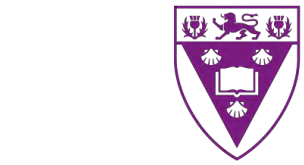


The expression and evaluation of CrpeNPV gp37 as a formulation additive for enhanced infectivity with CrleGV-SA and improved *Thaumatotibia leucotreta* control.

A thesis submitted in fulfilment of the

requirements for the degree of

MASTER OF SCIENCE

at

RHODES UNIVERSITY

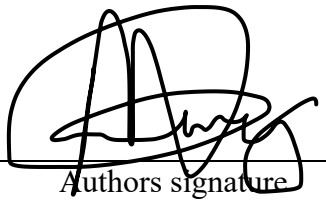
By

Naho Muleya

February 2024

Declaration

I, Naho Muleya (g22m0008) hereby declare that the thesis submitted is my own work. It is being submitted for the degree of Master of Science at Rhodes University. It has not been previously submitted for assessment of any degree at any other university or other body, organisation outside of the university.



Authors signature

09 February 2024

Dates

Abstract

Thaumatotibia leucotreta (Meyrick) (Lepidoptera: Tortricidae) is a significant pest native to Africa, causing damage to citrus and posing a threat to the export of fresh citrus in South Africa. Classified as a phytosanitary risk by several South African export markets, this pest necessitates effective control measures. Baculoviruses emerge as promising biological control agents against *T. leucotreta* due to their inherent safety and eco-friendly characteristics. Among these, Cryptophlebia leucotreta Granulovirus (CrleGV-SA) and Cryptophlebia peltastica Nucleopolyhedrovirus (CrpeNPV) stand out, both causing larval mortality upon infecting *T. leucotreta.* CrleGV-SA has been formulated into the products Cryptogran™, CryptoMax™ and Cryptex®, while CrpeNPV has been formulated into the product Multimax™. Both viruses are used in integrated pest management programmes to reduce fruit damage in agricultural fields, with CrleGV-SA having been employed against *T. leucotreta* for nearly 20 years in South Africa. However, these control options are limited by factors such as virulence and the slow speed of kill. This limitation can be addressed by exploiting potential synergistic relationships between baculoviruses infecting the same host. Previous studies have demonstrated that the truncated CpGV gp37 can enhance the infectivity of NPVs on other lepidopteran pests, such as *Spodoptera exigua* (Hübner). Although the mechanism behind this phenomenon remains unclear, it presents an opportunity to enhance the effectiveness of baculovirus-based management strategies. Notably, the genome of CrpeNPV encodes *gp37*, while CrleGV-SA lacks this gene. The potential interaction between CrleGV-SA and CrpeNPV gp37 remains unexplored. Therefore, investigating whether they exhibit synergistic or antagonistic effects is essential for optimising baculovirus-based management of *T. leucotreta*.

This study aims to express CrpeNPV gp37 in a bacterial system and then evaluate its effect on larval mortality when combined with CrleGV-SA in laboratory bioassays. The initial step involved extracting genomic DNA (gDNA) from occlusion bodies (OBs) of CrpeNPV. A modified Quick DNA Miniprep plus kit was utilised, which entailed pre-treatment with Na_2CO_3 followed by neutralisation with Tris-HCI before gDNA extraction using the kit. Subsequently, the concentration of the gDNA was estimated using a Nanodrop spectrophotometer. Oligonucleotides targeting the CrpeNPV *gp37* gene were designed for PCR amplification, with the gDNA serving as a template. The *gp37* amplicon was identified through agarose gel electrophoresis and then gel purified in preparation for cloning.

Secondly, the purified PCR product was cloned into the intermediate vector pJET1.2/blunt and then subcloned into the bacterial expression vector pCA528 through DNA ligation. The construction of recombinant plasmids (pJET-gp37 and pCA-gp37) was conducted and verified using Colony PCR, plasmid extraction, restriction enzyme analysis, and Sanger sequencing.

Thirdly, the recombinant protein (6×His-SUMO-gp37) was expressed and purified using Nickel affinity chromatography and analysed through SDS-PAGE and Western blot techniques. The expression of 6×His-SUMO-gp37 was carried out at both 25 °C and 18 °C. A time course induction study was conducted, inducing transformed cells for 0-, 3-, 5-, and 24-hours post induction (hpi). SDS-PAGE and Western blotting of samples collected at various time points revealed that 6×His-SUMO-gp37, approximately 42 kDa in size, was visible from 3 hpi, with maximal expression at 24 hpi. Solubility analysis of 6×His-SUMO-gp37 was performed at both temperatures, showing solubility at 18 °C but predominantly present in the insoluble fraction. The soluble protein was purified under native conditions, while the insoluble protein was purified under denaturing conditions. Despite being unable to elute 6×His-SUMO-gp37 under native conditions, successful elution was achieved under denaturing conditions, confirmed via Western blot analysis. No further experiments were conducted on the eluted 6×His-SUMO-gp37 under denaturing conditions.

Lastly, a preliminary surface dose bioassay was conducted to evaluate the efficacy of pelleted bacteria expressing 6×His-SUMO-gp37 in combination with CrleGV-SA against *T. leucotreta* neonates. Two lethal concentration doses of CrleGV-SA were prepared: a low concentration (2.96×10^4 OBs/mL) capable of killing 40 % of the *T. leucotreta* population, and a high concentration (2.96×10^5 OBs/mL) capable of killing 90 % of the population. The target protein, 6×His-SUMO-gp37, and the control, pCA528, were obtained by lysing the cells, centrifuging the samples, and collecting the insoluble fractions in pellet form. These fractions were then resuspended in PBS and used as treatments in combination with the prepared CrleGV-SA concentration doses. The concentration of the pellets was estimated using a Nanodrop spectrophotometer by measuring the absorbance at 280 nm. The bioassay results revealed that the combination of 100 µg/mL of pelleted bacteria expressing 6×His-SUMO-gp37 with CrleGV-SA had no effect on *T. leucotreta* larval mortality compared to CrleGV-SA alone. A one-way ANOVA was performed to assess differences among the virus treatment groups, concluding that no statistically significant differences were observed among the groups. The

experiments in this study provided valuable insights for future research, particularly in exploring the use of a protein-virus combination as a novel method for pest control.

Table of contents

Declaration ii

Abstract iii

Table of contents vi

List of figures xi

List of tables xv

List of abbreviations xvi

Research outputs xix

Acknowledgements xx

Chapter 1 1

General introduction and literature review 1

1.1 Importance of Citrus in South Africa 1

1.2 Insect pests and the threat to crop production 1

1.3 Management of *Thaumatotibia leucotreta* 2

1.3.1 Integrated pest management of *Thaumatotibia leucotreta* 2

1.3.2 Baculoviruses: their classification, morphology, genome structure and biology 3

1.3.2.1 Taxonomy and nomenclature 3

1.3.2.2 Morphology and classification 4

1.3.2.3 Infectious life cycle of baculoviruses 5

1.3.2.4 Genomic composition and Structural/Non-Structural protein encoded by Baculoviruses 6

1.4 Baculoviruses and their role as biological control agents 8

1.4.1 Introduction to baculoviruses-based biocontrol 8

1.4.2 *Betabaculovirus cryleucotretae,* Cryptophlebia leucotreta Granulovirus 9

1.4.3 *Alphabaculovirus Crypeltasticae*, Cryptophlebia peltastica Nucleopolyhedrovirus 10

1.4.4 Challenges associated with baculovirus-based biopesticides 10
1.5 Overcoming challenges associated with baculovirus-based biopesticides 11
1.5.1 Evaluating mixtures of baculoviruses in South Africa 11
1.5.2 Infectivity proteins targeting the peritrophic membrane 12
1.5.3 The use of gp37 an additive for the potential improvement of infectivity 13
1.6 Motivation 13
1.7 Aim and objectives 14
Overview of chapters 15
Chapter 2 16
Oligonucleotide design and PCR amplification of the *gp37* coding sequence from CrpeNPV genomic DNA in preparation for cloning 16
2.1 Introduction 16
2.2 Material and methods 18
2.2.1 pCA528 vector and CrpeNPV genome 18
2.2.2 Oligonucleotide design for PCR amplification of *gp37* 18
2.2.3 Orientation of the gene in the genome 18
2.2.4 Genomic DNA extraction 19
2.2.5 Agarose gel electrophoresis 20
2.2.6 PCR amplification and gel purification 20
2.3 Results 21
2.3.1 The pCA528 expression vector 21
2.3.2 Oligonucleotides design 22
2.3.3 Genomic DNA extraction 23
2.3.4 PCR amplification and gel purification of *gp37* 24
2.4 Discussion 24
Chapter 3 26
Cloning of the *gp37* gene into pJET1.2/blunt and subsequent sub-cloning into pCA528 for protein expression 26

3.1 Introduction 26
3.2 Methods and materials 27
3.2.1 Cloning of *gp37* into pJET1.2/blunt 27
3.2.1.1 *In silico* construction of pJET-gp37 plasmid 27
3.2.1.2 Ligation of *gp37* into pJET1.2/blunt 27
3.2.1.3 Transformation of pJET-gp37 28
3.2.1.4 Colony PCR and plasmid extraction 28
3.2.1.5 Restriction enzyme analysis 28
3.2.1.6 Gel purification of *gp37* product 29
3.2.2 Sub-cloning of gp37 into pCA528 29
3.2.2.1 Target vector preparation 29
3.2.2.2 *In silico* construction of the pCA-gp37 map 29
3.2.2.3 Ligation of *gp37* into pCA528 and bacterial transformation 29
3.2.2.4 Colony PCR and plasmid extraction 30
3.2.2.5 Restriction enzyme analysis 30
3.2.2.6 Sanger sequencing 30
3.3 Results 31
3.3.1 Cloning of *gp37* into pJET1.2/blunt 31
3.3.1.1 *In silico* pJET-gp37 plasmid map 31
3.3.1.2 Confirmation of pJET-gp37 32
3.3.2 Sub-cloning *gp37* into pCA528 33
3.3.2.1 Target vector preparation 33
3.3.2.2 *In silico* pCA-gp37 map 34
3.3.2.3 Colony PCR and plasmid extraction 35
3.3.2.4 Restriction enzyme analysis 36
3.3.2.5 Sanger sequence analysis of pCA-gp37 37
3.4 Discussion 39

Chapter 4 41

Expression and purification of 6×His-SUMO-gp37 in Rosetta (DE3) *Escherichia coli* competent cells 41

4.1 Introduction 41

4.2 Methods and materials 43

4.2.1 Competent cell preparation 43

4.2.2 Transformation of Rosetta (DE3) *E. coli* competent cells 43

4.2.3 The expression of 6×His-SUMO-gp37 in Rosetta (DE3) *E. coli* cells 44

4.2.4 SDS-PAGE analysis 44

4.2.5 Western blot analysis 45

4.2.7 Purification of 6×His-SUMO-gp37 and SUMO-GFP using Ni-NTA affinity chromatography 46

4.2.7.1 Green Fluorescent Protein (GFP) 46

4.2.7.2 Protein purification of 6×His-SUMO-gp37 and SUMO-GFP under native conditions 46

4.2.7.3 Protein purification of 6×His-SUMO-gp37 and SUMO-GFP under denaturing conditions 46

4.3 Results 47

4.3.1 Expression of 6×His-SUMO-gp37 at 25 °C and 18 °C 47

4.3.2 Solubility analysis of 6×His-SUMO-gp37 under native conditions 49

4.3.3 Purification of 6×His-SUMO-gp37 and SUMO-GFP using affinity chromatography 50

4.4 Discussion 52

Chapter 5 57

Evaluation of 6×His-SUMO-gp37 on *Thaumatotibia leucotreta* larval mortality when combined with CrleGV-SA in laboratory biological assay 57

5.1 Introduction 57

5.2 Methods and materials 58

5.2.1 6×His-SUMO-gp37 and pCA528 (empty vector) preparations....58
5.2.2 Acquisition of neonates....59
5.2.3 Surface-dose biological assay....59
5.2.4 Statistical analysis....60
5.3 Results....61
5.3.1 Solubility analysis of 6×His-SUMO-gp37 and pCA528 (empty vector)....61
5.3.2 Biological activity of CrleGV-SA in combination with 6×His-SUMO-gp37 against *T. leucotreta* neonates....62
5.3.3 One-way ANOVA statistical analysis....63
5.4 Discussion....64
Chapter 6....68
General discussion....68
6.1 Thesis overview....68
6.2 Genome analysis and cloning of CrpeNPV *gp37*....69
6.3 The expression and purification of the recombinant protein, CrpeNPV gp37....70
6.4 Biological assay using 6×His-SUMO-gp37....72
6.5 Potential future research....73
6.6 Conclusion....75
References....76

List of figures

Chapter 1:

Figure 1.1. The morphology of baculovirus displaying the A) Occlusion derived virus, B) Budded virus, C) Nucleopolyhedrovirus occlusion bodies, D) Granulovirus occlusion bodies (Jukes, 2018).....5

