Heckel, D.G. 2001. Identification of a gene associated with Bt resistance in *Heliothis virescens*. *Science* 293(5531), pp. 857–860. doi: 10.1126/science.1060949.

Galitsky, N., Cody, V., Wojtczak, A., Ghosh, D., Luft, J.R., Pangborn, W. and English, L. 2001. Structure of the insecticidal bacterial  $\delta$ -endotoxin Cry3Bb1 of *Bacillus thuringiensis*. *Acta Crystallographica Section D: Biological Crystallography* 57(8), pp. 1101–1109. doi: 10.1107/S09074444901008186.

Gazit, E., La Rocca, P., Sansom, M.S.P. and Shai, Y. 1998. The structure and organization within the membrane of the helices composing the pore-forming domain of *Bacillus thuringiensis*  $\delta$ -endotoxin are consistent with an “umbrella-like” structure of the pore. *Proceedings of the National Academy of Sciences* 95(21), pp. 12289–12294. doi: 10.1073/pnas.95.21.12289.

Gevorkov, Y. et al. 2019. XGANDALF - Extended gradient descent algorithm for lattice finding. *Acta Crystallographica Section A: Foundations and Advances* 75(5), pp. 694–704. doi: 10.1107/S2053273319010593.

Gómez, I. et al. 2006. Specific epitopes of domains II and III of *Bacillus thuringiensis* Cry1Ab toxin involved in the sequential interaction with cadherin and aminopeptidase-N receptors in *Manduca sexta*. *The Journal of biological chemistry* 281(45), pp. 34032–34039. doi: 10.1074/JBC.M604721200.

Gómez, I., Sánchez, J., Miranda, R., Bravo, A. and Soberón, M. 2002. Cadherin-like receptor binding facilitates proteolytic cleavage of helix  $\alpha$ -1 in domain I and oligomer pre-pore formation of *Bacillus thuringiensis* Cry1Ab toxin. *FEBS Letters* 513(2–3), pp. 242–246. doi: 10.1016/S0014-5793(02)02321-9.

Grochulski, P., Masson, L., Borisova, S., Pusztai-Carey, M., Schwartz, J.L., Brousseau, R. and Cygler, M. 1995. *Bacillus thuringiensis* CryIA(a) insecticidal toxin: Crystal structure and channel formation. *Journal of Molecular Biology* 254(3), pp. 447–464. doi: 10.1006/jmbi.1995.0630.

Grum, V.L., Li, D., MacDonald, R.I. and Mondragón, A. 1999. Structures of two repeats of spectrin suggest models of flexibility. *Cell* 98(4), pp. 523–535. doi: 10.1016/S0092-8674(00)81980-7.

Guo, Q., Ding, L., Gao, Y., Niu, Y. and Dai, X. 2021. Cys183 and Cys258 in Cry49Aa toxin from *Lysinibacillus sphaericus* are essential for toxicity to *Culex quinquefasciatus* larvae. *Archives of Microbiology* 203, pp. 4587–4592. doi: 10.1007/s00203-021-02436-x.

Guo, Q.Y., Hu, X.M., Cai, Q.X., Yan, J.P. and Yuan, Z.M. 2016. Interaction of *Lysinibacillus sphaericus* Cry48Aa/Cry49Aa toxin with midgut brush-border membrane fractions from *Culex quinquefasciatus* larvae. *Insect molecular biology* 25(2), pp. 163–170. doi: 10.1111/imb.12209.

Guo, S. et al. 2009. Crystal structure of *Bacillus thuringiensis* Cry8Ea1: An insecticidal toxin toxic to underground pests, the larvae of *Holotrichia parallela*. 168(2), pp. 259–266. doi: 10.1016/J.JSB.2009.07.004.

Haider, M.Z. and Ellar, D.J. 1989. Mechanism of action of *Bacillus thuringiensis* insecticidal  $\delta$ -endotoxin: interaction with phospholipid vesicles. *BBA - Biomembranes* 978(2), pp. 216–222. doi: 10.1016/0005-2736(89)90118-1.

- Hammel, M., Sfyroera, G., Ricklin, D., Magotti, P., Lambris, J.D. and Geisbrecht, B. V. 2007. A structural basis for complement inhibition by *Staphylococcus aureus*. *Nature Immunology* 8(4), pp. 430–437. doi: 10.1038/ni1450.
- Han, H. et al. 2021. The XBI BioLab for life science experiments at the European XFEL. *Journal of Applied Crystallography* 54(Pt 1), pp. 7–21. doi: 10.1107/S1600576720013989.
- Hazes, B. 1996. The (QxW)<sub>3</sub> domain: A flexible lectin scaffold. *Protein Science* 5(8), pp. 1490–1501. doi: 10.1002/pro.5560050805.
- He, J., Li, T. and Huang, S.Y. 2023. Improvement of cryo-EM maps by simultaneous local and non-local deep learning. *Nature Communications* 14(1), pp. 1–16. doi: 10.1038/s41467-023-39031-1.
- Heater, B.S., Yang, Z., Lee, M.M. and Chan, M.K. 2020. In Vivo Enzyme Entrapment in a Protein Crystal. *Journal of the American Chemical Society* 142(22), pp. 9879–9883. doi: 10.1021/jacs.9b13462.
- Hernández-Martínez, P. et al. 2010. Constitutive activation of the midgut response to *Bacillus thuringiensis* in Bt-resistant *Spodoptera exigua*. *PLoS ONE* 5(9), pp. 1–10. doi: 10.1371/journal.pone.0012795.
- Herrero, S., Gechev, T., Bakker, P.L., Moar, W.J. and de Maagd, R.A. 2005. *Bacillus thuringiensis* Cry1Ca-resistant *Spodoptera exigua* lacks expression of one of four Aminopeptidase N genes. *BMC Genomics* 6:96. doi: 10.1186/1471-2164-6-96.
- Hofte, H. and Whiteley, H.R. 1989. Insecticidal crystal proteins of *Bacillus thuringiensis*. *Microbiological Reviews* 53(2), pp. 242–255. doi: 10.1128/mr.53.2.242-255.1989.
- Hollingsworth, S.A. and Dror, R.O. 2018. Molecular Dynamics Simulation for All. *Neuron* 99(6), pp. 1129–1143. doi: 10.1016/j.neuron.2018.08.011.
- Holm, L. 2020. Using Dali for Protein Structure Comparison. *Methods in Molecular Biology* 2112, pp. 29–42. doi: 10.1007/978-1-0716-0270-6\_3.
- Holmes, K.C. and Monro, R.E. 1965. Studies on the structure of parasporal inclusions from *Bacillus thuringiensis*. *Journal of Molecular Biology* 14(2), pp. 572–581. doi: 10.1016/S0022-2836(65)80205-4.
- Holton, J.M. and Frankel, K.A. 2010. The minimum crystal size needed for a complete diffraction data set. *Acta Crystallographica Section D: Biological Crystallography* 66(4), pp. 393–408. doi: 10.1107/S0907444910007262.
- Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A. and Simmerling, C. 2006. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function and Genetics* 65(3), pp. 712–725. doi: 10.1002/prot.21123.

- Hou, Y., Xie, T., He, L., Tao, L. and Huang, J. 2023. Topological links in predicted protein complex structures reveal limitations of AlphaFold. *Communications Biology* 6(1), pp. 1–10. doi: 10.1038/s42003-023-05489-4.
- Huang, J. et al. 2016. CHARMM36m: An improved force field for folded and intrinsically disordered proteins. *Nature Methods* 14(1), pp. 71–73. doi: 10.1038/nmeth.4067.
- Hui, F., Scheib, U., Hu, Y., Sommer, R.J., Aroian, R. V. and Ghosh, P. 2012. Structure and glycolipid binding properties of the nematocidal protein Cry5B. *Biochemistry* 51(49), pp. 9911–9921. doi: 10.1021/bi301386q.
- Humphrey, W., Dalke, A. and Schulten, K. 1996. VMD: Visual molecular dynamics. *Journal of Molecular Graphics* 14(1), pp. 33–38. doi: 10.1016/0263-7855(96)00018-5.
- Iacovache, I. et al. 2006. A rivet model for channel formation by aerolysin-like pore-forming toxins. *EMBO Journal* 25(3), pp. 457–466. doi: 10.1038/sj.emboj.7600959.
- Iacovache, I., De Carlo, S., Cirauqui, N., Dal Peraro, M., Van Der Goot, F.G. and Zuber, B. 2016. Cryo-EM structure of aerolysin variants reveals a novel protein fold and the pore-formation process. *Nature Communications* 7(May), pp. 1–8. doi: 10.1038/ncomms12062.
- IPPC Secretariat. 2021. Scientific review of the impact of climate change on plant pests. In: *Scientific review of the impact of climate change on plant pests*. IPPC Secretariat. doi: 10.4060/cb4769en.
- Jacques, D.A. and Trehwella, J. 2010. Small-angle scattering for structural biology - Expanding the frontier while avoiding the pitfalls. *Protein Science* 19(4), pp. 642–657. doi: 10.1002/pro.351.
- Jandova, Z., Vargiu, A.V. and Bonvin, A.M.J.J. 2021. Native or Non-Native Protein-Protein Docking Models? Molecular Dynamics to the Rescue. *Journal of Chemical Theory and Computation* 17(9), pp. 5944–5954. doi: 10.1021/acs.jctc.1c00336.
- Jiang, D., Fan, J., Wang, X., Zhao, Y., Huang, B., Liu, J. and Zhang, X.C. 2012. Crystal structure of 1,3Gal43A, an exo- $\beta$ -1,3-galactanase from *Clostridium thermocellum*. *Journal of Structural Biology* 180(3), pp. 447–457. doi: 10.1016/j.jsb.2012.08.005.
- Jing, X., Yuan, Y., Wu, Y., Wu, D., Gong, P. and Gao, M. 2019. Crystal structure of *Bacillus thuringiensis* Cry7Ca1 toxin active against *Locusta migratoria manilensis*. *Protein Science* 28(3), pp. 609–619. doi: 10.1002/pro.3561.
- Jones, D.T. 1999. GenTHREADER: An efficient and reliable protein fold recognition method for genomic sequences. *Journal of Molecular Biology* 287(4), pp. 797–815. doi: 10.1006/jmbi.1999.2583.

