Copyright Statement

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

This thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognize the author's right to be identified as the author of this thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from their thesis.

General copyright and disclaimer

In addition to the above conditions, authors give their consent for the digital copy of their work to be used subject to the conditions specified on the Library Thesis Consent Form and Deposit Licence.
Expression, Characterisation, and Structural Investigation of *Cydia pomonella* Granulovirus Proteins

Guangmei Huang

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy in Biological Sciences

The University of Auckland

2019
I. Abstract

The infectious particles of baculovirus insect viruses are unique micron sized protein crystals called occlusion bodies. *Cydia pomonella* granulovirus (CpGV) occlusion bodies are small enough for cryo-electron microscopy and this method revealed that the surfaces of the crystals are coated with fibres. The three CpGV crystal surface proteins PEP-L, PEP-M, and PEP-S were investigated. Experiments using both insect and bacterial systems were carried out with the aim of producing proteins or complexes suitable for structural analysis. PEP-M and PEP-S form a complex PEP-M/PEP-S. PEP-L and the PEP-M/PEP-S complex both form oligomers.

The fibres on intact occlusion bodies were investigated using cryo-electron microscopy. The fibres are about 15 nm long and are attached to a base layer that is attached to the protein crystal.

A second part of the thesis describes the structural characterization of the *Cydia pomonella* granulovirus (CpGV) phosphatase protein PTP-2. Many viral genomes encode kinase and phosphatase enzymes to manipulate pathways controlled by phosphorylation events. The atomic structure of PTP-2 was determined in this study using X-ray crystallography to a resolution of 1.65 Å.
II. Acknowledgments

My four-year PhD study overseas has been an exceptional journey during which I experienced failure, success, frustration and joy. I would like to thank some people that have helped me along the way.

I would like to thank Dr Jason Busby for his kind help at the beginning of my PhD project. His excellent work ethic influenced me and has been very beneficial. I would like to thank Shinmei.Yeh and Martin Middleditch for their previous input on the project. I would also like to thank Dr Johannes Jehle for providing the CpGV samples, Ken Goldie for providing the first cryo-electron images of CpGV before the start of the project, and Michael Oliver for giving useful advice on the project. I am also indebted to Dr Jeremy Keown who carried out the SEC-MALS experiment and also collected X-ray data and helped with the atomic structure determination of the PTP-2 protein. I also thank Dr Adrian Turner for instruction and aid in electron microscopy, Dr David Goldstone for his encouragement, my co-supervisor Dr Richard Kingston for his suggestions on the project, and all the other members of the Structural Biology Laboratory for their input and company.

I'm grateful for financial support from the NZ Royal Society Marsden Fund and from the University of Auckland. Without this, it would have been impossible for me to have gone so far.

Mostly importantly, I would like to thank my supervisor, Associate Professor Peter Metcalf. The opportunity to work with him enabled me to study abroad, which had always been a dream of mine. His kindly support, patience, understanding, and guidance along the way have made my PhD study a most treasured experience.

Lastly, I would like to thank my parents, who raised us four siblings Lei Huang, me, Fengxian Huang, and Senlin Huang through hardship, who plough land in a remote village in China, and who have yet to see the world.
Co-Authorship Form

This form is to accompany the submission of any PhD that contains published or unpublished co-authored work. Please include one copy of this form for each co-authored work. Completed forms should be included in all copies of your thesis submitted for examination and library deposit (including digital deposit), following your thesis Acknowledgements. Co-authored works may be included in a thesis if the candidate has written all or the majority of the text and had their contribution confirmed by all co-authors as not less than 65%.

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Chapter 5 Crystal structure of Cydia pomonella PTP-2

<table>
<thead>
<tr>
<th>Nature of contribution by PhD candidate</th>
<th>Cloned, expressed, purified, and crystallized protein, analysed data, wrote the majority of the text.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extent of contribution by PhD candidate (%)</td>
<td>65%</td>
</tr>
</tbody>
</table>

CO-AUTHORS

<table>
<thead>
<tr>
<th>Name</th>
<th>Nature of Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Michael R. Oliver</td>
<td>planned and designed the research</td>
</tr>
<tr>
<td>Jeremy R. Keown</td>
<td>collected carried out the X-ray crystallographic structure determination</td>
</tr>
<tr>
<td>David C. Goldstone</td>
<td>helped with the structure determination</td>
</tr>
<tr>
<td>Peter Metcalf</td>
<td>analysed data and revised the majority of the text</td>
</tr>
</tbody>
</table>

Certification by Co-Authors

The undersigned hereby certify that:

- the above statement correctly reflects the nature and extent of the PhD candidate’s contribution to this work, and the nature of the contribution of each of the co-authors; and
- that the candidate wrote all or the majority of the text.

<table>
<thead>
<tr>
<th>Name</th>
<th>Signature</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Michael R. Oliver</td>
<td></td>
<td>11/12/2018</td>
</tr>
<tr>
<td>Jeremy R. Keown</td>
<td>JKeown</td>
<td>06/12/2018</td>
</tr>
<tr>
<td>David C. Goldstone</td>
<td></td>
<td>06/12/2018</td>
</tr>
<tr>
<td>Peter Metcalf</td>
<td></td>
<td>06/12/2018</td>
</tr>
</tbody>
</table>

Last updated: 28 November 2017
# III. Table of Contents

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Abstract</td>
<td>i</td>
</tr>
<tr>
<td>II</td>
<td>Acknowledgments</td>
<td>ii</td>
</tr>
<tr>
<td>III</td>
<td>Table of Contents</td>
<td>iv</td>
</tr>
<tr>
<td>IV</td>
<td>List of Figures</td>
<td>x</td>
</tr>
<tr>
<td>V</td>
<td>List of Tables</td>
<td>xiii</td>
</tr>
<tr>
<td>VI</td>
<td>Abbreviations</td>
<td>xv</td>
</tr>
</tbody>
</table>

## Chapter 1 Introduction

1.1 Baculovirus ................................................. 1
1.2 Life cycle and development .................................. 3
1.3 Baculovirus diversity and phylogeny .......................... 6
1.4 Genome and conserved proteins .................................. 7
1.5 Occlusion body matrix protein polyhedrin ......................... 8
1.6 Occlusion body envelope .......................................... 9
   1.6.1 Isolated envelope structures .............................. 11
   1.6.2 Occlusion body envelope components ...................... 13
      1.6.2.1 Carbohydrates ...................................... 13
      1.6.2.2 Polyhedral envelope protein ......................... 13
   1.6.3 Occlusion body envelope development ..................... 15
      1.6.3.1 Electron-dense spacers .............................. 15
      1.6.3.2 Fibrous structures .................................. 17
      1.6.3.3 Connections between spacers, fibrous structures and the envelope .......................... 17
   1.6.4 Occlusion body envelope function ......................... 18
      1.6.4.1 Occlusion body size and shape ..................... 18
      1.6.4.2 Occlusion body stability ........................... 19
1.7 Occlusion body envelope fibres ................................ 19
1.8 Research objectives and thesis outline ........................ 21
Chapter 2 Materials and Methods

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Bioinformatics</td>
<td>23</td>
</tr>
<tr>
<td>2.2 Sample preparation of <em>in vivo</em> grown granulovirus OBs.</td>
<td>23</td>
</tr>
<tr>
<td>2.3 Bacterial culture</td>
<td>24</td>
</tr>
<tr>
<td>2.3.1 <em>E.coli</em> Strains</td>
<td>24</td>
</tr>
<tr>
<td>2.3.2 Media</td>
<td>24</td>
</tr>
<tr>
<td>2.3.3 Antibiotic stock solutions and reagents</td>
<td>24</td>
</tr>
<tr>
<td>2.3.4 Procedure</td>
<td>25</td>
</tr>
<tr>
<td>2.4 Chemically competent cells</td>
<td>25</td>
</tr>
<tr>
<td>2.4.1 Reagents</td>
<td>25</td>
</tr>
<tr>
<td>2.4.2 Procedure</td>
<td>25</td>
</tr>
<tr>
<td>2.5 Electrocompetent cells</td>
<td>26</td>
</tr>
<tr>
<td>2.5.1 Reagents</td>
<td>26</td>
</tr>
<tr>
<td>2.5.2 Procedure</td>
<td>26</td>
</tr>
<tr>
<td>2.6 Genome DNA extraction from occlusion bodies</td>
<td>27</td>
</tr>
<tr>
<td>2.6.1 Reagents</td>
<td>27</td>
</tr>
<tr>
<td>2.6.2 Procedure</td>
<td>28</td>
</tr>
<tr>
<td>2.7 Gene cloning</td>
<td>29</td>
</tr>
<tr>
<td>2.7.1 PCR</td>
<td>29</td>
</tr>
<tr>
<td>2.7.1.1 Materials</td>
<td>29</td>
</tr>
<tr>
<td>2.7.1.2 Reaction system and program</td>
<td>29</td>
</tr>
<tr>
<td>2.7.2 Agarose gel electrophoresis</td>
<td>30</td>
</tr>
<tr>
<td>2.7.2.1 Reagents</td>
<td>30</td>
</tr>
<tr>
<td>2.7.2.2 Procedure</td>
<td>31</td>
</tr>
<tr>
<td>2.7.3 Agarose gel extraction</td>
<td>31</td>
</tr>
<tr>
<td>2.7.4 Restriction enzyme digestion</td>
<td>31</td>
</tr>
<tr>
<td>2.7.5 Ligation</td>
<td>32</td>
</tr>
</tbody>
</table>
2.7.6 Transformation of recombinant DNA into bacterial cloning strains .................33
2.7.7 Colony PCR ........................................................................................................33
2.7.8 Plasmid extraction ................................................................................................34
2.7.9 Sequencing ...........................................................................................................34
2.8 Transformation into bacterial expression strain ....................................................34
2.9 Small scale protein expression in bacteria ...............................................................34
2.10 Large scale protein expression in bacteria .............................................................35
2.11 Large scale SeMet-protein expression in bacteria ..................................................35
   2.11.1 Reagents ..........................................................................................................35
   2.11.2 Procedure .........................................................................................................37
2.12 Bacterial cell lysis ..................................................................................................38
   2.12.1 Cell lysis buffer ...............................................................................................38
   2.12.2 Procedure .........................................................................................................38
2.13 Insect cell culture ..................................................................................................39
   2.13.1 Sf9 cell culture procedure ..............................................................................39
   2.13.2 Sf9 cell counting .............................................................................................40
2.14 Baculovirus bacmid transposition .........................................................................40
2.15 Recombinant bacmid DNA isolation .....................................................................41
2.16 Insect cell transfection with recombinant bacmid DNA ......................................42
2.17 Baculoviral stock amplification ............................................................................43
2.18 Baculoviral stock storage .....................................................................................44
2.19 Protein expression in insect cells ..........................................................................44
   2.19.1 Small-scale protein expression in insect cells ................................................44
   2.19.2 Large-scale protein expression in insect cells ................................................44
   2.19.3 Aliquoting infected insect cells ....................................................................45
2.20 Insect cell lysis ......................................................................................................45
   2.20.1 Small scale insect cell lysis ............................................................................45
2.20.2 Large scale insect cell lysis .................................................................45
2.21 Pull-down experiment and sample preparations for SDS-PAGE .................................................................45
  2.21.1 Procedure for pull-down from *E. coli* culture ............................................46
  2.21.2 Procedure for pull-down from insect cell culture ........................................47
2.22 Protein purification ..........................................................................................48
  2.22.1 Immobilized metal affinity chromatography ..................................................48
  2.22.2 Recombinant tabacco etch virus protease cleavage ......................................48
  2.22.3 Size exclusion chromatography ...................................................................49
2.23 Protein concentration measurement ..................................................................50
2.24 SDS-PAGE .........................................................................................................50
  2.24.1 Gel recipe .....................................................................................................50
  2.24.2 Gel making procedure ..................................................................................51
  2.24.3 Gel run .........................................................................................................52
  2.24.4 Coomassie staining and destaining ...............................................................53
2.25 Protein characterization ......................................................................................53
  2.25.1 Native PAGE ...............................................................................................53
  2.25.2 Size exclusion chromatography-multi-angle static light scattering ................54
2.26 X-ray crystallography .......................................................................................54
  2.26.1 Robotic crystallization screening ...................................................................54
  2.26.2 Proteolysis during crystallization experiment ...............................................54
  2.26.3 Fine crystallization screening .......................................................................55
  2.26.4 X-ray diffraction ..........................................................................................55
2.27 Transmission electron microscopy ......................................................................55
  2.27.1 Negative stain electron microscopy .............................................................56
  2.27.2 Cryo-electron microscopy .............................................................................57
    2.27.2.1 Specimen cryo-cooling ...........................................................................57
    2.27.2.2 Specimen transfer to cryo-holder and into EM .....................................57
2.27.2.3 Cryo-electron microscopy alignments ........................................... 57
2.27.2.4 Cryo-electron microscopy image collection ..................................... 59
2.28 How to flash-freeze aliquots in liquid nitrogen ........................................ 60

Chapter 3 Expression of CpGV occlusion body surface proteins ...................... 61
3.1 Results ........................................................................................................ 63
3.1.1 Bioinformatics .......................................................................................... 63
3.1.2 CpGV-PEP-L ............................................................................................ 68
3.1.3 PEP-N domains of CpGV-PEP-M and CpGV-PEP-S ............................... 79
3.1.4 CpGV-PEP-L, PEP-M, PEP-S co-expression experiments .......................... 80
3.1.5 CpGV-PEP-L, PEP-M, PEP-S, polyhedrin co-expression .......................... 85
3.1.6 AcMNPV-PEP ........................................................................................... 86
3.2 Summary ........................................................................................................ 87

Chapter 4 Electron microscopy investigation of CpGV occlusion bodies .............. 90
4.1 Negative stain electron microscopy ............................................................... 90
4.2 Cryo-electron microscopy ............................................................................ 92
4.3 Summary ........................................................................................................ 98

Chapter 5 X-ray crystallographic structure determination for the CpGV phosphatase PTP-2 ................................................................. 100
5.1 Background ................................................................................................. 100
5.2 Cloning .......................................................................................................... 103
5.3 Expression and purification ......................................................................... 104
5.4 Crystallization ............................................................................................... 106
5.5 X-ray diffraction and structure solution ....................................................... 108
5.6 PTP-2 crystal structure and discussion ......................................................... 112

Chapter 6 Discussion ...................................................................................... 119
6.1 Introduction ................................................................................................. 119
6.2 The occlusion body surface and the fibres ................................................. 119
6.3 Characterization of occlusion body surface proteins ..................................................... 120
6.4 Cryo-electron microscopy study on occlusion body surface fibres .................................. 122
6.5 CpGV PTP-2 atomic structure ......................................................................................... 124

Appendices ................................................................................................................................. 126
Primer sequences ......................................................................................................................... 126
References ................................................................................................................................. 128
IV. List of Figures

Figure 1.1 Electron micrographs of alphabaculovirus OpMNPV occlusion bodies ..................1
Figure 1.2 Thin section electron micrograph of betabaculovirus PiGV occlusion bodies. ......2
Figure 1.3 Baculovirus life cycle. .................................................................................3
Figure 1.4 Baculovirus development in the cell nucleus. ..............................................5
Figure 1.5 Baculoviridae phylogenetic tree from core genes. ........................................6
Figure 1.6 Baculovirus polyhedrin protein atomic structure and organization. ..............9
Figure 1.7 Electron micrograph of baculovirus AcMNPV OB showing surface layer. .......10
Figure 1.8 Electron micrograph of baculovirus TnNPV OB showing surface layer.........11
Figure 1.9 Electron micrographs of envelope structures isolated from baculovirus PdMNPV OBs ..........................................................12
Figure 1.10 Immunogold labelled thin section electron micrograph of AcMNPV OB showing the location of PEP .................................................................14
Figure 1.11 Thin section electron micrograph of wild type AcMNPV and PEP-deletion OBs. ........................................................................................................14
Figure 1.12 Electron micrographs of wild type OpMNPV and PEP-deletion OBs............15
Figure 1.13 Thin section EM of AcMNPV infected cells showing OB, ES, and FS. .........16
Figure 1.14 Electron micrograph of OpMNPV OB of wild type, P10 deletion, and PEP deletion .................................................................18
Figure 1.15 Cryo-EM of CpGV baculovirus showing fibres extending from the OB surface. ........................................................................................................20
Figure 1.16 Thin section EM of TnNPV baculovirus showing fibres on PE. .................21
Figure 2.1 Incorporation of rTEV recognition sequence into constructs........................49
Figure 3.1 Previous study on CpGV occlusion body fibres from Metcalf group. ..........62
Figure 3.2 Schematic structure of PEP proteins. ..............................................................64
Figure 3.3 Multiple sequence alignment of AcMNPV-PEP and CpGV-PEPs .................65
Figure 3.4 Pairwise sequence alignment of AcMNPV-PEP and CpGV-PEP-L ...............66
Figure 3.5 Pairwise sequence alignment of CpGV-PEP-M and CpGV-PEP-S ...............67
Figure 3.6 Multiple sequence alignment of CpGV-PEP-L N-terminal domain, PEP-M and PEP-S ................................................................. 68
Figure 3.7 SDS-PAGE of CpGV-PEP-L expression in bacteria and insect cells ............... 70
Figure 3.8 SEC-MALS of CpGV-PEP-L in Tris-HCl pH 8.0 ........................................... 71
Figure 3.9 S200 size exclusion chromatography of CpGV-PEP-L in 100 mM Na2CO3 pH 11.8 ................................................................. 72
Figure 3.10 SEC-MALS of CpGV-PEP-L in 100 mM Na2CO3 pH 11.8 ....................... 73
Figure 3.11 Size exclusion chromatography of CpGV-PEP-L at different pH .......... 74
Figure 3.12 Native-PAGE of CpGV-PEP-L at different pH ........................................ 75
Figure 3.13 Ribbon diagram of CpGV-PEP-L-1-106 aa atomic structure .................... 76
Figure 3.14 Size exclusion chromatography of CpGV-PEP-L-1-300 ....................... 77
Figure 3.15 Size exclusion chromatography of CpGV-PEP-L-1-251 ....................... 78
Figure 3.16 Size exclusion chromatography of CpGV-PEP-L-1-211 ....................... 79
Figure 3.17 Size exclusion chromatography of CpGV-PEP-S-1-132 ....................... 80
Figure 3.18 SDS-PAGE analysis of PEP-M/PEP-S complex .................................... 82
Figure 3.19 Negative stain EM image of PEP-M/PEP-S aggregate ......................... 83
Figure 3.20 Size exclusion chromatography of MBP-PEP-M/PEP-S complex .......... 84
Figure 3.21 Size exclusion chromatography of MBP-PEP-S/PEP-M complex .......... 85
Figure 3.22 Diagram of all the PEP constructs investigated in this study .................. 89
Figure 4.1 Negative stain EM image of CpGV occlusion body .............................. 91
Figure 4.2 Non-stain EM images of CpGV occlusion bodies ............................... 91
Figure 4.3 Cryo-EM image of CpGV occlusion bodies showing thick ice ............. 93
Figure 4.4 Cryo-EM image of CpGV occlusion bodies showing crystalline ice ...... 94
Figure 4.5 Cryo-EM image of CpGV occlusion bodies showing thin, vitreous ice .......... 95
Figure 4.6 Cryo-EM image of CpGV occlusion bodies showing fibres .................. 96
Figure 4.7 Cryo-EM image of DTT treated CpGV occlusion bodies showing fibres still attached ................................................................. 98
Figure 5.1 Predicted protein kinases and phosphatases derived from virus genomes .... 101
Figure 5.2 BLAST search using *Cydia pomonella* granulovirus PTP-2 protein sequence against the NCBI refseq protein database.................................................................103

Figure 5.3 S200 size exclusion chromatography of native PTP-2........................................105

Figure 5.4 S75 size exclusion chromatography of SeMet PTP-2........................................106

Figure 5.5 Crystals of native PTP-2......................................................................................107

Figure 5.6 Crystals of SeMet PTP-2......................................................................................108

Figure 5.7 SeMet PTP-2 diffraction image (1° oscillation)......................................................109

Figure 5.8 Native PTP-2 diffraction image (1° oscillation)......................................................110

Figure 5.9 Ribbon diagram of CpGV PTP-2. ........................................................................112

Figure 5.10 Electron density for the C-terminal region of PTP-2............................................113

Figure 5.11 Superimposition of PTP-2 and its homologue 1YN9...........................................115

Figure 5.12 Structure based PTP-2 multiple sequence alignment .........................................116

Figure 5.13 Rooted phylogenetic tree for the PTP-2 NCBI protein cluster 2945255.............117

Figure 6.1 HHpred results for proteins in protein cluster 2945044.......................................120

Figure 6.2 Protein structures corresponding to the HHpred results for protein cluster 2945044.........................................................................................................................121

Figure 6.3 Combination of fibres from EM micrographs and the predicted triple coil structure of fibres showing their dimensions.................................................................123
V. List of Tables

Table 1.1 Baculovirus conserved proteins. ......................................................................................8
Table 2.1 Web based bioinformatics programs used in this study. .............................................23
Table 2.2 Alkaline buffer of sodium carbonate-bicarbonate mixtures. ......................................28
Table 2.3 PCR reaction system. ..................................................................................................30
Table 2.4 PCR thermocycling program. .......................................................................................30
Table 2.5 Double enzyme digestion system. ................................................................................32
Table 2.6 Ligation system. ..........................................................................................................33
Table 2.7 Cell lysis buffer ..........................................................................................................38
Table 2.8 Four trials of transfection experiment for each bacmid DNA construct.......................43
Table 2.10 Protein purification buffers. .......................................................................................50
Table 2.11 Resolving gel recipe for 12 pieces of gel. ..................................................................51
Table 2.12 Stacking gel recipe for 4% for 12 pieces of gel. ..........................................................51
Table 2.13 10 × SDS running buffer .........................................................................................52
Table 2.14 4 × SDS loading buffer. ............................................................................................52
Table 2.15 Coomassie staining solution ......................................................................................53
Table 2.16 Coomassie destaining solution ..................................................................................53
Table 2.17 X-ray diffraction data collection technical details at the Australian Synchrotron. 55
Table 3.1 PEP proteins. ..............................................................................................................63
Table 3.2 PEP proteins in protein clusters. ..................................................................................64
Table 3.3 Constructs of CpGV-PEP-L ..........................................................................................69
Table 3.4 Four constructs of the PEP-N domains of CpGV-PEP-M and CpGV-PEP-S ..........80
Table 3.5 Seven constructs of CpGV-PEP-L, PEP-M, PEP-S co-expression. .............................82
Table 3.6 One construct of CpGV-PEP-L, PEP-M, PEP-S, polyhedrin co-expression .................86
Table 3.7 Ten constructs of AcMNPV-PEP tested .......................................................................87
Table 5.1 Predicted protein kinases and phosphatases in AcMNPV and CpGV genomes .......102
Table 5.2 PTP-2 data collection and refinement statistics ...........................................................111
### VI. Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A&lt;sub&gt;280&lt;/sub&gt;</td>
<td>Absorbance at 280 nm</td>
</tr>
<tr>
<td>AcMNPV</td>
<td><em>Autographa californica</em> multiple nucleopolyhedrovirus</td>
</tr>
<tr>
<td>BLAST</td>
<td>Basic local alignment search tool</td>
</tr>
<tr>
<td>BSA</td>
<td>Bovine serum albumin</td>
</tr>
<tr>
<td>BV</td>
<td>Budded virus</td>
</tr>
<tr>
<td>CpGV</td>
<td><em>Cydia pomonella</em> (codlin moth) granulovirus</td>
</tr>
<tr>
<td>C-terminus</td>
<td>Carboxy-terminus</td>
</tr>
<tr>
<td>DMF</td>
<td>Dimethylformamide</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>DTT</td>
<td>1,4-Dithiothreitol</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>EM</td>
<td>Electron microscopy</td>
</tr>
<tr>
<td>GV</td>
<td>Granulovirus or granulosis virus</td>
</tr>
<tr>
<td>HEPES</td>
<td>4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid</td>
</tr>
<tr>
<td>IEC</td>
<td>Ion exchange chromatography</td>
</tr>
<tr>
<td>IMAC</td>
<td>Immobilised metal affinity chromatography</td>
</tr>
<tr>
<td>IPTG</td>
<td>Isopropyl β-D-1-thiogalactopyranoside</td>
</tr>
<tr>
<td>kDa</td>
<td>kilo Dalton</td>
</tr>
<tr>
<td>LB</td>
<td>Lysogeny broth</td>
</tr>
<tr>
<td>LC-MS</td>
<td>Liquid chromatography-mass spectrometry</td>
</tr>
<tr>
<td>LN&lt;sub&gt;2&lt;/sub&gt;</td>
<td>Liquid nitrogen</td>
</tr>
<tr>
<td>MBP</td>
<td>Maltose binding protein</td>
</tr>
<tr>
<td>MNPV</td>
<td>Multiple nucleopolyhedrovirus</td>
</tr>
<tr>
<td>MOI</td>
<td>Multiplicity of infection</td>
</tr>
<tr>
<td>MQ-H&lt;sub&gt;2&lt;/sub&gt;O</td>
<td>Milli-Q filtered water</td>
</tr>
<tr>
<td>MW</td>
<td>Molecular weight</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance</td>
</tr>
<tr>
<td>NPV</td>
<td>Nucleopolyhedrovirus or nucleopolyhedrosis virus</td>
</tr>
<tr>
<td>N-terminus</td>
<td>Amino-terminus</td>
</tr>
<tr>
<td>OB</td>
<td>Occlusion body</td>
</tr>
<tr>
<td>OD&lt;sub&gt;600&lt;/sub&gt;</td>
<td>Optical density at 600 nm</td>
</tr>
<tr>
<td>ODV</td>
<td>Occlusion derived virus</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>OpMNPV</td>
<td><em>Orgyia pseudotsugata</em> multiple nucleopolyhedrovirus</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PdNPV</td>
<td><em>Porthetria dispar</em> (gypsy moth) nucleopolyhedrovirus</td>
</tr>
<tr>
<td>PEG</td>
<td>Polyethylene glycol</td>
</tr>
<tr>
<td>PiGV</td>
<td><em>Plodia interpunctella</em> granulovirus</td>
</tr>
<tr>
<td>rTEV</td>
<td>Recombinant tobacco etch virus protease</td>
</tr>
<tr>
<td>RT</td>
<td>Room temperature</td>
</tr>
<tr>
<td>SDS</td>
<td>Sodium dodecylsulfate</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>Sodium dodecylsulfate polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>SEC</td>
<td>Size exclusion chromatography</td>
</tr>
<tr>
<td>SEC-MALS</td>
<td>Size exclusion chromatography–multi-angle static light scattering</td>
</tr>
<tr>
<td>SEM</td>
<td>Scanning electron microscopy</td>
</tr>
<tr>
<td>SeMet</td>
<td>Selenomethionine</td>
</tr>
<tr>
<td>Sf9</td>
<td><em>Spodoptera frugiperda</em> cell line IPLB-Sf9</td>
</tr>
<tr>
<td>SNPV</td>
<td>Single nucleopolyhedrovirus</td>
</tr>
<tr>
<td>TAE</td>
<td>Tris-acetate-EDTA</td>
</tr>
<tr>
<td>TCEP</td>
<td>Tris-(2-carboxy-ethyl) phosphine hydrochloride</td>
</tr>
<tr>
<td>TE</td>
<td>Tris-EDTA</td>
</tr>
<tr>
<td>TEM</td>
<td>Transmission electron microscopy</td>
</tr>
<tr>
<td>TEMED</td>
<td>Tetramethylethylenediamine</td>
</tr>
<tr>
<td>Tris</td>
<td>Tris-(hydroxymethyl) aminomethane</td>
</tr>
<tr>
<td>UV</td>
<td>Ultraviolet light</td>
</tr>
</tbody>
</table>
Chapter 1 Introduction