Figure 1.2. The schematic representation illustrates the primary and secondary infectious cycles of the baculovirus, demonstrating the transmission route through an infected host (Adapted from Williams et al., 2017). 6

Chapter 2:

Figure 2.1. Schematic representation depicting the position of the gene in the vector. **A**: Linear representation of the CrpeNPV genome and the orientation of the *gp37* gene, **B**: PCR product (*gp37*) and pCA528 expression vector. 19

Figure 2.2. The annotated pCA528 expression vector. Abbreviations used: *KanR*- Kanamycin resistance gene, SUMO- Small ubiquitin-related modifier tag, 6×His- Polyhistidine tag, RBS: Ribosome binding site, Ori- Origin of replication, bom- Basis of mobility, rop- Rop protein, lacI- *Lac* repressor, and MCS- Multiple cloning site. 22

Figure 2.3. Agarose gel image of genomic DNA extracted from CrpeNPV OBs. Lane L: GeneRuler 1 Kb plus DNA ladder and lane 1: CrpeNPV gDNA. 23

Figure 2.4. Agarose gel image showing the amplification and gel purification of the CrpeNPV *gp37* gene. Lane L:1 Kb plus GeneRuler DNA ladder, lane 1: NTC, lane 2: PCR amplicon (*gp37*), and lane 3: Purified *gp37* product 24

Chapter 3:

Figure 3.1. *In silico* construction of the pJET-gp37 map. Abbreviation used: AmpR- Ampicillin resistance gene, MCS- Multiple cloning site, *Eco47IR*- Lethal gene and Ori- origin of replication. 31

Figure 3.2. Agarose gel images showing colony PCR, plasmid extraction, restriction enzyme analysis of pJET-gp37 and gel purification of the *gp37* fragment. A: Colony PCR, lane 1: NTC, lane 2-5: *gp37* PCR product (colonies 1-4). B: Plasmid extraction, lane 1-4: Recombinant pJET-gp37. C: Restriction enzyme analysis of plasmid 2 (colony 2), lane 1:

Undigested pJET-gp37, lane 2: Single digest with BamHI, lane 3: Single digest with XhoI, and lane:4 Double digest with XhoI and BamHI. D: Gel purification, lane 1: Purified *gp37* fragment. For all panels, lane L: GeneRuler 1 Kb plus DNA ladder. 33

Figure 3.3. Agarose gel images showing plasmid extraction, restriction enzyme analysis and gel purification of pCA528. A: Plasmid extraction, lane 1-4: pCA528 DNA. B: Restriction enzyme analysis, lane 1: Undigested pCA528, and lane 2: Double digest with BamHI and XhoI. C: Gel purification, lane 1: Purified pCA528 backbone. For all panels, lane L: GeneRuler 1 Kb plus DNA ladder. 34

Figure 3.4. *In silico* construction of the pCA-gp37 map. Abbreviation used: RBS- Ribosome binding site, SUMO- Small ubiquitin-like modifier, 6×His- Polyhistidine tag, LacI- *Lac* repressor, *gp37*- The insert, 6×His-SUMO-gp37- Recombinant protein, KanR- Kanamycin resistance gene and Ori- Origin of replication. 35

Figure 3.5. Agarose gel images showing colony PCR and plasmid extraction of the pCA-gp37. A: Colony PCR, lane 1-3: *gp37* PCR product (colonies 1-3), and lane 4: NTC. B: Plasmid extraction, lane 1-3: pCA-gp37 from colonies 1-3. For both panels, lane l: GeneRuler 1 Kb plus DNA ladder.....36

Figure 3.6. Agarose gel image showing the restriction enzyme analysis of the pCA-gp37. Lane L: GeneRuler 1 Kb plus DNA ladder, lane 1: Undigested pCA-gp37 DNA, lane 2: Single digest with BamHI, lane 3: Single digest with XhoI, and lane 4: Double digest with BamHI and XhoI.....36

Figure 3.7. Sequence alignment of the regions flanking the *gp37* insert in pCA-gp37. The pCA-gp37 was used as a reference sequence, Sequence 1 and 2 are the forward and reverse sequences generated following Sanger sequencing of pCA-gp37. A) shows N-terminus of the recombinant protein, B) shows the central region of the recombinant protein, and C) shows the C-terminus of the recombinant protein. The translated amino acid sequence is shown from the start and to the stop codons in blue, with the 6×His tag in purple, the SUMO tag in red, the enzyme recognition sites of BamHI and XhoI in yellow, and the recombinant protein denoted as 6×His-SUMO-gp37 in orange. 38

Chapter 4:

Figure 4.1. Expression of 6×His-SUMO-gp37 at two different temperatures. Panel A1 shows SDS-PAGE analysis at 25 °C, while Panel B1 shows SDS-PAGE analysis at 18 °C. Panel A2 and B2 show Western blot analyses. Lane M represents the protein ladder in

kilodaltons (kDa), lanes 1,3,5, and 7 represent samples taken at 0-, 3-, 5- and 24- hpi (empty vector control), and lanes 2, 4, 6 and 8 represent samples taken at 0-, 3-, 5- and 24 hpi (6×His-SUMO-gp37), respectively.....49

Figure 4.2. Solubility analysis of 6×His-SUMO-gp37 under native conditions (cells lysed with lysozyme) at two different temperatures. Panel A1 shows SDS-PAGE analysis of proteins expressed at 25 °C, while Panel B1 shows SDS-PAGE analysis of proteins expressed at 18 °C. Panel A2 and B2 show Western blot analyses of proteins expressed at 25 and 18 °C, respectively. Lane M represents the protein marker in kilodaltons (kDa), lane W represents the whole lysate (total protein), lane P represents the pellet fraction (insoluble after centrifugation), and lane S represents the supernatant fraction (soluble after centrifugation).....50

Figure 4.3. Purification of 6×His-SUMO-gp37 and SUMO-GFP under native (cells lysed with lysozyme) and denaturing conditions (cells lysed with Urea). SDS-PAGE analysis under native conditions of A) 6×His-SUMO-gp37 and B) SUMO-GFP. SDS-PAGE analysis under denaturing condition of C1) 6×His-SUMO-gp37 and D) SUMO-GFP. Panel C2 shows Western blot analysis of 6×His-SUMO-gp37 under denaturing conditions. Lane M represents the Page Ruler pre-stained protein marker in kilodalton (kDa), lane W represents the whole lysate (total protein), lane P represents the pellet fraction (insoluble after centrifugation), lane S represents the supernatant fraction (soluble after centrifugation). Lane FT represents the flow-through fraction, lane W1 represents the wash fraction, and lanes E1-E3 represent the eluates, respectively.....52

Chapter 5:

Figure 5.1. Solubility analysis of 6×His-SUMO-gp37 and pCA528. Panel A shows SDS-PAGE analysis, and Panel B shows Western blot analysis, respectively. S1 indicates the solubility analysis of 6×His-SUMO-gp37, while S2 indicates the solubility analysis of pCA528 (empty vector). Lane M represents the protein marker in kilodaltons (kDa), lane W represents the whole lysate (total protein), lane P represents the pellet fraction (insoluble after centrifugation), and lane S represents the supernatant fraction (soluble after centrifugation).....61

Figure 5.2. Evaluation of the effect of 6×His-SUMO-gp37 on CrleGV-SA against *T. leucotreta* A) low (2.96×10^4 OBs/mL) and B) High (2.96×10^5 OBs/mL) concentrations. Treatments

include CrleGV-SA alone (CrleGV), CrleGV-SA combined with bacterial extracts expressing 6×His-SUMO-gp37 (CrleGV + gp37), CrleGV-SA combined with bacterial extracts transformed with the pCA528 vector (CrleGV + pCA528), and ddH_2O control (water). Different letters (a and b) indicate significant differences between treatments. 64

List of tables

Chapter 1:

Table 1.1. Core genes conserved among members of family *Baculoviridae* (Rohrmann, 2019; van Oers et al., 2023) 7

Chapter 2:

Table 2.1. PCR set up. 21

Table 2.2. Oligonucleotides targeting *gp37* gene in CrpeNPV genome. Sequences for the restriction sites are underlined, the *gp37* target sequence is shown in bold, and the additional thymine nitrogenous base pair is marked in red. 23

Chapter 4:

Table 4.1. Description of wash buffers used in denaturing purification experiments. 47

Chapter 5:

Table 5.1. List of the four treatments used for biological assays at two concentrations of CrleGV-SA OBs. The descriptions of the treatments were abbreviated as shown under the two concentration doses of CrleGV-SA. 60

Table 5.2. The average percentage mortality and standard deviation (SD) for each dose-response biological assay replicate. 62

List of abbreviations

AcMNPV - Autographa californica multiple nucleopolyhedrovirus

AGE - Agarose gel electrophoresis

AgMNPV - Anticarsia gemmatalis multiple nucleopolyhedrovirus

AgseGV - Agrotis segetum granulovirus

AgseNPV - Agrotis segetum nucleopolyhedrovirus

AmpR - Ampicillin resistance gene

ANOVA - Analysis of variance

Bp - Base pair

BLAST - Basic Local Alignment Search Tool

BVs - Budded virions

CE - Controlled environment

$CaCl_2$ - Calcium chloride

CpGV - Cydia pomonella granulovirus

CpGV-M - Cydia pomonella granulovirus, Mexican isolate

CrleGV-CV3 - Cryptophlebia leucotreta granulovirus, Cape Verde isolate

CrleGV-SA - Cryptophlebia leucotreta granulovirus, South African isolate

CrpeNPV - Cryptophlebia pelstastica nucleopolyhedrovirus

CTAB - Cetramethyl amonium bromide

DALRRD - Department of Agriculture, Land reform, and Rural development

DdH_2O - Double distilled water

DNA - Deoxyribonucleic acid

dsDNA - Double-strand deoxyribonucleic acid

EDTA - Ethylenediaminetetraacetic acid

g - Gram

gDNA - Genomic deoxyribonucleic acid

GV - Granuloviruses

hpi - Hour post induction

ICTV - International committees on Taxonomy of viruses

IPM - Integrated Pest Management

IPTG - Isopropyl ß-D-1-thiogalactopyranoside

KanR - Kanamycin resistance gene

kDa - Kilodalton

LC_{50} - Lethal concentration (50 %)

LC_{90} - Lethal concentration (90 %)

LT_{50} - Lethal Time (50 %)

M - Molar

MabrNPV - Mamestra brassicae nucleopolyhedrovirus

mg - Milligram

$MgCL_2$ - Magnesium chloride

MNPV - Multiple nucleopolyhedroviruses

µL - Microlitre

µM - Micromolar

mM - Millimolar

MCS - Multiple cloning site

Na_2CO_3 - Sodium carbonate

NaCl - Sodium chloride

NTC - No-template control

NPV - Nucleopolyhedrovirus

OBs - Occlusion bodies

ODV - Occlusion derived virion

ORF - Open reading frame

PCR - Polymerase chain reaction

Pif - *Per os* infectivity factor

PlxyGV - Plutella xylostella granulovirus

PsunGV - Pseudaletia unipunctata granulovirus

SA - South Africa

SDS-PAGE - Sodium dodecyl-sulphate polyacrylamide gel electrophoresis

SeMNPV - Spodoptera exigua multiple nucleopolyhedrovirus

SNPs - Single nucleotide polymorphisms

SNPV - Single nucleopolyhedrovirus

SpliNPV - Spodoptera littoralis nucleopolyhedrovirus

SUMO - Small ubiquitin-like modifier

TAE - Tris base, acetic acid and EDTA

TnGV - Trichoplusia ni granulovirus

V - Volts

v/v - Volume per volume

w/v - Weight per volume

× g - Times gravity

°C - Degrees Celsius

% - Percentage

Research outputs

Muleya N., Jukes M., Hill, M., Moore, S., Knox C. (2023). The expression and evaluation of CrpeNPV gp37 as a formulation additive for enhanced infectivity with CrleGV and improved *Thaumatotibia leucotreta* control. *23rd Congress of the Entomological Society of Southern Africa.* 11-14 July 2023. Stellenbosch University, Western Cape.

Acknowledgements

I would like to acknowledge the following people for their support and assistance in the completion of this work.

- My supervisor, Prof. Caroline Knox, and co-supervisors, Dr Michael Jukes, Prof. Martin Hill, and Prof. Sean Moore, I thank you for all the advice, support, assistance, and patience throughout my Master's journey. I am grateful; without your guidance, none of this would have been possible.
- Ntando Mkhwanazi for providing *T. leucotreta* eggs whenever I needed them.
- Citrus Research International (CRI), the National Research foundation (NRF), and the Centre for biological Control at Rhodes University, for the opportunity, financial support, and resource.
- Siviwe Tole for her support, laughter, and all the wonderful conversations
- My lab mates for all the support, assistances and creating a workable environment in the lab.
- A special thanks to my family, Patrick Muleya, Hilda Muleya, Haho Muleya, Asakhakhi Muleya, for their encouragement and prayers.