- Jones, G.W., Nielsen-Leroux, C., Yang, Y., Yuan, Z., Dumas, V.F., Gomes Monnerat, R. and Berry, C. 2007. A new Cry toxin with a unique two-component dependency from *Bacillus sphaericus*. *The FASEB Journal* 21(14), pp. 4112–4120. doi: 10.1096/fj.07-8913com.
- Jones, G.W., Wirth, M.C., Monnerat, R.G. and Berry, C. 2008. The Cry48Aa-Cry49Aa binary toxin from *Bacillus sphaericus* exhibits highly restricted target specificity. *Environmental Microbiology* 10(9), pp. 2418–2424. doi: 10.1111/j.1462-2920.2008.01667.x.
- Jumper, J. et al. 2021a. Applying and improving AlphaFold at CASP14. *Proteins: Structure, Function and Bioinformatics* 89(12), pp. 1711–1721. doi: 10.1002/prot.26257.
- Jumper, J. et al. 2021b. Highly accurate protein structure prediction with AlphaFold. *Nature* 596(7873), pp. 583–589. doi: 10.1038/s41586-021-03819-2.
- Jurat-Fuentes, J.L. et al. 2011. Reduced levels of membrane-bound alkaline phosphatase are common to lepidopteran strains resistant to Cry toxins from *Bacillus thuringiensis*. *PLoS ONE* 6(3), p. e17606. doi: 10.1371/journal.pone.0017606.
- Jurat-Fuentes, J.L., Gahan, L.J., Gould, F.L., Heckel, D.G. and Adang, M.J. 2004. The HevCaLP protein mediates binding specificity of the Cry1A class of *Bacillus thuringiensis* toxins in *Heliothis virescens*. *Biochemistry* 43(44), pp. 14299–14305. doi: 10.1021/bi048500i.
- Jurat-Fuentes, J.L., Gould, F.L. and Adang, M.J. 2002. 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. *Applied and Environmental Microbiology* 68(11), pp. 5711–5717. doi: 10.1128/AEM.68.11.5711-5717.2002.
- Jurrus, E. et al. 2018. Improvements to the APBS biomolecular solvation software suite. *Protein Science* 27(1), pp. 112–128. doi: 10.1002/pro.3280.
- Kantardjieff, K.A. and Rupp, B. 2003. Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Science* 12(9), pp. 1865–1871. doi: 10.1110/ps.0350503.
- Karaca, E., Prévost, C. and Sacquin-Mora, S. 2022. Modeling the Dynamics of Protein–Protein Interfaces, How and Why? *Molecules* 27(6), p. 1841. doi: 10.3390/molecules27061841.
- Karch, S., Monteny, N., Jullien, J.L., Sinègre, G. and Coz, J. 1990. Control of *Culex pipiens* by *Bacillus sphaericus* and role of nontarget arthropods in its recycling. *Journal of the American Mosquito Control Association* 6(1), pp. 47–54.

- Kelker, M.S. et al. 2014. Structural and biophysical characterization of *Bacillus thuringiensis* insecticidal proteins Cry34Ab1 and Cry35Ab1. *PLoS ONE* 9(11), p. 112555. doi: 10.1371/journal.pone.0112555.
- Keller, M. et al. 1996. Digestion of  $\delta$ -Endotoxin by gut proteases may explain reduced sensitivity of advanced instar larvae of *Spodoptera littoralis* to CryIc. *Insect Biochemistry and Molecular Biology* 26(4), pp. 365–373. doi: 10.1016/0965-1748(95)00102-6.
- Knoška, J. et al. 2020. Ultracompact 3D microfluidics for time-resolved structural biology. *Nature Communications* 11(1), p. 657. doi: 10.1038/s41467-020-14434-6.
- Knowles, B.H. and Ellar, D.J. 1987. Colloid-osmotic lysis is a general feature of the mechanism of action of *Bacillus thuringiensis*  $\delta$ -endotoxins with different insect specificity. *BBA - General Subjects* 924(3), pp. 509–518. doi: 10.1016/0304-4165(87)90167-X.
- Knowles, B.H., Thomas, W.E. and Ellar, D.J. 1984. Lectin-like binding of *Bacillus thuringiensis* var. kurstaki lepidopteran-specific toxin is an initial step in insecticidal action. *FEBS Letters* 168(2), pp. 197–202. doi: 10.1016/0014-5793(84)80245-8.
- Kozakov, D. et al. 2017. The ClusPro web server for protein-protein docking. *Nature Protocols* 12(2), pp. 255–278. doi: 10.1038/nprot.2016.169.
- Kozakov, D., Brenke, R., Comeau, S.R. and Vajda, S. 2006. PIPER: An FFT-based protein docking program with pairwise potentials. *Proteins: Structure, Function and Genetics* 65(2), pp. 392–406. doi: 10.1002/prot.21117.
- Krissinel, E. 2015. Stock-based detection of protein oligomeric states in jsPISA. *Nucleic Acids Research* 43, pp. W314–W319. doi: 10.1093/nar/gkv314.
- Krissinel, E. and Henrick, K. 2007. Inference of Macromolecular Assemblies from Crystalline State. *Journal of Molecular Biology* 372(3), pp. 774–797. doi: 10.1016/j.jmb.2007.05.022.
- Kryshtafovych, A. et al. 2023. New prediction categories in CASP15. *Proteins: Structure, Function and Bioinformatics* 91(12), pp. 1550–1557. doi: 10.1002/prot.26515.
- Kryshtafovych, A., Schwede, T., Topf, M., Fidelis, K. and Mout, J. 2021. Critical assessment of methods of protein structure prediction (CASP)—Round XIV. *Proteins: Structure, Function and Bioinformatics* 89(12), pp. 1607–1617. doi: 10.1002/prot.26237.
- Kufareva, I. and Abagyan, R. 2012. Methods of protein structure comparison. *Methods in Molecular Biology* 857, pp. 231–257. doi: 10.1007/978-1-61779-588-6\_10.

- Kühlbrandt, W. and D'imprima, E. 2021. Current limitations to high-resolution structure determination by single-particle cryoEM. *Quarterly Reviews of Biophysics* 54, pp. 1–15. doi: 10.1017/S0033583521000020.
- Lacey, L.A. 2007. Bacillus thuringiensis serovariety Israelensis and bacillus sphaericus for mosquito control. *Journal of the American Mosquito Control Association* 23(sp2), pp. 133–163. doi: 10.2987/8756-971x(2007)23[133:btsiab]2.0.co;2.
- Lacomel, C.J., Dunstone, M.A. and Spicer, B.A. 2021. Branching out the aerolysin, ETX/MTX-2 and Toxin\_10 family of pore forming proteins. *Journal of Invertebrate Pathology* 186, p. 107570. doi: 10.1016/j.jip.2021.107570.
- Laemmli, U.K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227(5259), pp. 680–685. doi: 10.1038/227680a0.
- Lawrence, M.C. and Colman, P.M. 1993. Shape complementarity at protein/protein interfaces. *Journal of molecular biology* 234(4), pp. 946–950. doi: 10.1006/JMBI.1993.1648.
- Lee, C. Te et al. 2015. The opportunistic marine pathogen *Vibrio parahaemolyticus* becomes virulent by acquiring a plasmid that expresses a deadly toxin. *Proceedings of the National Academy of Sciences* 112(34), pp. 10798–10803. doi: 10.1073/pnas.1503129112.
- Lee, M.K., Curtiss, A., Alcantara, E. and Dean, D.H. 1996. Synergistic effect of the Bacillus thuringiensis toxins CryIAa and CryIAC on the gypsy moth, *Lymantria dispar*. *Applied and Environmental Microbiology* 62(2), pp. 583–586. doi: 10.1128/aem.62.2.583-586.1996.
- Lee, M.K., You, T.H., Gould, F.L. and Dean, D.H. 1999. Identification of residues in domain III of Bacillus thuringiensis Cry1Ac toxin that affect binding and toxicity. *Applied and Environmental Microbiology* 65(10), pp. 4513–4520. doi: 10.1128/aem.65.10.4513-4520.1999.
- Lekakarn, H., Promdonkoy, B. and Boonserm, P. 2015. Interaction of Lysinibacillus sphaericus binary toxin with mosquito larval gut cells: Binding and internalization. *Journal of Invertebrate Pathology* 132, pp. 125–131. doi: 10.1016/j.jip.2015.09.010.
- Lensink, M.F. et al. 2019. Blind prediction of homo- and hetero-protein complexes: The CASP13-CAPRI experiment. *Proteins: Structure, Function and Bioinformatics* 87(12), pp. 1200–1221. doi: 10.1002/prot.25838.
- Lensink, M.F., Velankar, S., Baek, M., Heo, L., Seok, C. and Wodak, S.J. 2018. The challenge of modeling protein assemblies: the CASP12-CAPRI experiment. *Proteins: Structure, Function and Bioinformatics* 86, pp. 257–273. doi: 10.1002/prot.25419.