1.1 Baculovirus

Baculoviruses are a family (Baculoviridae) of insect viruses\(^1\). Baculoviruses have been found in three orders of insects: Lepidoptera, Hymenoptera, and Diptera\(^2\). Historically, diseases now known to be caused by baculoviruses had long been associated with the economically important insect *Bombyx mori* (silkworm, Lepidoptera). Electron microscopy investigations since 1945 revealed that the virions are rod shaped and are embedded in crystalline occlusion bodies\(^2,3\). These occlusion bodies (OBs) are stable infective form of the virus, resisting harsh conditions in the environment. OBs are dissolved in the alkaline midgut of insect after feeding larvae ingest OBs, releasing virions to initiate infection. The crystalline OBs are 0.2 \(\mu\)m up to 1 \(\mu\)m in size (Figure 1.1, Figure 1.2), and there can be multiple virions (Figure 1.1) or single virion embedded in OBs\(^4-6\) (Figure 1.2).

![Figure 1.1 Electron micrographs of alphabaculovirus OpMNPV occlusion bodies.](image)

(A) Scanning electron micrograph of OpMNPV occlusion bodies. Figure reproduced from Gross et al\(^6\). (B) Thin section transmission electron micrograph of OpMNPV virions embedded in an occlusion body. Figure reproduced from Hughes et al\(^5\).
As a natural pathogen of insects, baculoviruses have been used in controlling agricultural and forest pest insects. For example, baculovirus AgMNPV (Anticarsia gemmatalis multiple nucleopolyhedrovirus) has been used for control of the soybean pest caterpillar Anticarsia gemmatalis in Brazil, baculovirus HeardNPV (Helicoverpa armigera nucleopolyhedrovirus) has been used for control of the cotton pest Helicoverpa armigera in China, and baculovirus CpGV (Cydia pomonella granulovirus) for control of the insect Cydia pomonella on pear and apple crops in North America and Europe.

Another application of baculovirus is the widely used Baculovirus Expression Vector System (BEVS) for the production of recombinant proteins. The system has been particularly successful for producing eukaryotic proteins which requires post-translational modifications for folding and activity. Until 2014, of the 62 eukaryotic integral membrane proteins for which crystal structures had been solved, 35 were produced in insect cells using BEVS technology, with the remaining being 4 in E.coli, 20 in yeast, and 3 in mammalian cells. The system exploits the high level expression of the baculovirus protein that forms the occlusion bodies but which is not essential for baculovirus reproduction. The development of the bacmid system, has far efficiently improved the baculovirus-insect cell expression technology. Bacmid is an artificial baculovirus genome that can be propagated in both E.coli...
and insect cells, and that allow efficient generation of recombinant baculovirus through Tn7 transposase mediated site specific recombiniation and by blue/white colony screening.

1.2 Life cycle and development

The life cycle of baculovirus involves two types of virions: occlusion-derived virions, and budded virions. Occlusion-derived virions (ODVs) are embedded in occlusion bodies (OBs) and stay stable outside host insect. OBs are dissolved in the alkaline condition of insect midgut cells after ingestion and release ODVs, which initiate infection of insect. BV cause systemic infection throughout the insect (Figure 1.3). ODV and BV have different proteins comprising their envelopes, consistent with their different infection tissues. ODV infect midgut epithelial cells much more efficiently than BV, and BV are much more efficient at infecting cultured insect cells.
Life cycle and development

Introduction

Figure 1.3 Baculovirus life cycle. (A) OBs are ingested by an insect and dissolved in the alkaline condition of midgut cells, following which ODV are released and initiate infection. (B) BV buds out of the cell and initiate systemic infection throughout the insect. (C) More BV are produced. (D) Late in infection ODV and OBs are produced, and the cell dies releasing the OBs. Figure reproduced from Rohrmann².

The sequence of events in baculovirus replication was determined from thin section electron microscopy of infected larvae, first described by Arnott and Smith⁴,²², and later by Harrap²³. Baculovirus DNA replication begins 6 hours after infection, BVs start to bud from cells at about 12 hours, and OBs first appear at about 20 hours after infection²⁴,²⁵.

Naked nucleocapsids are first formed within virogenic stroma, a distinct DNA rich region that forms inside the nucleus. The naked nucleocapsids then begin to acquire their bilayer membranes. Subsequently, the occlusion body matrix protein begins to deposit on the outer membrane of nucleocapsids and grows until the occlusion body is complete. At the periphery of mature occlusion body there is a layer, called the occlusion body envelope and the formation of this layer marks the termination of virus replication²³ (Figure 1.4).
Figure 1.4 Baculovirus development in the cell nucleus. (A) Nucleocapsids formation from the DNA rich virogenic stroma inside the nucleus. VS, virogenic stroma; N, nucleocapsids; EVB, membrane-bound virion; NM, nuclear membrane; CM, cell membrane. (B) OB formation on the outer membrane of nucleocapsids. Arrow indicate the deposition of the OB matrix protein. (C) A mature OB. Arrow indicate a layer at the periphery of the mature OB. Figure reproduced from Harrap\textsuperscript{23}. 
1.3 Baculovirus diversity and phylogeny

Baculoviruses were originally divided into two taxonomical groups based on the morphology of the OBs: nuclear polyhedrosis viruses (NPVs) and granuloviruses (GVs). With the advent of molecular biology and the availability of whole genome DNA sequences, the Baculoviridae has been divided into four genera: alphabaculovirus (Lepidoptera specific NPV), betabaculovirus (Lepidoptera specific GV), gemmabaculovirus (Hymenoptera specific NPV), and deltabaculovirus (Diptera specific NPV)\textsuperscript{26,27} (Figure 1.5). The widely studied AcMNPV (Autographa californica multiple nucleopolyhedrovirus) is the alphabaculovirus type species and CpGV (Cydia pomonella granulovirus) is the betabaculovirus type species.

![Baculoviridae phylogenetic tree from core genes](image)

\textit{Figure 1.5 Baculoviridae phylogenetic tree from core genes}. Figure reproduced from Jehle et al\textsuperscript{26}.
1.4 Genome and conserved proteins

Baculovirus genomes are circular double-stranded DNA sequences ranging in size of 80 to 180 kbp. A 2011 study listed 57 sequenced baculovirus genomes with the predicted number of ORFs in the range of 90 to 183\textsuperscript{28}. 38 ORFs are common in all these baculovirus genomes\textsuperscript{28,29} (Table 1.1).

<table>
<thead>
<tr>
<th>AcMNPV ORF</th>
<th>CpGV ORF</th>
<th>Gene name</th>
<th>Description</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>41</td>
<td>Lef-2</td>
<td>DNA polymerase associated factor</td>
<td>OB</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>Polyhedrin/Granulin</td>
<td>Crystal matrix protein</td>
<td>OB</td>
</tr>
<tr>
<td>14</td>
<td>74</td>
<td>Lef-1</td>
<td>DNA polymerase associated factor</td>
<td>OB</td>
</tr>
<tr>
<td>22</td>
<td>48</td>
<td>PIF-1</td>
<td><em>Per os</em> infection factor</td>
<td>ODV envelope</td>
</tr>
<tr>
<td>40</td>
<td>60</td>
<td>P47</td>
<td>RNA polymerase subunit</td>
<td>Capsid</td>
</tr>
<tr>
<td>50</td>
<td>131</td>
<td>Lef-8</td>
<td>RNA polymerase subunit</td>
<td>Capsid</td>
</tr>
<tr>
<td>53</td>
<td>134</td>
<td>Ac53</td>
<td>U-box/RING-like domain</td>
<td>Capsid</td>
</tr>
<tr>
<td>54</td>
<td>138</td>
<td>Vp1054</td>
<td>Nucleocapsid protein</td>
<td>Capsid</td>
</tr>
<tr>
<td>62</td>
<td>117</td>
<td>Lef-9</td>
<td>RNA polymerase subunit</td>
<td>Capsid</td>
</tr>
<tr>
<td>65</td>
<td>111</td>
<td>DNA pol</td>
<td>DNA polymerase</td>
<td>Capsid</td>
</tr>
<tr>
<td>66</td>
<td>112</td>
<td>Desnup</td>
<td>Nucleocapsid protein</td>
<td>Capsid</td>
</tr>
<tr>
<td>68</td>
<td>114</td>
<td>PIF-6</td>
<td><em>Per os</em> infectivity factor</td>
<td>ODV envelope</td>
</tr>
<tr>
<td>77</td>
<td>106</td>
<td>Vlf1</td>
<td>Very late gene expression factor</td>
<td>Capsid</td>
</tr>
<tr>
<td>78</td>
<td>105</td>
<td>Ac78</td>
<td>Unknown function</td>
<td>Capsid</td>
</tr>
<tr>
<td>80</td>
<td>104</td>
<td>Gp41</td>
<td>tegument protein</td>
<td>Capsid</td>
</tr>
<tr>
<td>81</td>
<td>103</td>
<td>Ac81</td>
<td>Unknown function</td>
<td>Capsid</td>
</tr>
<tr>
<td>83</td>
<td>101</td>
<td>Vp95 (Vp91)</td>
<td>Capsid protein</td>
<td>Capsid</td>
</tr>
<tr>
<td>89</td>
<td>96</td>
<td>Vp39</td>
<td>Major capsid protein</td>
<td>Capsid</td>
</tr>
<tr>
<td>90</td>
<td>95</td>
<td>Lef-4</td>
<td>RNA polymerase subunit/ capping</td>
<td>Capsid</td>
</tr>
<tr>
<td>92</td>
<td>93</td>
<td>P33</td>
<td>Sulfahydryl oxidase</td>
<td>Capsid</td>
</tr>
<tr>
<td>93</td>
<td>92</td>
<td>P18</td>
<td>Unknown function</td>
<td>Capsid</td>
</tr>
<tr>
<td>94</td>
<td>91</td>
<td>Odv-e25</td>
<td>Unknown function</td>
<td>Capsid</td>
</tr>
<tr>
<td>95</td>
<td>106</td>
<td>P143</td>
<td>DNA helicase</td>
<td>Capsid</td>
</tr>
<tr>
<td>96</td>
<td>89</td>
<td>PIF-4</td>
<td><em>Per os</em> infectivity factor</td>
<td>ODV envelope</td>
</tr>
<tr>
<td>98</td>
<td>88</td>
<td>38K</td>
<td>Capsid protein</td>
<td>Capsid</td>
</tr>
<tr>
<td>99</td>
<td>87</td>
<td>Lef-5</td>
<td>Transcription initiation factor</td>
<td>Capsid</td>
</tr>
<tr>
<td>100</td>
<td>86</td>
<td>P6.9</td>
<td>DNA condensing protein</td>
<td>Capsid</td>
</tr>
<tr>
<td>101</td>
<td>85</td>
<td>C42</td>
<td>Unknown function</td>
<td>Capsid</td>
</tr>
<tr>
<td>103</td>
<td>83</td>
<td>P48</td>
<td>Unknown function</td>
<td>Capsid</td>
</tr>
</tbody>
</table>
Occlusion body matrix protein polyhedrin

## Introduction

<table>
<thead>
<tr>
<th>ORF No.</th>
<th>AcMNPV ORF No.</th>
<th>Gene Name</th>
<th>Function</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>109</td>
<td>55</td>
<td>Odv-ec43</td>
<td>Unknown function</td>
<td>ODV envelope</td>
</tr>
<tr>
<td>115</td>
<td>35</td>
<td>PIF-3</td>
<td>Per os infectivity factor</td>
<td>ODV envelope</td>
</tr>
<tr>
<td>119</td>
<td>48</td>
<td>PIF-2</td>
<td>Per os infectivity factor</td>
<td>ODV envelope</td>
</tr>
<tr>
<td>133</td>
<td>125</td>
<td>Alk exo</td>
<td>Alkaline exonuclease</td>
<td></td>
</tr>
<tr>
<td>138</td>
<td>60</td>
<td>P74</td>
<td>Per os infectivity factor</td>
<td>ODV envelope</td>
</tr>
<tr>
<td>142</td>
<td>15</td>
<td>49K</td>
<td>Unknown function</td>
<td></td>
</tr>
<tr>
<td>143</td>
<td>14</td>
<td>Odv-e18</td>
<td>Unknown function</td>
<td>ODV envelope</td>
</tr>
<tr>
<td>144</td>
<td>97</td>
<td>Odv-e27</td>
<td>Unknown function</td>
<td>ODV envelope</td>
</tr>
<tr>
<td>148</td>
<td>18</td>
<td>Odv-e56</td>
<td>Per os infectivity factor</td>
<td>ODV envelope</td>
</tr>
</tbody>
</table>

### Table 1.1 Baculovirus conserved proteins.

The table lists 38 ORFs for which homologues occur in sequenced baculovirus genomes. The columns list the ORF numbers for AcMNPV and CpGV, the gene name, and the function and location of the corresponding protein. Table adapted from Miele et al and Garavaglia et al.²⁸,²⁹

#### 1.5 Occlusion body matrix protein polyhedrin

The crystalline OB of baculovirus protects viruses embedded within it from harsh environmental conditions, enabling them to remain viable in the environment. The OB are composed of a single protein, called polyhedrin with molecular weight of 28 kDa.³⁰–³² The protein forms crystals in vivo when polyhedrin is expressed in insect cells, but does not crystallize from protein obtained by dissolving OBs. Using novel micro-crystallographic techniques the structure of polyhedrin was determined.³²,³³ The individual polyhedrin protein has a central compact sandwich domain. In the crystal it forms trimers, four of which are tightly associated into a tetrahedral cluster, the basic building block of OBs (Figure 1.6). Two tetrahedral clusters, consisting of 24 polyhedrin molecules, constitute the 103 Å crystallographic unit cell that repeats about 200 times in each direction to form a typical 2-μm crystal.³²,³³
1.6 Occlusion body envelope

The existence of a layer at the surface of OBs has been reported frequently\textsuperscript{4,22,23,34–36} (Figure 1.7, Figure 1.8). The layer is called polyhedral envelope (PE) or calyx\textsuperscript{6,36–42}.  

\textbf{Figure 1.6 Baculovirus polyhedrin protein atomic structure and organization.} (A) Polyhedrin protein. N-terminus to C-terminus of the molecule is denoted in blue-to-red gradient of colour. (B) Polyhedrin protein trimeric organization. (C) Polyhedrin protein tetrahedral cluster. Figure reproduced from Coulibaly et al\textsuperscript{32}.
Figure 1.7 Electron micrograph of baculovirus AcMNPV OB showing surface layer. Arrows indicate the surface layer. M, polyhedral envelope; OB, occlusion body. Scale bars represent 0.1 µm. Figure reproduced from Zuidema et al16.
Figure 1.8 Electron micrograph of baculovirus TnNPV OB showing surface layer. Arrow indicates the surface layer. Scale bar represents 0.5 µm. Figure reproduced from Summers et al\textsuperscript{22}.

1.6.1 Isolated envelope structures

Two early studies reported collapsed envelope structures remained after dissolving OB in alkaline solution\textsuperscript{23,43} (Figure 1.9). The structures consist of a layer with local hexagonal symmetry. The hexagonal subunit has a central stain filled cavity about 60 Å in diameter.
Figure 1.9 Electron micrographs of envelope structures isolated from baculovirus PdMNPV OBs. (A) Sample shadowed with gold and palladium. OBs are dissolved in alkaline solution and fractions are separated. (PM) envelope. (B) Sample negative-stained with uranyl acetate. Inset is an enlarged region showing a hexagonal array of hollow structures each of ring-like subunits each with a central darker cavity. Figure reproduced from Harrap. 


1.6.2 Occlusion body envelope components

1.6.2.1 Carbohydrates

The polyhedral envelope (PE) surrounding OBs does not appear to be a lipid bilayer when viewed in thin section. It has been reported to contain carbohydrates, mainly hexose and pentose, which were suggested to be components of a polysaccharide primary structure of the envelope\(^{37}\).

1.6.2.2 Polyhedral envelope protein

The protein, polyhedral envelope protein (PEP), was found to be associated with the polyhedral envelope in AcMNPV\(^{38}\). It could be released from OBs by treatment with reducing agents and was found to be phosphorylated. Some portion of polyhedrin protein was also found to be involved with the polyhedral envelope fraction.

Electron microscopy of immuno-gold labelled AcMNPV OBs was used to show that the PEP is located at the surface of OB\(^{39}\) (Figure 1.10). OBs produced from PEP-deletion virions did not show a visible envelope\(^{36}\) (Figure 1.11).

A similar polyhedral envelope protein was detected in OpMNPV\(^{39}\). It shares a 58% identity in amino acids with the PEP of AcMNPV, and it was also found to be associated with the polyhedral envelope of OpMNPV\(^{39}\). Deletion of this protein causes the OB to have a rough, pitted surface compared to wild type OBs, which has a smooth surface\(^{6}\) (Figure 1.12).

In CpGV, the type specific betabaculovirus, there are three ORFs, ORF22, ORF20, and ORF23\(^{44}\) which all have PEP N-terminal domain predicted by sequence alignment with the PEP of OpMNPV. ORF22 also has predicted PEP C-terminal domain. This contrasts to the fact that there is a single PEP in AcMNPV and OpMNPV, which belong to alphabaculovirus. Here we term ORF22, ORF20, and ORF23 CpGV-PEP-L, CpGV-PEP-M, and CpGV-PEP-S respectively.
Figure 1.10 Immunogold labelled thin section electron micrograph of AcMNPV OB showing the location of PEP. Section of insect cells infected with AcMNPV was treated with gold labelled anti-PEP antibodies. Arrows indicate polyhedral envelope. P, OB; V, virions. Scale bar represents 0.5 µm. Figure reproduced from Lent et al\textsuperscript{40}.

Figure 1.11 Thin section electron micrograph of wild type AcMNPV and PEP-deletion OBs. (A) Insect cells infected with wild type AcMNPV. (B) Insect cells infected with recombinant AcMNPV where the PEP gene had been deleted. Note there is a change at the OB surface between wild type and PEP deletion images. Scale bars represent 0.1 µm. Figure reproduced from Zuidema et al\textsuperscript{36}.
1.6.3 Occlusion body envelope development

Electron-dense spacers (ES) (Figure 1.13) and fibrous structures (FS) are characteristic structures observed in electron micrographs of cells infected with baculovirus and have frequently been reported to be associated with the development of polyhedral envelope (PE).

1.6.3.1 Electron-dense spacers

ES structures occur in cells infected with wild type or polyhedrin-deletion baculovirus, thus precluding the possibility of polyhedrin making up these spacers\(^4\). When PEP of AcMNPV was knocked out, not only the polyhedral envelope was absent (Figure 1.11, Figure 1.13), but the electron-dense spacers were also absent\(^4\) (Figure 1.13), suggesting PEP is also a structural component of the electron-dense spacers, as well as being an integral part of the polyhedral envelope. Protein A-gold immuno-electron microscopy was used to confirm that PEP was located in the electron-dense spacers, as well as in the polyhedral envelope\(^4\).
Figure 1.13 Thin section EM of AcMNPV infected cells showing OB, ES, and FS. (A) Cells infected with wild-type AcMNPV. N, nucleus; C, cytoplasm; P, OB; FS, fibrous structures; ES, electron-dense spacers. (B) Cells infected with recombinant PEP-minus AcMNPV. (C) Cells infected with recombinant P10-minus AcMNPV (for P10, see next section). Note the PEP deletion eliminates ES, but does not affect FS. P10 deletion eliminates FS, but not ES. Scale bar represents 2 µm. Figure reproduced from Lent et al.40.
1.6.3.2 Fibrous structures

The presence of the fibrous structures in infected cells was independent of polyhedrin expression, so the fibrous structures are not composed of polyhedrin\textsuperscript{45}. Deletion of P10 eliminated these fibrous structures, suggesting P10 is a component of the fibrous structures\textsuperscript{6,46,47} (Figure 1.13). Immunogold labelling experiment was used to confirm that P10 was located almost exclusively over the fibrous structures\textsuperscript{45}.

The P10 protein, just like polyhedrin, is expressed in large quantities\textsuperscript{48}. Because fibrous structures are also abundant, P10 may be the main component of these structures.

1.6.3.3 Connections between spacers, fibrous structures and the envelope

It was thought that the electron-dense spacers could be nascent polyhedral envelope or the polymerized excess PEP\textsuperscript{49}.

While the absence of fibrous structures did not seem to affect the generation of electron-dense spacers (Figure 1.13), they do affect the polyhedral envelope. Two groups reported that the P10-minus baculovirus OBs have normal envelope\textsuperscript{6,40}. Two other groups reported that P10-minus baculovirus OBs have abnormal envelopes\textsuperscript{46,47}. However, even when the polyhedral envelope was produced in P10 protein–deleted recombinants, envelope appeared deteriorated, as surface of the OBs was pitted\textsuperscript{6} (Figure 1.14), similar to OBs produced by PEP-minus baculovirus. This suggests that the P10 protein, if not the main structural component of the polyhedral envelope, is also involved in the morphogenesis of the polyhedral envelope.

In CpGV, the P10 homologue is absent\textsuperscript{44}. The largest PEP of its three PEPs, although much larger than P10 (329 vs 137 amino acids), shares a number of motifs with P10 and is 30\% identical to AcMNPV P10, suggesting P10 in CpGV could have been fused to one of its PEPs and also suggesting the possibility of P10 being implicated in the OB morphogenesis.
1.6.4 Occlusion body envelope function

1.6.4.1 Occlusion body size and shape

The morphology of OB is likely mostly determined by the amino acid sequence of the OB matrix protein polyhedrin, as a single mutation in AcMNPV polyhedrin (AcMNPV G25D) resulted in much larger and cuboidal OBs of 5-10 µm\(^3\), compared to wild type polyhedron-shaped OBs of typically 1 µm. In another case, OBs of cuboidal shape were also found exclusively in an insect infected with a AdorGV baculovirus, and the genome sequencing of the baculovirus revealed no mutation in the polyhedron protein but rather some mutations in the polyhedral envelope protein\(^5\), suggesting PEP also plays a role in the morphogenesis of
OBs. Unusually shaped OBs including elongated OBs, compound OBs, and agglomerated OBs were found in infected insects together with normal shaped OB\textsuperscript{34,52}.