Chapter 1

General introduction and literature review

1.1 Importance of Citrus in South Africa

South Africa is the world's second-largest citrus exporter, constituting 10 % of global exports, particularly excelling in citrus fruits such as lemons, and limes (Chisoro-Dube & Roberts, 2021). The industry's export earnings have significantly increased over the past two decades, driven by elevated production and prices. Citrus fruits played a crucial role in the export growth from 2010, witnessing an almost fourfold increase from US$202 million to US$730 million by 2020. Notably, 66 % of citrus production is exported as fresh fruit (Chisoro-Dube & Roberts, 2021). The citrus sector plays a vital role in job creation, spanning activities across the entire value chain, including cultivation, packhouses, marketing, logistics, and upstream in nurseries and input supply (DALRRD, 2022). Agriculture as a whole contributes significantly to economic development, supporting household food security. It experienced notable growth, with a 13.4 % year-on-year increase in 2020 and an 8.3 % growth in 2021. Employment figures in the sector remained stable, reaching 868,000 in the fourth quarter of 2021, indicating consistent performance over the past few years (DALRRD, 2022). Over the past decade, employment in citrus production has grown by more than 50 %. Taking all factors into consideration, the citrus value chain is estimated to have accounted for approximately 250,000 jobs in 2020, with ongoing growth projections (Chisoro-Dube & Roberts, 2021; Idamokoro et al., 2022). The citrus industry in South Africa emerges not only as a major global player in exports but also as a significant contributor to the country's employment and economic growth.

1.2 Insect pests and the threat to crop production

Agricultural pests cause substantial economic losses to South African crops annually, making it essential to identify potential future threats for effective biosecurity. The False Codling Moth (FCM), *Thaumatotibia leucotreta* (Meyrick) (Lepidoptera: Tortricidae), is one of the most economically damaging citrus pests in South Africa. *Thaumatotibia leucotreta*, native to sub-Saharan Africa, has demonstrated its effectiveness as an invader (Moore, 2021). This pest, which directly feeds on citrus fruit, inflicts considerable damage. More significantly, it poses a phytosanitary risk in multiple export markets, including the European Union (EU) and the United States of America (USA), potentially leading to reduced overall revenues (Hatting et

al., 2019; Moore et al., 2015; Moore, 2021). *Thaumatotibia leucotreta* targets various cultivated hosts, including citrus fruit, stone fruit, avocados, pomegranates, persimmons, macadamia nuts, and peppers (Malan et al., 2018; Moore, 2021). *Thaumatotibia leucotreta*'s destructive feeding habits and fruit penetration capabilities make it a substantial threat to crop production (Love et al., 2014; Moore, 2021). The larva's tunnelling leads to substantial damage and renders fruit unmarketable, causing significant economic losses and impacting export opportunities due to quarantine regulations (Hatting et al., 2019; Love et al., 2014; Moore, 2021). Managing *T. leucotreta* poses challenges due to its propensity to develop resistance to commonly used insecticides. Confronting the challenges presented by this persistent and destructive pest is imperative to secure the future sustainability of South Africa's agricultural sector.

1.3 Management of *Thaumatotibia leucotreta*

1.3.1 Integrated pest management of *Thaumatotibia leucotreta*

Controlling *T. leucotreta* involves implementing an integrated pest management (IPM) strategy that combines various methods to effectively address the pest. These methods include the use of chemical pesticides, pheromone traps, mating disruption techniques, sterile insect technique, orchard and post-harvest sanitation, as well as the deployment of biological control agents such as entomopathogenic fungi, baculoviruses, and parasitoids (Coombes et al., 2016; Hatting et al., 2019; Knox et al., 2015; Moore, 2021). Chemical control methods are often favoured due to their immediate efficacy against *T. leucotreta* (Nicolopoulou-Stamati et al., 2016). These pesticides demonstrate noticeable effects shortly after application, making them an appealing option for pest management (Furlong et al., 2013; Lacey et al., 2015). Nevertheless, concerns are rising regarding the detrimental effects of chemical pesticides, such as health risks, environmental harm, and the development of insect resistance (Furlong et al., 2013; Lacey et al., 2015; Nicolopoulou-Stamati et al., 2016). As a result, there is an increasing interest in biological pest control as an alternative method, aiming to decrease reliance on chemical pesticides while maintaining effective pest management (Moore & Jukes, 2023). Baculoviruses play a significant role in the biological control options available for managing *T. leucotreta* (Moore et al., 2018). Baculoviruses, notably CrleGV-SA and CrpeNPV, currently stand as significant biological control options against *T. leucotreta* (Marsberg et al., 2017; Moore et al., 2004). The CrleGV-SA isolate has been registered and commercially available for more than 20 years (Moore et al., 2015; van der Merwe et al., 2017), while the CrpeNPV isolate is now registered for commercial use in South Africa beginning in 2023. However, ongoing

exploration for novel baculovirus isolates and species is underway to expand biopesticide options. The identification and characterisation of new isolates offers potential opportunities for innovative biopesticides, which, when integrated into IPM programmes, can improve *T. leucotreta* management. Continuous bioprospecting efforts are crucial for discovering effective baculovirus isolates, holding the promise of developing new, sustainable, and environmentally friendly biopesticides for pest control (Opoku-Debrah et al., 2013, 2016). Baculoviruses play a major role in the IPM programme for controlling *T. leucotreta* and are highly relevant to this project.

1.3.2 Baculoviruses: their classification, morphology, genome structure and biology

1.3.2.1 Taxonomy and nomenclature

Baculoviruses are pathogens that specifically infect arthropods (Rohrmann, 2019). They have been found to infect over 600 insect species, including those belonging to the Lepidoptera, Hymenoptera, and Diptera orders (Herniou et al., 2003). The *Baculoviridae* family (Class *Naldaviricetes*, Order *Lefavirales*) is divided into four genera, determined by phylogenetic analysis, genomic structure, and morphological characteristics (van Oers et al., 2023). Lepidopteran-infecting viruses are classified in the genera *Alphabaculovirus* and *Betabaculovirus,* which includes Nucleopolyhedrovirus (NPVs) and Granulovirus (GVs), respectively. Viruses that infect Hymenoptera and Diptera are classified in the genera *Gammabaculovirus* and *Deltabaculovirus*, respectively (Harrison et al., 2018; van Oers et al., 2023). A phylogenetic analysis using amino acid sequences of core genes divides *Alphabaculovirus* into two branches: Group I and Group II. Similarly, the *Betabaculovirus* genus forms two distinct clusters, labelled as clades a and b (Harrison et al., 2018; Miele et al., 2011).

In 2021, the International Committee on Taxonomy of Viruses (ICTV) approved a proposal to adopt a binomial virus species naming system based on Linnaeus' method (van Oers et al., 2023). According to this system, a species name should consist of two distinct word components separated by a space. The first component, beginning with a capital letter, must correspond to the genus name. The second component should lack any suffixes specific to higher taxa. The complete species name, including both components, is to be italicised (van Oers et al., 2023). For example, Cryptophlebia leucotreta Granulovirus is now classified as *Betabaculovirus cryleucotretae.*

1.3.2.2 Morphology and classification

In the replication cycle of baculoviruses, two primary viral forms are produced: occlusion-derived virus (ODV) and budded virus (BV) (Rohrmann, 2019). The ODV initiates primary infections in the host's midgut, enclosed in crystalline occlusion bodies (OBs) for protection against environmental factors. The BV facilitates the virus's spread throughout the host (Rohrmann, 2019). The ODVs contain either a single nucleocapsid or multiple nucleocapsid within their envelope (Figure 1.1A), while the BVs contain a single nucleocapsid within their envelope that buds from the membrane of the larvae (Figure 1.1B). The OBs are classified into two major divisions: granular in granuloviruses (GV) or polyhedral in nucleopolyhedroviruses (NPV) in shape. The NPV OBs are formed in the nucleus with a polyhedral structure primarily composed of crystalline polyhedra, encapsulating ODVs (Rohrmann, 2019). The polyhedra range in diameter from 0.6 to 2 μm. The NPV OBs consist of multiple ODVs. Each ODV may contain either a single nucleocapsid, known as single-NPV (SNPV) or multiple nucleocapsids, referred to as multiple-NPV (MNPV) (Figure 1.1C). The GVs produce ovicylindrical OBs knowns as granules. These granules contain a single ODV characterised by a single nucleocapsid surrounded by the crystalline protein granulin (Figure 1.1D). The GV OBs are located in the nuclear-cytoplasmic interface via rupture of nuclear membranes (Rohrmann, 2019). The granules have a diameter of approximately 0.2 to 0.4 μm, and each nucleocapsid carries a single viral genome. Notably, OBs exhibit high stability and can endure harsh environmental conditions, enabling ODVs to maintain infectivity for prolonged durations.

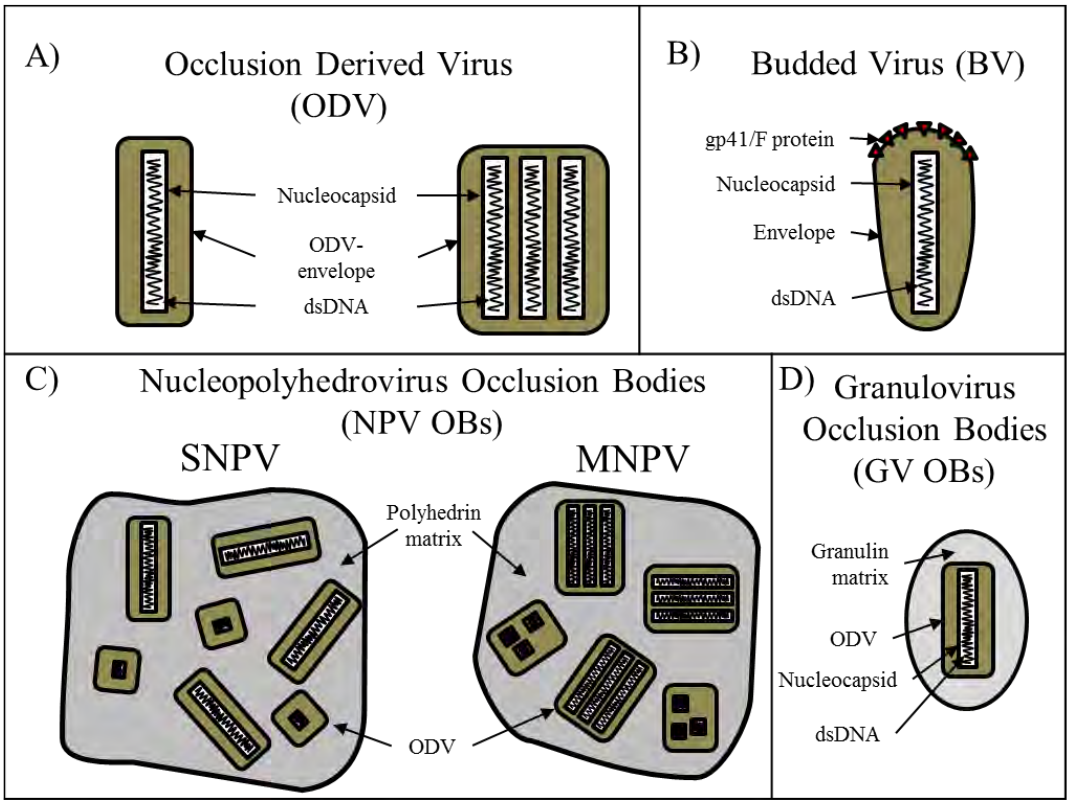


Figure 1.1. The morphology of baculovirus displaying the A) Occlusion derived virus, B) Budded virus, C) Nucleopolyhedrovirus occlusion bodies, D) Granulovirus occlusion bodies (Jukes, 2018).

1.3.2.3 Infectious life cycle of baculoviruses

The baculoviruses comprise two stages in their life cycle: the primary and secondary infection (Rohrmann, 2019). The primary infection initiates when susceptible larvae ingest contaminated foliage from the environment (Figure 1.2A, and B). The high alkalinity in the midgut triggers the breakdown of the crystalline lattice encasing the ODVs, facilitating their passage through the peritrophic membrane (PM) (Figure 1.2C). Proteins embedded in the viral envelope aid in creating pores large enough for the virions to traverse through (Hunter-Fujita et al., 1998; Rohrmann, 2019). The released ODVs infect and fuse with the microvilli of midgut epithelial cells (Figure 1.2D), releasing nucleocapsids into the cytoplasm. These nucleocapsids are subsequently transported to the nucleus, where they uncoat and initiate viral replication, leading to the formation of BVs (Figure 1.2E, and F). The BVs enter cells through endocytosis and replicate within the nucleus (Figure 1.2G). They then spread throughout the insect's haemolymph and tracheal system, disseminating the infection to other cells and tissues, thereby causing a secondary systemic infection. Infected cells initiate the production of nucleocapsids, which acquire an envelope and associated envelope proteins, leading to the formation of ODVs

that are occluded into OBs (Figure 1.2H, and I). In the later stages of infection, the host's tissues accumulate millions of OBs containing virions (Figure 1.2J), which are discharged when the larva's body liquefies (Figure 1.2K), releasing additional OBs that contaminate foliage, and initiate further cycles of horizontal transmission (Williams et al., 2017). Baculovirus are excellent candidates for biological pesticides because of their life cycle. This process not only eliminates the target pest but also produces additional virus particles capable of causing secondary infections in the field.