- Lensink, M.F., Velankar, S. and Wodak, S.J. 2017. Modeling protein–protein and protein–peptide complexes: CAPRI 6th edition. *Proteins: Structure, Function and Bioinformatics* 85(3), pp. 359–377. doi: 10.1002/prot.25215.
- Li, J., Carroll, J. and Ellar, D.J. 1991. Crystal structure of insecticidal  $\delta$ -endotoxin from *Bacillus thuringiensis* at 2.5 Å resolution. 353(6347), pp. 815–821. doi: 10.1038/353815a0.
- Li, J., Wang, L., Kotaka, M., Lee, M.M. and Chan, M.K. 2022. Insights from the Structure of an Active Form of *Bacillus thuringiensis* Cry5B. *Toxins* 14(12), p. 823. doi: 10.3390/toxins14120823.
- Li, R. et al. 2019. The conserved cysteine residues in *Bacillus thuringiensis* Cry1Ac protoxin are not essential for the bipyramidal crystal formation. *Journal of Invertebrate Pathology* 163, pp. 82–85. doi: 10.1016/j.jip.2019.03.009.
- Likitvivatanavong, S., Katzenmeier, G. and Angsuthanasombat, C. 2006. Asn183 in  $\alpha 5$  is essential for oligomerisation and toxicity of the *Bacillus thuringiensis* Cry4Ba toxin. *Archives of Biochemistry and Biophysics* 445(1), pp. 46–55. doi: 10.1016/j.abb.2005.11.007.
- Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J.L., Dror, R.O. and Shaw, D.E. 2010. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Structure, Function and Bioinformatics* 78(8), pp. 1950–1958. doi: 10.1002/prot.22711.
- Lüthy, P. and Ebersold, H.R. 1981. The entomocidal toxins of *Bacillus thuringiensis*. 13(2), pp. 257–283. doi: 10.1016/0163-7258(81)90003-6.
- Lwande, O.W., Obanda, V., Lindström, A., Ahlm, C., Evander, M., Näslund, J. and Bucht, G. 2020. Globe-Trotting *Aedes aegypti* and *Aedes albopictus*: Risk Factors for Arbovirus Pandemics. *Vector-Borne and Zoonotic Diseases* 20(2), pp. 71–81. doi: 10.1089/vbz.2019.2486.
- Ma, G., Rahman, M.M., Grant, W., Schmidt, O. and Asgari, S. 2012. Insect tolerance to the crystal toxins Cry1Ac and Cry2Ab is mediated by the binding of monomeric toxin to lipophorin glycolipids causing oligomerization and sequestration reactions. *Developmental and Comparative Immunology* 37(1), pp. 184–192. doi: 10.1016/j.dci.2011.08.017.
- De Maagd, R.A., Bravo, A. and Crickmore, N. 2001. How *Bacillus thuringiensis* has evolved specific toxins to colonize the insect world. *Trends in Genetics* 17(4), pp. 193–199. doi: 10.1016/S0168-9525(01)02237-5.
- Mancheño, J.M., Tateno, H., Goldstein, I.J., Martínez-Ripoll, M. and Hermoso, J.A. 2005. Structural analysis of the *Laetiporus sulphureus* hemolytic pore-forming lectin in complex with sugars. *Journal of Biological Chemistry* 280(17), pp. 17251–17259. doi: 10.1074/jbc.M413933200.

- Mancuso, A.P. et al. 2019. The single particles, clusters and biomolecules and serial femtosecond crystallography instrument of the european XFEL: Initial installation. *Journal of Synchrotron Radiation* 26(3), pp. 660–676. doi: 10.1107/S1600577519003308.
- Mariani, V. et al. 2016. OnDA: Online data analysis and feedback for serial X-ray imaging. *Journal of Applied Crystallography* 49(3), pp. 1073–1080. doi: 10.1107/S1600576716007469.
- Marze, N.A., Roy Burman, S.S., Sheffler, W. and Gray, J.J. 2018. Efficient flexible backbone protein-protein docking for challenging targets. *Bioinformatics* 34(20), pp. 3461–3469. doi: 10.1093/bioinformatics/bty355.
- Masson, L., Lu, Y.J., Mazza, A., Brousseau, R. and Adang, M.J. 1995. The CryIA(c) receptor purified from *Munduca sexta* displays multiple specificities. *Journal of Biological Chemistry* 270(35), pp. 20309–20315. doi: 10.1074/jbc.270.35.20309.
- McCammon, J.A., Gelin, B.R. and Karplus, M. 1977. Dynamics of folded proteins. *Nature* 267(5612), pp. 585–590. doi: 10.1038/267585a0.
- McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C. and Read, R.J. 2007. Phaser crystallographic software. *Journal of Applied Crystallography* 40(4), pp. 658–674. doi: 10.1107/S0021889807021206.
- De Melo, J.V. et al. 2009. Cytopathological effects of *Bacillus sphaericus* Cry48Aa/Cry49Aa toxin on binary toxin-susceptible and -resistant *Culex quinquefasciatus* larvae. *Applied and Environmental Microbiology* 75(14), pp. 4782–4789. doi: 10.1128/AEM.00811-09.
- Meng, E.C., Goddard, T.D., Pettersen, E.F., Couch, G.S., Pearson, Z.J., Morris, J.H. and Ferrin, T.E. 2023. UCSF ChimeraX: Tools for structure building and analysis. *Protein Science* 32(11), p. e4792. doi: 10.1002/pro.4792.
- Mészáros, B. et al. 2023. Recent breakthroughs in computational structural biology harnessing the power of sequences and structures. *Current Opinion in Structural Biology* 80, p. 102608. doi: 10.1016/j.sbi.2023.102608.
- Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S. and Steinegger, M. 2022. ColabFold: making protein folding accessible to all. *Nature Methods* 19(6), pp. 679–682. doi: 10.1038/s41592-022-01488-1.
- Moar, W.J., Berry, C. and Narva, K.E. 2017. The structure/function of new insecticidal proteins and regulatory challenges for commercialization. *Journal of Invertebrate Pathology* 142, pp. 1–4. doi: 10.1016/j.jip.2017.02.001.
- Morin, S. et al. 2003. Three cadherin alleles associated with resistance to *Bacillus thuringiensis* in pink bollworm. *Proceedings of the National Academy of Sciences* 100(9), pp. 5004–5009. doi: 10.1073/pnas.0831036100.

- Morse, R.J., Yamamoto, T. and Stroud, R.M. 2001. Structure of Cry2Aa suggests an unexpected receptor binding epitope. *9*(5), pp. 409–417. doi: 10.1016/S0969-2126(01)00601-3.
- Mortuza, S.M., Zheng, W., Zhang, C., Li, Y., Pearce, R. and Zhang, Y. 2021. Improving fragment-based ab initio protein structure assembly using low-accuracy contact-map predictions. *Nature Communications* 12(1), pp. 1–12. doi: 10.1038/s41467-021-25316-w.
- Moult, J., Pedersen, J.T., Judson, R. and Fidelis, K. 1995. A large-scale experiment to assess protein structure prediction methods. *Proteins: Structure, Function, and Bioinformatics* 23(3), pp. ii–iv. doi: 10.1002/prot.340230303.
- Muñoz-Garay, C. et al. 2009. Characterization of the mechanism of action of the genetically modified Cry1AbMod toxin that is active against Cry1Ab-resistant insects. *Biochimica et Biophysica Acta - Biomembranes* 1788(10), pp. 2229–2237. doi: 10.1016/j.bbamem.2009.06.014.
- Muñoz-Garay, C., Sánchez, J., Darszon, A., De Maagd, R.A., Bakker, P., Soberón, M. and Bravo, A. 2006. Permeability changes of *Manduca sexta* midgut brush border membranes induced by oligomeric structures of different cry toxins. *Journal of Membrane Biology* 212(1), pp. 61–68. doi: 10.1007/s00232-006-0003-8.
- Murshudov, G.N. et al. 2011. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D: Biological Crystallography* 67(4), pp. 355–367. doi: 10.1107/S0907444911001314.
- Nakamura, T., Tonozuka, T., Ide, A., Yuzawa, T., Oguma, K. and Nishikawa, A. 2008. Sugar-binding Sites of the HA1 Subcomponent of Clostridium botulinum Type C Progenitor Toxin. *Journal of Molecular Biology* 376(3), pp. 854–867. doi: 10.1016/j.jmb.2007.12.031.
- Nannenga, B.L. and Gonen, T. 2019. The cryo-EM method microcrystal electron diffraction (MicroED). *Nature Methods* 16(5), pp. 369–379. doi: 10.1038/s41592-019-0395-x.
- do Nascimento, N.A. et al. 2017. N-glycosylation influences the catalytic activity of mosquito  $\alpha$ -glucosidases associated with susceptibility or refractoriness to *Lysinibacillus sphaericus*. *Insect Biochemistry and Molecular Biology* 81, pp. 62–71. doi: 10.1016/j.ibmb.2016.12.009.
- Nielsen-Leroux, C., Pasquier, F., Charles, J.F., Sinégre, G., Gaven, B. and Pasteur, N. 1997. Resistance to *Bacillus sphaericus* involves different mechanisms in *Culex pipiens* (Diptera: Culicidae) larvae. *Journal of medical entomology* 34(3), pp. 321–327. doi: 10.1093/JMEDENT/34.3.321.

Nielsen-Leroux, C., Pasteur, N., Prêtre, J., Charles, J.F., Ben Sheikh, H. and Chevillon, C. 2002. High resistance to *Bacillus sphaericus* binary toxin in *Culex pipiens* (Diptera: Culicidae): The complex situation of west Mediterranean countries. *Journal of Medical Entomology* 39(5), pp. 729–735. doi: 10.1603/0022-2585-39.5.729.

Nielsen-Leroux, C. and Charles, J. -F. 1992. Binding of *Bacillus sphaericus* binary toxin to a specific receptor on midgut brush-border membranes from mosquito larvae. *European Journal of Biochemistry* 210(2), pp. 585–590. doi: 10.1111/j.1432-1033.1992.tb17458.x.