1.6.4.2 Occlusion body stability

It has also been suggested that the polyhedral envelope stabilizes OB\textsuperscript{6,36,53}, preventing them from fusion or aggregation during formation\textsuperscript{6}.

1.7 Occlusion body envelope fibres

Outside at the surface of OB, fibres have been found from baculovirus CpGV by Ken Goldie (collaborator, Biozentrum, University of Basel, Switzerland, unpublished results) (Figure 1.15). The first cryo-electron micrographs of CpGV revealed details of the envelope. The occlusion body crystalline matrix is surrounded by a uniform about 4 nm thick inner layer. Closely placed fibres extend about 15 nm from the inner layer and form the outer surface of the OB. These fibre structures at the surface of PE has hardly been noticed and was only obscurely reported once from baculovirus TnNPV\textsuperscript{22} (Figure 1.16).
Figure 1.15 Cryo-EM of CpGV baculovirus showing fibres extending from the OB surface. (Arrow 1) inner layer. (Arrow 2) fibres at the surface of the OB. Scale bar 100 nm. Image from K. Goldie.
Figure 1.16 Thin section EM of TnNPV baculovirus showing fibres on PE. Arrows indicate the fibres on the surface of OBs. Scale bars unit is µm. Figure reproduced from Summers et al.22.

1.8 Research objectives and thesis outline

Previous studies of our group discovered that the *Cydia pomonella* granulovirus have fibres extending from the OB surface. These fibres are about 15 nm long, and extend from an inner layer attached to the crystalline matrix of polyhedrin protein that surrounds the embedded virion.

Prior to the start of the project, other members of our group investigated the fibres by isolating them and analysing the components by mass spectrometry. The results showed that the fibres might contain all the three polyhedral envelope proteins (PEPs), the CpGV-PEP-L, CpGV-PEP-M, and CpGV-PEP-S, but the result is still preliminary. The atomic structure of the fibres is yet to be determined.
In project one I will investigate the polyhedral envelope proteins of the type specific betabaculovirus CpGV and the type specific alphabaculovirus AcMNPV, i.e., the CpGV-PEP-L, CpGV-PEP-M, CpGV-PEP-S, and the AcMNPV-PEP, by expressing and purifying the recombinant PEPs, with the aim to characterise the proteins, to reconstitute the fibre, and to crystalize the proteins. These results are presented in Chapter 3.

I will also investigate the structure of the intact fibres using cryo-electron microscopy. This work included studying the fibres attached to OBs and experiments to produce isolated fibres more suitable for high resolution analysis. These results are presented in Chapter 4.

Chapter 5 describes the expression, purification, crystallization, and X-ray crystallographic structure determination of the CpGV phosphatase protein PTP-2.
Chapter 2 Materials and Methods

2.1 Bioinformatics

Nucleotide sequences of genes were retrieved from GenBank database. Amino acid sequences of proteins were obtained from UniProt database. The Web based computer programs used in the study are listed in Table 2.1.

<table>
<thead>
<tr>
<th>Program</th>
<th>Purpose</th>
<th>Web link</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLAST</td>
<td>Searches of homologous genes and homologous proteins against Genebank</td>
<td><a href="https://blast.ncbi.nlm.nih.gov/Blast.cgi">https://blast.ncbi.nlm.nih.gov/Blast.cgi</a></td>
</tr>
<tr>
<td></td>
<td>database</td>
<td></td>
</tr>
<tr>
<td>Clustal Omega</td>
<td>Multiple sequence alignment to find conserved amino acids</td>
<td><a href="https://www.ebi.ac.uk/Tools/msa/clustalo/">https://www.ebi.ac.uk/Tools/msa/clustalo/</a></td>
</tr>
<tr>
<td>EMBOSS Needle</td>
<td>Pairwise sequence alignment to find percentage of identity and similarity</td>
<td><a href="https://www.ebi.ac.uk/Tools/psa/emboss_needle/">https://www.ebi.ac.uk/Tools/psa/emboss_needle/</a></td>
</tr>
<tr>
<td>ProtParam</td>
<td>Computation of physical and chemical parameters for a given protein</td>
<td><a href="https://web.expasy.org/protparam/">https://web.expasy.org/protparam/</a></td>
</tr>
<tr>
<td>PSIPRED</td>
<td>Prediction of protein secondary structure</td>
<td><a href="http://bioinf.cs.ucl.ac.uk/psipred/">http://bioinf.cs.ucl.ac.uk/psipred/</a></td>
</tr>
</tbody>
</table>

Table 2.1 Web based bioinformatics programs used in this study.

2.2 Sample preparation of in vivo grown granulovirus OBs.

The CpGV occlusion bodies were a gift from J.A. Jehle, and were purified as follows. OBs of CpGV (Mexican strain) were produced by infection of fourth instar larvae of C. pomonella and purified as follows. OBs from raw larval homogenates or from commercial preparations were filtered through cotton wool and repeatedly washed by centrifugation (18,000 g, 30 min) followed by resuspending the pellet in water. The concentrated OB suspension was then purified using a discontinuous glycerol gradient 30–80% (vol/vol). The OB band was collected, resuspended in water, and pelleted again (3,200 × g, 45 min). If the required purity was not achieved, the glycerol gradient purification was repeated. The collected OBs were
washed three by centrifugation (3,200 g, 45 min) followed by resuspending the pellet in water. The OB suspension was stored at −18 °C until use.

2.3 Bacterial culture

2.3.1 E.coli Strains

The DH5α (Invitrogen) strain was used for cloning in this study. The BL21 (DE3) (Novagen), Rosetta™ (DE3) (Novagen), LOBSTER (Kerafast), Shuffle® (NEB) strains were used for protein expression.

2.3.2 Media

1. Lysogeny Broth (LB).

LB medium was used for cultivation of bacteria cells and expression of proteins. LB was purchased as a pre-mixed base powder (Lennox L broth base, Invitrogen) and was mixed with water according to the manufacturer’s directions. The final formulation in 1 liter medium contained 10 g peptone, 5 g yeast extract, and 5 g NaCl. The medium was autoclaved and cooled before appropriate antibiotics were added.

2. LB agar.

LB agar was used for growing bacterial colonies and selection of recombinant clones. LB agar was made by adding 15 g agar (Invitrogen) in 1 liter LB medium. It was autoclaved and cooled to about 60°C before appropriate antibiotics were added.

2.3.3 Antibiotic stock solutions and reagents

Antibiotics stock solutions were prepared in Milli-Q™-H2O, filter-sterilized, and stored at -20 °C. They were diluted in 1:1000 ratio when used. Chloramphenicol and Tetracycline were dissolved in organic liquids (see below) and need not filtering.

Ampicilin: 100 mg/mL

Kanamycin: 50 mg/mL
Gentamicin: 7 mg/mL

Chloramphenicol: 34 mg/mL in 70% ethanol

Tetracycline: 10 mg/mL in 70% ethanol

2.3.4 Procedure

Transformed competent cells were spread on a LB agar plate containing appropriate antibiotics, and were grown at 37°C overnight. The next morning, a single colony was picked and used to inoculate 4 mL LB medium with the appropriate antibiotics. This inoculated LB medium was grown at 37°C overnight. Aliquots from the overnight culture was preserved in 10% glycerol as a stock (mixing 800 µL of the culture with 200 µL 20% glycerol) and stored at -80°C.

The remaining overnight culture was used either for plasmid extraction or for scaled up culture for protein expression (See protein expression section).

2.4 Chemically competent cells

2.4.1 Reagents

The following amounts of reagents were for 100 aliquots of competent cells, 50 µL per aliquot. These were autoclaved and chilled on ice prior to use.

MgCl₂: 0.1M, 50 mL

CaCl₂: 0.1M, 50 mL

CaCl₂: 0.1M, 50 mL (10% glycerol)

2.4.2 Procedure

1. 100 mL of cell culture was grown until OD₆₀₀ reached 0.3-0.5.

2. The cell culture was chilled on ice for about 10 min, divided into two 50 mL Falcon tubes, and were centrifuged at 3000 g at 4°C for 5 min.
3. The culture medium was poured off or aspirated and the cell pellet was gently resuspended in 10 mL of ice-cold 0.1M MgCl₂ for each Falcon tube. The resuspension was centrifuged at 3000 g at 4℃ for 2 min.

4. The supernatant was poured off or aspirated and the cell pellet was gently resuspended in 10 mL of ice-cold 0.1M CaCl₂ for each Falcon tube.

5. The resuspension was put on ice for 15-30 min and then was centrifuged at 3000 g at 4℃ for 1 min.

6. The supernatant was poured off or aspirated and the cell pellet was gently resuspended in 2.5 mL of ice-cold 0.1M CaCl₂ (10% glycerol) for each Falcon tube.

7. The resuspensions from the two Falcon tubes were aliquoted to 100 Eppendorf tubes, 50 µL per tube.

8. The aliquots were snap-frozen in liquid nitrogen.

9. The aliquots were used immediately or stored at -80℃ for later use.

2.5 Electrocompetent cells

2.5.1 Reagents

The following amounts of reagents were for 50-100 aliquots of competent cells, 50 µL per aliquot. These were autoclaved and chilled on ice prior to use.

Milli-Q® water: 50 mL

Glycerol solution: 20% (v/v) containing 1.5% (w/v) mannitol, 50 mL

2.5.2 Procedure

1. 100 mL of cell culture was grown until OD₆₀₀ reached 0.5-0.7.

2. The cell culture was chilled on ice for about 10 min, divided into two 50 mL Falcon tubes, and were centrifuged at 3000 g at 4℃ for 5 min.
3. The supernatant was discarded and the cell pellet was gently suspended in 10 mL of ice-cold Milli-Q® water for each tube.

4. A distinct layer of glycerol solution was made beneath the resuspension by drawing about 12 mL of the glycerol solution into a disposable plastic pipette and placing the tip of the pipette to the base of the resuspension and slowly dispensing 10 mL of the glycerol solution. The last 2 mL was not dispensed to avoid disturbing the density interface which blocks the culture medium and ions while the bacteria was forced through it by centrifugation.

5. The tubes were centrifuged at 3000 g at 4°C for 5 min, with slow acceleration and deceleration.

6. The culture medium and the glycerol layer following behind, were removed by aspiration.

7. The cell pellet was resuspended in 1-2 mL of fresh glycerol solution (20% (v/v) glycerol solution, 1.5 % (w/v) mannitol). 1 mL was used for each tube if the starting OD$_{600}$ = 0.5; 2 mL for each tube if OD$_{600}$ = 0.7.

8. The resuspensions from the two Falcon tubes were aliquoted to 50-100 Eppendorf tubes, 50 µL per tube.

9. The aliquots were snap-frozen in liquid nitrogen.

10. The aliquots were used immediately or stored at -80°C for later use.

2.6 Genome DNA extraction from occlusion bodies

2.6.1 Reagents

Alkaline buffer: a sodium carbonate-bicarbonate buffer was made according to the method described by Delory et al$^{56}$. Mixing varying volumes of the two gives different pH values, as listed in Table 2.2.
Genome DNA extraction from occlusion bodies

Materials and Methods

<table>
<thead>
<tr>
<th>0.1 M Na$_2$CO$_3$ (mL)</th>
<th>0.1 M NaHCO$_3$ (mL)</th>
<th>pH (20°C)</th>
<th>pH (37°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9</td>
<td>9.16</td>
<td>8.77</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>9.40</td>
<td>9.12</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>9.51</td>
<td>9.40</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>9.78</td>
<td>9.50</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>9.90</td>
<td>9.72</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>10.14</td>
<td>9.90</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>10.28</td>
<td>10.08</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>10.53</td>
<td>10.28</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>10.83</td>
<td>10.57</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>11.8</td>
<td>Not tested</td>
</tr>
</tbody>
</table>

Table 2.2 Alkaline buffer of sodium carbonate-bicarbonate mixtures. Table adapted from Delory et al\textsuperscript{56}.

proteinase K: 20 mg/mL

Phenol

NaOAc: 3 M

70% ethanol

Milli-Q water: autoclaved

2.6.2 Procedure

1. 100 µL of CpGV occlusion body suspension was pelleted at 10,000 g at R.T. for 2 min. The pellet was resuspended and lysed in 250 µL of alkaline buffer, pH 10.9.

2. The lysed solution was added to with 2.5 µL of proteinase K (20 mg/mL) and incubated at 50°C for 2 h.

3. Then the solution was added to with 500 µL of phenol, gently inverted a few times, and centrifuged at 15,000 g at R.T. for 10 min. The upper layer (water layer containing genomic DNA) was aspirated gently to a new Eppendorf tube without disturbing the water-phenol interface.
4. This phenol extraction step was repeated once.

5. The upper water layer was transferred to another new tube. 10% volume of NaOAc (3M) and 2 volumes of absolute ethanol was added to the water solution. The solution was put at -20°C for 1 h to overnight to precipitate the genomic DNA.

6. The precipitated DNA was centrifuged at 15,000 g at R.T. for 10 min. The supernatant was removed, and the DNA precipitation was resuspended in 70% ethanol and centrifuged at 15,000 g at R.T. for 10 min. The 70% ethanol was removed.

7. This 70% ethanol wash step was repeated once. The 70% ethanol was removed.

8. The DNA was air-dried on the bench for about 10 min (it was not over-dried for the DNA would be very hard to dissolve again).

9. The genomic DNA was dissolved in 40 µL of autoclaved Milli-Q water and was stored at -20°C.

2.7 Gene cloning
Gene cloning was done in this sequence: amplification of target genes by polymerase chain reaction (PCR), PCR product purification by agarose gel electrophoresis, restricted digestion of PCR product and plasmid, ligation, transformation into cloning strain, plasmid extraction and gene sequencing. The recombinant plasmid was then used to transform bacterial expression strains or to transpose baculovirus bacmid for downstream protein expression.

2.7.1 PCR

2.7.1.1 Materials
The PCR machine (Eppendorf), DNA Polymerase and other essential components were supplied in a commercial kit (TaKaRa).

2.7.1.2 Reaction system and program
Amplification for each gene was duplicated to retain enough amount of PCR product after agarose gel extraction prior to ligation step.
The reaction system was set up in a thin-walled PCR tube (Table 2.3), and the program was set up as outlined in Table 2.4.

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume (µL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ddH2O</td>
<td>34</td>
</tr>
<tr>
<td>5 × PrimeSTAR buffer (Mg2+ plus)</td>
<td>10</td>
</tr>
<tr>
<td>dNTP mixture (2.5 mM each)</td>
<td>4</td>
</tr>
<tr>
<td>Primer 1 (100 µM)</td>
<td>0.5</td>
</tr>
<tr>
<td>Primer 2 (100 µM)</td>
<td>0.5</td>
</tr>
<tr>
<td>Gene template</td>
<td>0.5</td>
</tr>
<tr>
<td>PrimeSTAR HS DNA Polymerase</td>
<td>0.5</td>
</tr>
<tr>
<td>Total</td>
<td>50</td>
</tr>
</tbody>
</table>

Table 2.3 PCR reaction system.

<table>
<thead>
<tr>
<th>Step</th>
<th>Temperature</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial denaturation</td>
<td>98°C</td>
<td>30 seconds</td>
</tr>
<tr>
<td>30 cycles of denaturation, annealing, and extension</td>
<td>98°C, 60°C, 72°C</td>
<td>10 seconds, 5 seconds, 1 minute / kb</td>
</tr>
<tr>
<td>Final extension</td>
<td>72°C</td>
<td>10 minutes</td>
</tr>
<tr>
<td>Hold</td>
<td>10°C</td>
<td>Until use</td>
</tr>
</tbody>
</table>

Table 2.4 PCR thermocycling program.

2.7.2 Agarose gel electrophoresis

Agarose gel electrophoresis was used to visualize or extract DNA products.

2.7.2.1 Reagents

10 × TAE buffer: 0.4 M Tris-acetate, 0.1 M EDTA, pH 8.0
Gene cloning

Materials and Methods

6 × DNA loading dye: 0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol FF, 30% (v/v) glycerol.

2.7.2.2 Procedure

1 % (w/v) agarose (Bio-Rad) was dissolved in 1 × TAE buffer by microwaving. The solution was poured into a gel apparatus. 5 µL Sybr® Safe DNA gel stain (Bio-Rad) was added to the gel apparatus and stirred by a pipet tip to evenly spread the stain before the gel set. After the gel set in about 30 min at R.T., the gel apparatus including the gel was placed in an electrophoresis apparatus (Hoefer HE 33) and covered with 1 × TAE buffer. DNA samples were mixed with 6 × DNA loading dye and loaded into wells. 1 Kb Plus DNA ladder (Invitrogen) was also loaded into an individual well to allow estimation of DNA fragment sizes. Gels were run at 100 V for about 20 min and the DNA was visualized using a Gel Doc™ UV transillumination system (Bio-Rad).

2.7.3 Agarose gel extraction

DNA samples were run in 1 × TAE buffer in an electrophoresis apparatus in the manner described in section 2.6.2 (above). The DNA bands were then revealed under long wavelength UV light (360 nm UV light) in a Gel Doc™ UV transillumination system (Bio-Rad) and were excised using a scalpel blade under the UV light. When cutting the gel, protections against any potential harm were taken by wearing goggles, gloves and long-sleeve lab coat to avoid any part of the body being exposed to the UV light.

Cut DNA bands were extracted from the agarose gel using Gel Extraction Kit (Qiagen) following the manufacturer’s instructions, with the modification that the TE buffer in the elution step was replaced by autoclaved Milli-Q water for wider downstream application. In the elution step, duplicates of each DNA sample were applied to one column membrane and ended in one elution of 40 µL for concentration.

2.7.4 Restriction enzyme digestion

Double enzyme digestion was carried out on the agarose gel recovered DNA samples to generate sticky ends on the DNA fragments for easy ligation. Restriction enzymes were
Gene cloning

Materials and Methods

purchased from Roche or NEB. Reaction was done in a water bath or on a heat block at 37°C for 3-4 hours. Reaction system was set up as outlined in Table 2.5.

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume (µL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ddH2O</td>
<td>6</td>
</tr>
<tr>
<td>10 × buffer</td>
<td>5</td>
</tr>
<tr>
<td>Gene or plasmid</td>
<td>35</td>
</tr>
<tr>
<td>Enzyme 1</td>
<td>2</td>
</tr>
<tr>
<td>Enzyme 2</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>50</td>
</tr>
</tbody>
</table>

Table 2.5 Double enzyme digestion system.

The products went through again agarose gel purification before subsequent ligation step. The purified DNA samples were used for ligation and the remaining was stored at -20°C freezer for future use.

2.7.5 Ligation

Ligation was carried out in a 5 µL system to save the digested plasmid for more times’ use on the same double enzyme digestion. T4 DNA ligase was purchased from TaKaRa company. Ligation was done at 18°C for 1 h in a 5 µL system (Table 2.6).

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume (µL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 × T4 DNA ligase buffer</td>
<td>0.5</td>
</tr>
<tr>
<td>T4 DNA ligase</td>
<td>0.5</td>
</tr>
</tbody>
</table>

32
Plasmid 3
Gene 1
Total 5

Table 2.6 Ligation system.

2.7.6 Transformation of recombinant DNA into bacterial cloning strains

5 µL of ligation product, or 2 µL of plasmid, was mixed into 50 µL of chemically competent DH5α cells. The cells were put on ice for 30 min without shaking. The cells were then heated at 42°C for 45 s in a water bath, and were chilled on ice for 2 min after heat-shock. 800 µL of LB medium (without antibiotics) was pipetted into the mixture and the mixture was incubated at 37°C on a shaker with 180 rpm for 1 h. For plasmid transformation, 10 µL-50 µL of cell culture was spread on a LB agar plate with appropriate antibiotics. For ligation product transformation, this 800 µL cell culture was centrifuged and resuspended in 50 µL medium and was all spread on a LB agar plate with appropriate antibiotics. The LB agar plate was put at 37°C overnight to let the colonies grow.

2.7.7 Colony PCR

Colony PCR was done using colonies from LB agar plate as PCR templates to save the bacteria culture and plasmid extraction steps before a positive colony had been confirmed.

A colony was picked up using a pipet tip to a PCR tube containing 20 µL LB medium. The medium in the PCR tube was stirred by the pipet tip to disperse the bacteria in the colony. 5-10 colonies were prepared in this manner and screened by PCR. 1 µL of the LB medium containing the colony was used as the template for a PCR reaction, and the remaining of the LB medium was kept at 4°C to later inoculate 4 mL of LB medium with appropriate antibiotics if the colony was found to be positive by PCR and agarose gel electrophoresis.

The colony PCR products then went through agarose gel electrophoresis for visualization.
2.7.8 Plasmid extraction

4 mL of LB medium with appropriate antibiotics was inoculated with either the bacteria stock or the LB medium containing the positive colony. The medium was then cultured at 37°C on a shaker with 180 rpm overnight. Plasmid was extracted using MiniPrep Plasmid Purification Kit (Qiagen) following the manufacturer’s instructions, with the modification that the TE buffer supplied was replaced by autoclaved Milli-Q water for wider downstream applications. The plasmid was stored at -20°C.

2.7.9 Sequencing

10 µL of plasmid was sent for in-house sequencing to confirm there is no mutation in the inserted gene sequence.

2.8 Transformation into bacterial expression strain

Transformation of recombinant plasmid into bacterial expression strain was done in the same manner as transforming into cloning strain (Section 2.6.6), except that bacterial expression strains were used instead of cloning strains.

2.9 Small scale protein expression in bacteria

A little of the frozen stock from the surface was scraped with a sterile pipet tip and was used to inoculate 100 mL LB medium with appropriate antibiotics in 500 mL flask. The flask was covered with aluminium foil. This was grown at 37°C on a shaker at 180 rpm for 3-4 hours until OD600 reached 0.6-0.8. The culture was cooled at 18°C for 30 min before IPTG was added to a final concentration of 0.25 mM. This continued to be grown at 18°C on a shaker at 180 rpm for 16-20 hours for recombinant protein to be expressed. The cells were harvested by centrifugation at 4000 g at 4°C for 10 min and the pellets were either used immediately or frozen at -20°C for later use.
2.10 Large scale protein expression in bacteria

Procedure for protein large expression was similar to that for protein test expression, with the difference of including the scale-up step from 100 mL culture to 500 mL culture. Specifically, it was as follows.

A little of the frozen stock from the surface was scraped with a sterile pipet tip and was used to inoculate 100 mL LB medium with appropriate antibiotics in 500 mL flask. The flask was covered with aluminium foil. This was grown at 37°C on a shaker at 180 rpm overnight. The next morning, 5 mL overnight culture was used to inoculate 500 mL LB medium with appropriate antibiotics in 2-L baffled or unbaffled flask covered with aluminium foil. The medium was grown at 37°C on a shaker at 180 rpm for 3-4 hours until OD_{600} reached 0.6-0.8. The culture was cooled at 18°C for 30 min before IPTG was added to a final concentration of 0.25 mM. This continued to be grown at 18°C on a shaker at 180 rpm for 16-20 hours for recombinant protein to be expressed. The cells were harvested by centrifugation at 4000 g at 4°C for 10 min and the pellets were either used immediately or frozen at -20°C for later use.

2.11 Large scale SeMet-protein expression in bacteria

Bacterial cells expressing SeMet protein were cultured in a different medium than LB medium.

2.11.1 Reagents

1. 700 mL 1× M9 salt solution was made in a 2-L baffled or unbaffled flask:

Milli-Q water: 700 mL

\( \text{Na}_2\text{HPO}_4\cdot12\text{H}_2\text{O} \) (358.1 Da): 10.64 g

\( \text{KH}_2\text{PO}_4 \) (136 Da): 2.4 g

\( \text{NaCl} \) (58.5 Da): 0.4 g

\( \text{NH}_4\text{Cl} \) (53.49 Da): 0.8 g
The solution was autoclaved.

2. 20 % (w/v) glucose solution (for 10 2-L flasks):

Glucose: 100 g or Glucose·H₂O 110 g

Less than 500 mL of Milli-Q water was added to make a final solution of 500 mL. The solution was filtered through 0.22 µm membrane. The solution was stored at R.T.

3. Vitamin B1 (Thiamine) solution (for 80 2-L flasks):

0.12 g of vitamin B1 (Thiamine) was dissolved in 40 mL Milli-Q water. The solution was filtered through a 0.22 µm syringe filter. The solution was stored at -20°C.

4. MgSO₄: 1 M, 40 mL

5. 10,000 × metal solution:

CaCl₂: 1 M, 2 mL

ZnSO₄: 1 M, 1 mL

CoCl₂: 0.2 M, 1 mL

NiCl₂: 0.2 M, 1 mL

H₃BO₃: 0.1 M, 2 mL

93 mL of Milli-Q water was added to make a final solution of 100 mL. The solution was autoclaved and stored at R.T.