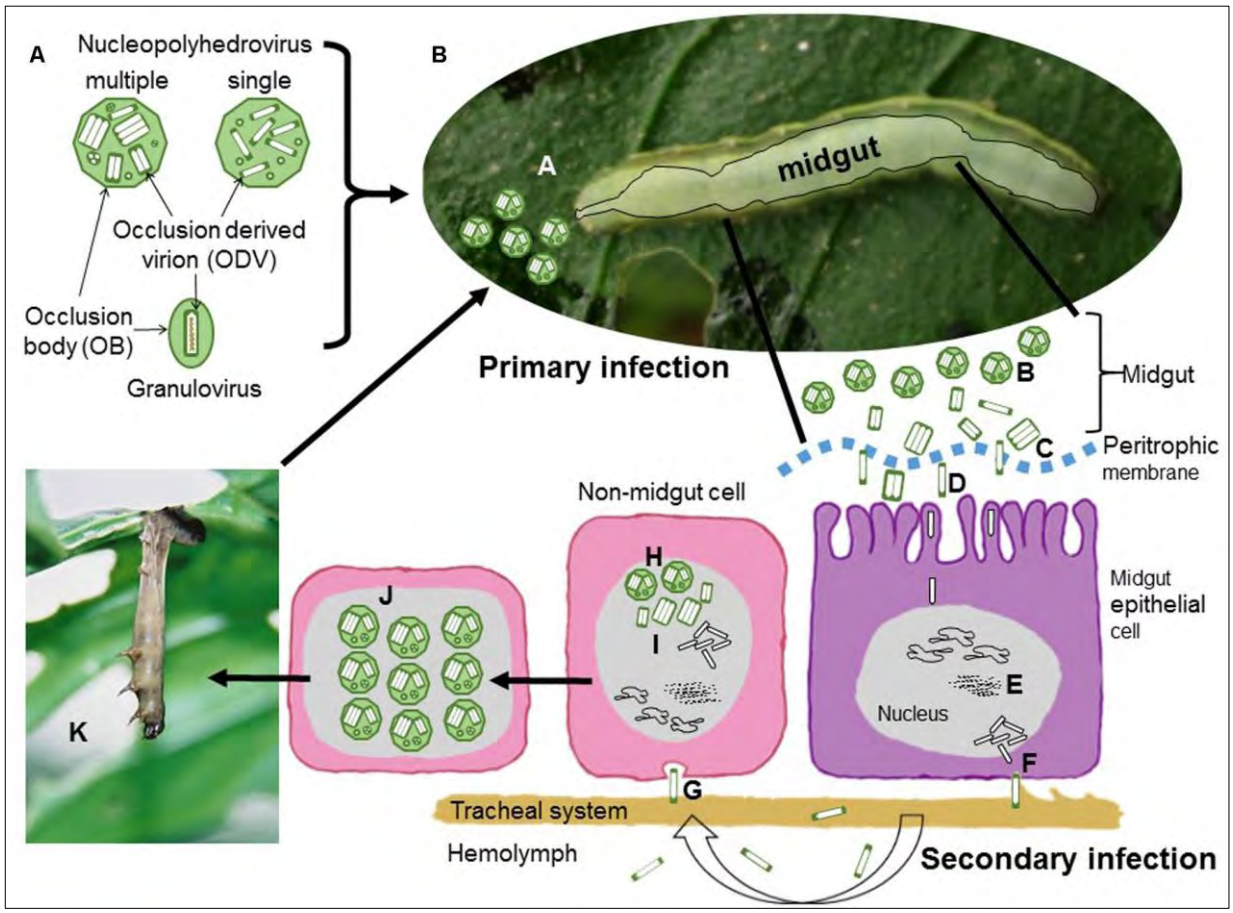


Figure 1.2. The schematic representation illustrates the primary and secondary infectious cycles of the baculovirus, demonstrating the transmission route through an infected host (Adapted from Williams et al., 2017).

1.3.2.4 Genomic composition and Structural/Non-Structural protein encoded by Baculoviruses

Baculoviruses possess circular double-stranded DNA genomes ranging from 80 to 180 kbp, containing 100 to 200 protein-encoding open reading frames (ORFs) that are closely spaced and occur in either orientation (van Oers & Vlak, 2007). There can be significant variation in

ORF content and order between species, and most genomes include regions with short repeats. Despite gene content variations among different baculoviruses, all sequenced genomes have a common set of conserved genes called core genes (Garavaglia et al., 2012; Javed et al., 2017). These genes are classified into functional groups such as replication, transcription, and structural elements (Rohrmann, 2019). Across the family, there are 37 conserved core genes present in the genomes of all members (Table 1.1).

Table 1.1. Core genes conserved among members of family *Baculoviridae* (Rohrmann, 2019; van Oers et al., 2023)

Per os infectivity factor (PIF)	Viral transcription complex
Pif0/pif4	*Late expression factors 4*
Pif1	*Late expression factors 8*
Pif2	*Late expression factors 9*
Pif3	*Late expression factor 5*
Pif5/pdv-es6	*Very late factor 1*
Pif6	*P47*
Pif8/up91	
Nucleocapsid protein	**DNA replication**
38k	*DNApol*
Vp37	*Helicase*
P6.9	*Alkaline nuclease*
Ac66	*Lef 1*
Ac109	*Lef 2*
Ac142	
Ac144	
Gp41	
ODV envelope 18	
ODV envelope 25	
Vp39 capsid	
Vp91	
Vp1053	
Unknown functions	**Sulfhydryl oxidase**
Ac81	*P33*
Ac78	
Ac93	
Ac101	
Ac103	

Several proteins in baculoviruses are associated with OBs (Rohrmann, 2019). Polyhedrin and Granulin, major proteins in OBs, are closely related and serve as structural elements in NPVs and GVs respectively (Rohrmann, 2019). Occlusion bodies are associated with viral enhancing factors, also known as infectivity proteins (Hara et al., 1976; Tanada et al., 1973), such as enhancin, which belongs to a class of metalloproteinases encoded by certain lepidopteran NPVs

and GVs, speculated to aid baculovirus infections by breaking down the peritrophic membrane (Harrison, 2015; Lepore et al., 1996; Rohrmann, 2019). *Per os* infectivity factors (*PIFs*) represent another group of proteins associated with the ODV envelope, crucial for midgut infection. Initially these were identified for their role in infecting insects, PIFs were later found to be non-essential for infecting cultured cells (Boogaard et al., 2018; Wang et al., 2017). The *PIFs* associated with the ODV-envelope aid viral entry into midgut cells (Boogaard et al., 2018). In Autographa californica multiple nucleopolyhedrovirus (AcMNPV), PIFs combine to form a protein complex, indicating collaboration in ODV entry (Peng et al., 2010, 2012). The presence of a PIF complex and conservation in baculovirus genomes indicate collaborative mediation of oral infectivity (Boogaard et al., 2018). All identified PIF proteins are produced by core genes of baculoviruses, showing significant similarity across all sequenced baculovirus genomes within the four genera of the *Baculoviridae* family (Rohrmann, 2019). In viral entry, ODVs initially encounter the PM, a barrier composed of chitin, mucopolysaccharides, and proteins, shielding midgut cells from gut material and restricting microorganism entry (Hegedus et al., 2009; Rohrmann, 2019). Another viral enhancing factor associated with the OBs is the glycoprotein, gp37, which has been reported to aid in the ODV passage across the PM and fusion with epithelial cells (Gross et al., 1993; Liu et al., 2000, 2011; 2019; Rohrmann, 2019). Despite variations, most baculovirus species lack enhancin or gp37, raising uncertainties about the peritrophic membrane's role in impeding ODV passage to midgut cells and whether PIFs contribute to overcoming this obstacle. PIF8 (P95) is currently the only PIF with experimental evidence supporting its involvement in facilitating ODVs passing through the PM (Javed et al., 2017; Zhu et al., 2013). The role of PIFs in this process remains unclear, but understanding it is crucial for comprehending additional steps required for successful midgut epithelium infection beyond ODV binding and fusion (Rohrmann, 2019).

1.4 Baculoviruses and their role as biological control agents

1.4.1 Introduction to baculoviruses-based biocontrol

Baculoviruses are highly effective agents for biological pest control against insect crop pests, possessing advantageous qualities such as a rich research history, proven safety, and environmental friendliness (Harrison & Hoover, 2012; Mudgal et al., 2013; Rohrmann, 2019). Additionally, the practicality of using them as biological pesticides is enhanced by the existence of well-developed mass production systems for many baculoviruses, enabling large-scale manufacturing (Grzywacz & Moore, 2017). These factors have facilitated the development of

a flourishing commercial production of baculovirus insecticides in various regions worldwide, including America, Europe, Asia, Australasia, and Africa (Copping, 2004; Lacey et al., 2015; Moore & Jukes, 2023). As biopesticides, baculovirus insecticides are viewed as promising alternatives to decrease reliance on chemical pesticides, primarily due to safety concerns or insect resistance (Furlong et al., 2013; Glare et al., 2012; Lacey et al., 2015). This study aims to utilise two baculoviruses, Cryptophlebia leucotreta Granulovirus (CrleGV) and Cryptophlebia peltastica Nucleopolyhedrovirus (CrpeNPV), as components of biological management strategies against *T. leucotreta.* The prevalence, effectiveness, and diversity of baculovirus utilisation for controlling lepidopteran pests on citrus in South Africa and worldwide are expected to increase in the future.

1.4.2 *Betabaculovirus cryleucotretae,* Cryptophlebia leucotreta Granulovirus

The CrleGV was initially identified in infected *T. leucotreta* larvae found in the Ivory Coast (Angélini et al., 1965). Additional isolates of CrleGV, including one from Cape Verde (CrleGV-CV3) and several from South Africa (CrleGV-SA), were subsequently isolated and genetically characterised (Opoku-Debrah et al., 2013). The CrleGV-SA was introduced as a biopesticide in South Africa in 1998 and underwent testing in numerous field trials beginning in 2000 (Moore et al., 2004). In 2003, the first CrleGV product, Cryptogran™ (River Bioscience, South Africa), was registered in South Africa for the control of *T. leucotreta*. Since then, another South African isolate, CryptoMax™ (River Bioscience, South Africa), and several distinct isolates, Cryptex®, and Gratham® (both Andermatt Biocontrol, Switzerland) (Kessler & Zingg, 2008), have been registered for field use as commercial products. The extensively tested CrleGV-SA has undergone over 50 field trials and has been commercially applied on tens of thousands of hectares annually for nearly two decades (Moore et al., 2004; Moore et al., 2011, 2015). According to trial results, applications of CrleGV-SA consistently led to substantial reductions in fruit infestation compared to untreated controls. The highest reduction reported was 92 %, while the lowest was 27 % (Moore et al., 2015). Despite the variability in suppression levels, it highlights the significance of an IPM programme that utilises diverse technologies to gradually diminish pest levels over time (Moore, 2021). Additionally, several trials showed that incorporating an adjuvant like molasses significantly improve treatments efficacy, often achieving control levels similar to chemical alternatives and, in some instances, surpassing certain chemicals (Moore et al., 2015). The CrleGV-SA has proven its efficacy and genetic stability for over 15 years as part of the IPM programme for controlling *T. leucotreta*

in South Africa (Moore, 2021; van der Merwe et al., 2017). However, it is imperative to continually enhance this programme and tackle emerging challenges.

1.4.3 *Alphabaculovirus Crypeltasticae*, Cryptophlebia peltastica Nucleopolyhedrovirus

Cryptophlebia peltastica (Meyrick), commonly known as the litchi moth, belongs to the Lepidoptera family Tortricidae and is a major pest of litchi in several countries across the Sub-Saharan and Indian Ocean regions, including Mauritius, South Africa, and Réunion Island (Manrakhan et al., 2008). The CrpeNPV is a recently discovered *alphabaculovirus* that has undergone genetic and biological characterisation, demonstrating its distinct nature and its ability to infect hosts beyond its homologous host (Marsberg, 2016; Marsberg et al., 2017, 2018). This virus is known to combat *Cydia pomonella* (Linnaeus) (Lepidoptera: Tortricidae) infestations in pome fruit, as well as *T. leucotreta* and *Thaumatotibia batrachopa* (Meyrick) (Lepidoptera: Tortricidae) infestations in macadamia orchards, and *T. leucotreta* and *C. peltastica* infestations in litchi crops (Marsberg, 2016; Marsberg et al., 2017). Examination of the virus using transmission electron microscopy showed the presence of polyhedral OBs containing numerous individually enveloped nucleocapsids (Marsberg et al., 2017). Currently, CrpeNPV is undergoing approval processes as a biopesticide for the use in Europe to control various pests, including *C. pomonella* and *Grapholita molesta* (Busck) (Lepidoptera: Tortricidae) on pome fruit and stone fruit (Moore & Jukes, 2023). It has also recently received approval for use and is commercially available in South Africa under the name MultiMax™ (River Bioscience, South Africa). Preliminary field trials with CrpeNPV against *T. leucotreta* in citrus have been conducted, using rates ranging from 5×10^{11} to 5×10^{13} OBs/ha (Hatting et al., 2019). In one case, a single application reduced infestation by over 90 %, showing great promise (Hatting et al., 2019; Marsberg et al., 2017). The CrpeNPV has proven to be an effective biopesticide not only against *T. leucotreta* but also against other citrus pests.