Nielsen-Leroux, C., Charles, J.F., Thiéry, I. and Georghiou, G.P. 1995. Resistance in a Laboratory Population of *Culex quinquefasciatus* (Diptera: Culicidae) to *Bacillus Sphaericus* Binary Toxin is Due to a Change in the Receptor on Midgut Brush-Border Membranes. *European Journal of Biochemistry* 228(1), pp. 206–210. doi: 10.1111/j.1432-1033.1995.tb20251.x.

Núñez-Ramírez, R., Huesa, J., Bel, Y., Ferré, J., Casino, P. and Arias-Palomo, E. 2020. Molecular architecture and activation of the insecticidal protein Vip3Aa from *Bacillus thuringiensis*. *Nature Communications* 2020 11:1 11(1), pp. 1–9. doi: 10.1038/s41467-020-17758-5.

Oberthuer, D. et al. 2017. Double-flow focused liquid injector for efficient serial femtosecond crystallography. *Scientific Reports* 7(1), p. 44628. doi: 10.1038/srep44628.

Oei, C., Hindley, J. and Berry, C. 1992. Binding of purified *Bacillus sphaericus* binary toxin and its deletion derivatives to *Culex quinquefasciatus* gut: Elucidation of functional binding domains. *Journal of General Microbiology* 138(7), pp. 1515–1526. doi: 10.1099/00221287-138-7-1515.

Oliveira, C.M.F., Silva-Filha, M.H., Nielsen-Leroux, C., Pei, G., Yuan, Z. and Regis, L. 2004. Inheritance and Mechanism of Resistance to *Bacillus sphaericus* in *Culex quinquefasciatus* (Diptera: Culicidae) from China and Brazil. *Journal of Medical Entomology* 41(1), pp. 58–64. doi: 10.1603/0022-2585-41.1.58.

Opota, O., Charles, J.F., Warot, S., Pauron, D. and Darboux, I. 2008. Identification and characterization of the receptor for the *Bacillus sphaericus* binary toxin in the malaria vector mosquito, *Anopheles gambiae*. *Comparative Biochemistry and Physiology - B Biochemistry and Molecular Biology* 149(3), pp. 419–427. doi: 10.1016/j.cbpb.2007.11.002.

Opota, O., Gauthier, N.C., Doye, A., Berry, C., Gounon, P., Lemichez, E. and Pauron, D. 2011. *Bacillus sphaericus* binary toxin elicits host cell autophagy as a response to intoxication. *PLoS ONE* 6(2), p. e14682. doi: 10.1371/journal.pone.0014682.

Oppert, B., Kramer, K.J., Beeman, R.W., Johnson, D. and McGaughey, W.H. 1997. Proteinase-mediated insect resistance to *Bacillus thuringiensis* toxins. *Journal of Biological Chemistry* 272(38), pp. 23473–23476. doi: 10.1074/jbc.272.38.23473.

Ounjai, P., Unger, V.M., Sigworth, F.J. and Angsuthanasombat, C. 2007. Two conformational states of the membrane-associated *Bacillus thuringiensis* Cry4Ba  $\delta$ -endotoxin complex revealed by electron crystallography: Implications for toxin-pore formation. *Biochemical and Biophysical Research Communications* 361(4), pp. 890–895. doi: 10.1016/j.bbrc.2007.07.086.

Pacheco, S. et al. 2009. Domain II loop 3 of *Bacillus thuringiensis* Cry1Ab toxin is involved in a “Ping Pong” binding mechanism with *Manduca sexta* aminopeptidase-N and cadherin receptors. *Journal of Biological Chemistry* 284(47), pp. 32750–32757. doi: 10.1074/jbc.M109.024968.

Pak, M.A. et al. 2023. Using AlphaFold to predict the impact of single mutations on protein stability and function. *PLoS ONE* 18(3), p. e0282689. doi: 10.1371/journal.pone.0282689.

Palma, L., Muñoz, D., Berry, C., Murillo, J., Caballero, P. and Caballero, P. 2014. *Bacillus thuringiensis* toxins: An overview of their biocidal activity. *Toxins* 6(12), pp. 3296–3325. doi: 10.3390/toxins6123296.

Pan, A.C., Jacobson, D., Yatsenko, K., Sritharan, D., Weinreich, T.M. and Shaw, D.E. 2019. Atomic-level characterization of protein–protein association. *Proceedings of the National Academy of Sciences* 116(10), pp. 4244–4249. doi: 10.1073/pnas.1815431116.

Pandey, S. et al. 2020. Time-resolved serial femtosecond crystallography at the European XFEL. *Nature Methods* 17(1), pp. 73–78. doi: 10.1038/s41592-019-0628-z.

Panneerselvam, S., Mishra, R., Berry, C., Crickmore, N. and Bonning, B.C. 2022. BPPRC database: A web-based tool to access and analyse bacterial pesticidal proteins. *Database* 2022, pp. 1–6. doi: 10.1093/database/baac022.

Pardo-López, L., Gómez, I., Rausell, C., Sánchez, J., Soberón, M. and Bravo, A. 2006. Structural changes of the Cry1Ac oligomeric pre-pore from *Bacillus thuringiensis* induced by N-acetylgalactosamine facilitates toxin membrane insertion. *Biochemistry* 45(34), pp. 10329–10336. doi: 10.1021/bi060297z.

Pardo-López, L., Soberón, M. and Bravo, A. 2013. *Bacillus thuringiensis* insecticidal three-domain Cry toxins: mode of action, insect resistance and consequences for crop protection. *FEMS Microbiology Reviews* 37(1), pp. 3–22. doi: 10.1111/J.1574-6976.2012.00341.X.

Park, H.W., Bideshi, D.K., Johnson, J.J. and Federici, B.A. 1999. Differential enhancement of Cry2A versus Cry11A yields in *Bacillus thuringiensis* by use of the cry3A STAB mRNA sequence. *FEMS microbiology letters* 181(2), pp. 319–327. doi: 10.1111/J.1574-6968.1999.TB08862.X.

Park, Y., Hua, G., Taylor, M.D. and Adang, M.J. 2014. A coleopteran cadherin fragment synergizes toxicity of *Bacillus thuringiensis* toxins Cry3Aa, Cry3Bb, and Cry8Ca against lesser mealworm, *Alphitobius diaperinus* (Coleoptera: Tenebrionidae). *Journal of Invertebrate Pathology* 123, pp. 1–5. doi: 10.1016/j.jip.2014.08.008.

Parker, M.W., Buckley, J.T., Postma, J.P.M., Tucker, A.D., Leonard, K., Pattus, F. and Tsernoglou, D. 1994. Structure of the *Aeromonas* toxin proaerolysin in its water-soluble and membrane-channel states. *Nature* 367(6460), pp. 292–295. doi: 10.1038/367292a0.

Pathak, V.M. et al. 2022. Current status of pesticide effects on environment, human health and it's eco-friendly management as bioremediation: A comprehensive review. *Frontiers in Microbiology* 13, p. 962619. doi: 10.3389/fmicb.2022.962619.

Peña-Cardena, A., Grande, R., Sánchez, J., Tabashnik, B.E., Bravo, A., Soberón, M. and Gómez, I. 2018. The C-terminal protoxin region of *Bacillus thuringiensis* Cry1Ab toxin has a functional role in binding to GPI-anchored receptors in the insect midgut. *Journal of Biological Chemistry* 293(52), pp. 20263–20272. doi: 10.1074/jbc.RA118.005101.

Peng, D. et al. 2011. Single cysteine substitution in *Bacillus thuringiensis* Cry7Ba1 improves the crystal solubility and produces toxicity to *Plutella xylostella* larvae. *Environmental Microbiology* 13(10), pp. 2820–2831. doi: 10.1111/j.1462-2920.2011.02557.x.

Pérez, C., Muñoz-Garay, C., Portugal, L.C., Sánchez, J., Gill, S.S., Soberón, M. and Bravo, A. 2007. *Bacillus thuringiensis* ssp. *israelensis* Cyt1Aa enhances activity of Cry11Aa toxin by facilitating the formation of a pre-pore oligomeric structure. 9(12), pp. 2931–2937. doi: 10.1111/J.1462-5822.2007.01007.X.

Pierce, B.G., Wiehe, K., Hwang, H., Kim, B.H., Vreven, T. and Weng, Z. 2014. ZDOCK server: Interactive docking prediction of protein-protein complexes and symmetric multimers. *Bioinformatics* 30(12), pp. 1771–1773. doi: 10.1093/bioinformatics/btu097.

Pigott, C.R. and Ellar, D.J. 2007. Role of Receptors in *Bacillus thuringiensis* Crystal Toxin Activity. *Microbiology and Molecular Biology Reviews* 71(2), pp. 255–281.

Potterton, L. et al. 2018. CCP4i2: The new graphical user interface to the CCP4 program suite. *Acta Crystallographica Section D: Structural Biology* 74(2), pp. 68–84. doi: 10.1107/S2059798317016035.