6. FeCl₃ solution:

0.27 g FeCl₃-6H₂O was dissolved in 10 mL Milli-Q water containing 100 µL of 37% (w/w) (i.e., 12 M) HCl. The solution was stored at R.T. and protected from light.

7. 6 amino acids mixture solution (for 12.5 2-L flasks):

Lysine: 1 g

Phenylalanine: 1 g
Large scale SeMet-protein expression in bacteria

Materials and Methods

Threonine: 1 g

Isoleucine: 0.5 g

Leucine: 0.5 g

Valine: 0.5 g

96 mL Milli-Q water was added to the beaker to make a final solution of 100 mL. The dissolution of these amino acids was accelerated on a magnetic stirrer (care was taken not to turn on the heat function of the stirrer). The solution was divided into two 50-mL Falcon tubes and stored at -20°C.

8. L-selenomethionine (L-SeMet) solution (for 25 2-L flasks):

1 g of L-selenomethionine was added to 50 mL Milli-Q water. The dissolution of the amino acid was accelerated on a magnetic stirrer (care was taken not to turn on the heat function of the stirrer). The solution was stored at -20°C.

2.11.2 Procedure

1. A little amount of BL21 (DE3) frozen stock containing the recombinant plasmid of interest was scraped from the surface with a sterile pipet tip and was used to inoculate 50 mL LB medium with appropriate antibiotics and was grown at 37°C on a shaker at 180 rpm overnight.

2. The next morning, 5 mL overnight culture was added to the 700 mL 1 × M9 salt solution in the 2-L baffled or unbaffled flask. The following items were also added to the M9 salt solution: 50 mL of 20 % (w/v) glucose solution, 500 µL of Vitamin B1 (Thiamine) solution, 750 µL of MgSO₄ solution, 37.5 µL of metal solution, 37.5 µL of FeCl₃ solution, and appropriate antibiotics (i.e., 750 µL of 50 mg/mL Kanamycin or 750 µL of 100 mg/mL Ampicillin).

3. This medium was incubated at 37°C on a shaker at 180 rpm for 6-8 hours until OD₆₀₀ reached 0.6-0.8.
4. The culture was cooled at 18°C for 30 min and then the following the items were added to the culture: 8 mL of 6 amino acids mixture solution, and 2 mL of L-selenomethionine (L-SeMet) solution.

5. The culture was incubated at 18°C on a shaker at 180 rpm for 15 min to allow absorption of these added amino acids.

6. 200 µL of 1 M IPTG was added to the culture to allow protein induction and the culture was grown at 18°C on a shaker at 180 rpm overnight for recombinant protein to be expressed.

7. Next morning, the cells were harvested by centrifugation at 4000 g at 4°C for 10 min and the pellets were either used immediately or frozen at -20°C for later use.

### 2.12 Bacterial cell lysis

Bacterial cell lysis was performed using sonication (Qsonica sonicators) method in this study.

#### 2.12.1 Cell lysis buffer

<table>
<thead>
<tr>
<th>Buffer Type</th>
<th>Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>HEPES-NaOH lysis buffer</td>
<td>50 mM HEPES-NaOH pH 7.0, 200 mM NaCl, 20 mM imidazole.</td>
</tr>
<tr>
<td>Tris-HCl lysis buffer</td>
<td>20 mM Tris-HCl pH 8.0, 200 mM NaCl, 20 mM imidazole.</td>
</tr>
<tr>
<td>Alkaline lysis buffer</td>
<td>0.1 M Na₂CO₃-0.1 M NaHCO₃ pH 10.8, 200 mM NaCl, 20 mM imidazole.</td>
</tr>
</tbody>
</table>

Table 2.7 Cell lysis buffer.

#### 2.12.2 Procedure

1. Small scale cell culture sonication

100 mL of cell culture was divided into 2 50-mL Falcon tubes and centrifuged at 4000 rpm, 4 °C for 10 mins. The pellets were resuspended in a total of 10 mL of lysis buffer. The suspension was added to with lysozyme to final concentration of 1 mg/mL and put on ice and
water mixture for 15 min. The suspension was sonicated while it was being put on ice and water using the following parameters: large sonicator tip, 20-30W, 1 s on / 3 s off per round, 90 rounds.

2. Large scale cell culture sonication

Cell pellets of $8 \times 500$ mL culture was resuspended in 100 mL of lysis buffer. The suspension was added to with lysozyme to final concentration of 1 mg/mL and put on ice and water mixture for 15 min. The suspension was divided into $3 \times 50$ mL Falcon tubes. The suspension in each Falcon tube was sonicated while it was being put on ice and water using the following parameters: large sonicator tip, about 30W, 1 s on / 3 s off per round, 180 rounds.

2.13 Insect cell culture

2.13.1 Sf9 cell culture procedure

*Spodoptera frugiperda* Sf9 (Invitrogen) cells were cultured in serum-free medium Sf-900 II SFM (Invitrogen) in suspension culture. To allow optimal oxygen diffusion, cell culture volume occupied 1/4 or 1/5 volume of the Erlenmeyer shaker flask with loosened cap.

As a cell seed stock, 25 mL cell culture was grown in a 100 mL shaker flask at 27 °C on a shaker at 120 rpm with a starting cell density of $0.5 \times 10^6$ cells/mL. The culture was subcultured every 4 days into a new flask when it should have reached a cell density of $4 \times 10^6$ cells/mL. Antibiotics or antimycotics are not specifically recommended for SFM culture. If needed, the following antibiotics and antmycotic could be used: Penicilin-Streptomycin (Thermo Fisher Scientific), with a working concentration of 100 U/mL of penicillin and 100 µg/mL of streptomycin, Amphotericin B (Thermo Fisher Scientific) with a working concentration of 0.25 µg/mL.

Large volume cell culture, i.e., 250 mL and 500 mL, was grown in 1 L and 2 L Erlenmeyer shaker flasks in the same manner as the seed stock described above.
2.13.2 Sf9 cell counting

Cell counting during the cell growth days and at the subculture day was done in a 35 mm-diameter dish using light microscope by estimating the approximate cell confluence percentage per mL of culture in the dish. This cell counting method utilizes the fact that a 35 mm-diameter dish accommodates about $2.0 \times 10^6$ cells at around 100% confluence for a monolayer culture. Specifically, cell counting was carried out as follows.

1 mL of SFM medium was added to a 35 mm-diameter dish, and 500 µL of cell culture was aspirated into the medium. The dish was shaken side to side a few times to let the cells evenly distribute. Then the dish was left still for about 10 minutes to let the cells settle. By then the cells should have become a monolayer. If this monolayer has a 100% confluence, then the cell density being counted is $2.0 \times 10^6 \div 0.5 = 4.0 \times 10^6$ cells/mL. If the monolayer has less than 100% confluence, say it has X% confluence, then the cell density being counted can be estimated as $2.0 \times 10^6 \times X\% \div 0.5 = 4.0 \times X\% \times 10^6$ cells/mL. Or another round of cell counting can be done by adjustively adding more cell culture to a new dish containing 1 mL of SFM medium, to make the confluence as near 100% as possible. This method avoids using the hemocytometer and has the advantage of not only easily getting an idea of cell density, but at the same time being able to examine cell morphology and viability, i.e., whether the cells are clumped or not and whether there was bacteria or fungi contamination in the cell culture.

2.14 Baculovirus bacmid transposition

The transposition was done in a similar manner as the transformation, except with the following modifications. 2 µL donor plasmid was used to mix with 50 µL chemically competent E.coli DH10Bac cells. The cells were put on ice for 30 min without shaking. The cells were then heated at 42°C for 45 s in a water bath, and were put on ice for 2 min after the heat-shock. 800 µL of LB medium (without antibiotics) was pipetted into the mixture and the mixture was incubated at 37°C on a shaker at 180 rpm for 4 h. The transformed DH10Bac cells were spread on the LB agar plate containing three antibiotics (tetracycline, gentamicin, kanamycin) where 10 µL of 1M IPTG and 40 µL of 20 mg/mL X-gal (dissolved in DMF) had been mixed and spread enabling white-blue colony screening. For each transposition, three dilution gradients (5 µL, 10 µL, 50 µL) of the transformed DH10Bac cells were spread on three LB agar plates to make easy the selection of white colonies. The LB agar plates were
Recombinant bacmid DNA isolation

incubated at 37 °C for 24–48 h because the colonies grew slower and white colonies were not discernible from blue ones prior to 24 h.

A white colony was picked up and grown in LB medium containing the three antibiotics overnight. The next morning, the culture was streaked on a new LB agar plate containing the three antibiotics and IPTG and X-gal, to ensure the white phenotype of the selected colony. A white colony on the streaked LB agar plate was picked to inoculate 3 mL of LB medium containing the three antibiotics and this was grown at 37°C for 24 h. The DH10Bac cell stock containing the transposed bacmid was made by mixing 800 µL of the culture with 200 µL 50% glycerol and stored at -80°C.

2.15 Recombinant bacmid DNA isolation

Bacmid DNA isolation was performed in a similar manner to plasmid extraction using MiniPrep Plasmid Purification Kit (Qiagen), except with one change: instead of using membrane spin column devised for plasmid purification step, bacmid DNA was purified by isopropanol precipitation because bacmid is too large (about 130 kbp) to pass through the column membrane. Specifically, bacmid DNA was isolated in the procedure as below.

1. A little of frozen DH10Bac cell stock contain the recombinant bacmid was inoculated to 3 mL LB medium containing the three antibiotics and was grown at 37°C on a shaker at 180 rpm for 24 h.

2. The culture was centrifuged at 12,000 g for 2 min and the supernatant was discarded. The pellet was treated with Solution I, Solution II, Solution III reagents in the same way as instructed in the plasmid extraction kit.

3. The mixture was centrifuged at 14000g-15000g for 10-15 min. Then the supernatant (about 700 µL) was gently transfered to a new Eppendorf tube containing 800 µL absolute isopropanol, without disturbing the white precipitate material. This was mixed a few times by gently inverting the microcentrifuge tube.

4. The mixture was placed on ice for 5-10min. At this stage, the sample could be stored at -20 °C overnight.

5. The sample was centrifuged at 14000g-15000g for 15min at room temperature.
Insect cell transfection with recombinant bacmid DNA

6 The supernatant was removed. 500 µL 70% ethanol was added to the tube to wash the pellet by inverting a few times.

7. The sample was centrifuged at 14000g-15000g for 5 min at R.T.

8. Step 6 was repeated.

9. The supernatant was removed as much as possible.

10. The pellet was air dried for about 10 min (it was not over-dried for the DNA would be very hard to dissolve again). The DNA was dissolved in 40 µL TE buffer (10mM Tris-HCl, 1mM EDTA, pH8.0).

11. The DNA concentration was about 1000 ng/µL determined by Nanodrop spectrophotometer (Thermo Fisher Scientific). The DNA was stored at 4°C for near future use (within a week) or at -20 for long term storage.

2.16 Insect cell transfection with recombinant bacmid DNA

Insect cell transfection was done following Invitrogen’s BEVS protocol with modifications as follows. For each bacmid DNA construct, four trials of transfection experiments were carried out at the same time, with each round being a combination of varying Sf9 cell number and DNA amount, to enhance the transfection probability because the number of cells and the amount of DNA were found to be the two pivotal factors for the success of transfection. With the cell density of the cell culture to seed Sf9 cells from being about 4 × 10^6 cells/mL and the bacmid DNA concentration in the purified bacmid sample being about 1000 ng/µL, the combinations of Sf9 cell number and DNA amount for the four trials for transfection of each bacmid DNA construct are listed in Table 2.8.
### Materials and Methods

#### 2.17 Baculoviral stock amplification

Baculoviral stock amplification was done according to the instructions in BEVS from Invitrogen, with special consideration given on the use of multiplicity of infection (MOI) for virus amplification. The MOI for virus amplification is much lower than the MOI for protein expression and needs to be between 0.01-0.1. If the MOI was too low, no sufficient amount of virus would be produced; if the MOI was higher than 0.1, defective viruses with deleted heterogeneous genes would be produced\(^{57}\). This is also explicitly pointed out in Invitrogen’s BEVS manual that deleterious viruses would accumulate if higher than 0.1 MOI was used.

P1 viral stock generally has a titre at the magnitude of \(10^7\) pfu/mL. In this study, it was assumed to be in the range of \(1 \times 10^7 - 9 \times 10^7\) pfu/mL. From P1 viral stock, P2 virus stock was amplified in the following manner. Sf9 cells were seeded to 50 mL SFM medium with a starting cell density of \(0.5 \times 10^6\) cells/mL and grown at 27 °C on a shaker at 120 rpm. The next day, 5 µL of P1 viral stock was added to this 50 mL cell culture. 3 days post infection, symptoms of infection should be observed by examining a sample of the cells in the light microscope, and the cell culture was harvested on this day and centrifuged at 500g for a few

<table>
<thead>
<tr>
<th>Transfection trials for each bacmid DNA construct</th>
<th>Insect cell volume (µL)</th>
<th>Insect cell amount</th>
<th>Bacmid volume (µL)</th>
<th>Bacmid amount (µg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trial 1</td>
<td>100</td>
<td>(4 \times 10^5)</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Trial 2</td>
<td>100</td>
<td>(4 \times 10^5)</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>Trial 3</td>
<td>150</td>
<td>(6 \times 10^5)</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Trial 4</td>
<td>150</td>
<td>(6 \times 10^5)</td>
<td>50</td>
<td>50</td>
</tr>
</tbody>
</table>

Table 2.8 Four trials of transfection experiment for each bacmid DNA construct.

Symptoms of successful transfection, such as enlarged cell nucleus, swollen cell shape and increased contrast at the periphery of the infected cell, appear 4-7 days post transfection.

The medium containing the virus was harvested then and clarified by centrifuging at 500g for a few minutes. The supernatant was labelled as P1 viral stock (virus passage 1 stock). Sterile-filtered BSA was added to the virus to a final concentration of 0.5% (w/v) and the virus was stored at 4°C, protected from light.
minutes. The clarified supernatant contained the amplified virus and it was labelled as P2 viral stock. P2 viral stock amplified in this way generally had a titre at the magnitude of $10^8$ pfu/mL. In this study, it was assumed to be $1 \times 10^8$ pfu/mL for subsequent infection of insect cells and recombinant proteins expression.

**2.18 Baculoviral stock storage**

Viral stock was protected with the addition of sterile-filtered BSA to a final concentration of 0.5% (w/v), and was stored at 4°C and -80°C, protected from light.

**2.19 Protein expression in insect cells**

Insect Sf9 cells were infected with recombinant baculoviral stock for protein expression according to the instructions from the Invitrogen’s BEVS manual. MOI for insect cell infection was kept between 1-10.

**2.19.1 Small-scale protein expression in insect cells**

Small-scale expression was done in 20 mL of Sf9 cell culture. Cells were seeded at a starting cell density of $0.5 \times 10^6$ cells/mL and grown at 27 °C on a shaker at 120 rpm for two days. Then 2 mL of P2 viral stock was added to this 20 mL cell culture. 2 days post infection the cells should have showed infection symptoms that were examined in the light microscope and the cells were harvested. Cells were centrifuged at 4000 rpm for 10 mins and the pellet was stored at -20°C until use.

**2.19.2 Large-scale protein expression in insect cells**

Large-scale expression was carried out in 250 mL and 500 mL culture using 1 L and 2 L Erlenmeyer shaker flasks in the same manner as for small-scale expression described above, using accordingly increased amount of P2 viral stock. Cells were harvested and centrifuged at 4000 rpm for 10 mins and the pellet was stored at -20°C until use.
2.19.3 Aliquoting infected insect cells

When the infected insect cells were going to be harvested, insect cells were aliquoted directly from the culture, not from resuspension of the centrifuged pellet, because resuspending the centrifuged insect cell pellet would break insect cells and release proteins including the expressed recombinant protein, i.e., insect cell pellets are never resuspended until the batch is being subjected to cell lysis and protein purification.

2.20 Insect cell lysis

Insect cell lysis was performed using sonication (Qsonica sonicators) method in this study. The cell lysis buffer was the same as for bacterial cell lysis.

2.20.1 Small scale insect cell lysis

3 mL of cell culture was pelleted in two 1.5-mL Eppendorf tubes by centrifuging at 12,000 rpm at 4°C for 1 min. The pellets from the two tubes were resuspended in 1 mL lysis buffer. The resuspension was sonicated using the following parameters while the tube was put on ice and water: small sonicator tip, 5-10 W, 1s on/3s off per round, 30 rounds. The sample was checked in the light microscope to make sure that the cells were broken.

2.20.2 Large scale insect cell lysis

Cell pellet of 400 mL culture was resuspended in 40 mL lysis buffer in a 50-mL Falcon tube. The resuspension was sonicated using the following parameters while the tube was put on ice and water: large sonicator tip, about 30 W, 1s on/3s off per round, 90 rounds. The sample was checked in the light microscope to make sure that the cells were broken.

2.21 Pull-down experiment and sample preparations for SDS-PAGE

A small-scale pull-down experiment was carried out to test for the solubility of the expressed polyhistidine-tagged target protein, and the purity and yield of the protein to the stage of immobilized metal affinity chromatography (IMAC). Procedures of pull-down for *E. coli* culture and for insect cell culture differed in terms of sonication parameters and the amount of
cell culture used and are described below. Samples at certain steps were prepared for SDS-PAGE (Section 2.23).

2.21.1 Procedure for pull-down from *E. coli* culture

1. 100 mL of *E. coli* culture to express target protein was prepared.

2. SDS-PAGE sample was prepared before IPTG induction. Just before IPTG induction, 400 µL of cell culture was aspirated and pelleted. The pellet was reuspended in 20 µL lysis buffer and mixed with 20 µL 2 × SDS loading buffer. The mixture was heated at 90-100 °C for 30 minutes. This was the sample ‘total cell lysate before induction’ and served as a control for the target protein expression.

3. SDS-PAGE sample was prepared after IPTG induction. After IPTG induction, 200 µL of cell culture was aspirated and pelleted. The pellet was reuspended in 20 µL lysis buffer and mixed with 20 µL 2 × SDS loading buffer. The mixture was heated at 90-100 °C for 30 minutes. This was the sample ‘total cell lysate after induction’ and used to compare with the control to see if the target protein was expressed.

4. The induced 100 mL cell culture was divided into 2 × 50 mL Falcon tubes and centrifuged at 4000 rpm, 4°C for 10 mins. The pellets were resuspended in a total of 10 mL lysis buffer. The suspension was added to with lysozyme to final concentration of 1 mg/mL and put on ice and water mixture for 15 min. The suspension was sonicated while it was being put on ice and water using the following parameters: large sonicator tip, 20-30W, 1 s on / 3 s off per round, 90 rounds.

5. The cell lysate was divided into 6 × 1.5 mL Eppendorf tubes and centrifuged at 12,000 g at 4°C for 10 min in a bench top centrifuge.

6. SDS-PAGE sample was prepared after centrifugation. Sample ‘supernatant’: 20 µL of supernatant was mixed with 20 µL 2 × SDS loading buffer and heated at 90-100 °C for 10 minutes. Sample ‘precipitate’: 20 µL of precipitate was mixed with 20 µL 2 × SDS loading buffer and heated at 90-100 °C for 10 minutes.
7. The supernatant was pooled to a 10-mL Falcon tube and 100 µL of lysis buffer-pre-equilibrated nickel-resin slurry was added to the tube (This 100 µL nickel-resin slurry contained 50% lysis buffer).

8. The 10-mL Falcon tube was inverted in an end-to-top fashion on a inverter for at least 30 min at 4°C to incubate the target protein with the nickel resin.

9. 10-mL Falcon tube was put on ice for about 10 min and the nickel resin was settled down by gravity.

10. The supernatant was removed and the nickel resin was transferred to new 1.5-mL Eppendorf tube. The resin was washed with lysis buffer for 3 times, 1 mL per time.

11. SDS-PAGE sample was prepared. Sample ‘flow through’: 20 µL of supernatant was mixed with 20 µL 2 × SDS loading buffer and heated at 90-100 ºC for 10 minutes. Sample ‘wash’: 20 µL of wash was mixed with 20 µL 2 × SDS loading buffer and heated at 90-100 ºC for 10 minutes.

12. The last wash was removed from the nickel resin and 100 µL of elution buffer was added to the nickel resin. The resin was pipetted gently and centrifuged at 12,000 g at 4°C for 10 min. The supernatant was the elution, and expected to contain the recombinant protein at reasonable purity.

13. Protein concentration in the elution was measured and SDS-PAGE sample ‘elution’ was prepared.

**2.21.2 Procedure for pull-down from insect cell culture**

Procedure for pull-down from insect cell culture was basically the same as for *E. coli* culture described above, except that the initial insect cell culture was 3 mL, the lysis buffer was 1 mL, sonication parameters were different (see section 2.19 for insect cell lysis), 40 µL of nickel resin slurry was used and the incubation was in a 1.5-mL Eppendorf tube.
2.22 Protein purification

After bacterial cells or insect cells had been lysed (see section 2.11, section 2.19) and clarified by centrifugation at 16,000 g at 4°C for 30 min, recombinant protein was purified in two steps: immobilized metal affinity chromatography (IMAC) and size exclusion chromatography (SEC).

2.22.1 Immobilized metal affinity chromatography

The clarified cell lysate was filtered through a 0.22 µm syringe filter before it was applied to Immobilized metal affinity chromatography (IMAC). IMAC was carried out in a 5-mL pre-packed nickel-resin column using a peristaltic pump. The column was pre-equilibrated by cell lysis buffer for 5 column volumes before the cell lysate was loaded to the column. After the cell lysate was loaded to the column, the column was washed with 20 column volumes of IMAC wash buffer (Table 2.9), and then the protein was eluted with 3 column volumes of IMAC elution buffer (Table 2.9).

In cases where the clarified cell lysate was hard to be filtered through 0.22 µm syringe filters, it was first filtered through 0.8 µm membranes. Alternatively, the IMAC purification was done using batch/gravity-flow column (Bio-Rad) and nickel-charged resin (Qiagen) where the cell lysate was first incubated with the nickel resin rotating at 4°C for at least 30 min and then the resin was gently pelleted and transferred to the empty gravity-flow column, followed by wash and elution.

2.22.2 Recombinant tabacco etch virus protease cleavage

A recognition sequence for the recombinant tabacco etch virus protease (rTEV protease)\textsuperscript{58} was incorporated in certain protein constructs (Figure 2.1). Upstream of the rTEV protease recognition sequence was the 6-histidine tag in the plasmid used (for example, pETDuet-1). After purification from IMAC, protein solution was mixed with rTEV protease at a ratio of 1:20 (w/w) of rTEV to target protein. This was transferred to a dialysis tubing (Spectrum laboratories) and the tubing was equilibrated against SEC buffer at 4°C overnight with gentle stirring. The following day, the solution was applied to IMAC again and the flow through was collected, which contained the his-tag removed protein. The uncleaved protein, the
contaminations that bound non-specifically to the IMAC column which had been eluted from the first IMAC, and the rTEV itself, would bind to the IMAC column.

![Diagram](image)

**Figure 2.1 Incorporation of rTEV recognition sequence into constructs.** (Restriction site) restriction enzyme digestion site, used for double enzyme restriction digestion. (Gap) the region to add 0-2 bp of nucleotides according to the specific plasmid used to make the reading frame of the rTEV recognition sequence unshifted from the start codon of the plasmid. rTEV cleavage position is indicated by the red triangle. (Spacer) the three amino acid linker between rTEV recognition sequence and target gene sequence. (Gene-specific) gene sequence region used for forward and reverse primer Figure adapted from reference 58 and re-drawn using OpenOffice Drawing.

**2.22.3 Size exclusion chromatography**

Protein solution volumes were reduced to about 2.5 mL by concentration using a centrifugal concentrator (GE). This concentrated 2.5 mL protein was loaded to Highload 16/60 Superdex 75 or 200 prep grade size exclusion columns using a 5-mL injection loop. Before loading the protein into the SEC column, the SEC column was pre-equilibrated with at least one column volume of 0.22 µm filtered Milli-Q water and then with at least one column volume of SEC buffer (Table 2.9), and the concentrated protein sample was filtered through a 0.22 µm centrifugal filter (Merck Millipore).

The column was run at a flow rate of 1 mL/min and 1 mL fractions were collected and analysed by SDS-PAGE. Then appropriate fractions were pooled and concentrated to high concentration for protein characterization and crystallization. The concentrated protein stock was aliquoted, snap-frozen and stored at -80°C.
### Protein purification buffers using Tris-HCl pH 8.0:
- IMAC wash buffer: 20 mM Tris-HCl pH 8.0, 200 mM NaCl, 20 mM imidazole.
- IMAC elution buffer: 20 mM Tris-HCl pH 8.0, 200 mM NaCl, 250 mM imidazole.
- SEC buffer: 20 mM Tris-HCl pH 8.0, 200 mM NaCl.

### Protein purification buffers using HEPES-NaOH pH 7.0:
- IMAC wash buffer: 50 mM HEPES-NaOH pH 7.0, 200 mM NaCl, 20 mM imidazole.
- IMAC elution buffer: 50 mM HEPES-NaOH pH 7.0, 200 mM NaCl, 250 mM imidazole.
- SEC buffer: 50 mM HEPES-NaOH pH 7.0, 200 mM NaCl.