1.4.4 Challenges associated with baculovirus-based biopesticides

The challenges of developing and applying baculoviruses include concerns about insect resistance, virulence, and a slow speed of kill (Moore & Jukes, 2019). A major challenge is the potential development of resistance in target insects to commercial products. Notably, resistance has been identified in European populations of *C. pomonella* against a commercially formulated baculovirus (CpGV-M) isolated in Mexico (Asser-Kaiser et al., 2007; Jehle et al., 2017). Studies on resistance in *C. pomonella* have identified the locus involved and genetic differences in isolates to overcome this resistance (Eberle & Jehle, 2006; Gebhardt et al., 2014;

Jehle et al., 2017). Laboratory experiments with *Adoxophyes honmai* (Yasuda) (Lepidoptera: Tortricidae) also revealed the potential for resistance development, showing a significant decrease in virulence after generations of exposure to Adoxophyes honmai NPV (Nakai et al., 2017). Resistance occurred in both primary and secondary infection routes and proved stable even after ceasing virus exposure. Field populations exhibited faster resistance development than laboratory colonies, emphasising the importance of novel control options (Briese & Mende, 1983). Although no resistance has been recorded for CrleGV-SA in field applications, continuous development of control measures is essential to mitigate potential resistance and ensure the long-term viability of the citrus industry in South Africa.

Baculovirus-based biopesticides face challenges related to their effectiveness in terms of virulence and speed of kill against specific hosts. Baculoviruses exhibit a relatively slow speed of kill, taking days or weeks, compared to most chemical insecticides, which act within hours (typically less than a day) (Beas-Catena et al., 2014; Venette et al., 2003). The delayed action of the virus remains problematic for crops with thinner skins, such as avocados and grapes (Venette et al., 2003). An additional consequence of short-term crop loss is that it can lead to the perception that baculoviruses are ineffective. This perception can discourage farmers from using baculoviruses (Moore & Jukes, 2019). Opoku-Debrah et al. (2016) studied the virulence of seven CrleGV isolates on diverse populations of *T. leucotreta.* The results showed that certain isolates demonstrated enhanced virulence against particular populations, suggesting that the discovery of new isolates could enhance biological control and mitigate resistance development to CrleGV isolates.

1.5 Overcoming challenges associated with baculovirus-based biopesticides

1.5.1 Evaluating mixtures of baculoviruses in South Africa

Mixed infections of two different baculoviruses in nature, infecting the same host, have been studied to comprehend virus interactions, their impact on susceptible insects, and their insecticidal consequences (Ferrelli & Salvador, 2023). In South Africa, CrpeNPV has been evaluated for potential use alongside CrleGV-SA. Initial studies indicate that the mixture exhibits greater virulence against *T. leucotreta* compared to either virus alone (Jukes, 2018). However, the impact of simultaneous infection by both viruses on the speed of kill was unknown. In a study by Taylor (2021), the combination of CrpeNPV and CrleGV-SA demonstrated an enhancement in the lethal concentration against *T. leucotreta* neonates. However, it led to a decrease in the speed of kill compared to using either virus alone. The

studies suggest that the slowed speed of kill may be attributed to the viruses competing for resources within the host insect (Taylor, 2021; Wennmann et al., 2015). Although baculovirus mixtures have shown a synergistic effect in enhancing insecticidal activity, the exact reason for this enhancement remains unclear. One potential explanation is that one virus releases viral enhancing factors that the other lacks (Biedma et al., 2015). For example, CrpeNPV encodes gp37, which is absent in CrleGV-SA. Therefore, there is potential to exploit such proteins as a means of enhancing baculovirus infectivity.

1.5.2 Infectivity proteins targeting the peritrophic membrane

Baculoviruses envelopes are known to be associated with proteins that enhance the infectivity of other baculoviruses (Rohrmann, 2019). Several proteins have been identified in association with OBs in baculoviruses. An interesting observation reported in the literature is that specific baculovirus proteins, such as enhancin and gp37, can enhance the infection of another baculovirus (Liu et al., 2011; 2019; Ricarte-Bermejo et al., 2021; Rohrmann, 2019; Yang et al., 2017). *Enhancin* genes are more commonly observed in GVs than in NPVs, despite being found in both types of viruses (Ferrelli & Salvador, 2023). Most GVs that exhibit increased NPV infections are notably found within the phylogenetic clade "a." (Miele et al., 2011). Studies involving intact OBs or protein extracts containing enhancin, along with the expression and purification of enhancin in heterologous systems, has revealed two main functions: enhancing the permeability of the PM and promoting the fusion of virions with midgut cells (Hukuhara & Zhu, 1989; Lepore et al., 1996; Tanada, 1985; Wang & Granados, 1997). The GV enhancins are typically found within OBs, while NPV enhancin are commonly located in ODV envelopes (Slavicek, 2012; Slavicek & Popham, 2005). Other specific proteins, such as Cydia pomonella GV gp37, have been identified to possess enhancin-like activities, altering the PM structure, and enhancing infectivity (Liu et al., 2011; 2019). The glycoprotein, gp37, has been associated with certain baculovirus OBs or BVs (Gross et al., 1993; Liu et al., 2000). In instances where GVs lack *enhancin* or *gp37* genes but display synergistic effects, other factors are proposed to act as enhancers. For example, Epinotia aporema GV and Spodoptera ornithogalli GV, devoid of both *gp37* and *enhancin* genes, exhibit enhancement effects, indicating the presence of different enhancing factors (Barrera et al., 2021; Biedma et al., 2015; Masson et al., 2019). Epinotia aporema GV, belonging to clade b GV, enhances the virulence of Anticarsia gemmatalis MNPV in *A. gemmatalis* (Masson et al., 2019). Similarly, Spodoptera ornithogalli GV has been shown to synergise with Spodoptera ornithogalli NPV action on *S. ornithogalli* (Guenée) (Lepidoptera: Noctuidae) (Biedma et al., 2015). Other proteins have

been identified that disrupt the PM, facilitating virion entry. The baculoviral ODV-E66, a conserved lepidopteran baculovirus protein, plays a crucial in primary infection by breaking down chondroitin sulphate in the PM. Recent studies have highlighted its role in facilitating the passage of ODVs across the PM during oral infection by Helicoverpa armigera NPV (Hou et al., 2019; Sugiura et al., 2011; Wang & Hu, 2019). Another protein, chitinase, has been implicated in disrupting the chitinous PM in AcMNPV OBs, thus facilitating ODV passage (Hawtin et al., 1995; Ishimwe et al., 2015). The continual discovery of genes associated with virulence and speed of kill will help improve appropriate baculovirus isolates for use in biological control.

1.5.3 The use of gp37 an additive for the potential improvement of infectivity

The glycoprotein, gp37, has been reported to enhance the infection of another baculovirus (Liu et al., 2011; 2019; Ricarte-Bermejo et al., 2021; Yang et al., 2017). The gp37 shares structural similarities and amino acid sequence identity (approximately 30-40 %) with the protein fusolin found in Entomopoxviruses. Both gp37 and fusolin feature five highly conserved domains, including a chitin-binding domain (Cheng et al., 2001). Fusolin has been demonstrated to enhance the infectivity of NPVs and Entomopoxviruses, suggesting a potential similar function for gp37 (Cheng et al., 2001). Homologues of gp37 have been discovered to contain a chitin-binding domain in Spodoptera litura NPV, enabling them to bind to chitin (Li et al., 2003). In a study by Liu et al. (2011), a truncated *gp37* gene from CpGV was cloned into a suitable vector and expressed in *Escherichia coli* cells. The gene was truncated to remove signal sequences and transmembrane helices at the N-terminus to ensure maximal expression and solubility of gp37 in *E. coli*. When combined with either SeMNPV or AcMNPV OBs and evaluated against third-instar *S. exigua* larvae by droplet-feeding bioassays, a 13.98- and 20.20-fold improvement in LC_{50} was calculated, respectively. In a similar study by Yang et al. (2017), recombinant AcMNPV viruses incorporating truncated gp37 into OBs exhibited 3- to 5-fold lower median lethal concentrations (LC_{50}) compared to the control virus, suggesting enhanced baculovirus infectivity (Yang et al., 2017). Existing South African baculoviruses provide a rich bioresource library, with two isolates, CpGV-SA (Motsoeneng et al., 2019) and CrpeNPV (Marsberg et al., 2017), each containing the *gp37* gene.

1.6 Motivation

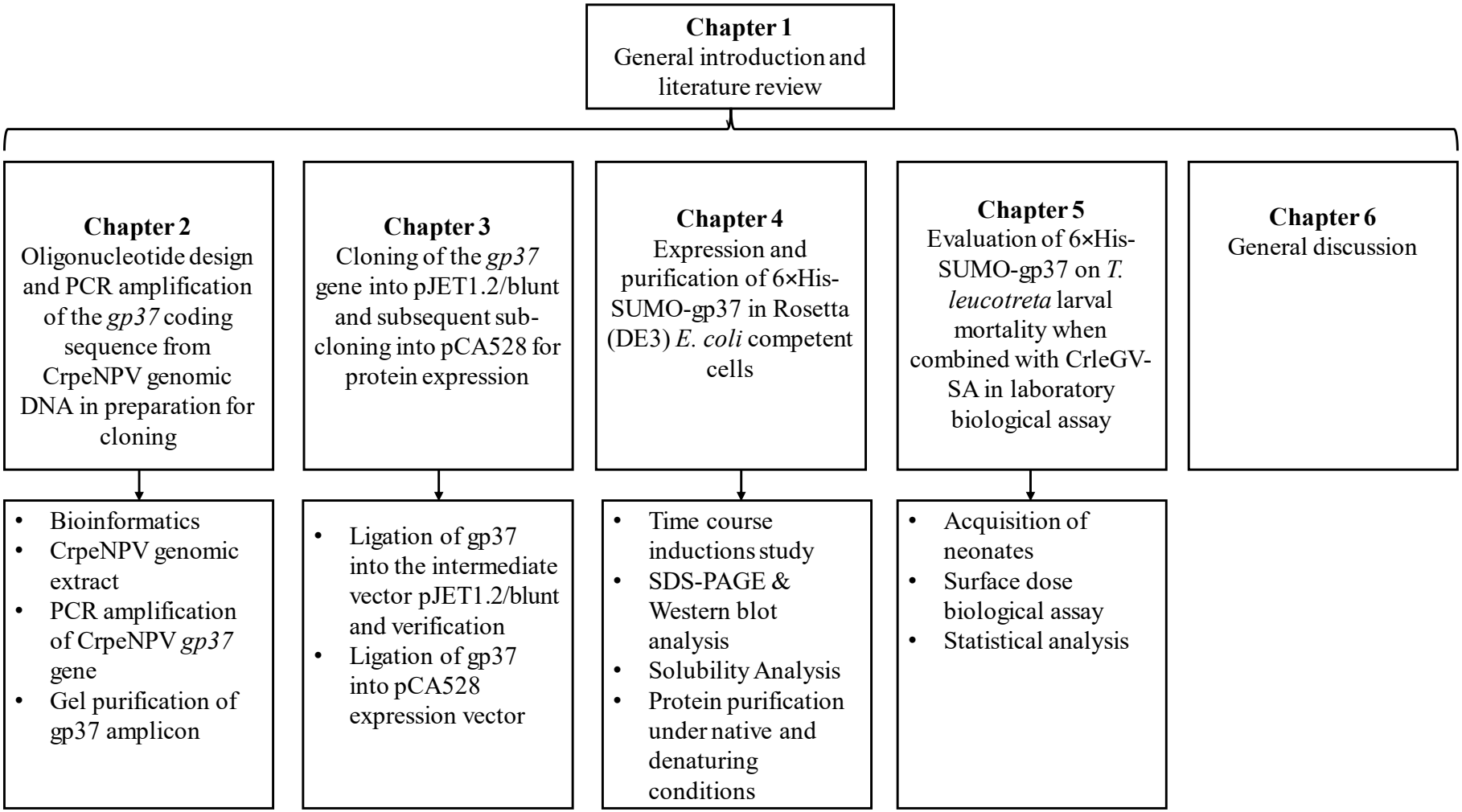
Thaumatotibia leucotreta is a significant pest that inflicts damage on fruit crops in South Africa. Two baculoviruses, CrleGV-SA and CrpeNPV, infect *T. leucotreta* larvae, resulting in

larval mortality. Both viruses are utilised in integrated pest management programmes to reduce fruit damage in agricultural fields, with CrleGV-SA being used against *T. leucotreta* for nearly twenty years in South Africa. However, these control options are limited by factors such as virulence and the slow speed of kill, which can be addressed by exploiting potential synergistic relationships between baculoviruses infecting the same host. Mixed infection of CrleGV-SA and CrpeNPV against *T. leucotreta* larvae were found to improve the lethal concentration against *T. leucotreta* neonates but resulted in a delay in the speed of kill compared to the use of either virus separately (Taylor, 2021). Speculation suggests that this delay could be due to competition between the two viruses for resources in the insect. Baculovirus proteins, such as gp37, play a crucial role in initiating primary infections by aiding the penetration of virions through the PM. The CrpeNPV genome encodes *gp37*, unlike CrleGV-SA. Previous studies indicate that bacterially expressed and purified CpGV gp37 notably enhances the infectivity of NPVs and the lethality of *Bacillus thuringiensis* in *S. exigua* larvae (Liu et al., 2011). However, this has never been tested on CrleGV-SA, so it is unknown whether CrleGV-SA and CrpeNPV gp37 will show a synergistic effect in terms of dose mortality in *T. leucotreta*. It is anticipated that the inclusion of this protein in baculovirus formulation may lead to increased infection of CrleGV-SA, thereby improving the control of *T. leucotreta* in the field.