- Priest, F.G., Ebdrup, L., Zahner, V. and Carter, P.E. 1997. Distribution and characterization of mosquitocidal toxin genes in some strains of *Bacillus sphaericus*. *Applied and Environmental Microbiology* 63(4), pp. 1195–1198. doi: 10.1128/aem.63.4.1195-1198.1997.
- Pulido, D., Briggs, D.C., Hua, J. and Hohenester, E. 2017. Crystallographic analysis of the laminin  $\beta 2$  short arm reveals how the LF domain is inserted into a regular array of LE domains. *Matrix Biology* 57–58, pp. 204–212. doi: 10.1016/j.matbio.2016.06.006.
- Puntheeranurak, T., Stroh, C., Zhu, R., Angsuthanasombat, C. and Hinterdorfer, P. 2005. Structure and distribution of the *Bacillus thuringiensis* Cry4Ba toxin in lipid membranes. In: *Ultramicroscopy*. Ultramicroscopy, pp. 115–124. doi: 10.1016/j.ultramic.2005.06.026.
- Radom, F., Plückthun, A. and Paci, E. 2018. Assessment of ab initio models of protein complexes by molecular dynamics. *PLoS Computational Biology* 14(6), p. e1006182. doi: 10.1371/journal.pcbi.1006182.
- Rajagopal, R., Sivakumar, S., Agrawal, N., Malhotra, P. and Bhatnagar, R.K. 2002. Silencing of midgut aminopeptidase N of *Spodoptera litura* by double-stranded RNA establishes its role as *Bacillus thuringiensis* toxin receptor. *Journal of Biological Chemistry* 277(49), pp. 46849–46851. doi: 10.1074/jbc.C200523200.
- Rajamohan, F., Hussain, S.R.A., Cotrill, J.A., Gould, F. and Dean, D.H. 1996. Mutations at domain II, loop 3, of *Bacillus thuringiensis* CryIAa and CryIAb  $\delta$ -endotoxins suggest loop 3 is involved in initial binding to lepidopteran midguts. *Journal of Biological Chemistry* 271(41), pp. 25220–25226. doi: 10.1074/jbc.271.41.25220.
- Ramírez-Aportela, E., López-Blanco, J.R. and Chacón, P. 2016. FRODOCK 2.0: Fast protein-protein docking server. *Bioinformatics* 32(15), pp. 2386–2388. doi: 10.1093/bioinformatics/btw141.
- Rausell, C., García-Robles, I., Sánchez, J., Muñoz-Garay, C., Martínez-Ramírez, A.C., Real, M.D. and Bravo, A. 2004. Role of toxin activation on binding and pore formation activity of the *Bacillus thuringiensis* Cry3 toxins in membranes of *Leptinotarsa decemlineata* (Say). 1660(1–2), pp. 99–105.
- Regis, L., Silva-Filha, M.H., Nielsen-LeRoux, C. and Charles, J.F. 2001. Bacteriological larvicides of dipteran disease vectors. *Trends in Parasitology* 17(8), pp. 377–380. doi: 10.1016/S1471-4922(01)01953-5.
- Rohou, A. and Grigorieff, N. 2015. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *Journal of Structural Biology* 192(2), pp. 216–221. doi: 10.1016/j.jsb.2015.08.008.

- Romão, T.P., de-Melo-Neto, O.P. and Silva-Filha, M.H.N.L. 2011. The N-terminal third of the BinB subunit from the *Bacillus sphaericus* binary toxin is sufficient for its interaction with midgut receptors in *Culex quinquefasciatus*. *FEMS Microbiology Letters* 321(2), pp. 167–174. doi: 10.1111/j.1574-6968.2011.02325.x.
- Roy, S., Ghosh, P., Bandyopadhyay, A. and Basu, S. 2022. Capturing a Crucial ‘Disorder-to-Order Transition’ at the Heart of the Coronavirus Molecular Pathology—Triggered by Highly Persistent, Interchangeable Salt-Bridges. *Vaccines* 10(2), p. 301. doi: 10.3390/vaccines10020301.
- Rukmini, V., Reddy, C.Y. and Venkateswerlu, G. 2000. *Bacillus thuringiensis* crystal  $\delta$ -endotoxin: Role of proteases in the conversion of protoxin to toxin. *Biochimie* 82(2), pp. 109–116. doi: 10.1016/S0300-9084(00)00355-2.
- Sakano, T., Mahamood, M.I., Yamashita, T. and Fujitani, H. 2016. Molecular dynamics analysis to evaluate docking pose prediction. *Biophysics and physicobiology* 13(0), pp. 181–194. doi: 10.2142/biophysico.13.0\_181.
- Sangadala, S., Azadi, P., Carlson, R. and Adang, M.J. 2001. Carbohydrate analyses of *manduca sexta* aminopeptidase n, co-purifying neutral lipids and their functional interactions with *Bacillus thuringiensis* cry1ac toxin. *Insect Biochemistry and Molecular Biology* 32(1), pp. 97–107. doi: 10.1016/S0965-1748(01)00086-8.
- Sawaya, M.R. et al. 2014. Protein crystal structure obtained at 2.9 Å resolution from injecting bacterial cells into an X-ray free-electron laser beam. *Proceedings of the National Academy of Sciences* 111(35), pp. 12769–12774. doi: 10.1073/pnas.1413456111.
- Schnepf, E. et al. 1998. *Bacillus thuringiensis* and Its Pesticidal Crystal Proteins. *Microbiology and Molecular Biology Reviews* 62(3), pp. 775–806. doi: 10.1128/mnbr.62.3.775-806.1998.
- Schnepf, H.E. et al. 2005. Characterization of Cry34/Cry35 binary insecticidal proteins from diverse *Bacillus thuringiensis* strain collections. *Applied and Environmental Microbiology* 71(4), pp. 1765–1774. doi: 10.1128/AEM.71.4.1765-1774.2005.
- Schwartz, J.L. et al. 2001. Permeabilization of model lipid membranes by *Bacillus sphaericus* mosquito-cidal binary toxin and its individual components. *Journal of Membrane Biology* 184(2), pp. 171–183. doi: 10.1007/s00232-001-0086-1.
- Schwartz, J.L., Lu, Y.J., Söhnlein, P., Brousseau, R., Laprade, R., Masson, L. and Adang, M.J. 1997. Ion channels formed in planar lipid bilayers by *Bacillus thuringiensis* toxins in the presence of *Manduca sexta* midgut receptors. *FEBS Letters* 412(2), pp. 270–276. doi: 10.1016/S0014-5793(97)00801-6.

- Sharma, M., Aswal, V.K., Kumar, V. and Chidambaram, R. 2020. Small-angle neutron scattering studies suggest the mechanism of BinAB protein internalization. *IUCrJ* 7(2), pp. 166–172. doi: 10.1107/S2052252519017159.
- Sharma, M., Gupta, G.D. and Kumar, V. 2018. Receptor protein of *Lysinibacillus sphaericus* mosquito-larvicidal toxin displays amyloamylase activity. *Insect Biochemistry and Molecular Biology* 93, pp. 37–46. doi: 10.1016/j.ibmb.2017.12.002.
- Sharma, M. and Kumar, V. 2019. Crystal structure of BinAB toxin receptor (Cqm1) protein and molecular dynamics simulations reveal the role of unique Ca(II) ion. *International Journal of Biological Macromolecules* 140, pp. 1315–1325. doi: 10.1016/J.IJBIOMAC.2019.08.126.
- Sharma, P., Nain, V., Lakhanpaul, S. and Kumar, P.A. 2010. Synergistic activity between *Bacillus thuringiensis* Cry1Ab and Cry1Ac toxins against maize stem borer (*Chilo partellus* Swinhoe). *Letters in Applied Microbiology* 51(1), pp. 42–47. doi: 10.1111/j.1472-765X.2010.02856.x.
- Shrestha, R.B., Jakka, S.R.K. and Gassmann, A.J. 2018. Response of Cry3Bb1-resistant western corn rootworm (Coleoptera: Chrysomelidae) to Bt maize and soil insecticide. *Journal of Applied Entomology* 142(10), pp. 937–946. doi: 10.1111/jen.12505.
- Shu, C. et al. 2009. Characterization of two novel cry8 genes from *Bacillus thuringiensis* strain BT185. *Current Microbiology* 58(4), pp. 389–392. doi: 10.1007/s00284-008-9338-y.
- Shu, C. et al. 2020. Characterization of Two Novel *Bacillus thuringiensis* Cry8 Toxins Reveal Differential Specificity of Protoxins or Activated Toxins against Chrysomeloidea Coleopteran Superfamily. *Toxins* 2020, Vol. 12, Page 642 12(10), p. 642. doi: 10.3390/TOXINS12100642.
- Shu, C., Liu, R., Wang, R., Zhang, J., Feng, S., Huang, D. and Song, F. 2007. Improving toxicity of *Bacillus thuringiensis* strain contains the cry8Ca gene specific to *Anomala corpulenta* larvae. *Current Microbiology* 55(6), pp. 492–496. doi: 10.1007/s00284-007-9018-3.
- Silva-Filha, M.H.N.L. et al. 2021. Bacterial toxins active against mosquitoes: Mode of action and resistance. *Toxins* 13(8), p. 523. doi: 10.3390/toxins13080523.
- Silva-Filha, M.H.N.L., Berry, C. and Regis, L. 2014. *Lysinibacillus sphaericus*: Toxins and Mode of Action, Applications for Mosquito Control and Resistance Management. In: *Advances in Insect Physiology*. pp. 89–176.

Silva-Filha, M.H.N.L., Nielsen-Leroux, C. and Jean-François, C. 1999. Identification of the receptor for *Bacillus sphaericus* crystal toxin in the brush border membrane of the mosquito *Culex pipiens* (Diptera: Culicidae). *Insect Biochemistry and Molecular Biology* 29(8), pp. 711–721. doi: 10.1016/S0965-1748(99)00047-8.

Silva-Filha, M.H.N.L. and Peixoto, C.A. 2003. Immunocytochemical localization of the *Bacillus sphaericus* binary toxin components in *Culex quinquefasciatus* (Diptera: Culicidae) larvae midgut. *Pesticide Biochemistry and Physiology* 77(3), pp. 138–146. doi: 10.1016/j.pestbp.2003.07.002.

Silva-Filha, M.H.N.L., Regis, L., Nielsen-Leroux, C. and Charles, J.F. 1995. Low-level resistance to *Bacillus sphaericus* in a field-treated population of *Culex quinquefasciatus* (Diptera: Culicidae). *Journal of Economic Entomology* 88(3), pp. 525–530. doi: 10.1093/jee/88.3.525.

Singkhamanan, K., Promdonkoy, B., Chaisri, U. and Boonserm, P. 2010. Identification of amino acids required for receptor binding and toxicity of the *Bacillus sphaericus* binary toxin. *FEMS Microbiology Letters* 303(1), pp. 84–91. doi: 10.1111/j.1574-6968.2009.01865.x.