### Protein purification buffers using Na₂CO₃-NaHCO₃ pH 10.8:
- IMAC wash buffer: 0.1 M Na₂CO₃-0.1 M NaHCO₃ pH 10.8, 200 mM NaCl, 20 mM imidazole.
- IMAC elution buffer: 0.1 M Na₂CO₃-0.1 M NaHCO₃ pH 10.8, 200 mM NaCl, 250 mM imidazole.
- SEC buffer: 0.1 M Na₂CO₃-0.1 M NaHCO₃ pH 10.8, 200 mM NaCl.

Table 2.9 Protein purification buffers.

## 2.23 Protein concentration measurement

Protein concentration was measured on a Nanodrop spectrophotometer (Thermo Fisher Scientific) by measuring absorbance at 280 nm using a calculated extinction coefficient calculated from ProtParam program (Table 2.1).

## 2.24 SDS-PAGE

### 2.24.1 Gel recipe

Discontinuous SDS-PAGE consisting of a resolving gel at the bottom (Table 2.10) and a stacking gel on top (Table 2.11) was made.
## Materials and Methods

### Resolving gel recipe for 12 pieces of gel

<table>
<thead>
<tr>
<th>Ingredient</th>
<th>10% Gel</th>
<th>12% Gel</th>
<th>15% Gel</th>
<th>18% Gel</th>
</tr>
</thead>
<tbody>
<tr>
<td>30% acrylamide</td>
<td>18 mL</td>
<td>24 mL</td>
<td>30 mL</td>
<td>36 mL</td>
</tr>
<tr>
<td>1.5M Tris-HCl pH 8.8</td>
<td>15 mL</td>
<td>15 mL</td>
<td>15 mL</td>
<td>15 mL</td>
</tr>
<tr>
<td>10% SDS (w/v)</td>
<td>0.6 mL</td>
<td>0.6 mL</td>
<td>0.6 mL</td>
<td>0.6 mL</td>
</tr>
<tr>
<td>Milli-Q water</td>
<td>26 mL</td>
<td>20 mL</td>
<td>14 mL</td>
<td>8 mL</td>
</tr>
<tr>
<td>10% (w/v) APS</td>
<td>600 µL</td>
<td>600 µL</td>
<td>600 µL</td>
<td>600 µL</td>
</tr>
<tr>
<td>TEMED</td>
<td>60 µL</td>
<td>60 µL</td>
<td>60 µL</td>
<td>60 µL</td>
</tr>
</tbody>
</table>

Table 2.10 Resolving gel recipe for 12 pieces of gel.

### Stacking gel recipe for 4% for 12 pieces of gel

<table>
<thead>
<tr>
<th>Ingredient</th>
<th>Volume for 12 pieces of 4% stacking gel</th>
</tr>
</thead>
<tbody>
<tr>
<td>30% acrylamide</td>
<td>6.6 mL</td>
</tr>
<tr>
<td>0.5 M Tris-HCl pH6.8</td>
<td>12.6 mL</td>
</tr>
<tr>
<td>10% SDS (w/v)</td>
<td>0.5 mL</td>
</tr>
<tr>
<td>Milli-Q water</td>
<td>30 mL</td>
</tr>
<tr>
<td>Bromophenol blue</td>
<td>1 drop</td>
</tr>
<tr>
<td>10% (w/v) APS</td>
<td>800 µL</td>
</tr>
<tr>
<td>TEMED</td>
<td>100 µL</td>
</tr>
</tbody>
</table>

Table 2.11 Stacking gel recipe for 4% for 12 pieces of gel.

### 2.24.2 Gel making procedure

A suitable percent resolving gel was selected and the ingredients were added to a beaker. 600 µL of 10% (w/v) APS and 60 µL of TEMED were added as the last step because the gel would set quickly in the presence of these agents. After these two agents were added, the solution was pipetted several times and poured into the gel caster, up to about 2 cm below the bottom of the comb to leave space for the stacking gel. 200 µL isopropanol was overlaid on top of the resolving gel to prevent dehydration.

30 minutes later, the resolving gel polymerized, and the stacking gel was made. The stacking gel ingredients were added to a beaker. The isopropanol on top of the resolving gels was drained using paper towels. Then 800 µL 10% (w/v) APS and 100 µL of TEMED were
added. The solution was pipetted several times and pipetted on top of the resolving gels. Then combs were added. The gels set in 30 min and ready to be used.

### 2.24.3 Gel run

Gels were run in 1 × SDS running buffer in the SDS-PAGE apparatus under constant current of 30 mA per gel for about 30 min. 10 µL of sample was loaded to well of a 15-well gel and 20 µL for a 10-well gel.

Table 2.12 lists recipes for SDS running buffer and Table 2.13 for SDS loading buffer.

<table>
<thead>
<tr>
<th>Ingredient</th>
<th>Amount for 10 × SDS running buffer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tris base</td>
<td>30.3 g</td>
</tr>
<tr>
<td>Glycine</td>
<td>144.4 g</td>
</tr>
<tr>
<td>SDS</td>
<td>10 g</td>
</tr>
<tr>
<td>Milli-Q water</td>
<td>Filled to 1 litre</td>
</tr>
</tbody>
</table>

Table 2.12 10 × SDS running buffer.

The following 4 × SDS loading buffer didn’t contain reducing agent such as DTT or β-Mercaptoethanol. It was added fresh when needed when 1 × SDS loading buffer was prepared from this stock.

<table>
<thead>
<tr>
<th>Ingredient</th>
<th>Amount for 4 × SDS loading buffer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stacking buffer</td>
<td>56% (v/v)</td>
</tr>
<tr>
<td>SDS</td>
<td>8% (w/v)</td>
</tr>
<tr>
<td>Glycerol</td>
<td>40% (v/v)</td>
</tr>
<tr>
<td>Bromophenol blue</td>
<td>0.4% (w/v)</td>
</tr>
</tbody>
</table>

Table 2.13 4 × SDS loading buffer.
2.24.4 Coomassie staining and destaining

Gel running was stopped when the bromophenol blue had reached the bottom of the resolving gel. The gels were disassembled from the apparatus and put into Coomassie staining solution (Table 2.14). Both staining and destaining (Table 2.15) process could be accelerated by microwaving for 10-20 seconds.

Coomassie blue R-250 was dissolved in pure methanol first, then acetic acid and water were added.

<table>
<thead>
<tr>
<th>Coomassie staining solution ingredient</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methanol</td>
<td>50% (v/v)</td>
</tr>
<tr>
<td>Coomassie blue R-250</td>
<td>0.25% (w/v)</td>
</tr>
<tr>
<td>Acetic acid</td>
<td>10% (v/v)</td>
</tr>
<tr>
<td>Milli-Q water</td>
<td>40% (v/v)</td>
</tr>
</tbody>
</table>

Table 2.14 Coomassie staining solution

<table>
<thead>
<tr>
<th>Coomassie destaining solution ingredient</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methanol</td>
<td>50% (v/v)</td>
</tr>
<tr>
<td>Acetic acid</td>
<td>10% (v/v)</td>
</tr>
<tr>
<td>Milli-Q water</td>
<td>40% (v/v)</td>
</tr>
</tbody>
</table>

Table 2.15 Coomassie destaining solution

2.25 Protein characterization

2.25.1 Native PAGE

Native PAGE was performed in a similar manner to SDS-PAGE, with SDS absent from all solutions. Protein samples to be run had native loading buffer (Below) added, and were not heated.
4 × native loading buffer: 40% (v/v) glycerol, 40% native running buffer, a small amount of bromophenol blue.

2.25.2 Size exclusion chromatography-multi-angle static light scattering

Size exclusion chromatography with multi-angle static light scattering (SEC-MALS) was used to measure masses of the protein molecules. An analytical Superdex 200 10/300 tricorn column (GE) in line with MALS (miniDAWN TREOS, Wyatt technology) had the protein sample run at a flow rate of 0.5 mL/min using SEC buffer, and the MALS measured the molar mass and size of the molecules at each elution peak.

2.26 X-ray crystallography

2.26.1 Robotic crystallization screening

Initial screening of protein crystallization conditions was carried out using liquid-handling robotics by sitting-drop vapor-diffusion. Proteins were concentrated to an appropriate concentration (usually 10 mg/mL-20 mg/mL, if attainable) and over 500 conditions were routinely screened. 80 µL of a condition from crystallization screen kits (Molecular Dimensions) was transferred to each reservoir of a 96-well crystallization plate (Intelli-Plate 96, Art Robins Instruments) by a MultiPROBE II HT EX liquid handling robot (Perkin-Elmer). 100 nL of protein was combined with 100 nL of a condition to set up a droplet of 200 nL by the Cartesian HoneyBee nanolitre dispensing robot. The plates were then sealed with clear adhesive films (Crystal Clear Sealing Film, Hampton Research) and incubated at 18°C.

2.26.2 Proteolysis during crystallization experiment

Protein crystallization can be inhibited by the presence of flexible loops on the surface of proteins. In situ proteolysis has been reported to facilitate crystallization by removing the flexible loops and leaving the compact core of a protein59.

α-chymotrypsin (Sigma-Aldrich) and trypsin (Sigma-Aldrich) were used for in situ proteolysis. α-chymotrypsin was dissolved at 2 mg/mL in 1 mM HCl, 2 mM CaCl₂. The stock
was added to protein sample at a ratio of 1:100 (w/w), and crystallization trials were set up immediately by robotic screening as described above. Trypsin was dissolved at 1.5 mg/mL in 1 mM HCl, 2 mM CaCl\(_2\). The stock was added to protein sample at a ratio of 1:1000 (w/w), and crystallization trials were set up immediately by robotic screening as described above.

### 2.26.3 Fine crystallization screening

Once a promising crystallization condition had been identified, fine grid screens around this condition were performed in a 48-well plate or 24-well plate with varying precipitant concentration and pH.

### 2.26.4 X-ray diffraction

X-ray diffraction data were collected at Australian Synchrotron, Victoria, Australia, using the Blue-Ice control system on the MX1 and MX2 beamlines (Table 2.16).

<table>
<thead>
<tr>
<th></th>
<th>MX1</th>
<th>MX2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collection temperature</td>
<td>100 K</td>
<td>100 K</td>
</tr>
<tr>
<td>Radiation type</td>
<td>synchrotron</td>
<td>synchrotron</td>
</tr>
<tr>
<td>Source</td>
<td>bending magnet</td>
<td>In-vacuum undulator</td>
</tr>
<tr>
<td>Beam size</td>
<td>130 × 90 µm</td>
<td>37 × 32 µm</td>
</tr>
<tr>
<td>Flux at 12.6 keV</td>
<td>1.5 × 10^{11} photons / s</td>
<td>4 × 10^{12} photons / s</td>
</tr>
<tr>
<td>Monochromator</td>
<td>silicon double crystal</td>
<td>silicon double crystal</td>
</tr>
<tr>
<td>Detector type</td>
<td>CCD</td>
<td>CCD</td>
</tr>
<tr>
<td>Detector brand</td>
<td>ADSC quantum 210r</td>
<td>ADSC quantum 315r</td>
</tr>
</tbody>
</table>

Table 2.16 X-ray diffraction data collection technical details at the Australian Synchrotron.

### 2.27 Transmission electron microscopy

Transmission electron microscope (FEI Tecnai T12) was used to study the whole CpGV occlusion bodies in this study by negative staining and cryo-EM. To obtain high quality images and to avoid introducing artefacts to the images, a standard alignment procedure was...
always done before collecting negative stain images. For cryo-EM, another two cryo-EM alignment procedures were carried out in addition to the standard alignment. These ‘Search-Exposure’ alignment and the ‘Search-Focus’ alignment procedures are described in this section.

2.27.1 Negative stain electron microscopy

1. The specimen was mixed with 2% uranyl acetate and drained by blot paper.

2. The specimen was loaded into the specimen holder.

3. The specimen holder was loaded into the Campustage of the EM. Care was taken during this step by making the following moves. The pin on the holder was aligned with the mark on the Campustage at the 5 o’clock position, and the holder was slowly inserted into the Campustage until a stop was felt. During this insertion, the red light on the Campustage would turn on and the pump would be heard starting to work. The holder was gently held and a moment was waited until the vacuum sucked in the holder a little. Then the ‘Specimen holder-Single Tilt’ option on the controlling screen was selected and the ‘Confirm’ button was clicked. A moment was waited until the red light on the Campustage had turned off and the Vacuum status on the screen changed from ‘Air Lock’ to ‘Gun/Col Vacuum’. Then the holder was slowly rotated counterclockwise from 5 o’clock to 12 o’clock. After this rotation, the holder would be slowly sucked in automatically, and it was gently held and guided during the holder sliding in.

4. The standard alignments were always performed following the manufacturer’s instructions before sample was looked for and image collection. The standard alignments included condenser aperture centering, condenser lens astigmatism correction (these two were about the illumination behaviour which could be easily noticed if there was something wrong), eucentric height adjustment, objective aperture centering (normally there was no need to mechanically change this one), objective lens astigmatism correction (it was hard to do objective lens astigmatism correction using Live FFT at low magnification of 4400 × as the Thon rings are not obvious, and easier at high magnification of 15000 ×), beam tilt pivot point X alignment, beam tilt pivot point Y alignment, and rotation centering alignment at high magnification > 50 k ×.
5. Focus condition was adjusted. Images were usually taken at defocus = 0 µm, -2 µm, -4 µm, -8 µm conditions.

2.27.2 Cryo-electron microscopy

Cryo-EM experiments comprised of 4 parts, as described below.

2.27.2.1 Specimen cryo-cooling

Specimen for cryo-EM was prepared by plunge-freezing it in liquid ethane using Vitrobot (FEI) according to the manufacturer’s instructions.

2.27.2.2 Specimen transfer to cryo-holder and into EM

The cryo-specimen was transferred to the cryo-holder in a work station, with the specimen all the way being submerged in liquid nitrogen. The cryo-holder was then transferred to the Campustage of EM following specific instructions to retain most of liquid nitrogen in the cryo-holder during this transfer process.

2.27.2.3 Cryo-electron microscopy alignments

The cryo-EM alignments consist of the standard alignments, the ‘Search-Exposure’ alignment, and the ‘Search-Focus’ alignment. The latter two alignments are special for cryo-EM experiment.

1. The standard alignments.

a). The ‘low dose’ button was clicked, and now it was in ‘Search’ mode.

b). The ‘magnification’ was lowered to < 4400 ×.

c). The ‘high tension’ button was clicked. The ‘filament’ button was then clicked. Then the ‘Gun column valve’ button was clicked.

d). The shutter on the cryo-holder was pulled open now.

e). One window of the grid needed to be used in order to do the standard alignments. The standard alignments included condenser aperture centering, condenser lens astigmatism
correction (these two were about the illumination behaviour which could be easily noticed if there was something wrong), eucentric height adjustment, objective aperture centering (normally there was no need to mechanically change this one), objective lens astigmatism correction (it’s hard to do objective lens astigmatism correction using Live FFT because at low magnification of 4400 × the Thon rings are not obvious, and at high magnification of 15000 × the feature will be radiation-damaged. The only way to do this for cryo-EM is do it optically, i.e., using naked eyes, and an empty hole in the window was needed. After finding an empty hole, it was defocused (i.e., underfocus) by turning the ‘focus’ knob, and an inner white Fresnel ring around the hole should be seen. This Fresnel ring should be even. If it was not even and seemed stretched in some direction, then the objective lens astigmatism could be corrected until the ring became even), beam tilt pivot point X alignment, beam tilt pivot point Y alignment, and rotation centering alignment at high magnification > 50 k ×.

The same window was used to set up another two alignments special for cryo-EM: the ‘Search-Exposure’ alignment and the ‘Search-Focus’ alignment.

2. The ‘Search-Exposure’ alignment.

a). The beam stopper was put in and an image was taken of it to tell which part of the beam stopper was lying right at the center of the image taken.

b). In ‘Search’ mode, the magnification was adjusted to 4400 ×. An obvious feature was selected and it was moved to the area of the image center-occupying part of the beam stopper, by moving the stage.

c). In ‘Exposure’ mode, the magnification was adjusted to 21000 ×. That same feature in ‘Search’ mode was used as the reference for alignment and it was moved to the area of the image center-occupying part of the beam stopper, by moving the stage.

d). Now it was switched back to the ‘Search’ mode, to check if the feature had moved from the area of the image center-occupying part of the beam stopper (it probably had.). The feature was moved to the area of the image center-occupying part of the beam stopper again, but this time not by moving the stage, but by turning MF-X and MF-Y knobs, which in the ‘Search’ mode controlled image shift rather than stage movement.
e). Now it was switched forth to the ‘Exposure’ mode, to check if the feature had moved from the area of the image center-occupying part of the beam stopper (it probably had not.). If it had not moved, then the alignment between ‘Search’-‘Exposure’ was completed; if it had slightly moved, the feature was moved to the area of the image center-occupying part of the beam stopper by moving the stage, and another round of switching back to ‘Search’ mode and forth to ‘Exposure’ mode should be well enough.

3. The ‘Search-Focus’ alignment.

a). In ‘Search’ mode, the magnification was adjusted to 4400 ×. An obvious feature was selected and it was moved to the area of the image center-occupying part of the beam stopper, by moving the stage.

b). In ‘Focus’ mode, the magnification was adjusted to the range of low to 42000 × and the view of the window was adjusted to the edge of the 3rd hole from the hole with that feature by turning MF-X and MF-Y knobs, which in the ‘Focus’ mode controlled focus angle and focus distance rather than stage movement. Then the ‘intensity’ button was turned to have a smaller illumination area.

These 2 holes, i.e., the target hole that was observed in ‘Search’ mode and the focused hole whose edge was focused in ‘Focus’ mode, were 4-hole distance away with 2 holes in between. This was to avoid any radiation damage to the specimen of the target hole while the focus to the focused hole was being adjusted. After adjusting focus to the focused hole, whose plain could be practically thought to be the same as the target hole because there were only 2 holes in between and therefore their focus could be thought the same, an image of the target hole could be taken by clicking the ‘Expose’ button in ‘Focus’ mode using the parameters set in the ‘Exposure’ mode.

By now the ‘Search-Focus’ alignment has been completed.

2.27.2.4 Cryo-electron microscopy image collection

a). In ‘Search’ mode, in the same window that had been used for adjusting focus, a new hole was found.

b). In ‘Focus’ mode, the ice at the edge of the hole was melted by turning the ‘intensity’ button to have a central beam spot.
c). In ‘Focus’ mode, the at-focus condition was found by adjusting the ‘Focus’ knob. (the at-focus condition is that when there is minimum contrast at the edge of the hole; a white inner ring ((Fresnel fringe)) denotes an under-focused condition, a thicker outer ring ((Fresnel fringe)) denotes an over-focused condition and the at-focus condition is in between these two scenarios).

d). The ‘reset defocus’ button was pressed. Now the defocus = 0 µm, meaning it was at-focus (practically, one window area could be thought to be at the same focus condition because one window area is very small and should be very even, although there are a few holes in one window, so this at-focus adjustment could be done once and used for the same window).

e). In ‘Focus’ mode, an image of the target hole was taken by clicking the ‘Expose’ button. The image was taken at defocus = 0 µm. The defocus could also be changed to -2 µm or -4 µm to take images of the same hole or new holes. Usually under-focused images look significant regarding to the details of the specimen.

f). In ‘Search’ mode, another interesting hole was found. Then the ‘Focus’ mode was switched to and another image of the interesting hole was taken by clicking the ‘Expose’ button.

Many images could be taken from different holes in the same window.

g). Now another window of the same grid was moved to and step a) to f) was repeated to take as many images as needed.

### 2.28 How to flash-freeze aliquots in liquid nitrogen

Snap-freezing cell culture aliquots or protein aliquots in liquid nitrogen was performed with special care to allow gas to escape from the tubes to avoid explosion. This could be achieved by having tubes sit on a foam plate with only the bottom parts of the tubes to be immersed in the liquid nitrogen, preventing liquid nitrogen from permeating into the chamber of the tubes from the caps which would otherwise end up with severe popping up of the tube caps when the confined liquid nitrogen becomes gas and most times the caps fly away so fiercely that its impact is like an explosion. Another way to safely snap freeze the tubes is to pierce through each cap using a needle before the snap-freezing so the gas would dissipate through this outlet.
Chapter 3 Expression of CpGV occlusion body surface proteins

Prior to the start of this project, previous studies carried out by members of our research group had shown that CpGV occlusion body surface proteins CpGV-PEP-L, CpGV-PEP-M, and CpGV-PEP-S are all present in the occlusion body surface and form fibres (Figure 3.1).

The occlusion bodies were treated with reducing agent DTT, and the isolated fractions were resolved on SDS-PAGE. The protein bands revealed on SDS-PAGE were excised and sent for identification by mass spectrometry. Three CpGV occlusion body surface proteins PEP-L, PEP-M, and PEP-S were identified. Polyhedrin protein was also found to be involved in two of the three bands. There are discrepancies between the expected positions of the three PEP proteins and the three fragments from the fibres solution, and in each single band there are more than one protein identified.
How to flash-freeze aliquots in liquid nitrogen

Expression of CpGV occlusion body surface proteins

Figure 3.1 Previous study on CpGV occlusion body fibres from Metcalf group. (Left) Cryo-EM of CpGV occlusion bodies before and after DTT treatment. Note there are fibres on the surface of CpGV occlusion body before DTT treatment, and the fibres are absent from the occlusion bodies after DTT treatment. Cryo-EM was performed by A. Turner. (Middle) SDS-PAGE of the proteins after DTT treatment on CpGV occlusion bodies. M, molecular weight marker; A, B, C are protein bands from the solution where detached fibres were concentrated; coloured triangles indicate the expected positions of PEP-L, PEP-M, and PEP-S. Note there are significant discrepancies between the expected positions of the three PEP proteins and the three fragments from the fibres solution. SDS-PAGE and DTT treatment experiment was carried out by S.Yeh. (Right) Proteins identified from the three bands on the SDS-PAGE by LC-MS. A, B, C correspond to the respective bands on the SDS-PAGE in the middle. Note in each single band there are more than one protein identified. LC-MS was carried out by M. Middleditch.

This chapter describes expression experiments for the three CpGV occlusion body surface proteins PEP-L, PEP-M, PEP-S, and for the single AcMNPV occlusion body surface protein PEP. The aim was to produce soluble protein suitable for crystallization trials and electron microscopy. Several full length and truncated constructs were expressed both in *E.coli* and in insect cells and combinations of the three envelope proteins were co-expressed in the two systems. Complex formation for various combinations of PEP-L, PEP-M and PEP-S proteins was investigated.
3.1 Results

3.1.1 Bioinformatics

The basic biophysical properties of the three CpGV-PEP proteins and for AcMNPV-PEP are shown in Table 3.1. Conserved domains of these PEP proteins were predicted using BLAST tool against GenBank database and used to make the schematic structure of the PEP proteins shown in Figure 3.2.

All of the PEP proteins have PEP-N domain. CpGV-PEP-L and AcMNPV-PEP have an additional PEP-C domain (Figure 3.2). CpGV-PEP-L and AcMNPV-PEP belong to the same protein cluster, while CpGV-PEP-M and CpGV-PEP-S belong to two other different protein clusters respectively (Table 3.2). There are only a few amino acids conserved in these four PEP proteins (Figure 3.3). Despite both having two domains, CpGV-PEP-L and AcMNPV-PEP are not similar at the amino acid sequence level, with an overall identity of 14.9% and a similarity of 21.6% based on a global pairwise sequence alignment using EMBOSS Needle program (Figure 3.4). CpGV-PEP-M and CpGV-PEP-S, the two PEP proteins that have only one domain of PEP-N, however, share an identity of 29% and a similarity of 43% in their N-terminal regions, although they have a low overall identity of 15.5% and an overall similarity of 23.5% (Figure 3.5).

<table>
<thead>
<tr>
<th>Name</th>
<th>ORF</th>
<th>Length (aa)</th>
<th>MW</th>
<th>NCBI Accession Number</th>
<th>UniProt ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>AcMNPV-PEP</td>
<td>ORF131</td>
<td>252 aa</td>
<td>29.1 kDa</td>
<td>NP_054161</td>
<td>P24728</td>
</tr>
<tr>
<td>CpGV-PEP-L</td>
<td>ORF22</td>
<td>347 aa</td>
<td>37.2 kDa</td>
<td>NP_148806</td>
<td>A0A097P0I8</td>
</tr>
<tr>
<td>CpGV-PEP-M</td>
<td>ORF20</td>
<td>235 aa</td>
<td>26.2 kDa</td>
<td>NP_148804</td>
<td>A0A097P1S2</td>
</tr>
<tr>
<td>CpGV-PEP-S</td>
<td>ORF23</td>
<td>152 aa</td>
<td>16.9 kDa</td>
<td>NP_148807</td>
<td>A0A097P0K3</td>
</tr>
</tbody>
</table>

Table 3.1 PEP proteins.
Figure 3.2 Schematic structure of PEP proteins. Numbers below each sequence indicate the approximate location of the domain in the sequence. Characters in between the two domains in AcMNPV-PEP denote the amino acids repeats in this region. Drawn using OpenOffice Drawing.