1.7 Aim and objectives

The overall aim of this study was to express the CrpeNPV gp37 in a bacterial system. The specific objectives were: i) to use Benchling to design and test oligonucleotides for PCR amplification of the CrpeNPV *gp37* gene in preparation for cloning, ii) to clone *gp37* into the intermediate vector, pJET1.2/blunt, iii) to sub-clone *gp37* into the bacterial expression vector, pCA528, iv) to express and purify recombinant protein using Nickel column affinity chromatography, and analyse it using SDS-PAGE and Western blot analysis, and v) to determine the effect of the recombinant protein on the biological activity of CrleGV-SA against *T. leucotreta* neonate in laboratory biological assays.

Overview of chapters



Chapter 2

Oligonucleotide design and PCR amplification of the *gp37* coding sequence from CrpeNPV genomic DNA in preparation for cloning

2.1 Introduction

Baculovirus genomes encode various structural and non-structural proteins necessary for their lifecycle within the host (Rohrmann, 2019). Certain baculovirus proteins, such as enhancin and gp37, have been found to enhance the infectivity of other baculoviruses (Granados et al., 2001; Lei et al., 2020; Liu et al., 2000, 2011; Peng et al., 1999; Slavicek, 2012). The protein gp37, has been reported to facilitate the passage of ODVs across the peritrophic membrane (PM) (Liu et al., 2019). Interestingly, bacterially expressed gp37 encoded by CpGV has been shown to significantly enhance the infectivity of NPVs (Liu et al., 2011, 2019), as previously described in Chapter 1, section 1.5.2. This study aims to clone the CrpeNPV *gp37* gene and express the recombinant protein in a bacterial system. The CrpeNPV isolate was genetically and biologically characterised by Marsberg et al. (2017), and the first complete genome sequence for the CrleGV-SA isolate was described by van der Merwe et al. (2017). The CrpeNPV encodes *gp37,* consisting of 252 amino acids and 759 base pair (bp). Notably, the *gp37* gene was found to be in the reverse orientation within the CrpeNPV genome. The CrpeNPV genome spans 115,728 bp (Marsberg et al., 2017), and the complete genome sequence is accessible on GenBank (Accession: NC_055500.1). The *gp37* coding sequence was obtained from the complete genome of CrpeNPV (Accession no: YP_010086923).

This chapter focuses on the design of oligonucleotides specifically targeting the *gp37* gene in the genome of CrpeNPV for PCR amplification. Numerous software tools are employed for *in silico* oligonucleotide design, including FastPCR software (Kalendar, 2021), and Primer-BLAST (Ye et al., 2012). Oligonucleotides targeting *gp37* coding sequence were designed using the Benchling software, a cloud-based platform designed for life science research and development (Benchling, 2023). The software provides numerous benefits, including enhanced efficiency, accuracy, seamless integration, flexibility, optimisation and, notably, a user-friendly interface. Importantly, it ensures that the designed oligonucleotides contain the necessary characteristics such as length (18 - 30 bp), melting temperature (50 - 60 °C), Guanine-Cytosine contents (40 - 60 %) and sequence specificity.

It's crucial to use highly sensitive and reliable extraction methods for genomic DNA when conducting PCR to amplify the target gene. Several extraction methods have been used for their efficiency in isolating viral gDNA, with the CTAB method commonly used (Opoku-Debrah et al., 2013). Several studies have reported the efficiency of this method for extracting gDNA from baculoviruses (Jukes, 2018; Mela, 2022; Taylor, 2021; van der Merwe et al., 2017). The CTAB method effectively isolates gDNA from baculoviruses by dissolving OBs in an alkaline solution, although it does have potential drawbacks such as a prolonged two-day procedure. As an alternative extraction method for comparison, the Quick-DNA Miniprep Plus Kit (Zymo Research, CA, USA), known for its speed and effectiveness in gDNA extraction, was employed. A protocol described by Mela (2022) was effectively used to modify this kit. This protocol involved the use of sodium carbonate to dissolve the polyhedron matrix and Tris-base for pH neutralisation before using the DNA extraction kit. It has demonstrated its speed and efficiency in extracting the baculovirus DNA genome (Mela, 2022).

Once the gDNA has been extracted, and oligonucleotides targeting the gene of interest have been designed, DNA amplification of the CrpeNPV *gp37* gene can be performed using PCR. This method allows the amplification of the desired DNA segment (Mullis, 1990). This application includes special features, such as *Taq* polymerase, which operates at the high temperatures required to denature the DNA and make the nucleotides accessible for the oligonucleotides to anneal to them (Saiki et al., 1988). The annealing temperature ensures that only oligonucleotides with the exact sequence will anneal and allow replication at the desired site in the gDNA sequence. After PCR amplification, the resulting PCR product can undergo gel purification using a gel purification kit as a step before cloning. Gel purification is crucial because it enables the purification of a single type of DNA, particularly in cases of suspected DNA contamination, such as primer dimer and nonspecific amplification.

The chosen expression vector for this study was pCA528 (WISP08-174, DNASU plasmid repository), which is 5597 bp in size. This plasmid contains key components, including the T7 promoter, an N-terminal 6×His-SUMO tag, a kanamycin resistance gene, and sequencing primers for both forward and reverse. The presence of the polyhistidine tag facilitates the affinity purification of the fusion protein. It has been observed that the SUMO tag positively impacts the protein's solubility, although the exact mechanism behind this solubility enhancement remains unclear (Kuo et al., 2014; Marblestone et al., 2006). The presence of an antibiotic resistance gene within a plasmid allows for the identification and selection of

transformed cells in a culture that includes a particular antibiotic, such as kanamycin, distinguishing them from cells lacking this plasmid (Bennett, 2008).

The aim of this chapter was to design oligonucleotides for the PCR amplification of the *gp37* coding sequence from the genome of CrpeNPV in preparation for cloning. The specific objectives were: i) designing forward and reverse oligonucleotides that target the *gp37* coding sequence, ii) extracting the gDNA of CrpeNPV, and iii) performing PCR amplification of *gp37* and obtaining the PCR product via gel purification in preparation for cloning into the target vector.

2.2 Material and methods

2.2.1 pCA528 vector and CrpeNPV genome

The DNA sequence of pCA528 was obtained from DNASU plasmid repository (Seiler et al., 2014). The pCA528 DNA, kindly provided by Matthias Mayer (Heidelberg University, Germany), was used as a target vector for expressing the recombinant protein. The pCA528 vector features a start codon at its N-terminus, and the *gp37* gene possesses a stop codon at its C-terminus. The complete genome sequence of CrpeNPV was downloaded from GenBank (Accession no: NC_055500). The *gp37* coding sequence was obtained from the complete genome of CrpeNPV (Accession no: YP_010086923).

2.2.2 Oligonucleotide design for PCR amplification of *gp37*

Oligonucleotides targeting the CrpeNPV *gp37* gene were designed using Benchling software (Benchling, 2023). This process involved several steps, including accessing the DNA sequence, importing these into Benchling, identifying suitable oligonucleotide sequences, incorporating restriction sequences, verifying pair selection, checking oligonucleotide specificity, via *in silico* PCR, optimising the oligonucleotides, and finally, sending the designed oligonucleotides for synthesis. The designed oligonucleotides were synthesised by Inqaba Biotechnical Industries (Pty) Ltd (South Africa).

2.2.3 Orientation of the gene in the genome

The schematic representation below shows the full genome of CrpeNPV in a linear conformation and the position of the *gp37* gene within the genome. In the genome of CrpeNPV, the *gp37* gene is situated in the reverse orientation relative to the *polyhendrin* gene, which is typically used as a reference frame. As a result, the forward oligonucleotide was designed at the gene's endpoint, and the reverse oligonucleotide at the starting point (Figure 2.1A). The

PCR product was then reversed in orientation to ligate the gene into the pCA528 expression vector (Figure 2.1B).

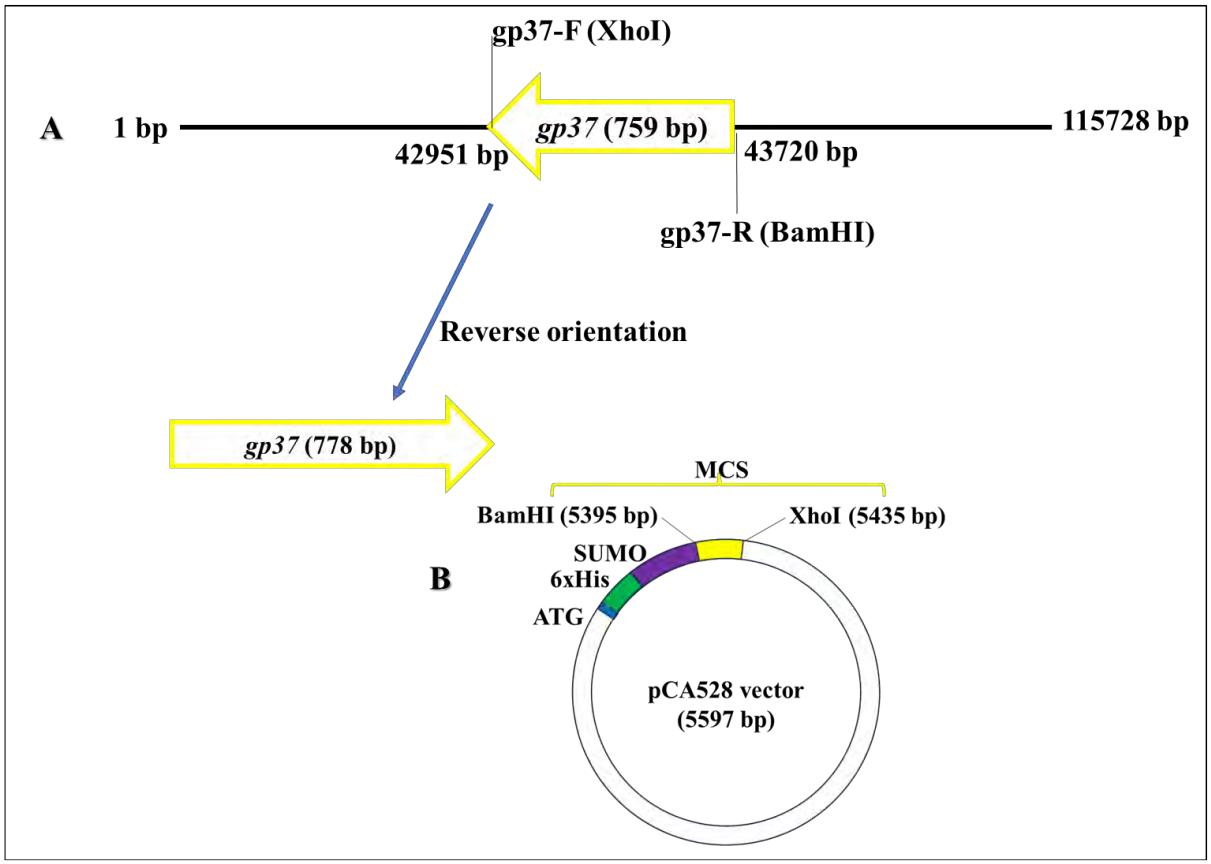


Figure 2.1. Schematic representation depicting the position of the gene in the vector. **A**: Linear representation of the CrpeNPV genome and the orientation of the *gp37* gene, **B**: PCR product (*gp37*) and pCA528 expression vector.