Singkhamanan, K., Promdonkoy, B., Sriksirin, T. and Boonserm, P. 2013. Amino acid residues in the N-terminal region of the BinB subunit of *Lysinibacillus sphaericus* binary toxin play a critical role during receptor binding and membrane insertion. *Journal of Invertebrate Pathology* 114(1), pp. 65–70. doi: 10.1016/j.jip.2013.05.008.

Sinka, M.E., Pironon, S., Massey, N.C., Longbottom, J., Hemingway, J., Moyes, C.L. and Willis, K.J. 2020. A new malaria vector in Africa: Predicting the expansion range of *Anopheles stephensi* and identifying the urban populations at risk. *Proceedings of the National Academy of Sciences* 117(40), pp. 24900–24908. doi: 10.1073/pnas.2003976117.

Skendžić, S., Zovko, M., Živković, I.P., Lešić, V. and Lemić, D. 2021. The impact of climate change on agricultural insect pests. *Insects* 12(5), p. 440. doi: 10.3390/insects12050440.

Smedley, D.P. and Ellar, D.J. 1996. Mutagenesis of three surface-exposed loops of a *Bacillus thuringiensis* insecticidal toxin reveals residues important for toxicity, receptor recognition and possibly membrane insertion. *Microbiology* 142(7), pp. 1617–1624. doi: 10.1099/13500872-142-7-1617.

Smith, C.A. and Kortemme, T. 2008. Backrub-Like Backbone Simulation Recapitulates Natural Protein Conformational Variability and Improves Mutant Side-Chain Prediction. *Journal of Molecular Biology* 380(4), pp. 742–756. doi: 10.1016/j.jmb.2008.05.023.

Soberón, M., Gill, S.S. and Bravo, A. 2009. Signaling versus punching hole: How do *Bacillus thuringiensis* toxins kill insect midgut cells? *Cellular and Molecular Life Sciences* 66(8), pp. 1337–1349. doi: 10.1007/s00018-008-8330-9.

- Soberón, M., Pardo-López, L., López, I., Gómez, I., Tabashnik, B.E. and Bravo, A. 2007. Engineering modified Bt toxins to counter insect resistance. *Science* 318(5856), pp. 1640–1642. doi: 10.1126/science.1146453.
- Srisucharitpanit, K., Yao, M., Promdonkoy, B., Chimnaronk, S., Tanaka, I. and Boonserm, P. 2014. Crystal structure of BinB: A receptor binding component of the binary toxin from *Lysinibacillus sphaericus*. *Proteins: Structure, Function and Bioinformatics* 82(10), pp. 2703–2712. doi: 10.1002/prot.24636.
- Stevenson, H.P. et al. 2016. Transmission electron microscopy for the evaluation and optimization of crystal growth. *Acta Crystallographica Section D: Structural Biology* 72(5), pp. 603–615. doi: 10.1107/S2059798316001546.
- Stillinger, F.H. and Rahman, A. 1974. Improved simulation of liquid water by molecular dynamics. *The Journal of Chemical Physics* 60(4), pp. 1528–1532. doi: 10.1063/1.1681229.
- Sugase, K., Dyson, H.J. and Wright, P.E. 2007. Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* 447(7147), pp. 1021–1025. doi: 10.1038/nature05858.
- Sun, Q., Heater, B.S., Li, T.L., Ye, W., Guo, Z. and Chan, M.K. 2022. Cry3Aa\*SpyCatcher Fusion Crystals Produced in Bacteria as Scaffolds for Multienzyme Coimmobilization. *Bioconjugate Chemistry* 33(2), pp. 386–396. doi: 10.1021/acs.bioconjchem.2c00003.
- Szczesny, P., Iacovache, I., Muszewska, A., Ginalski, K., van der Goot, F.G. and Grynberg, M. 2011. Extending the aerolysin family: From bacteria to vertebrates. *PLoS ONE* 6(6), p. e20349. doi: 10.1371/journal.pone.0020349.
- Tangsongcharoen, C., Chomanee, N., Promdonkoy, B. and Boonserm, P. 2015. *Lysinibacillus sphaericus* binary toxin induces apoptosis in susceptible *Culex quinquefasciatus* larvae. *Journal of Invertebrate Pathology* 128, pp. 57–63. doi: 10.1016/j.jip.2015.04.008.
- Tangsongcharoen, C., Jupatanakul, N., Promdonkoy, B., Dimopoulos, G. and Boonserm, P. 2017. Molecular analysis of *Culex quinquefasciatus* larvae responses to *Lysinibacillus sphaericus* bin toxin. *PLoS ONE* 12(4), p. e0175473. doi: 10.1371/journal.pone.0175473.
- Tenboer, J. et al. 2014. Time-resolved serial crystallography captures high-resolution intermediates of photoactive yellow protein. *Science* 346(6214), pp. 1242–1246. doi: 10.1126/science.1259357.
- Terwilliger, T.C. et al. 2007. Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallographica Section D: Biological Crystallography* 64(1), pp. 61–69. doi: 10.1107/S090744490705024X.

- Tetreau, G. et al. 2020. Serial femtosecond crystallography on in vivo-grown crystals drives elucidation of mosquitocidal Cyt1Aa bioactivation cascade. *Nature Communications* 11(1), p. 1153. doi: 10.1038/s41467-020-14894-w.
- Tetreau, G. 2021a. Can (We make) bacillus thuringiensis crystallize more than its toxins? *Toxins* 13(7), p. 441. doi: 10.3390/toxins13070441.
- Tetreau, G. 2021b. How does *Bacillus thuringiensis* crystallize such a large diversity of toxins? *Toxins* 13(7), p. 443. doi: 10.3390/toxins13070443.
- Tetreau, G. et al. 2022. De novo determination of mosquitocidal Cry11Aa and Cry11Ba structures from naturally-occurring nanocrystals. *Nature Comm* 13, p. 4376. doi: 10.1101/2021.12.15.472578.
- Thomson, M.C. and Stanberry, L.R. 2022. Climate Change and Vectorborne Diseases. *New England Journal of Medicine* 387(21), pp. 1969–1978. doi: 10.1056/nejmra2200092.
- Tian, C. et al. 2020. Ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *Journal of Chemical Theory and Computation* 16(1), pp. 528–552. doi: 10.1021/acs.jctc.9b00591.
- Tiewsiri, K. and Wang, P. 2011. Differential alteration of two aminopeptidases N associated with resistance to *Bacillus thuringiensis* toxin Cry1Ac in cabbage looper. *Proceedings of the National Academy of Sciences* 108(34), pp. 14037–14042. doi: 10.1073/pnas.1102555108.
- Tigue, N.J., Jacoby, J. and Ellar, D.J. 2001. The  $\alpha$ -Helix 4 Residue, Asn135, Is Involved in the Oligomerization of Cry1Ac1 and Cry1Ab5 *Bacillus thuringiensis* Toxins. *Applied and Environmental Microbiology* 67(12), pp. 5715–5720. doi: 10.1128/AEM.67.12.5715-5720.2001.
- Tuntitippawan, T., Boonserm, P., Katzenmeier, G. and Angsuthanasombat, C. 2005. Targeted mutagenesis of loop residues in the receptor-binding domain of the *Bacillus thuringiensis* Cry4Ba toxin affects larvicidal activity. *FEMS Microbiology Letters* 242(2), pp. 325–332. doi: 10.1016/j.femsle.2004.11.026.
- Tuukkanen, A.T., Kleywegt, G.J. and Svergun, D.I. 2016. Resolution of ab initio shapes determined from small-angle scattering. *International Union of Crystallography* 3(6), pp. 440–447. doi: 10.1107/S2052252516016018.
- Vadlamudi, R.K., Weber, E., Ji, I., Ji, T.H. and Bulla, L.A. 1995. Cloning and expression of a receptor for an insecticidal toxin of *Bacillus thuringiensis*. *Journal of Biological Chemistry* 270(10), pp. 5490–5494. doi: 10.1074/jbc.270.10.5490.
- Vakser, I.A. 2014. Protein-protein docking: From interaction to interactome. *Biophysical Journal* 107(8), pp. 1785–1793. doi: 10.1016/j.bpj.2014.08.033.

- Varadi, M. et al. 2022. AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research* 50(D1), pp. D439–D444. doi: 10.1093/nar/gkab1061.
- Vié, V. et al. 2001. Lipid-induced pore formation of the *Bacillus thuringiensis* Cry1Aa insecticidal toxin. *Journal of Membrane Biology* 180(3), pp. 195–203. doi: 10.1007/s002320010070.
- De Vries, S.J., Schindler, C.E.M., Chauvot De Beauchêne, I. and Zacharias, M. 2015. A web interface for easy flexible protein-protein docking with ATTRACT. *Biophysical Journal* 108(3), pp. 462–465. doi: 10.1016/j.bpj.2014.12.015.
- Wang, C., Schueler-Furman, O. and Baker, D. 2005. Improved side-chain modeling for protein–protein docking. *Protein Science* 14(5), pp. 1328–1339. doi: 10.1110/ps.041222905.
- Waterhouse, A. et al. 2018. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Research* 46(W1), pp. W296–W303. doi: 10.1093/nar/gky427.
- Watson, J.L. et al. 2023. De novo design of protein structure and function with RFdiffusion. *Nature* 620(7976), pp. 1089–1100. doi: 10.1038/s41586-023-06415-8.
- White, T.A. et al. 2016. Recent developments in CrystFEL. *Journal of Applied Crystallography* 49(2), pp. 680–689. doi: 10.1107/S1600576716004751.
- White, T.A., Kirian, R.A., Martin, A. V., Aquila, A., Nass, K., Barty, A. and Chapman, H.N. 2012. CrystFEL: A software suite for snapshot serial crystallography. *Journal of Applied Crystallography* 45(2), pp. 335–341. doi: 10.1107/S0021889812002312.
- Wiedorn, M.O. et al. 2018a. Megahertz serial crystallography. *Nature Communications* 9(1), p. 4025. doi: 10.1038/s41467-018-06156-7.
- Wiedorn, M.O. et al. 2018b. Rapid sample delivery for megahertz serial crystallography at X-ray FELs. *IUCrJ* 5(5), pp. 574–584. doi: 10.1107/S2052252518008369.
- Williamson, L.J. et al. 2023. Structure of the *Lysinibacillus sphaericus* Tpp49Aa1 pesticidal protein elucidated from natural crystals using MHz-SFX. *Proceedings of the National Academy of Sciences* 120(49), p. e2203241120. doi: 10.1073/PNAS.2203241120.
- Wilson, C.J., Choy, W.Y. and Karttunen, M. 2022. AlphaFold2: A Role for Disordered Protein/Region Prediction? *International Journal of Molecular Sciences* 23(9), p. 23. doi: 10.3390/ijms23094591.