<table>
<thead>
<tr>
<th>Protein cluster ID</th>
<th>Number of proteins in cluster</th>
<th>ORFs in type-specific baculoviruses</th>
<th>Name used in thesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>2945044</td>
<td>48</td>
<td>C. pomonella baculovirus ORF 22</td>
<td>CpGV-PEP-L</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A. californica baculovirus ORF 131</td>
<td>AcMNPV-PEP</td>
</tr>
<tr>
<td>2511597</td>
<td>2</td>
<td>C. pomonella baculovirus ORF 20</td>
<td>CpGV-PEP-M</td>
</tr>
<tr>
<td>2511017</td>
<td>12</td>
<td>C. pomonella baculovirus ORF 23</td>
<td>CpGV-PEP-S</td>
</tr>
</tbody>
</table>

Table 3.2 PEP proteins in protein clusters.
Results

Expression of CpGV occlusion body surface proteins

Figure 3.3 Multiple sequence alignment of AcMNPV-PEP and CpGV-PEPs. Sequences are coloured based on conservation, from dark blue (completely conserved) to white (not conserved). Sequences were aligned using Clustal Omega and displayed using JalView.
Figure 3.4 Pairwise sequence alignment of AcMNPV-PEP and CpGV-PEP-L. Sequences are coloured based on conservation, from dark blue (completely conserved) to white (not conserved). Sequences are aligned using EMBOSS Needle and displayed using JalView.
Figure 3.5 Pairwise sequence alignment of CpGV-PEP-M and CpGV-PEP-S. Alignment shows high sequence identity in the N-terminal region. Sequences are coloured based on conservation, from dark blue (completely conserved) to white (not conserved). Sequences are aligned using EMBOSS Needle and displayed using JalView.
Figure 3.6 Multiple sequence alignment of CpGV-PEP-L N-terminal domain, PEP-M and PEP-S. Sequences are coloured based on conservation, from dark blue (completely conserved) to white (not conserved). Sequences were aligned using Clustal Omega and displayed using JalView.

3.1.2 CpGV-PEP-L

13 constructs of CpGV-PEP-L were made and tested with the aim of producing fragments suitable for structural characterization. The results are summarized in Table 3.3.
In addition to the full-length proteins, I made a series of truncations. The decision on where to truncate these proteins was based on their domain and their secondary structure prediction. The rationale was to retain complete α-helices or β-strands at the boundaries of predicted domains. This was the rationale for the PEP-N domain for example, which was crystallized and its structure solved by previous group member Dr Jason Busby. The PEP-N domain was predicted as amino acids 1-111 from NCBI sequence, but its secondary structure prediction

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1-347 (full- length)</td>
<td>Insect</td>
<td>Yes</td>
<td>IMAC, SEC</td>
<td>Aggregates</td>
<td>Forward-CpGV-PEP-L-BamHI, Reverse-CpGV-PEP-L-EcoRI</td>
<td>pFastBacHTb</td>
</tr>
<tr>
<td>1-347 (full- length)</td>
<td>E.coli</td>
<td>Yes</td>
<td>IMAC, SEC</td>
<td>Aggregates. In pH 11, it breaks down to a mixture</td>
<td>CpGV-PEP-L-EcoRI-Duet, CpGV-PEP-L-XhoI-Duet</td>
<td>pETDuet</td>
</tr>
<tr>
<td>1-300</td>
<td>E.coli</td>
<td>Yes</td>
<td>IMAC, SEC</td>
<td>Oligomerizes</td>
<td>CpGV-PEP-L-EcoRI-Duet, CpGV-L-300-XhoI-Duet</td>
<td>pETDuet</td>
</tr>
<tr>
<td>MBP-1-300</td>
<td>E.coli</td>
<td>Yes</td>
<td>IMAC, SEC</td>
<td>Oligomerizes</td>
<td>MBP-3aa-Cp300aa-Fo-1, MBP-3aa-Cp300aa-Re-1, MBP-3aa-Cp300aa-Fo-2, MBP-3aa-Cp300aa-Re-2</td>
<td>pETDuet</td>
</tr>
<tr>
<td>1-296</td>
<td>E.coli</td>
<td>Yes</td>
<td>No</td>
<td>N/A</td>
<td>CpGV-PEP-L-EcoRI-Duet, CpGV-L-296-XhoI-Duet</td>
<td>pETDuet</td>
</tr>
<tr>
<td>1-282</td>
<td>E.coli</td>
<td>Yes</td>
<td>No</td>
<td>N/A</td>
<td>CpGV-PEP-L-EcoRI-Duet, CpGV-L-282-XhoI-Duet</td>
<td>pETDuet</td>
</tr>
<tr>
<td>1-277</td>
<td>E.coli</td>
<td>Yes</td>
<td>No</td>
<td>N/A</td>
<td>CpGV-PEP-L-EcoRI-Duet, CpGV-L-300-XhoI-Duet</td>
<td>pETDuet</td>
</tr>
<tr>
<td>1-269</td>
<td>E.coli</td>
<td>Yes</td>
<td>No</td>
<td>N/A</td>
<td>CpGV-PEP-L-EcoRI-Duet, CpGV-L-269-XhoI-Duet</td>
<td>pETDuet</td>
</tr>
<tr>
<td>1-251</td>
<td>E.coli</td>
<td>Yes</td>
<td>IMAC, SEC</td>
<td>Oligomerizes</td>
<td>CpGV-PEP-L-EcoRI-Duet, CpGV-L-251-XhoI-Duet</td>
<td>pETDuet</td>
</tr>
<tr>
<td>1-211</td>
<td>E.coli</td>
<td>Yes</td>
<td>IMAC, SEC</td>
<td>Oligomerizes</td>
<td>CpGV-PEP-L-EcoRI-Duet, CpGV-L-211-XhoI-Duet</td>
<td>pETDuet</td>
</tr>
<tr>
<td>1-170</td>
<td>E.coli</td>
<td>Yes</td>
<td>IMAC</td>
<td>Precipitates Immediately after IMAC</td>
<td>CpGV-PEP-L-EcoRI-Duet, CpGV-L-170-XhoI-Duet</td>
<td>pETDuet</td>
</tr>
<tr>
<td>1-106</td>
<td>E.coli</td>
<td>Yes</td>
<td>IMAC, SEC</td>
<td>1.4Å structure by Busby</td>
<td>PEP-L-1-106-Forward, PEP-L-1-106-Reverse</td>
<td>pDEST17</td>
</tr>
<tr>
<td>141-300</td>
<td>E.coli</td>
<td>Yes</td>
<td>IMAC, SEC</td>
<td>Aggregates</td>
<td>PEP-L-141aa-EcoRI, PEP-L-300-XhoI-Duet</td>
<td>pETDuet</td>
</tr>
</tbody>
</table>

Table 3.3 Constructs of CpGV-PEP-L. (IMAC) immobilized metal affinity chromatography. (SEC) size exclusion chromatography. (pFastBacHTb) a vector for insect expression system. (pETDuet) an expression vector for E.coli system. Constructs covering amino acids 1-296, 1-282, 1-277, and 1-269 were not purified. The longer truncation 1-300 and the shorter 1-251 both produced non-homogeneously oligomerized material, and the intermediate truncations were therefore abandoned.
suggested a coil structure from 107 to 140; Dr Busby truncated the protein at amino acid 106 producing the construct that was successfully crystallized. My rationale for making truncations was the same.

The CpGV-PEP-L (i.e., the full-length protein) expression was tested in both the bacterial expression system and insect expression system (i.e., the BEVS system), to find out the optimal condition for CpGV-PEP-L production. Five bacterial strains Rosetta, Shuffle, Chaperon 3, Chaperon 4, and LOBSTR were tested. Two temperatures 37°C and 18°C were tested. The bacterial strain LOBSTR was found to produce the most abundant expression of CpGV-PEP-L at 18°C (Figure 3.7). The LOBSTR strain has been engineered to reduce background binding to the IMAC purification step by eliminating two major E.coli contaminants on the basis of BL21(DE3). For insect cell expression, Sf9 cell was routinely used.

**Figure 3.7 SDS-PAGE of CpGV-PEP-L expression in bacteria and insect cells.** For bacterial expression, Five bacterial strains were tested at two temperature 37°C and 18°C. (M) molecular weight marker (kDa); (Contr.) strain Rosetta before IPTG induction used as a control; (Ros.) strain Rosetta; (Shu.) strain SHuffle; (C3) strain Chaperon 3; (C4) strain Chaperon 4; (LOB.) strain LOBSTR. All the strains were induced by 0.5 mM IPTG except Ros*, which was strain Rosetta and induced by 0.25 mM IPTG. Total cell lysate was run on 12% SDS-PAGE. CpGV-PEP-L was indicated by a red triangle. For insect cell expression, Sf9 cell was routinely used. (Sf9-C) Sf9 cells with AcMNPV-PEP expression used as a control; (Sf9-CpGV-PEPs) Sf9 cells with CpGV-PEP-L expression. CpGV-PEP-L was indicated by a red triangle.
CpGV-PEP-L was expressed in *E. coli* strain LOBSTR and purified by IMAC and SEC using Tris-HCl pH 8.0 buffer (see Chapter 2). However, at this pH, the protein aggregates and forms a mixture of irregular oligomers with molecular weight ranging from 100 kDa to over 1000 kDa, as shown by SEC-MALS (Figure 3.8). The insect cell expression system was also used to produce CpGV-PEP-L. However, the protein produced also aggregated in Tris-HCl pH 8.0 buffer.

![SEC-MALS of CpGV-PEP-L in Tris-HCl pH 8.0.](image)

**Figure 3.8 SEC-MALS of CpGV-PEP-L in Tris-HCl pH 8.0.** SEC-MALS was run in the buffer of 20 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM TCEP, 3 mM azide. Protein concentration used was 2.0 mg/mL. Inset is an SDS-PAGE showing the purity of the protein used for SEC-MALS.

Homogeneous protein is required for crystallization and characterization (e.g., by SAXS). A series of experiments were carried out to test the effect of different agents on the aggregation of CpGV-PEP-L. These included changing the salt concentration (0 mM (NH₄)₂SO₄ to 800 mM (NH₄)₂SO₄), adding extra reducing agent (1 mM TCEP to 10 mM TCEP, because CpGV-PEP-L has 5 cysteines), and testing different pH conditions. Varying salt concentration and reducing agent concentration did not produce any obvious change in aggregation (data not
shown). However, when purified in a high pH buffer of 100 mM Na$_2$CO$_3$ pH 11.8, 150 mM NaCl, size exclusion chromatography (SEC) revealed a more homogeneous dispersion of the protein by showing a more symmetrical peak with a narrower elution width (Figure 3.9). Sodium carbonate buffer is a buffer commonly used in purifying alkaline proteins and is stable for up to 6 months$^{56}$. SEC-MALS further showed that the aggregation of the CpGV-PEP-L broke down to a mixture of dimer to tetramer in 100 mM Na$_2$CO$_3$ pH 11.8 (Figure 3.10). Four different concentration of CpGV-PEP-L were run on SEC-MALS and did not show any obvious shifting of the profiles of the oligomers.

Figure 3.9 S200 size exclusion chromatography of CpGV-PEP-L in 100 mM Na$_2$CO$_3$ pH 11.8. Fraction numbers are shown in red. Selected fractions were analyzed by SDS-PAGE. (M) molecular weight marker (kDa).
Results

Expression of CpGV occlusion body surface proteins

Figure 3.10 SEC-MALS of CpGV-PEP-L in 100 mM Na₂CO₃ pH 11.8. Four concentrations of CpGV-PEP-L were run. Inset is a lane of an SDS-PAGE showing the purity of the sample used for SEC-MALS.

The pH dependence of CpGV-PEP-L oligomerization is further reflected by comparing the size exclusion chromatography of the protein that had been purified in different buffers of different pH (Figure 3.11). The peak shifts to later elution volumes in relation to the increase of the pH of the buffer used to purify the protein and to run the SEC in, suggesting the oligomerization attenuates as the pH becomes higher. This is also revealed on Native-PAGE of the protein in different pH conditions (Figure 3.12). As the pH lowered, more and more protein appeared at the top of the stacking gel and didn’t even get to enter the stacking gel, suggesting the oligomers got bigger and bigger as the pH lowered. In MES pH 6.0 buffer and HEPES pH 7.0 buffer, the protein precipitated so there was no protein showing on the gel, and probably too much aggregation caused the precipitation. There were several bands in each lane, indicating there were more than one species of the protein in solution. Only a little of de-oligomerization was observed as the protein was buffer-exchanged from Tris-HCl pH 8.0 to Na₂CO₃ pH 11.8 (data not shown), suggesting the pH dependence of oligomerization and de-oligomerization of CpGV-PEP-L is not completely reversible.
Figure 3.11 Size exclusion chromatography of CpGV-PEP-L at different pH.
Expression of CpGV occlusion body surface proteins

Results

Figure 3.12 Native-PAGE of CpGV-PEP-L at different pH. 10% Native-PAGE was run with protein in different buffers of different pH that are indicated on top of each lane. BSA was also run in a lane as a control.

Although up to this point CpGV-PEP-L has looked more homogeneous when purified in 100 mM Na₂CO₃ pH 11.8, it’s still a mixture of dimer to tetramer by nature. Despite this, extensive trials were still carried out to crystallize the protein, including using the protein purified in different pH, varying the protein concentration, setting up crystallization at different temperature, incorporating protease in the protein solution when setting up crystallization such as trypsin or α-chymotrypsin. No crystals of protein were observed.

Because of its being a mixture of dimer to tetramer by nature, SAXS could not be performed to characterize the protein in solution.

Since the CpGV-PEP-L full-length protein was not suitable for structural characterization, a number of truncations were made and tested (Table 3.3). None of these produced homogeneous protein (Figure 3.14 for truncation 1-300 aa, Figure 3.15 for truncation 1-251 aa, Figure 3.16 for truncation 1-211 aa), until they reached the N-terminal truncation of 1-106 aa.
aa of CpGV-PEP-L where the 1-106 aa truncation yielded a 1.4 Å crystal structure completed by Busby (PDB ID 4YE7). The N-terminal truncation (1-106 aa) of CpGV-PEP-L is comprised of a β-sheet layer which is made up of five anti-parallel β-strands (S1 to S5), surrounded by five α-helices (H1 to H5) (Figure 3.13).

Figure 3.13 Ribbon diagram of CpGV-PEP-L-1-106 aa atomic structure. A β-sheet layer made up of five anti-parallel β-strands (S1 to S5) surrounded by five α-helices (H1 to H5). Figure reproduced from Busby, PDB ID 4YE7.
Figure 3.14 Size exclusion chromatography of CpGV-PEP-L-1-300. SEC was run in 20 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM TCEP. Inset is a lane from an SDS-PAGE showing the purity of the sample used for SEC.
Figure 3.15 Size exclusion chromatography of CpGV-PEP-L-1-251 aa. SEC was run in 20 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM TCEP. Inset is an SDS-PAGE of the fractions indicated on top of each lane. (M) molecular weight marker (kDa).
Expression of CpGV occlusion body surface proteins

3.1.3 PEP-N domains of CpGV-PEP-M and CpGV-PEP-S

Since the PEP-N region of CpGV-PEP-L (the CpGV-PEP-L-1-106 aa truncation) had been crystallized, similar domains from CpGV-PEP-M and CpGV-PEP-S were expressed (Table 3.4). Of these, only CpGV-PEP-S-1-132 aa could be expressed and purified. It appeared as one main peak on size exclusion chromatography (Figure 3.17). Despite many trials, crystallization of this fragment was not successful. The construct was analysed by Native-PAGE (Figure 3.17, right inset). The PEP-S-1-132 band was smeared compared to the BSA control, suggesting more than one species in the protein solution and possibly explaining the difficulty of crystallization.

Figure 3.16 Size exclusion chromatography of CpGV-PEP-L-1-211. SEC was run in 20 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM TCEP. Inset is two lanes from an SDS-PAGE showing the purity of the sample at the peaks of SEC.
### Expression of CpGV occlusion body surface proteins

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1-131</td>
<td>E.coli</td>
<td>Yes</td>
<td>No</td>
<td>No expression</td>
<td>PEP-M-1-BamHI-Fo, PEP-M-131-EcoRI-Re</td>
<td>pETDuet</td>
</tr>
<tr>
<td>1-87</td>
<td>E.coli</td>
<td>Yes</td>
<td>No</td>
<td>No expression</td>
<td>PEP-M-1-87-EcoRI-Fo, PEP-M-1-87-XhoI-Re</td>
<td>pETDuet</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CpGV-PEP-S (aa)</th>
<th>System</th>
<th>Expression tested?</th>
<th>Purified?</th>
<th>Comment</th>
<th>Primers</th>
<th>Plasmid</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-132</td>
<td>E.coli</td>
<td>Yes</td>
<td>Yes</td>
<td>Mono-dispersed</td>
<td>PEP-S-1-Fo, PEP-S-132-Reverse</td>
<td>pETDuet</td>
</tr>
<tr>
<td>1-88</td>
<td>E.coli</td>
<td>Yes</td>
<td>No</td>
<td>Insoluble</td>
<td>PEP-S-1-88-EcoRI-Fo, PEP-S-1-88-XhoI-Re</td>
<td>pETDuet</td>
</tr>
</tbody>
</table>

**Table 3.4** Four constructs of the PEP-N domains of CpGV-PEP-M and CpGV-PEP-S.

![Size exclusion chromatography of CpGV-PEP-S-1-132 aa](image)

**Figure 3.17** Size exclusion chromatography of CpGV-PEP-S-1-132 aa. Left inset is an SDS-PAGE of the fractions from SEC indicated on top of each lane. Right inset is a Native-PAGE of the protein pooled from the peak of SEC. BSA was used as a reference.

#### 3.1.4 CpGV-PEP-L, PEP-M, PEP-S co-expression experiments

It is not clear which CpGV-PEP protein or proteins constitute the CpGV occlusion body envelope. It could be a single PEP protein of them, or two of them, or all of the three as a complex that comprised the envelope. Co-expression of various PEP proteins combinations
was also tried (Table 3.5). It was found that PEP-L could not be co-purified with PEP-M or PEP-S, while PEP-M/PEP-S was always co-purified as a complex regardless of being expressed in insect cells or *E.coli* (Figure 3.18). However, this PEP-M/PEP-S complex aggregated and started to precipitate when its concentration was above 1 mg/mL regardless of being produced in insect cells or *E.coli* (Figure 3.19). Adding Maltose-binding protein (MBP) tag to the PEP-M/PEP-S complex to enhance the complex’s solubility was also tried. Maltose-binding protein (MBP) has previously been shown to enhance solubility of fusion proteins, and has been employed in the University of Auckland Structural Biology group widely for both solubilisation and X-ray crystallography. MBP has 374 aa, and is 41 kDa. PEP-M and PEP-S still formed a complex when fused with MBP tag at the PEP-M end or at the PEP-S end (Figure 3.20, Figure 3.21), and its solubility was indeed increased, as it could be concentrated to a high concentration. However, the MBP-PEP-M/PEP-S or PEP-M/MBP-PEP-S complex was not homogeneous, as revealed by SEC and Native-PAGE (Figure 3.20, Figure 3.21). Despite this observation, many trials were still done to crystallize the protein complex, including varying the protein concentration, setting up crystallization at different temperature, incorporating protease in the protein solution when setting up crystallization such as trypsin or α-chymotrypsin. No protein crystals were observed.

<table>
<thead>
<tr>
<th>CpGV-PEPs (aa)</th>
<th>System</th>
<th>Expression tested?</th>
<th>Purified?</th>
<th>Comment</th>
<th>Primers</th>
<th>Plasmids</th>
</tr>
</thead>
<tbody>
<tr>
<td>His-L + M + S</td>
<td>Insect</td>
<td>Yes</td>
<td>IMAC, SEC</td>
<td>Only L was pulled down by IMAC. L oligomerizes.</td>
<td>Co-infection by pFastBacHTb, His-L and Multibac_L-M-S. Made by Busby.</td>
<td>pFastBacHTb, Multibac</td>
</tr>
<tr>
<td>L+ His-M + S</td>
<td>Insect</td>
<td>Yes</td>
<td>IMAC, SEC</td>
<td>Only M and S were pulled down by IMAC. This complex precipitates when concentration is &gt;1 mg/mL.</td>
<td>Co-infection by pFastBacHTb, His-M and Multibac_L-M-S. Made by Busby.</td>
<td>pFastBacHTb, Multibac</td>
</tr>
<tr>
<td>His+M+S</td>
<td><em>E.coli</em></td>
<td>Yes</td>
<td>IMAC, SEC</td>
<td>M and S were pulled down by IMAC. This complex precipitates when concentration is &gt;1 mg/mL.</td>
<td>orf23-BamHI-Fo, Orf23-EcoRI-Re, orf20-Forward-1, orf20-Reverse-1, orf20-Forward-2, orf20-Reverse-2</td>
<td>pETDuet</td>
</tr>
<tr>
<td>His-M + MBP-S</td>
<td><em>E.coli</em></td>
<td>Yes</td>
<td>IMAC, SEC</td>
<td>M and MBP-S were pulled down by IMAC. This complex can be concentrated, but not homogeneous in SEC.</td>
<td>MBP-3aa-orf23-Fo-1, MBP-3aa-orf23-Re-1, MBP-3aa-orf23-Fo-2, MBP-3aa-orf23-Re-2, orf20-BamHI-Fo, Orf20-EcoRI-Re</td>
<td>pETDuet</td>
</tr>
</tbody>
</table>
Results

Expression of CpGV occlusion body surface proteins

<table>
<thead>
<tr>
<th>Construct</th>
<th>E.coli</th>
<th>IMAC, SEC</th>
<th>Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>His-S + MBP-M</td>
<td>Yes</td>
<td>IMAC, SEC</td>
<td>S and MBP-M were pulled down by IMAC. This complex can be concentrated, but not homogeneous in SEC.</td>
</tr>
<tr>
<td>His-M-1-131 + S-1-132</td>
<td>Yes</td>
<td>No</td>
<td>No expression</td>
</tr>
<tr>
<td>His-M-1-87 + S-1-88</td>
<td>Yes</td>
<td>No</td>
<td>No expression</td>
</tr>
</tbody>
</table>

Table 3.5 Seven constructs of CpGV-PEP-L, PEP-M, PEP-S co-expression.

Figure 3.18 SDS-PAGE analysis of PEP-M/PEP-S complex. Infection of Sf9 cells by virus amplified from different batches produced the same result shown in different lanes. Expected positions of PEP-L, PEP-M, and PEP-S are indicated by coloured triangles. (M) molecular weight marker (kDa). Similar data was seen for protein complexes expressed in E.coli.
Figure 3.19 Negative stain EM image of PEP-M/PEP-S aggregate. The red arrows denote the aggregates. Protein complex was expressed in insect cells.
Figure 3.20 Size exclusion chromatography of MBP-PEP-M/PEP-S complex. Left inset shows an SDS-PAGE of the selected SEC fractions. MBP-PEP-M and His-tagged PEP-S on the SDS-PAGE are indicated by red triangles. Right inset is a Native-PAGE of the sample from the peak of the SEC. The smeared band on the Native-PAGE suggested the sample was not homogeneous. (M) molecular weight marker (kDa); (input) SEC input of sample purified by IMAC. MBP is 41 kDa, PEP-M is 26 kDa, PEP-S is 17 kDa. These bands are consistent with their expected positions.
Figure 3.21 Size exclusion chromatography of MBP-PEP-S/PEP-M complex. Left inset is an SDS-PAGE of the SEC fractions indicated on top of the lanes. MBP-PEP-S and His-tagged PEP-M on the SDS-PAGE are indicated by red triangles. Right inset is a Native-PAGE of the sample from the peak of the SEC. Notice the smeared band on the Native-PAGE, suggesting the sample was not homogeneous. (M) molecular weight marker (kDa); (input) SEC input of sample purified by IMAC. MPB is 41 kDa, PEP-M is 26 kDa, PEP-S is 17 kDa. These bands are consistent with their expected positions.

### 3.1.5 CpGV-PEP-L, PEP-M, PEP-S, polyhedrin co-expression

The possibility that the three CpGV-PEP proteins (CpGV-PEP-L, PEP-M, PEP-S) could be attached to the surface of CpGV-polyhedrin protein crystals when all four proteins were co-expressed with CpGV-polyhedrin in insect cells was investigated. One co-expression construct of these four proteins had been made by colleague Busby (Table 3.6). CpGV-polyhedrin crystals did form within Sf9 cells as when polyhedrin protein was expressed alone\(^{32,33}\). However, it turned out to be very difficult to purify these *in vivo* crystals without using 1% SDS, which would in turn denature proteins that were attached to the crystals, if any. Without purified crystals, it would be impossible to discern which proteins had been attached to the crystals, if any, because there would be too many bands on SDS-PAGE gel in addition to the polyhedrin protein band. Another approach of dissolving the crystals first and then pull down the associated proteins by IMAC was also tried. Alkaline condition dissolved
the crystals, but the alkaline condition at the same time also dissociated the PEP-M/PEP-S complex (data not shown). Therefore, this co-expression was not investigated in depth. Future work could be undertaken to purify these crystals via other approaches such as density gradient ultracentrifugation.