2.2.4 Genomic DNA extraction

The CrpeNPV OBs were provided by Rhodes University, South Africa. A protocol described by Mela (2022) was effectively employed to pre-treat the CrpeNPV OBs. A 100 µL sample of CrpeNPV OBs was mixed with 45 µL of 1 M Na_2CO_3 and incubated at 37 °C for 30 minutes to solubilise the polyhedron matrix. The sample was then treated with 60 µL 1 M pH 6.8 Tris-HCL to neutralise the pH for DNA extraction (Mela, 2022). The gDNA of CrpeNPV was extracted using the Quick-DNA Miniprep Plus Kit (Zymo Research, CA, USA) according to the manufacturer's instruction and eluted with 35 µL to concentrate the DNA. The concentration was measured using a Nanodrop® 2000 spectrophotometer (Thermo Fisher Scientific, USA) in triplicate. The gDNA was analysed by 1 % agarose gel electrophoresis (AGE).

2.2.5 Agarose gel electrophoresis

Agarose gels (1 % [w/v]) were prepared using LE agarose (Benchmark Scientific, USA), 1×TAE buffer (40 mM Tris base; 20 mM glacial acetic acid; 1 mM ethylenediaminetetraacetic acid (EDTA) pH 8) was prepared and heated in a microwave for 1-2 minutes until the agarose was completely dissolved. A 5 µL volume of the gDNA was prepared with 6× loading dye and analysed using 1 % AGE stained with ethidium bromide (0.4 µg/mL), run at 90 V for 30 minutes. The GeneRuler 1 Kb plus DNA ladder (Thermo Fisher Scientific, USA) was used to estimate the size. Agarose gels were visualised using the ChemiDoc™ XRS+ (Bio-Rad, USA) with Image Lab™ (Bio-Rad, USA) software.

2.2.6 PCR amplification and gel purification

A reaction mixture of 25 µL was prepared based on the instructions provided in Table 2.1. Genomic DNA extracted from CrpeNPV OBs was used as template in the PCR. A no-template control (NTC) was performed, where no template DNA was used as a control reaction (Table 2.1). A SimpliAmp™ Thermal cycler (Thermo Fisher Scientific, USA) was used to carry out the PCR amplification. The PCR cycling parameters involved an initial denaturation step at 95 °C for 3 minutes, followed by 30 cycles of denaturation at 95 °C for 1 minute, annealing at 57 °C for 30 seconds, extension at 72 °C for 1 minute, and a final elongation step at 72 °C for 3 minutes. The PCR product was analysed by 1 % AGE as described in section 2.2.5. The *gp37* amplicon was purified from the agarose gel using the GeneJET™ PCR Purification Kit (Thermo Fisher Scientific, USA), following the manufacturer's instructions, and analysed by 1 % AGE as described in section 2.2.5.

Table 2.1. PCR set up.

Reagent	Sample	NTC
2× Taq master mix	12.5 µL	12.5 µL
gp37-F (10 µM)	1 µL	1 µL
gp37-R (10 µM)	1 µL	1 µL
Template (3.23 ng/µL)	3 µL	-
Double distilled water	7.5 µL	10.5 µL
Total	25 µL	25 µL

2.3 Results

2.3.1 The pCA528 expression vector

Figure 2.2 illustrates the structure of the target bacterial expression vector, pCA528, as generated using SnapGene® software. This vector comprises several key components: *KanR*, which generates aminoglycoside phosphotransferase and confers resistance to kanamycin in bacteria; Ori, ensuring high copy number through the ColE1/pMB1/pBR322/pUC origin of replication; Rop, maintaining plasmids at a low copy number; and LacI, which binds to the *lac* operator to inhibit transcription in *E. coli*. This inhibition can be relieved by adding lactose or isopropyl-β-D-thiogalactopyranoside (IPTG) (QIAexpressionist, 2002). Additionally, the vector includes a T7 promoter, providing a promoter for bacteriophage T7 RNA polymerase; RBS, an efficient ribosome binding site from bacteriophage T7 gene (Olins & Rangwala, 1989); 6×His affinity tag: aiding in detecting the protein of interest (QIAexpressionist, 2002), SUMO, a cleavable ubiquitin-like protein tag known to enhance the solubility of the protein (Guerrero et al., 2015; Kuo et al., 2014; Marblestone et al., 2006); T7 terminator, allowing transcription termination for bacteriophage T7 RNA polymerase (QIAexpressionist, 2002); and an MCS containing unique restriction enzyme recognition sites for gene cloning. Analysis of the MCS of pCA528 indicates a single BamHI digestion site at position 5395 and an XhoI digestion site at position 5435 (Figure 2.2).

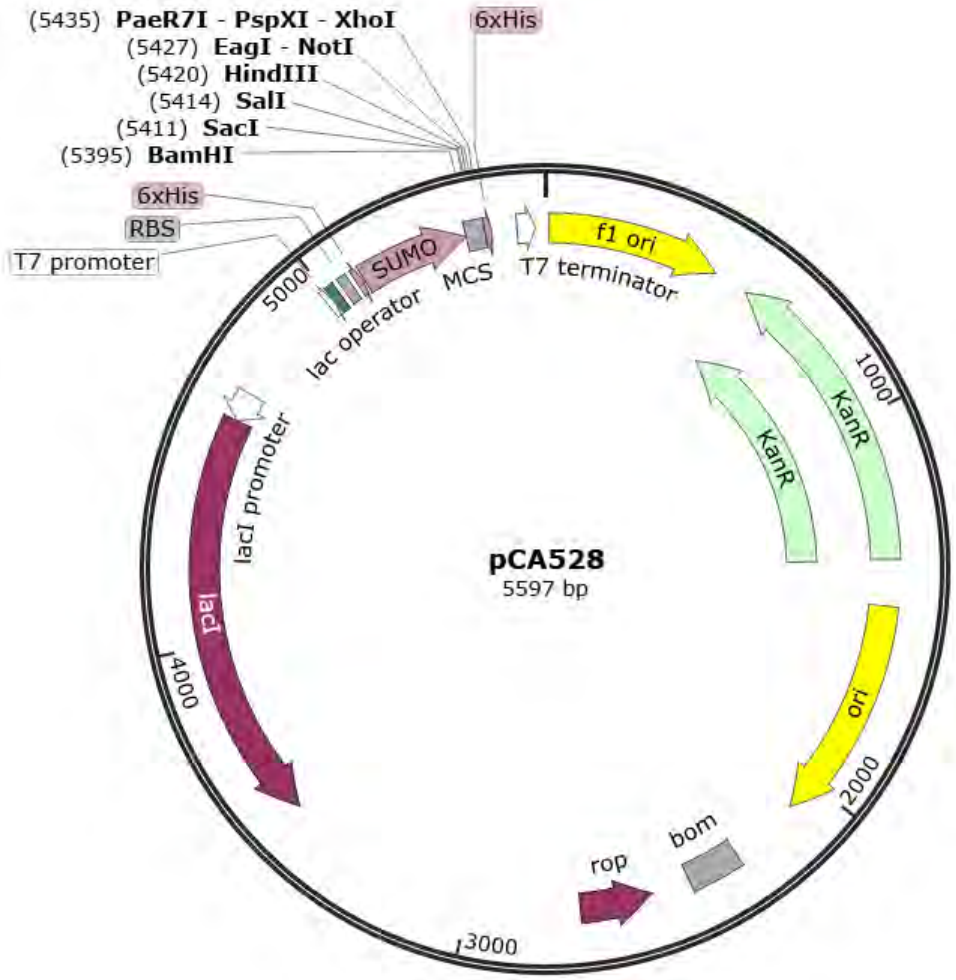


Figure 2.2. The annotated pCA528 expression vector. Abbreviations used: *KanR*- Kanamycin resistance gene, SUMO- Small ubiquitin-related modifier tag, 6×His- Polyhistidine tag, RBS: Ribosome binding site, Ori- Origin of replication, bom- Basis of mobility, rop- Rop protein, lacI- *Lac* repressor, and MCS- Multiple cloning site.

2.3.2 Oligonucleotides design

Oligonucleotides were designed to target the CrpeNPV *gp37* gene. A XhoI restriction site was incorporated into the forward oligonucleotide and a BamHI restriction site was incorporated in the reserve oligonucleotide (Table 2.2). The forward oligonucleotide was 34 bp in length and the reverse oligonucleotide was 35 bp in length. These oligonucleotides are presented in the 5'- 3' directions. Additionally, A-tailing nucleotides (AAA) were added to the oligonucleotides. To ensure in-frame cloning of the *gp37* coding sequence into the pCA528 expression vector with the 6×His-SUMO tag, a thymine nitrogenous base pair, marked in red, was included within the reverse oligonucleotide (Table 2.2). The melting temperature of the nucleotide sequences for the forward and reverse oligonucleotides was 62 °C and 55 °C, respectively

(Table 2.2). The GC content of the forward and reverse oligonucleotides was 41 % and 22 %, respectively (Table 2.2).

Table 2.2. Oligonucleotides targeting *gp37* gene in CrpeNPV genome. Sequences for the restriction sites are underlined, the *gp37* target sequence is shown in bold, and the additional thymine nitrogenous base pair is marked in red.

Names	Sequences 5'-3'	Restriction enzyme	Tm (°C)	GC %
gp37-F	AAACTCGAG**TTAAAACTCATAGTGTGTACGGTCG**	XhoI	62	41
gp37-R	AAAGGATCC**TATGATTTTAATTAATTTTGCTATAG**	BamHI	55	22

2.3.3 Genomic DNA extraction

Genomic DNA was extracted from CrpeNPV OBs and analysed by 1 % AGE (Figure 2.3). Figure 2.3 below shows a band in lane 1 running above 20,000 bp. The expected size of CrpeNPV genome was 115,728 bp. Due to limitations in gel resolution, the exact size of the DNA could not be accurately determined, primarily because of its large size.

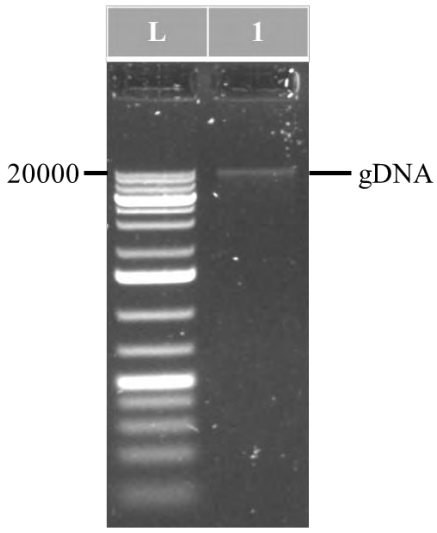


Figure 2.3. Agarose gel image of genomic DNA extracted from CrpeNPV OBs. Lane L: GeneRuler 1 Kb plus DNA ladder and lane 1: CrpeNPV gDNA.

2.3.4 PCR amplification and gel purification of *gp37*

The *gp37* gene was amplified by PCR (Figure 2.4). The PCR product was detected running between 1000 and 700 bp at approximately 800 bp. The expected size of the PCR fragment was 778 bp. The agarose gel indicates the presence of a band at a higher molecular weight in lane 2, suggesting non-specific binding. The PCR product was recovered by gel purification as shown in lane 3. As expected, no band was detected in the NTC in lane 1.

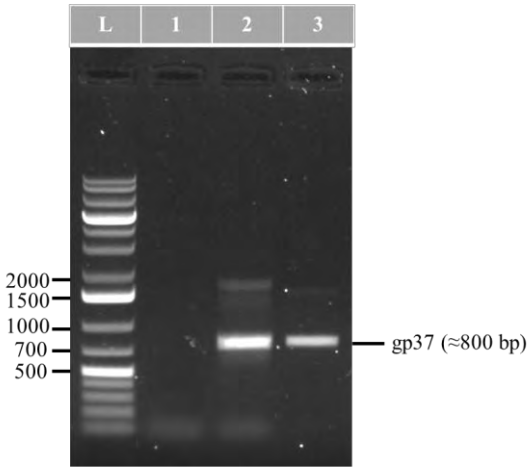


Figure 2.4. Agarose gel image showing the amplification and gel purification of the CrpeNPV *gp37* gene. Lane L:1 Kb plus GeneRuler DNA ladder, lane 1: NTC, lane 2: PCR amplicon (*gp37*), and lane 3: Purified *gp37* product

2.4 Discussion

The aim of this chapter was to design and test oligonucleotides targeting the CrpeNPV *gp37* gene in preparation for cloning. The first objective was to design oligonucleotides that target the *gp37* coding sequence using the Benchling software (Benchling, 2023). To achieve this, the complete genome sequence of CrpeNPV was analysed to locate the position of the *gp37* coding sequence *in silico*. The position of *gp37* in the genome of CrpeNPV was identified, and oligonucleotides were designed to meet specific criteria, including length, melting temperatures and GC content. The length of an oligonucleotide directly correlates with its specificity, stability, and efficiency (Domingues, 2017). Each oligonucleotide contained either a BamHI or XhoI restriction site, targeting the *gp37* gene within the CrpeNPV genome. These restriction sites were incorporated into the oligonucleotides to facilitate downstream ligation into the pCA528 expression vector.