- Wlodawer, A., Minor, W., Dauter, Z. and Jaskolski, M. 2013. Protein crystallography for aspiring crystallographers or how to avoid pitfalls and traps in macromolecular structure determination. *FEBS Journal* 280(22), pp. 5705–5736. doi: 10.1111/febs.12495.
- Xiao, Y. and Wu, K. 2019. Recent progress on the interaction between insects and *Bacillus thuringiensis* crops. *Philosophical Transactions of the Royal Society B: Biological Sciences* 374(1767). doi: 10.1098/rstb.2018.0316.
- Xu, D., Tsai, C.J. and Nussinov, R. 1997. Hydrogen bonds and salt bridges across protein-protein interfaces. *Protein Engineering* 10(9), pp. 999–1012. doi: 10.1093/protein/10.9.999.
- Xu, X., Yu, L. and Wu, Y. 2005. Disruption of a cadherin gene associated with resistance to Cry1Ac  $\delta$ -endotoxin of *Bacillus thuringiensis* in *Helicoverpa armigera*. *Applied and Environmental Microbiology* 71(2), pp. 948–954. doi: 10.1128/AEM.71.2.948-954.2005.
- Yamaguchi, T., Sahara, K., Bando, H. and Asano, S. ichiro. 2008. Discovery of a novel *Bacillus thuringiensis* Cry8D protein and the unique toxicity of the Cry8D-class proteins against scarab beetles. *Journal of Invertebrate Pathology* 99(3), pp. 257–262. doi: 10.1016/j.jip.2008.05.009.
- Yang, Y. et al. 2011. Down regulation of a gene for cadherin, but not alkaline phosphatase, associated with cry1ab resistance in the sugarcane borer *diatraea saccharalis*. *PLoS ONE* 6(10), p. e25783. doi: 10.1371/journal.pone.0025783.
- Yang, Z. et al. 2019. Targeted delivery of antimicrobial peptide by Cry protein crystal to treat intramacrophage infection. *Biomaterials* 217, p. 119286. doi: 10.1016/j.biomaterials.2019.119286.
- Yang, Z., Lee, M.M.M. and Chan, M.K. 2021. Efficient intracellular delivery of p53 protein by engineered protein crystals restores tumor suppressing function in vivo. *Biomaterials* 271, p. 120759. doi: 10.1016/j.biomaterials.2021.120759.
- Yefanov, O. et al. 2019. Evaluation of serial crystallographic structure determination within megahertz pulse trains. *Structural Dynamics* 6(6), p. 064702. doi: 10.1063/1.5124387.
- Yefanov, O., Mariani, V., Gati, C., White, T.A., Chapman, H.N. and Barty, A. 2015. Accurate determination of segmented X-ray detector geometry. *Optics Express* 23(22), p. 28459. doi: 10.1364/oe.23.028459.
- Yu, D., Chojnowski, G., Rosenthal, M. and Kosinski, J. 2023. AlphaPulldown- a python package for protein-protein interaction screens using AlphaFold-Multimer. *Bioinformatics* 39(1), p. btac749. doi: 10.1093/bioinformatics/btac749.

- Yuan, Z., Zhang, Y., Cai, Q. and Liu, E.Y. 2000. High-level field resistance to *Bacillus sphaericus* C3-41 in *Culex quinquefasciatus* from southern China. *Biocontrol Science and Technology* 10(1), pp. 41–49. doi: 10.1080/09583150029378.
- Yuan, Z.M., Pei, G.F., Regis, L., Nielsen-Leroux, C. and Cai, Q.X. 2003. Cross-resistance between strains of *Bacillus sphaericus* but not *B. thuringiensis israelensis* in colonies of the mosquito *Culex quinquefasciatus*. *Medical and Veterinary Entomology* 17(3), pp. 251–256. doi: 10.1046/j.1365-2915.2003.00429.x.
- Zhang, Q. et al. 2016. Recent Advances in Protein-Protein Docking. *Current Drug Targets* 17(14), pp. 1586–1594. doi: 10.2174/1389450117666160112112640.
- Zhang, S., Cheng, H., Gao, Y., Wang, G., Liang, G. and Wu, K. 2009. Mutation of an aminopeptidase N gene is associated with *Helicoverpa armigera* resistance to *Bacillus thuringiensis* Cry1Ac toxin. *Insect Biochemistry and Molecular Biology* 39(7), pp. 421–429. doi: 10.1016/j.ibmb.2009.04.003.
- Zhang, X., Candas, M., Griko, N.B., Rose-Young, L. and Bulla, L.A. 2005. Cytotoxicity of *Bacillus thuringiensis* Cry1Ab toxin depends on specific binding of the toxin to the cadherin receptor BT-R1 expressed in insect cells. *Cell Death and Differentiation* 12(11), pp. 1407–1416. doi: 10.1038/sj.cdd.4401675.
- Zhang, X., Candas, M., Griko, N.B., Taussig, R. and Bulla, L.A. 2006. A mechanism of cell death involving an adenylyl cyclase/PKA signaling pathway is induced by the Cry1Ab toxin of *Bacillus thuringiensis*. *Proceedings of the National Academy of Sciences of the United States of America* 103(26), pp. 9897–9902. doi: 10.1073/pnas.0604017103.
- Zhang, Y., Zheng, G., Tan, J., Li, C. and Cheng, L. 2013. Cloning and characterization of a novel cry8Ab1 gene from *Bacillus thuringiensis* strain B-JJX with specific toxicity to scarabaeid (Coleoptera: Scarabaeidae) larvae. *Microbiological Research* 168(8), pp. 512–517. doi: 10.1016/j.micres.2013.03.003.
- Van Zundert, G.C.P. et al. 2016. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *Journal of Molecular Biology* 428(4), pp. 720–725. doi: 10.1016/j.jmb.2015.09.014.

## 8. Appendices

### *Appendix 1 – Cry8Ba2 protein sequence*

1	MSPNNQNEYE	IIDATPSTSV	SNDSNRYPPFA	NEPTNALQNM	DYKDYLKMSA	50
51	GNVSEYPGSP	EVFLSEQDAV	KAAIDIVGKL	LTGLGVPFVG	PIVSLYTQLI	100
101	DILWPSKQKS	QWEIFMEQVE	ELINQKIAEY	ARNKALSELE	GLGNNYQLYL	150
151	TALEEWKENP	NGSRALRDVR	NRFEILDRLF	TQYMPSFRVT	NFEVPPFLTVY	200
201	TMAANLHLLL	LRDASIFGEE	WGLSTSTINN	YYNRQMKLTA	EYSDHCVKWY	250
251	ETGLAKLKGS	SAKQWIDYNQ	FRREMTLTVL	DVVALFSNYD	TRTYPLATTA	300
301	QLTREVYTDV	LGAVDVPNIG	SWYDKAPSFV	EIEKAAIRPP	HEFDYITGLT	350
351	VYTKKRSFTS	DRYMRYWAGH	QISYKTIGTS	STFTQMYGTN	QNLQSTSNFD	400
401	FTNYDIYKTL	SNDAVLLDIV	YPGYTYTFFG	MPETEFFMVN	QLNNTRKTLT	450
451	YKPASKDIID	RTRDSELELP	PETSGQPNYE	SYSHRLGHIT	FIYSSSTSTY	500
501	VPVFSWTHRS	ADLTNTVKSG	EITQIPGGKS	SYIGRNTYII	KGRGYTGGDL	550
551	LALDRIGSC	EFQMIFPESQ	RFRIRIRYAS	NETSYISLYG	LNQSGTLKFN	600
601	QTYSNKNEND	LTYNDFKYIE	YPRVISVNAS	SNIQRLSIGI	QTNTNLFILD	650
651	RIEFIPVDET	YEAETDLEAA	KKAVNALFTN	TKDGLQPGVT	DYEVNQAAANL	700
701	VECLSDDLYP	NEKRLLFDAV	REAKRLSEAR	NLLQDPDFQE	INGENGWTS	750
751	TGIEVIEGDA	VFKGRYLRLP	GAREIDTETY	PTYLYQKVEE	GVLKPYTRYR	800
801	LRGFVGSSQG	LEIYTIRHQT	NRIVKNVPDD	LLPDVPPVNN	DGRINRCSEQ	850
851	KYVNSRLEVE	NRSGEAHEFS	IPIDTGELDY	NENAGIWVGF	KITDPEGYAT	900
901	LGNLELVEEG	PLSGDALERL	QKEEQQWKIQ	MTRRREETDR	RYMASKQAVD	950
951	RLYADYQDQQ	LNPVVEITDL	TAAQDLIQSI	PYVYNEMFPE	IPGMNYTKFT	1000
1001	ELTDRLQQAW	GLYDQRNAIP	NGDYRNELSN	WNTTSGVNVQ	QINHTSVLVI	1050
1051	PNWNEQVSQK	FTVQPNQRYV	LRVTARKEGV	GNGYVSIRDG	GNQSETLTFS	1100
1101	ASDYDTNGMY	DTQASNTNGY	NTNSVYNDQT	GYITKTVTFI	PYTNQMWIEI	1150
1151	SETEGTFYIE	SVELIVDVE				1169