<table>
<thead>
<tr>
<th>CpGV-PEPs and polyhedrin</th>
<th>System</th>
<th>Expression tested?</th>
<th>Purified?</th>
<th>Comment</th>
<th>Primers</th>
<th>Plasmid</th>
</tr>
</thead>
<tbody>
<tr>
<td>L+M+S+polyhedrin</td>
<td>Insect</td>
<td>Yes</td>
<td>No</td>
<td>These in vivo crystals are difficult to purify without using 1% SDS.</td>
<td>Made by Busby.</td>
<td>Multibac</td>
</tr>
</tbody>
</table>

Table 3.6 One construct of CpGV-PEP-L, PEP-M, PEP-S, polyhedrin co-expression.

3.1.6 AcMNPV-PEP

Ten constructs of AcMNPV occlusion body envelope protein AcMNPV-PEP were made and tested (Table 3.7). Like CpGV-PEP-L, the AcMNPV-PEP full-length protein also aggregated (data not shown), and trials to produce and crystallize truncations of AcMNPV-PEP were unsuccessful.

<table>
<thead>
<tr>
<th>AcMNPV-PEP (aa)</th>
<th>System</th>
<th>Expression tested?</th>
<th>Purified?</th>
<th>Comment</th>
<th>Primers</th>
<th>Plasmid</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-252 (full-length)</td>
<td>Insect</td>
<td>Yes</td>
<td>IMAC, SEC</td>
<td>Aggregates</td>
<td>BamHI-AcPEP-Forward, EcoRI-AcPEP-Reverse</td>
<td>pFastBacHTb</td>
</tr>
<tr>
<td>1-252 (full-length)</td>
<td>E.coli</td>
<td>Yes</td>
<td>No</td>
<td>No expression</td>
<td>AcPEP-EcoRI-Duet, AcPEP-Xhol-Duet</td>
<td>pETDuet</td>
</tr>
<tr>
<td>MBP-1-252</td>
<td>E.coli</td>
<td>Yes</td>
<td>No</td>
<td>No expression</td>
<td>MBP-3aa-AcPEP-Forward-1, MBP-3aa-AcPEP-Reverse-1, MBP-3aa-AcPEP-Forward-2, MBP-3aa-AcPEP-Reverse-2</td>
<td>pETDuet</td>
</tr>
<tr>
<td>1-170</td>
<td>E.coli</td>
<td>Yes</td>
<td>No</td>
<td>No expression</td>
<td>AcPEP-1aa-Forward, AcPEP-170aa-Reverse</td>
<td>pETDuet</td>
</tr>
<tr>
<td>1-140</td>
<td>E.coli</td>
<td>Yes</td>
<td>No</td>
<td>Insoluble</td>
<td>AcPEP-1aa-Forward, AcPEP-140aa-Reverse</td>
<td>pETDuet</td>
</tr>
<tr>
<td>1-120</td>
<td>E.coli</td>
<td>Yes</td>
<td>No</td>
<td>Insoluble</td>
<td>AcPEP-1aa-Forward, AcPEP-120aa-Reverse</td>
<td>pETDuet</td>
</tr>
</tbody>
</table>
### Table 3.7 Ten constructs of AcMNPV-PEP tested.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Yes</th>
<th>IMAC</th>
<th>Reverse</th>
<th>pETDuet</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-99</td>
<td><em>E. coli</em></td>
<td></td>
<td>IMAC Precipitates immediately after IMAC</td>
<td>AcPEP-1aa-Forward, AcPEP-99aa-Reverse</td>
<td>pETDuet</td>
</tr>
<tr>
<td>1-81</td>
<td><em>E. coli</em></td>
<td></td>
<td>IMAC Precipitates immediately after IMAC</td>
<td>AcPEP-1aa-Forward, AcPEP-81aa-Reverse</td>
<td>pETDuet</td>
</tr>
<tr>
<td>1-73</td>
<td><em>E. coli</em></td>
<td>Yes</td>
<td>No</td>
<td>No expression</td>
<td>pETDuet</td>
</tr>
<tr>
<td>152-252</td>
<td><em>E. coli</em></td>
<td>Yes</td>
<td>IMAC, SEC Aggregates</td>
<td>AcPEP-151aa-Forward, AcPEP-252aa-Reverse</td>
<td>pETDuet</td>
</tr>
</tbody>
</table>

#### 3.2 Summary

34 constructs of CpGV and AcMNPV occlusion body surface proteins were investigated in this research (Figure 3.22, below). This included full length proteins and a series of truncations. There are 35 constructs in all, counting the shortest construct of the CpGV-PEP-L N-terminus region (1-106 aa) studied before.
Expression of CpGV occlusion body surface proteins

Summary
CpGV-PEP-L formed oligomers of dimer to tetramer in pH 11 condition, and aggregated more and more as the pH decreases. Expression of a series of truncations of CpGV PEPs or elongated protein fused with MBP tag did not result in protein fragment suitable for crystallization and X-ray crystallography. They oligomerized less, but still formed a mixture of different species in solution.

Co-expression of the three CpGV PEPs did not result in a complex of the three proteins as expected, but rather formed a complex of two PEPs the PEP-M/PEP-S complex. This suggests that the three CpGV PEPs might not work together. The complex of CpGV-PEP-M/PEP-S also oligomerized.

Co-expression of the three CpGV PEPs with polyhedrin protein did form in vivo crystals in Sf9 cells as when polyhedrin protein was expressed alone\textsuperscript{32-33}. The crystals were not easy to purify from insect Sf9 cells without using 1% SDS to remove cell contaminants. This co-expression of the four proteins was not investigated in depth. Future work could be undertaken to purify these crystals via other approaches such as density gradient ultracentrifugation.

There is only one PEP protein in baculovirus AcMNPV, but this AcMNPV-PEP also aggregated. It has not been tested whether AcMNPV-PEP will behave better in pH 11 condition, just like CpGV-PEP-L did. Future work could test this.

Crystallization trials for all these proteins and truncations were not successful.

The shortest truncation of CpGV-PEP-L, the N-terminus region (1-106 aa) yielded a crystal structure, which had been completed by Busby (PDB ID 4YE7).
Chapter 4 Electron microscopy investigation of CpGV occlusion bodies

This chapter describes investigations of the CpGV occlusion bodies by electron microscopy (EM) method, including negative stain EM and cryo-EM. The aim was to find the condition where the occlusion body surface fibres could be observed and then to take quality images of the fibres, to lay the groundwork for structural determination of the fibres in future work. This chapter also includes experiments that were carried out with the aim of imaging isolated fibres.

4.1 Negative stain electron microscopy

Negative stain EM of CpGV occlusion bodies was performed in the manner described in Chapter 2. In negative stain EM, Fibres on the surface of CpGV occlusion bodies could not be distinguished (Figure 4.1). Intact occlusion bodies could also be observed in non-stain EM (Figure 4.2), although this did not shed any light on the surface of the occlusion bodies. In conclusion, negative stain EM can aid in assessing the CpGV samples, but is not suitable for generating images with observable fibre structures.
Figure 4.1 Negative stain EM image of CpGV occlusion body. TMV was used as an internal control. TMV and CpGV occlusion body are indicated by magenta and red triangles. Scale bar 200 nm.

Figure 4.2 Non-stain EM images of CpGV occlusion bodies.
4.2 Cryo-electron microscopy

Cryo-EM was performed in the manner described in Chapter 2. As can be seen in Chapter 2 section 2.26.2, cryo-EM is a complicated experiment, and there are many steps of the experiment that can go wrong including condensing ethane to a good temperature of just above its melting point (around -170°C), vitrification, transferring vitrified sample into cryo-holder, and transferring the cryo-holder into EM. It needs two days for one round of cryo-EM experiment from the preparation of vacuumed cryo-holder, and if one step went wrong, the whole experiment would need to start over. At first, a lot of time was consumed to learn and practice this technique, and many problems were encountered such as no vitrified ice in the holes of the grid (data not shown), thick ice in the holes of grid (Figure 4.3), and crystalline ice (Figure 4.4). After many trials, images with thin, vitreous ice suitable for image collection were produced (Figure 4.5). In this condition, the fibres on the surface of CpGV occlusion body are shown clearly (Figure 4.6).
Figure 4.3 Cryo-EM image of CpGV occlusion bodies showing thick ice. Red triangles indicate radiation damaged CpGV occlusion bodies. Note the thick ice in the holes of the grid not suitable for cryo-EM study.
Figure 4.4 Cryo-EM image of CpGV occlusion bodies showing crystalline ice. Blue triangles indicate crystalline ice not suitable for image collection.
Figure 4.5 Cryo-EM image of CpGV occlusion bodies showing thin, vitreous ice. Red triangles indicate CpGV occlusion bodies. Blue triangle indicates either crystalline ice or blobs of ethane contaminants. Note the vitreous ice in the holes of the grid suitable for cryo-EM study.
Figure 4.6 shows the crystalline facets of occlusion bodies in polygonal outline. The fibres are about 15 nm long and are attached to the protein crystal.

Images of these (Figure 4.6) were not suitable for generating a 3-D reconstruction model of the fibre structure. Because intact CpGV occlusion body has a certain thickness and fibres along this thickness always overlapped each other when the image was taken. Second, the occlusion bodies did not spread evenly on a holy grid for cryo-EM, and this made it difficult to produce as many images as needed of the fibres on the intact CpGV occlusion bodies. This was also encountered by our EM technical support staff member Adrian Turner (personal communication).

Experiment to isolate the fibres from the CpGV occlusion bodies was done in the way the previous member of our group S.Yeh did (unpublished result), i.e., to treat CpGV occlusion
bodies in reducing agent DTT solution. If separate fibres could be obtained, they could be studied in cryo-EM. However, after treating CpGV occlusion bodies in 50 mM HEPES buffer pH 7.5 containing 10 mM DTT for 7 days at 4°C, the fibres were found to be still attached to the occlusion bodies (Figure 4.7). This contradicts our previous colleague’s result. One big discrepancy between the current experiment and our previous colleague’s experiment was that in the current experiment the reaction temperature was set at 4°C for 7 days, while in the previous colleague’s experiment it was carried out at 56°C for 1 hour. The reason a lower temperature was used in the current experiment was because the fibres were to be maintained in a native condition. However, if it’s the reducing agent DTT that was working for detachment of the fibres, then the temperature wouldn’t have been an affecting factor, as DTT is widely used at 4°C to treat proteins. On the other hand, if the fibres could be detached at 56°C, whether it’s still in a native conformation or not was still in question. The isolation of fibres from intact CpGV occlusion bodies could be further investigated in future work.
Figure 4.7 Cryo-EM image of DTT treated CpGV occlusion bodies showing fibres still attached. Blue triangle indicates either ethane contaminant or crystalline ice. Magenta triangle indicates TMV. Red triangle indicates CpGV occlusion body fibres. Notice the fibres are still attached to the occlusion bodies after DTT treatment.

4.3 Summary

Fibres surrounding intact occlusion bodies were observed by Cryo-EM. Images of these were not suitable for generating a 3-D reconstruction model of the fibre structure. Because intact CpGV occlusion body has a certain thickness and fibres along this dimension overlap each other when the image is taken. Second, the occlusion bodies did not spread evenly on a holy grid for cryo-EM, and this made it difficult to produce as many images as needed of the fibres on the intact CpGV occlusion bodies. This was also encountered by our EM technical support staff member Adrian Turner (personal communication).
An experiment to detach and isolate the fibres from OB was done. However, this experiment could not reproduce the result the previous member of our group S.Yeh had, i.e., fibres were not removed from the OB in this research using reducing agent DTT as the previous member of our group S.Yeh did. The different temperature in these two experiments might have caused the difference of the results. Future work could investigate the effect of DTT at different temperature on the detachment of fibres.
Chapter 5 X-ray crystallographic structure determination for the CpGV phosphatase PTP-2

5.1 Background

Protein kinases and phosphatases regulate cellular processes by adding or removing phosphates at specific sites on target proteins. These two enzyme classes together constitute about 3% of eukaryotic proteomes\(^6^2\). Approximately 30% of the human proteome is phosphorylated\(^6^3\) illustrating the extent and complexity of eukaryotic phosphorylation networks, which in addition to other functions control the cell cycle and signalling pathways\(^6^4\). Some large (>100Kbp) DNA virus genomes also encode protein kinases and phosphatases. Figure 5.1 represents all the predicted protein kinases and phosphatases derived from the complete virus genomes currently available at the NCBI protein clusters database (http://www.ncbi.nlm.nih.gov/proteinclusters). Protein kinases and phosphatases occur mainly in the *Chordopoxvirinae* and *Baculoviridae* virus families (i.e., poxviruses and baculoviruses). Poxviruses are predominantly vertebrate viruses and include significant human pathogens (e.g., the variola virus that causes smallpox). Baculoviruses are ubiquitous insect viruses and are utilized for the biocontrol of specific insect pests. They are divided into two main subfamilies, the *Alphabaculoviridae* and the *Betabaculoviridae*. The prototypical alphabaculovirus *Autographa californica* multicapsid nucleopolyhedrovirus (AcMNPV) is the best characterized baculovirus and is well known because it is widely used to express recombinant proteins in cultured insect cells. The AcMNPV genome encodes two predicted kinases and three predicted protein phosphatases (Table 5.1). The genome of the prototypical betabaculovirus *Cydia pomonella* granulovirus (CpGV) encodes one predicted protein kinase PK-1 and three predicted protein phosphatases. Both viruses have representatives of the most common kinase and phosphatase protein clusters (2744809 and 274817), but the other predicted protein kinases and phosphatases in the viral genomes are not shared between the two.
Background  X-ray crystallographic structure determination for the CpGV phosphatase PTP-2

Figure 5.1 Predicted protein kinases and phosphatases derived from virus genomes. Homology groups were obtained from NCBI protein clusters website (http://www.ncbi.nlm.nih.gov/proteinclusters) using the search string "viruses"[organism] together with either of the terms "kinase" or "phosphatase". Kinases and phosphatases with non-protein targets were omitted. Each segment of the pie charts is labelled with the NCBI proteincluster number and name together with the included virus families and PDB codes for atomic structures. Clusters representing less than 5% of the total are indicated by the light blue unlabelled segments. *Cydia pomonella* PTP-2 phosphatase belongs to the protein tyrosine phosphatase-2 proteincluster shown by the expanded green segment in B.
### Background

X-ray crystallographic structure determination for the CpGV phosphatase PTP-2

<table>
<thead>
<tr>
<th>Virus</th>
<th>Predicted function</th>
<th>Protein name</th>
<th>NCBI protein ID</th>
<th>NCBI protein cluster ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>AcMNPV</td>
<td>protein kinase</td>
<td>PK-1</td>
<td>NP_054039</td>
<td>2744809₁</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PK2</td>
<td>NP_054153</td>
<td>2509980</td>
</tr>
<tr>
<td></td>
<td>protein phosphatase</td>
<td>PTP</td>
<td>NP_054030</td>
<td>2509865</td>
</tr>
<tr>
<td></td>
<td></td>
<td>38K</td>
<td>NP_054128</td>
<td>2744817²</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NP_054062</td>
<td>2509896</td>
</tr>
<tr>
<td>CpGV</td>
<td>protein kinase</td>
<td>PK-1</td>
<td>NP_148787</td>
<td>2744809³</td>
</tr>
<tr>
<td></td>
<td>protein phosphatase</td>
<td>ORF88</td>
<td>NP_148872</td>
<td>2744817²</td>
</tr>
<tr>
<td></td>
<td>PTP-2</td>
<td>NP_148882</td>
<td>2945255</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PTP-3*</td>
<td>NP_148850</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 5.1 Predicted protein kinases and phosphatases in AcMNPV and CpGV genomes.** Data from [https://www.ncbi.nlm.nih.gov/proteinclusters](https://www.ncbi.nlm.nih.gov/proteinclusters). Superscripts indicate homologues in protein clusters which occur in both AcMNPV and CpGV. * Renamed here as PTP-3 because the NCBI names for NP_148882 and NP_148850 are both ‘PTP-2’.

Structures have been determined for the poxvirus phosphatases vaccinia VH1⁶⁵ and variola H1⁶⁶, and for the orf virus phosphatase OH1⁶⁷, which all belong to phosphatase protein cluster 2509785. There is a single structure in phosphatase protein cluster 2509865, that for the AcMNPV BVP phosphatase which can hydrolyze the 5’-phosphate from RNA and free nucleotide di- and tri-phosphates⁶⁸. No structures have yet been reported for the major class of viral phosphatase protein cluster 2744817, that comprise over a third of all predicted viral phosphatases.

Here we report the crystal structure of the phosphatase PTP-2 from *Cydia pomonella* granulovirus, the first structure of a member of the protein tyrosine phosphatase-2 protein cluster 2945255. A BLAST search with the *C. pomonella* PTP-2 amino acid sequence revealed that most homologues occur in insect baculoviruses and entomopoxviruses (Figure 5.2). The best functionally characterised viral PTP-2 phosphatase is that of the alphabaculovirus *Spodoptera exigua* multiple nucleopolyhedrovirus (SeMNPV), where expression of the ptp-2 gene induces apoptosis and increases yields of infectious occlusion bodies in infected larvae, compared to recombinant viruses without the ptp-2 gene⁶⁹.
Figure 5.2 BLAST search using *Cydia pomonella* granulovirus PTP-2 protein sequence against the NCBI *refseq* protein database. Horizontal coloured lines represent the 50 best matching protein sequence alignments listed in order below that for *C. pomonella* PTP-2, shown by the top red line. Except for *C. pomonella* PTP-3 the indigo lines all represent alphabaculovirus PTP-2 homologues. The green lines are mainly protein fragments from a variety of insects, but also from other species such as fish. The two longer green lines are full-length entomopox insect viral PTP-2 homologues. The vertical red lines represent the active site residues. The green box shows the region of PTP-2 that corresponds to the DSPc superfamily fold-type identified by the NCBI routines. All matches have E values < 1e-6.

5.2 Cloning

The ptp-2 gene was amplified from extracted CpGV genome by PCR using the ptp-2 forward primer (5'-CCGGAATTGATGTCGCCACGCGAATAC-3') and the ptp-2
reverse primer (5’- CCGCTCGAGCTATTCAACTCTAACCACCAACTTATTATC-3’). The PCR product was digested using EcoRI and XhoI, and ligated into pETDuet plasmid, including a His<sub>6</sub> tag directly upstream of the gene sequence. Clone was verified by sequencing.

### 5.3 Expression and purification

The recombinant plasmid was transformed into *Escherichia coli* BL21 (DE3) competent cells. The transformed cells were cultured in 2 L Erlenmeyer flasks with 100 µg/ml ampicillin at 37 °C to an OD<sub>600 nm</sub> of 0.6-0.8, and induced at 18°C with IPTG at a final concentration of 0.25 mM for 18 hours.

The cells were harvested by centrifugation and the pellets were resuspended in lysis buffer (20 mM Tris-HCl pH 8.0, 200 mM NaCl, 20 mM imidazole). Lysozyme was added to the resuspension to a final concentration of 1 mg/ml and incubated at 4°C for 15 min. The resuspension was then homogenized by sonication at 4°C using the procedure stated in Chapter 2. The cell lysate was clarified by centrifugation at 17,000 RCF at 4°C for 30 min. The supernatant was loaded to a Ni-IMAC column equilibrated with lysis buffer. The column was then washed with twenty column volumes of lysis buffer before the His<sub>6</sub> tagged PTP-2 protein was eluted in three column volumes of elution buffer (20 mM Tris-HCl pH 8.0, 200 mM NaCl, 250 mM imidazole). The protein was then concentrated and applied to a Superdex S200 size exclusion column pre-equilibrated in 20 mM Tris-HCl pH 8.0, 200 mM NaCl. Fractions containing PTP-2 were pooled and concentrated to 22 mg/ml for crystallization. SeMet protein was produced from bacteria grown in SeMet medium using procedure detailed in Chapter 2. SeMet protein was purified by Ni-IMAC column and a Superdex S75 size exclusion column, and concentrated to 40 mg/ml for crystallization. Both proteins were readily purified by IMAC and SEC (Figure 5.3, Figure 5.4), with 2 L bacteria culture yielding 22 mg native PTP-2 and 4 L bacteria culture yielding 80 mg SeMet PTP-2 after SEC. The purified proteins were aliquoted and flash-frozen in liquid nitrogen and stored at -80°C.
Figure 5.3 S200 size exclusion chromatography of native PTP-2. Red numbers are collected fractions. Inset is an SDS-PAGE for the selected fractions indicated on top of the gel. (M) molecular weight marker (kDa).
Crystallization

Initial crystallization condition was found in the condition of 0.1 M HEPES-NaOH pH 7.5, 200 mM CaCl$_2$, and 20% PEG 6000 (PACT premier kit, Molecular Dimensions), using sitting-drop vapour-diffusion method. Further drops were prepared by mixing 1 µl protein
solution with 1 µl reservoir solution and were equilibrated against 80 µl reservoir solution. Crystals appeared in 24 hours in 0.1 M HEPES-NaOH pH 7.5, 200 mM CaCl2, and 14%-20% PEG 6000 (Figure 5.5). SeMet PTP-2 was crystallized in the same condition as the native PTP-2 (Figure 5.6), with higher protein concentration (40 mg/mL of SeMet PTP-2) and at higher precipitant concentration (18%-26% PEG 6000). SeMet PTP-2 at 20 mg/mL did not yield good crystals and also not reproducible. Crystals were briefly soaked in a cryo-protectant solution of 15 % glycerol (substituting water in the reservoir solution with 15 % glycerol) prior to flash-cooling in liquid nitrogen.

Figure 5.5 Crystals of native PTP-2. Crystal form is cluster of thin plates.
5.5 X-ray diffraction and structure solution

Diffraction data was measured at the Australian Synchrotron using beamline MX1: Macromolecular Crystallography beamline. Diffraction data of SeMet crystal were collected to a resolution of 2.0 Å (Figure 5.7) and native crystal to a resolution of 1.65 Å (Figure 5.8). The data were processed using IMOSFLM\textsuperscript{70} and AIMLESS\textsuperscript{71}. The structure was solved by the SAD method using AutoSol\textsuperscript{72} which led to an initial model for the protein. This model was then used for molecular replacement with PHASER\textsuperscript{73} to solve the structure of the native protein structure. The model was then improved with iterative rounds of manual building and refining with Coot\textsuperscript{74} and subsequently refined with PHENIX\textsuperscript{72}. Data collection and refinement statistics are listed in Table 5.2. Atomic coordinates and structure factors have been deposited into the Protein Data Bank (www.pdb.org) with the PDB accession code 6I28.

**Figure 5.6 Crystals of SeMet PTP-2.** Crystal form is cluster of thin plates.
Figure 5.7 SeMet PTP-2 diffraction image (1° oscillation).
Figure 5.8 Native PTP-2 diffraction image (1° oscillation).
<table>
<thead>
<tr>
<th></th>
<th>SeMet PTP-2</th>
<th>Native PTP-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelength (Å)</td>
<td>0.9537</td>
<td>0.9537</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>54.7-2.20 (2.27-2.20)</td>
<td>33.7 -1.65 (1.68 -1.65)</td>
</tr>
<tr>
<td>Space group</td>
<td>I 2 2 2</td>
<td>I 2 2 2</td>
</tr>
<tr>
<td>Unit cell (Å, °)</td>
<td>64.4 65.97 97.56 90 90 90</td>
<td>64.6898 66.72 97.96 90 90 90</td>
</tr>
<tr>
<td>Total reflections</td>
<td>154401 (13430)</td>
<td>870135 (84879)</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>10898 (937)</td>
<td>25887 (2526)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>14.2 (14.3)</td>
<td>14.1 (14.2)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>100 (100)</td>
<td>100 (100)</td>
</tr>
<tr>
<td>Mean I/sigma(I)</td>
<td>10.4 (1.8)</td>
<td>11.2 (0.7)</td>
</tr>
<tr>
<td>Wilson B-factor (Å²)</td>
<td>22.1</td>
<td>20.7</td>
</tr>
<tr>
<td>R-merge</td>
<td>0.25 (1.71)</td>
<td>0.17 (3.9)</td>
</tr>
<tr>
<td>R-meas</td>
<td>0.27 (1.84)</td>
<td>0.18 (4.1)</td>
</tr>
<tr>
<td>R-pim</td>
<td>0.01 (0.67)</td>
<td>0.05 (1.1)</td>
</tr>
<tr>
<td>CC1/2</td>
<td>0.99 (0.68)</td>
<td>0.90 (0.40)</td>
</tr>
</tbody>
</table>

Refinement

| Reflections used in refinement | -             | 25883 (2526) |
| Reflections used for R-free   | 1307 (142)    |              |
| R-work                         | 0.19 (0.37)   |              |
| R-free                         | 0.22 (0.40)   |              |
| CC(work)                       | 0.94 (0.24)   |              |
| CC(free)                       | 0.87 (0.13)   |              |
| Number of non-hydrogen atoms  | 1445          |              |
| Macromolecules                 | 1328          |              |
| Ligands                        | 7             |              |
| Solvent                        | 110           |              |
| Protein residues               | 161           |              |
| RMSD(bonds)                    | 0.016         |              |
| RMSD(angles)                   | 1.29          |              |
| Ramachandran favored (%)       | 96.9          |              |
| Ramachandran allowed (%)       | 3.1           |              |
| Ramachandran outliers (%)      | 0             |              |
| Rotamer outliers (%)           | 2.1           |              |
| Clashscore                     | 3.4           |              |
| Average B-factor (Å²)          | 34.6          |              |
| Macromolecules                 | 33.7          |              |
| Ligands                        | 60.8          |              |
| Solvent                        | 43.45         |              |
| PDB ID                         | -             | 6I28         |

Table 5.2 PTP-2 data collection and refinement statistics.
5.6 PTP-2 crystal structure and discussion

The electron density revealed a single molecule in the asymmetric unit and the complete polypeptide was clearly interpretable allowing an unambiguous model to be fitted for all 161 amino acids (Figure 5.9, Figure 5.10). The refinement statistics and model parameters indicate that the model is well defined. No significant surface regions were found using the PISA server (www.ccp4.ac.uk/pisa) that form extended intermolecular contacts in the crystal.

Figure 5.9 Ribbon diagram of CpGV PTP-2. (Black labels), secondary structure elements. (Red labels), conserved active site residues. Figure prepared using PyMol®.
A search for structural homologues using PDBefold (www.ebi.ac.uk/msd-srv/ssm) revealed that the N-terminal ~130 amino acids of PTP-2 form a CC1 protein phosphatase fold, the most commonly occurring phosphatase fold[62]. This fold corresponds to the DSPc (dual specificity phosphatase, catalytic domain) architecture in the SMART catalogue (http://smart.embl.de/smart/do_annotation.pl?DOMAIN=SM00195), and is also evident from the results of a NCBI BLAST sequence homology search (Figure 5.2). The DSPc family is widely represented in major phyla and currently includes 125 structures and 5580 sequences.

The closest DSPc structure to PTP-2 is a human dual specificity phosphatase DUSP13b (PDB ID 2PQ5) that matches with an RMS deviation of 1.5 Å over 130 C-alpha positions with a sequence identity of 23%. The structures of viral protein phosphatases in NCBI protein clusters 2509785 and 2509865 (Figure 5.1) fit to PTP-2 with an RMS deviation of C-alpha positions of ~2.0 Å over ~130 residues when compared to the PTP-2 structure using gesamt[76]. The corresponding sequence identities are 15% and 17%. A superimposition of PTP-2 and its homologue 1YN9 (viral protein phosphatase in NCBI protein cluster 2509865) is shown in Figure 5.11.

In addition to the common fold, DSPc structures share a conserved active site architecture and catalytic mechanism[77]. The cysteine in the active site motif CX5R ‘P-loop’ (C97 in PTP-2) is located at the base of a positively charged pocket at the N-terminal end of helix H5 and forms
a phosphor-cysteine intermediate with the substrate phosphate. This intermediate is resolved by the neighbouring aspartate in the ‘D-loop’ (D64 in PTP-2).

PTP-2 in Figure 5.9 shows the fold together with side-chains of the five P-loop residues and the D-loop aspartate that have atoms within 5Å of the active site. The active site residues are also shown in the alignment of NCBI protein cluster 2945255 homologues (Figure 5.12) and the wider alignment of BLAST hits to PTP-2 shown in Figure 5.2. The regions of sequence homology extend well beyond the active site. PTP-2 shares the larger HCxxGxxR motif and other similarities to DUSPs (dual specificity phosphatases) and the active site pocket topography is similar to that of DUSP1478. The closest full length BLAST matches to PTP-2 are from alphabaculoviruses and entomopoxviruses, but not from other betabaculoviruses. The shorter matches shown in Figure 5.2 are mainly to fragments of insect proteins and these matches cover only the C-terminal part of the core of the molecule extending from strand S5 to H6 and including the active site P-loop but excluding the D-loop. The *Cydia pomonella* granulovirus genome encodes one of these shorter homologues, which we call here PTP-3. PTP-3 is 60% identical to PTP-2 suggesting it is functional despite the lack of a D-loop, having presumably been maintained in the genome for sufficient time to diverge from PTP-2. Interestingly, similar truncated versions of PTP-2 are not found in other baculoviruses.
Figure 5.11 Superimposition of PTP-2 and its homologue 1YN9. (Magenta), PTP-2. (Green), 1YN9. Figure prepared using PyMol. 

PTP-2 crystal structure and discussion  X-ray crystallographic structure determination for the CpGV phosphatase PTP-2
Figure 5.12 Structure based PTP-2 multiple sequence alignment. The header lists species names and NCBI codes for representative sequences selected from the NCBI protein cluster 2945255 (the whole cluster is shown in Figure 5.13). Coloured dots correspond to the amino acid numbers in the alignment below. The top row above the alignment is the index for C. pomonella PTP-2. The next row shows secondary structure elements from Figure 5.9. The row below the alignment indicates homology. Active site residues shown in Figure 5.9 are marked in red with the letter ‘a’. The structure based alignment was prepared using T-Coffee using the PTP-2 atomic structure.
PTP-2 and its viral homologues and PTP-3 all have C-terminal extensions beyond the conserved DSPc core structure (Figure 5.2). In PTP-2 this extension forms an anti-parallel beta sheet containing the two strands S6-S7 which extends from the molecule in a different configuration to the C-terminal extensions previously described for DUSP14 and DUSP18. The protein cluster 2945255 sequence alignment (Figure 5.12) shows that other alphabaculovirus PTP-2 homologues have either a similar C-terminal extension or an N-terminal extension, but not both. The only other betabaculovirus PTP-2 (Crytophlebia leucotreta granulovirus) in the protein cluster sequence alignment does not have extensions at either end of the molecule.

Given the complexity of phosphorylation networks in cells, it is likely that PTP-2 has a number of phosphorylated protein and non-protein substrates. It may also function in cells as an oligomer or in a complex with other proteins. Homologous viral phosphatases have already been described that are dimeric (e.g. OH1) or that have non-protein substrates. Future proteomics investigations can be expected to identify substrates and interacting partners. This

Figure 5.13 Rooted phylogenetic tree for the PTP-2 NCBI protein cluster 2945255. The root sequence on the left side of the figure is Cydia pomonella PTP-2 (NP_148882). Coloured sequences were included in the multiple sequence alignment shown in Figure 5.12.
structure, together with biochemical characterization of its activity on defined substrates will lead to a much better understanding of the role of phosphorylation networks in granulovirus infected cells.
Chapter 6 Discussion

6.1 Introduction

The goal of project one in this research was to characterize the crystalline occlusion body surface proteins of baculovirus, with the aim of understanding the crystalline occlusion body surface, and the fibres. The goal of project two in this research was to structurally characterize the PTP-2, a phosphatase of CpGV, to provide molecular structural information for understanding its function in baculovirus phosphorylation.

6.2 The occlusion body surface and the fibres

Fibres at the surface of crystalline occlusion bodies of baculovirus CpGV are novel structures. The baculovirus occlusion body surface was reported to contain mostly carbohydrates and the polyhedral envelope protein\textsuperscript{37,38}. However, there had not been any reports about the fibres at the occlusion body surface at that time and it was not known what made up the fibres.

Our group previously investigated the composition of the fibres by a combination of experiments including cryo-EM and protein identification by mass spectrometry. The results showed that all the three PEPs of CpGV are involved in making up the fibres. There are many remaining questions. The expected size positions of the three full length PEPs do not correspond to the bands on the SDS-PAGE, that were much lower than the expected PEPs. The explanation for this could be that the PEPs had been degraded by protease associated with OB\textsuperscript{30,80}, but then it gives rise to another question, which is, in each single band on the SDS-PAGE, two proteins were detected. This means either each band on the SDS-PAGE had actually been a mixture of two fragments degraded to the same size, or each single band had been a single fragment and the other protein detected from this band was a false positive response from mass spectrometry. If there was false positive response\textsuperscript{81,82} from mass spectrometry in that study, not all the three PEPs of CpGV might be the components of the fibres. Future work could investigate this again.
6.3 Characterization of occlusion body surface proteins

There are three PEPs proteins in CpGV and there is one PEP protein in AcMNPV. All these PEPs have an N-terminal PEP-N domain. CpGV-PEP-L and AcMNPV-PEP have an additional C-terminal PEP-C domain (Figure 3.2). The crystal structure of the truncated CpGV-PEP-L N-terminus region (1-106 aa) had been solved by Busby (PDB ID 4YE7).

CpGV-PEP-L and AcMNPV-PEP belong to the same protein cluster 2945044 (Table 3.2). Remote protein homology detection for proteins in the protein cluster 2945044 using program HHpred shows PEP-C domain is a structural homology of the reovirus fibre structure (PDB ID 6GAP) of a triple coil (Figure 6.1, Figure 6.2).

Based on the sequence-structure alignment (Figure 6.2), the boundary for the CpGV-PEP-C coiled-coli domain (Figure 3.2) should start from the amino acid 134 rather than the 161, and the boundary for the AcMNPV-PEP-C coiled-coli domain (Figure 3.2) begins at amino acid 115 rather than 167 as previously predicted in the literature.

Interestingly, the PEP-N structure (4YE7) was solved as a monomeric structure while the predicted coiled-coil structure is a triple coil. This suggests that in forming the triple coil structure of the fibres the PEP-C domain plays a significant role.

Figure 6.1 HHpred results for proteins in protein cluster 2945044.
Characterization of occlusion body surface proteins

Discussion

Figure 6.2 Protein structures corresponding to the HHpred results for protein cluster 2945044.

Coordinates from the CpGV_L N-terminal domain (4YE7) and the reovirus fibre (6GAP) PDB structures are shown in ribbon representation. The C-terminal helices (red) of three 4YE7 molecules were aligned with the 6GAP triple coil. CpGV has a 27 amino acid linker with three conserved cysteines between PEP-N and PEP-C domain, AcMNPV has a 20 amino acid SR repeat with two conserved cysteines located between the PEP-N and PEP-C domain. The overall length is 230Å. The figure was prepared using PyMol.

In this research, a number of constructs for the CpGV and AcMNPV full length occlusion body surface proteins and truncations were investigated (Figure 3.21).

CpGV-PEP-L formed oligomers. In particular, it formed a dimer to tetramer mixture in pH 11 condition. The oligomers aggregated more and more as the pH decreased. A series of truncations suggests that the oligomerization is caused by PEP-C region, as N-terminal truncations shorter and shorter (Figure 3.21) became less and less oligomerized (Table 3.3), with the shortest truncation of 1-106 aa having yielded a crystal structure (PDB ID 4YE7). C-terminal truncation was also investigated, and this region indeed aggregated.
Co-expression of the three CpGV PEPs did not result in a complex of the three proteins, but rather formed a complex of two PEPs, the PEP-M/PEP-S complex, which also oligomerized. This suggests that the three CpGV PEPs might not work together. Together with the HHpred prediction result (Figure 6.1, Figure 6.2), it implies CpGV-PEP-L alone or the CpGV-PEP-M/CpGV-PEP-S complex could have formed the fibre structure.

Expression of a series of truncations of CpGV PEPs or elongated protein fused with MBP tag did not result in protein fragments suitable for crystallization and X-ray crystallography. They oligomerized less, but still formed a mixture of different species in solution, and the crystallization trials were unsuccessful.

Co-expression of the three CpGV PEPs with polyhedrin protein formed crystals within Sf9 cells. The crystals were not easy to purify from insect Sf9 cells without using 1% SDS to remove cell debris. This co-expression of the four proteins was not investigated in depth. Future work could be undertaken to purify these crystals via other approaches such as density gradient ultracentrifugation. If purified crystals were available, they could be dissolved in alkaline solution and analysed on SDS-PAGE, to determine if the PEPs protein were associated with the surface of these in vivo crystals of polyhedrin. If they were, the crystals with the PEPs at the surface could be further studied by cryo-EM. If they were not, other components or proteins from CpGV genome, which was lacking in this four-protein co-expression, might have been needed to locate the PEPs to the surface of polyhedral envelope.

The single PEP protein in baculovirus AcMNPV also aggregated. It was not tested whether the AcMNPV-PEP would behave better in pH 11 condition, just like in the case of CpGV-PEP-L where it appeared as smaller oligomers of a mixture of dimer to tetramer. This could be tested in future work.

### 6.4 Cryo-electron microscopy study on occlusion body surface fibres

Fibres surrounding intact occlusion bodies were observed by Cryo-EM. Images of these were not suitable for generating a 3-D reconstruction model of the fibre structure. Because intact CpGV occlusion body has a certain thickness and fibres along this dimension overlap each other when the image is taken. Second, the occlusion bodies did not spread evenly on a holy grid for cryo-EM, and this made it difficult to produce as many images as needed of the fibres.
on the intact CpGV occlusion bodies. This was also encountered by our EM technical support staff member Adrian Turner (personal communication). An estimation of the dimensions of the fibres from the EM micrographs is consistent with the dimensions of the predicted coiled-coil structure of the fibres (Figure 6.3).

Figure 6.3 Combination of fibres from EM micrographs and the predicted triple coil structure of fibres showing their dimensions.

Detaching and isolating the fibres from OB is another way of getting fibres for cryo-EM study. However, two protocols need to be established for this scheme to work. First, a protocol needs to be set up where fibres can be detached without denaturing them. Although it was previously found in our group that reducing agent DTT could detach the fibres when the OBs were treated at 56°C (S.Yeh, unpublished result), this experiment was not reproduced in this research when the treating temperature was 4°C. Low temperature was not expected to
have an effect on the working ability of DTT, as DTT is widely used to treat proteins at 4°C. Future work could test the effect of temperature on the detachment of fibres, i.e., to test different temperatures, and also to test different temperatures with the use of DTT. The second problem is that a lot of raw materials of CpGV OB is needed, to isolate fibres from OB. The CpGV sample used in this study was a gift from Dr Johannes Jehle and was not much. Future work could consider purifying CpGV OBs purchased from companies before carrying out the fibre isolation experiment.

6.5 CpGV PTP-2 atomic structure

The crystallographic analysis of PTP-2 revealed a single molecule in the asymmetric unit. The model is well defined and unambiguously fitted for all 161 amino acids. The N-terminal ~130 amino acids of PTP-2 form a CC1 protein phosphatase fold, the most commonly occurring phosphatase fold. This fold corresponds to the DSPc (dual specificity phosphatase, catalytic domain) architecture in the SMART catalogue (http://smart.embl.de/smart/do_annotation.pl?DOMAIN=SM00195). The DSPc family is widely represented in major phyla and currently includes 125 structures and 5580 sequences.

In addition to the common fold, DSPc structures share a conserved active site architecture and catalytic mechanism. The cysteine in the active site motif CX5R ‘P-loop’ (C97 in PTP-2) is located at the base of a positively charged pocket at the N-terminal end of helix H5 and forms a phosphor-cysteine intermediate with the substrate phosphate. This intermediate is resolved by the neighbouring aspartate in the ‘D-loop’ (D64 in PTP-2). PTP-2 shares the larger HCxxGxxR motif and other similarities to DUSPs (dual specificity phosphatases) and the active site pocket topography is similar to that of DUSP14.

The closest full-length BLAST matches to PTP-2 are from alphabaculoviruses and entomopoxviruses, but not from other betabaculoviruses. The Cydia pomonella granulovirus genome encodes one shorter homologue of PTP-2, which we call here PTP-3. PTP-3 is 60% identical to PTP-2 suggesting it is functional despite the lack of a D-loop, having presumably been maintained in the genome for sufficient time to diverge from PTP-2. Interestingly, similar truncated versions of PTP-2 are not found in other baculoviruses.

Given the complexity of phosphorylation networks in cells, it is likely that PTP-2 has a number of phosphorylated protein and non-protein substrates. It may also function in cells as
an oligomer or in a complex with other proteins. Homologous viral phosphatases have already been described that are dimeric (e.g. OH1\textsuperscript{67}) or that have non-protein substrates\textsuperscript{68,67}. Clearly much remains to be learned about the role of PTP-2 in virally infected cells. Future proteomics investigations can be expected to identify substrates and interacting partners. This structure, together with biochemical characterization of its activity on defined substrates will lead to a much better understanding of the role of phosphorylation networks in granulovirus infected cells.
## Primer sequences

### BamHI-AcPEP-Forward
CCGGGATCCATGAAGCCGACGAATAACG

### EcoRI-AcPEP-Reverse
CCGGAAATTCTCAACGTGGTGCAGCAGGTTTG

### AcPEP-EcoRI-Duet
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### AcPEP-Xhol-Duet
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### MBP-3aa-AcPEP-Forward
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### MBP-3aa-AcPEP-Reverse
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### MBP-3aa-AcPEP-Forward-2
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### MBP-3aa-AcPEP-Reverse-2
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### AcPEP-1aa-Forward
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### AcPEP-73aa-Reverse
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### AcPEP-81aa-Reverse
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### AcPEP-99aa-Reverse
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### AcPEP-120aa-Reverse
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### AcPEP-140aa-Reverse
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### AcPEP-151aa-Reverse
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### Forward-CpGV-PepeL-BamHI
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### Reverse-CpGV-PepeL-EcoRI
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### CpGV-PepeL-EcoRI-Duet
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### CpGV-PepeL-Xhol-Duet
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### PEP-L-300-Xhol-Duet
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### PEP-L-296-Xhol-Duet
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### PEP-L-282-Xhol-Duet
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### PEP-L-277-Xhol-Duet
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### PEP-L-269-Xhol-Duet
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### PEP-L-251-Xhol-Duet
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### PEP-L-241-Xhol-Duet
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### PEP-L-230-Xhol-Duet
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### PEP-L-220-Xhol-Duet
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### PEP-L-210-Xhol-Duet
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### PEP-L-200-Xhol-Duet
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### PEP-L-190-Xhol-Duet
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### PEP-L-180-Xhol-Duet
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### PEP-L-170-Xhol-Duet
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### PEP-L-160-Xhol-Duet
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### PEP-L-150-Xhol-Duet
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### PEP-L-140-Xhol-Duet
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### PEP-L-130-Xhol-Duet
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### PEP-L-120-Xhol-Duet
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### PEP-L-110-Xhol-Duet
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### PEP-L-100-Xhol-Duet
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### PEP-L-90-Xhol-Duet
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### PEP-L-80-Xhol-Duet
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### PEP-L-70-Xhol-Duet
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### PEP-L-60-Xhol-Duet
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### PEP-L-50-Xhol-Duet
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### PEP-L-40-Xhol-Duet
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### PEP-L-30-Xhol-Duet
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### PEP-L-20-Xhol-Duet
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### PEP-L-10-Xhol-Duet
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### PEP-L-0-Xhol-Duet
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### PEP-L-100-Forward
GAAAGCTGGGTGTTACGCAATGTACGCG

### PEP-L-100-Reverse
GAAAGCTGGGTGTTACTTTTTCACTACTT

### MBP-3aa-Cp300aa-Forward
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### MBP-3aa-Cp300aa-Reverse
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### MBP-3aa-Cp300aa-Fo-2
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### MBP-3aa-Cp300aa-Re-2
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### MBP-3aa-Cp300aa-Fo-2
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### MBP-3aa-Cp300aa-Re-2
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### orf23-BamHI-Forward
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### orf23-EcoRI-Forward
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### orf20-Forward-1
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### orf20-Reverse-1
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### orf20-Forward-2
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG

### orf20-Reverse-2
CCGGAAATTCCAGATATCCTCAACGTGGTGCAGCAGGTTTG
<table>
<thead>
<tr>
<th>Primer sequences</th>
<th>Appendses</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>orf20-Reverse-2</strong></td>
<td>CGCTCGAGTCAAGTGTGTGATCTCCCTGTTCAATAG</td>
</tr>
<tr>
<td><strong>MBP-3aa-orf20-Fo-1</strong></td>
<td>CCGGAATTCGAGCTCGGCGCGCCTGCAGGTCGAC</td>
</tr>
<tr>
<td><strong>MBP-3aa-orf20-Re-1</strong></td>
<td>CTATTACGCGAGTATCCAGATTACGCATATGATATCTCCTTCTTATAC</td>
</tr>
<tr>
<td><strong>MBP-3aa-orf20-Fo-2</strong></td>
<td>ATGTGTGAATCTGAGGATACGGCGATAAAG</td>
</tr>
<tr>
<td><strong>MBP-3aa-orf20-Re-2</strong></td>
<td>CTTTAGAATCTCAGGGGAATAGGCGCAGTCTCCGCGGTC</td>
</tr>
<tr>
<td><strong>MBP-3aa-orf20-Fo-4</strong></td>
<td>ATGGCCCTCTTATCCCGCAGATTCAAA</td>
</tr>
<tr>
<td><strong>MBP-3aa-orf20-Re-4</strong></td>
<td>CCGCTCGAGTCAAGTGTGTGATCTCCCTGTTCAATAG</td>
</tr>
<tr>
<td><strong>MBP-3aa-orf23-Fo-1</strong></td>
<td>CCGGAATTCGAGCTCGGCGCGCCTGCAGGTCGAC</td>
</tr>
<tr>
<td><strong>MBP-3aa-orf23-Re-1</strong></td>
<td>ATGGCTGTGAATCTGAGGATACGGCGATAAAG</td>
</tr>
<tr>
<td><strong>MBP-3aa-orf23-Fo-2</strong></td>
<td>CCGCTCGAGTCAAGTGTGTGATCTCCCTGTTCAATAG</td>
</tr>
<tr>
<td><strong>MBP-3aa-orf23-Re-2</strong></td>
<td>CCGGAATTCGAGCTCGGCGCGCCTGCAGGTCGAC</td>
</tr>
<tr>
<td><strong>orf20-BamHI-Fo</strong></td>
<td>CCGGAATTCGAGCTCGGCGCGCCTGCAGGTCGAC</td>
</tr>
<tr>
<td><strong>orf20-BamHI-Re</strong></td>
<td>CCGGAATTCGAGCTCGGCGCGCCTGCAGGTCGAC</td>
</tr>
<tr>
<td><strong>PEP-M-1-87-BamHI-Fo</strong></td>
<td>CCGGAATTCGAGCTCGGCGCGCCTGCAGGTCGAC</td>
</tr>
<tr>
<td><strong>PEP-M-1-87-EcoRI-Re</strong></td>
<td>CCGGAATTCGAGCTCGGCGCGCCTGCAGGTCGAC</td>
</tr>
<tr>
<td><strong>PEP-S-1-88-Fo-1</strong></td>
<td>CCGGAATTCGAGCTCGGCGCGCCTGCAGGTCGAC</td>
</tr>
<tr>
<td><strong>PEP-S-1-88-Re-1</strong></td>
<td>AGGGTTGTGTGGAGCACATATGATATCTCCTTCTTATAC</td>
</tr>
<tr>
<td><strong>PEP-S-1-88-Fo-2</strong></td>
<td>CCGGAATTCGAGCTCGGCGCGCCTGCAGGTCGAC</td>
</tr>
<tr>
<td><strong>PEP-S-1-88-Re-2</strong></td>
<td>ATGTGCTCCACACAACCT</td>
</tr>
<tr>
<td><strong>PEP-M-1-87-BamHI-Fo</strong></td>
<td>CCGGAATTCGAGCTCGGCGCGCCTGCAGGTCGAC</td>
</tr>
<tr>
<td><strong>PEP-M-1-87-EcoRI-Re</strong></td>
<td>CCGGAATTCGAGCTCGGCGCGCCTGCAGGTCGAC</td>
</tr>
<tr>
<td><strong>PEP-S-1-132-Fo-2</strong></td>
<td>ATGTGCTCCACACAACCT</td>
</tr>
<tr>
<td><strong>PEP-S-1-132-Re-2</strong></td>
<td>CCGGAATTCGAGCTCGGCGCGCCTGCAGGTCGAC</td>
</tr>
<tr>
<td><strong>CpGV-tp1-2-F</strong></td>
<td>CCGGAATTCGAGCTCGGCGCGCCTGCAGGTCGAC</td>
</tr>
<tr>
<td><strong>CpGV-tp1-2-R</strong></td>
<td>CCGGAATTCGAGCTCGGCGCGCCTGCAGGTCGAC</td>
</tr>
</tbody>
</table>
References

   doi:10.1002/9780470015902.a0020712.pub2