## Appendix 2 – Cry48Aa1 protein sequence

1	MDINNNNEKE	IINSHLLPYS	LLKKYPIKSL	QSTNYKDWLN	LCQDFNKDIE	50
51	SYDLVTAVSS	GTIVVGTMLS	AIYAPALIAG	PIGIIGAI I I	SFGTLLP L L W	100
101	NESENNPKTT	WIEFIRMGEQ	LVDKTISQTV	FNILES Y L K D	LKVN L V D Y E K	150
151	AKQDWIELKK	QQLPGSPPST	KLRNAADIAH	QR L D S L H N K F	A E L N K F K V E P	200
201	YETILLPVYA	QAANLHLNLL	QQGAMFADQW	I E D K Y S S R N D	T F A G N S N D Y Q	250
251	NLLKSRTITY	INHIENTYQN	GLNYLWNQPE	M T W D I Y N E Y R	T K M T I T A L D L	300
301	MALFPFYNKE	LYDPTVGIKS	ELTREIFINT	P V E P H L H R Y F	K L S E T E E K L T	350
351	NNSDLFKWLT	SLKFRTLYQP	GFPFLIGNMN	S F T N T N G T Q L	I N N Q Q Q L W S F	400
401	PGTTENEKEL	FPSPANIDQV	TMYIYYGSGW	G I P E P I S T T I	N K L I F N H D K H	450
451	ELISEYDAGN	TNAPTRSLSL	GLPNHYLSCL	N S Y Y P L T A T T	D G M N K E E L K M	500
501	YSFGWTHNSV	DFLNEISKDK	ITQIPAVKAY	R L T S N S R V I K	G P S H I G G N L V	550
551	YLSSENSQMAL	TCRYTNSSPQ	EYKIRIRYAS	N R L N M G Q L F T	T F S S H Q F V L P	600
601	PTFNHFNIEQ	AKYEDYAYAE	FPESMSIRGN	L N S D I L L I L N	I L A G G E L L L D	650
651	KIEFIPLTQK	VKDNLEKEKI	DMLKNLTDSL	F N S P S K D T L K	I D S T D Y Q I D Q	700
701	IAFQIESINE	EINPQEKMEL	LDNIKYAKKL	N Q L R N L L Y S R	E S Q A Q I D W V T	750
751	SNDVSIYHGK	KPFNDYTLVM	SRTSSSLSEI	T A T N Y Q T Y I Y	K K I E E S K L K P	800
801	YTRYLVRGFI	SNSEDLEIFI	SRYENEIHTN	M N V H G D D D T L	L N S D I R Q N E C	850
851	ESKLPIIFDA	TSQYSLSPSR	TSGISNHSYY	N N G H Q S S C N D	T H I F S F S I D T	900
901	GEVDFNNYPG	IEILFKLSNT	NGYASISNLE	V I E E R L L T E E	E K R Q I I Q I E N	950
951	RWKAKKESQR	NETEKITTQA	QQAINSLFTD	T Q Y S N L K F E T	T K Q N I T E A N T	1000
1001	ILENIPYVYN	ALLPTEPGMN	FVLFNSFKDQ	I N K A H A L Y K M	R N L I K N G D F I	1050
1051	NDTKYWSIST	DVKLEKVNKE	TILVLSSWEA	Q A S Q Q I L V Q K	Q K R Y L L R V I A	1100
1101	KKEDMGRGNV	IISDCLNNIA	KIDFTPHDCN	M N H I Q N S S E F	I I K T I H F S P N	1150
1151	TEQVRIDIGQ	SDGVFKVESI	ELICVNY			1177

## Appendix 3 – Tpp49Aa1 protein sequence

1	MENQIKEEFN	KNNHGIPSDC	SCIKEGDDYN	SLTEINAKEF	SYCSPNMFNL	50
51	NLPEQSTRFQ	TIASIHSNNC	SFEILNNDPG	YIYGDSVDGE	CRIAVAHREL	100
101	GNGLERTGDD	RFLFIFYALD	NNNFIIANRH	DGFVLQFLIA	NGQGIVVSRE	150
151	YQPNIHQEFT	IQSINSDTFR	LHSRDTNTFA	TVCWAQFNSW	TKIVSRVDNP	200
201	GAPANLKHHR	SLLTDINMPQ	LPSLTPLQPL	PRLTELEDGG	LSPAQAPRAI	250
251	IGRTLIPCLF	VNDPVLRLN	RIKQSPYYVL	EHRQYWHRIW	TDIFTAGERR	300
301	EYREVTGINN	NAQNDMNKMI	NITIGADGPN	RLRFGNLSTP	FRQQIIDNSN	350
351	TLGSFANTNY	GTRTDIVNVF	NSEFHQVRYA	RFVKAYEYRL	TRADGSQVGT	400
401	PWVVLDRKEM	DLRTPHNMA	ITLENVKIDN	ADNSYDLSIW	KTPLKLLKDGK	450
451	IIIEHNHENS	K P Y Y N				464

#### Appendix 4 – *Tpp1Aa2* protein sequence

1	<b>MRNLDF</b> IDSF	IPTEGKYIRV	MDFYNSEYPF	CIHAPSAPNG	DIMTEICSRE	50
51	NNQYFIFFPT	DDGRVIIANR	HNGSVFTGEA	TSVVSDIYTG	SPLQFFREVK	100
101	RTMATYYLAI	QNPEASATDVR	ALEPHSHELP	SRLYYTNNIE	NNSNILISNK	150
151	EQIYLTLP	SLPENEQYPKTP	VLSGIDDIGP	NQSEKSIIGS	TLIPCIMVSD	200
201	FISLGERMKT	TPYYYVKHTQ	YWQSMWSALF	PPGSKETKTE	KSGITDTSQI	250
251	SMTDGINVSI	GADFGRLRFGN	KTFGIKGGFT	YDTKTQITNT	SQLLIETTYT	300
301	REYTNTEENFP	VRVTGYVLAS	EFTLHRSDGT	QVNTIPWVAL	NDNYTTIARY	350
351	PHFASEPLLG	NTKIITDDQN				400

Residues shown in bold were excluded from the clone used in this work.

#### Appendix 5 – *Tpp2Aa2* protein sequence

1	MCDSKDNSGV	SEKCGKKFTN	YPLNTTPTSL	NYNLPEISKK	FYNLKNKYSR	50
51	NGYGLSKTEF	PSSIENCPSN	EYSIMYDNKD	PRFLIRFLLD	DGRYIIADRD	100
101	DGEVFDEAPT	YLDNNNHPII	SRHYTGEEERQ	KFEQVGS GDY	ITGEQFFQFY	150
151	TQNKTRVLSN	CRALDSRTIL	LSTAKIFPIY	PPASETQLTA	FVNSSFYAAA	200
201	IPQLPQTSLL	ENIPEPTSLD	DSGVLPKDAV	RAVKGSALLP	CIIVHDPNLN	250
251	NSDKMKFNTY	YLLEYKEYWH	QLWSQIIPAH	QTVKIQERTG	ISEVVQNSMI	300
301	EDLNMYIGAD	FGMLFYFRSS	GFKEQITRGL	NRPLSQTTTQ	LGERVEEMEY	350
351	YNSNDLDVRY	VKYALAREFT	LKRVNGEIVK	NWVAVDYRLA	GIQSYPNAPI	400
401	TNPLTLTKHT	IIRCENSYDG	HIFKTPLIFK	NGEVIVKTNE	ELIPKINQ	448

#### Appendix 6 – *Cqm1* protein sequence

-3	MLAIREPDSK	DWYQHATFYQ	IYPRSFSDSN	GDGIGDLAGI	TSKMKYLADI	47
48	GIDATWLSPP	FKSPLKDFGY	DVSDFYDIQP	EYGNLTD FDK	LVEESHKNGI	97
98	KLMLDFIPNH	SSDQHEWFVK	SVLRDPEYSD	FYVWRPPATG	GGPPNNWISV	147
148	FGGSAWTYNQ	ARGEYYLHQF	TPQQPDLNYR	NPKVLAEMTK	MLFFWLDRGV	197
198	DGFRLDAINH	MFEDEQFRDE	PLSGWGQPGE	YDSL DHIYTK	DIPDVYDVVY	247
248	NWRDQMDKYS	AEKGRTIILM	TEAYSSIEGT	MLYYESADRK	RQGAHMPFNF	297
298	QLIYDFKKEQ	NAVGLKNSID	WWMNNMPARH	TPSWVAGSHD	HSRVASRVGL	347
348	DRVDQVMTLL	HTLPGTSITY	YGEEVAMQDF	KEAQQFDNRD	PNRTPMQWDS	397
398	STSAGFSTNT	NTWLRVHPDY	ARYNV DVMQK	NPQSTFH HFQ	HLTKLRRHRT	447
448	MQSGEYVHKT	VGTKVYALLR	ELRGEDSFLT	VLN MAGAEDT	VDLGDFVNLP	497
498	QKMRVEVAQP	NSKSKAGNEV	DIGKLTLPY	DSVVL RATVC		537

The pET28a expression vector comprised residues 20 – 557 of the *cqm1* gene. Throughout this work, the *Cqm1* sequence is numbered according to the PDB entry of the *Cqm1* crystal structure (PDB 6K5P), which, itself, was solved from an expression vector encoding residues 23 – 600 of the insect *cqm1* gene. Hence, our *Cqm1* sequence is numbered -3 – 537.

**Appendix 7 – Tpp2Aa2-Cqm1 single-particle cryoEM 2D-classifications and gold-standard Fourier shell correlation graph**