The second objective was to extract gDNA from purified CrpeNPV OBs. The expected size of the extracted DNA was 115,728 bp (Marsberg et al., 2018). However, the extracted DNA

produced a band greater than 20,000 bp, exceeding the gel's resolution capacity. The gDNA was extracted in preparation for PCR amplification. The concentration of the CrpeNPV gDNA was 3.23 ng/µL. The modified Quick-DNA miniprep plus kit protocol described by Mela (2022) proved to be an effective tool for extracting gDNA from baculoviruses.

The third objective was to PCR amplify and isolate the *gp37* gene from the CrpeNPV gDNA. PCR was conducted using the designed oligonucleotides at an annealing temperature of 57 °C. Both the gp37-F and the gp37-R oligonucleotides were designed to produce an amplicon with a size of 778 bp. PCR amplification was performed, and the resulting PCR amplicon was generated at approximately 800 bp. However, nonspecific bands were also observed, which can occur when oligonucleotides bind to unintended regions of the template DNA, resulting in the amplification of nonspecific products. This issue can be mitigated by optimising the annealing temperature, as oligonucleotides may fail to bind to the template if the annealing temperature is too low (Domingues, 2017). Alternatively, the target amplicon can be isolated and extracted from an agarose gel and purified. In this study, the latter option was chosen, which involved gel purifying the target amplicon to isolate it from nonspecific bands. The *gp37* purified product was recovered from the agarose gel in preparation for cloning.

In conclusion, oligonucleotides were designed to target the CrpeNPV *gp37* gene in preparation for PCR amplification. Genomic DNA was extracted from CrpeNPV OBs and used as a template for PCR amplification of *gp37*. The resulting PCR product was detected at the expected size and isolated for gel purification in preparation for cloning. The next chapter involves cloning the purified *gp37* product into pJET1.2/blunt and subcloning it into pCA528 expression vector through DNA ligation, followed by the verification of the constructed recombinant plasmids.

Chapter 3

Cloning of the *gp37* gene into pJET1.2/blunt and subsequent sub-cloning into pCA528 for protein expression

3.1 Introduction

In the previous chapter, oligonucleotides that target the CrpeNPV *gp37* gene were designed for PCR amplification to generate the *gp37* amplicon. The *gp37* amplicon was then gel-purified in preparation for cloning. In this study, pJET1.2/blunt served as an intermediate vector prior to sub-cloning into the target vector. It is a linearised cloning vector with a length of 2974 bp, featuring a T7 promoter, ampicillin resistance gene, multiple cloning site (MCS), and the lethal *eco47IR* gene, among other components. When the DNA insert is ligated into the MCS of pJET1.2/blunt, the lethal gene is disrupted (Daryaei et al., 2017; Nawawi et al., 2022). As a result, only bacterial cells with the recombinant plasmid can form colonies. In cases where no DNA insertion occurs, the lethal gene remains intact, leading to cell death and consequently, enabling positive selection of the recombinant plasmid (Reece-Hoyes & Walhout, 2018). Blue/white screening is unnecessary, and the vector is suitable for DNA fragments produced with either sticky ends or blunt ends, allowing for the cloning of inserts ranging from 6 bp to 10 Kb. Chemically competent *E. coli* host strains are commonly used for the propagation of recombinant plasmids (Casali, 2003). In this study, TOP10 *E. coli* competent cells were chosen as the model system for transforming the recombinant plasmids. These cells are derived from MC1061 and have a genotype similar to DH10B (Casadaban & Cohen, 1980; Durfee et al., 2008). Their cloning efficiency, versatility, and genotype made them preferred options for this study (Casali, 2003).

To confirm the presence of inserts in recombinant plasmids, it is necessary to conduct screening experiments such as colony PCR, restriction enzyme analysis and Sanger sequencing. In this study, colony PCR and restriction enzyme analysis were used to confirm the presence of the insert in both pJET1.2/blunt and pCA528. Colony PCR was conducted using oligonucleotides that flank the insert region on pJET1.2/blunt, and oligonucleotides designed for cloning, which target gp37 on pCA528. Following plasmid extraction, XhoI and BamHI restriction enzymes are employed to confirm insert presence. These enzymes facilitate both single and double digestion of recombinant plasmids, crucial for ensuring correct cohesive ends. Sanger sequencing was used to validate and verify the pCA528 construct. This technique is highly

effective in verifying the open reading frame (ORF) in the expression vector and detecting mutations, including single nucleotide polymorphisms (SNPs) and base pair insertions or deletions (indels) that may have occurred during PCR amplification and/or cloning (Behjati & Tarpey, 2013; Men et al., 2008; Sanger et al., 1977). Ensuring the accuracy of the inserted DNA sequence is essential in plasmid construction for subsequent applications such as protein expression.

The aim of this chapter was to construct a recombinant plasmid containing the *gp37* gene by DNA ligation in preparation for protein expression and purification. The specific objectives were: i) to clone *gp37* into pJET1.2/blunt and verify the recombinant plasmid, and ii) to sub-clone *gp37* into the pCA528 expression vector and verify the recombinant plasmid through colony PCR, restriction enzyme analysis, and Sanger sequencing.

3.2 Methods and materials

3.2.1 Cloning of *gp37* into pJET1.2/blunt

3.2.1.1 *In silico* construction of pJET-gp37 plasmid

In silico ligation of *gp37* into pJET1.2/blunt was conducted using SnapGene® software (GSL Biotech LLC, San Diego, CA), resulting in the pJET-gp37 map. The pJET1.2/blunt sequence was downloaded from the SnapGene website (https://www.snapgene.com). The *gp37* sequence was retrieved from the GenBank database (Accession no: YP_010086923).

3.2.1.2 Ligation of *gp37* into pJET1.2/blunt

The purified *gp37* PCR product was cloned into pJET1.2/blunt cloning vector using the CloneJET PCR cloning kit (Thermo Fisher Scientific, USA) following the manufacturer's guidelines. A ligation reaction was prepared with a total volume of 18 µL. The reaction consisted of 10 µL of 2× Reaction Buffer, 1 µL (13,92 ng/µL) of PCR product (*gp37*), 6 µL of double distilled water (ddH_2O) and 1 µL of DNA blunting enzyme. The ligation was incubated at 70 °C for 5 minutes to activate the DNA blunting enzyme. Subsequently, 1 µL of pJET1.2/blunt Cloning Vector (50 ng/µL) and 1 µL of T4 DNA Ligase were added to the mixture, which was vortexed, centrifuged for 3-5 seconds, and then incubated at room temperature (22 °C) for 5 minutes. The resulting recombinant plasmid was named pJET-gp37.

3.2.1.3 Transformation of pJET-gp37

Transformation was carried out by mixing 100 µL of TOP10 *E. coli* competent cells with the ligation reaction and cooled on ice for 20 minutes. The cells were then heat shocked at 42 °C for 30 seconds, followed by a cooling step on ice for 2 minutes. Subsequently, the cells were allowed to recover in 950 µL of Luria broth (LB) at 37 °C for 60 minutes and were then spread onto LA plates containing ampicillin (100 µg/mL) (LA-AMP plates). The plates were incubated overnight at 37 °C.

3.2.1.4 Colony PCR and plasmid extraction

Colony PCR was performed on four colonies following the protocol provided by the CloneJET PCR cloning kit (Thermo Fisher Scientific, USA) in accordance with the manufacturer's guidelines. Colony PCR reactions were prepared, with each containing 12.5 µL of 2× Taq Master Mix, 1 µL of a 10 µM pJET1.2 forward sequencing primer, 1 µL of a 10 µM pJET1.2 reverse sequencing primer, a single transformed colony containing template DNA, and 10.5 µL of ddH_2O, resulting in a total reaction volume of 25 µL. A control reaction (NTC) was set up as described above. A SimpliAmp™ Thermal cycler (Thermo Fisher Scientific, USA) was used to carry out the PCR amplification. The PCR cycle parameters included an initial denaturation step at 95 °C for 3 minutes, followed by 25 cycles consisting of denaturation at 94 °C for 30 seconds, annealing at 60 °C for 30 seconds, and elongation at 72 °C for 30 seconds. The PCR amplicons were analysed by 1 % AGE as described in section 2.2.5.

The same four colonies selected from the agar plate for colony PCR were inoculated into 5 mL of LB containing ampicillin (100 µg/mL) and incubated at 37 °C overnight. The recombinant pJET-gp37 vector was then extracted from the overnight cultures using the GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific, USA), following the manufacturer's provided guidelines. The extracted plasmids were analysed by 1 % AGE as described in section 2.2.5.

3.2.1.5 Restriction enzyme analysis

The restriction digestion reaction consisted of 4 µL of 2× Tango Buffer (Thermo Fisher Scientific, USA), 0.5 µL of BamHI (10 U/µL) (Thermo Fisher Scientific, USA), 0.5 µL of XhoI (10 U/µL) (Thermo Fisher Scientific, USA), 5 µL of plasmid DNA, and 10 µL ddH_2O, making a total volume of 20 µL. Both single and double digests were set up as described above. The reactions were centrifuged and then incubated at 37 °C for 1 hour and 30 minutes. The samples were analysed by 1 % AGE as described in section 2.2.5.

3.2.1.6 Gel purification of *gp37* product

Prior to sub-cloning into the target vector, the *gp37* insert was gel-purified following double digest with BamHI and XhoI using the GeneJET Purification Kit (Thermo Fisher Scientific, USA) following the manufacturer's guidelines, with the exception that the insert was eluted in 20 µL of elution buffer. The purified amplicon was analysed by 1 % AGE as described in section 2.2.5.

3.2.2 Sub-cloning of gp37 into pCA528

3.2.2.1 Target vector preparation

Glycerol stocks of pCA528 were streaked onto LA plates supplemented with kanamycin (50 µg/mL) and then incubated at 37 °C overnight. Subsequently, four distinct colonies were selected from the plate and inoculated in 5 mL of LB containing kanamycin (50 µg/mL) and left to grow overnight at 37 °C. The pCA528 plasmid was extracted from the bacterial culture using the GeneJET Gel Extraction kit (Thermo Fisher Scientific, USA), following the provided manufacturer's instructions. Plasmid DNA was analysed by 1 % AGE as previously described in section 2.2.5.

The pCA528 plasmid was then subjected to restriction digestion with both BamHI and XhoI to linearise the pCA528 in preparation for ligation with the insert. The digested pCA528 plasmid was analysed by 1 % AGE, as previously described in section 2.2.5. The pCA528 vector backbone was purified from the agarose gel using the GeneJET Purification Kit (Thermo Fisher Scientific, USA) following the manufacturer's instructions. The purified plasmid was analysed by 1 % AGE, as previously described in section 2.2.5.

3.2.2.2 *In silico* construction of the pCA-gp37 map

In silico ligation of *gp37* into pCA528 was conducted using Benchling software. The resulting sequence was used to create the pCA-gp37 map using SnapGene® software (GSL Biotech LLC, San Diego, CA). The pCA528 vector sequence was downloaded from the DNASU database (WISP08-174, DNASU plasmid repository). The *gp37* sequence was retrieved from the GenBank database (Accession no: YP_010086923).

3.2.2.3 Ligation of *gp37* into pCA528 and bacterial transformation

A ligation reaction was prepared with a total volume of 20 µL. The reaction included 2 µL of 10× ligase buffer, 1.5 µL (3.93 ng/µL) of vector DNA (pCA528), 8 µL (0.9 ng/µL) of insert

DNA (*gp37*), 1 µL of T4 DNA ligase (5 U/µL) (Thermo Fisher Scientific, USA), and 7.5 µL of ddH_2O. The ligation mixture was vortexed, centrifuged, and incubated at room temperature (±22 °C) for 2 hours. The resulting recombinant plasmid was named pCA-gp37. Transformation of TOP10 *E. coli* was carried out as described in section 3.2.1.3, and then spread onto LA plates containing kanamycin (50 µg/mL) (LA-KanR plates). The plates were incubated overnight at 37 °C. The successfully transformed bacteria were screened using colony PCR, restriction enzyme analysis, and Sanger sequencing.

3.2.2.4 Colony PCR and plasmid extraction

A colony PCR reaction was prepared with a total volume of 25 µL. The reaction included 12.5 µL of 2× Amplicon Taq Master Mix, 2 µL of a 10 µM gp37-F oligonucleotide, 2 µL of a 10 µM gp37-R oligonucleotide, a single transformed colony containing template DNA, and 8.5 µL of double distilled water. A control reaction (NTC) was set up in the same manner as described above. A SimpliAmp™ Thermal cycler was used to carry out the PCR amplification. The PCR cycle parameters included an initial denaturation step at 95 °C for 1 minute, followed by 30 cycles consisting of denaturation at 95 °C for 1 minute, annealing at 57 °C for 30 seconds, and elongation at 72 °C for 1 minute. This was followed by a final extension step at 72 °C for 3 minutes. The PCR amplicons were analysed by 1 % AGE, as described in Chapter 2, section 2.2.5. Subsequently, three distinct colonies were selected from the plate and inoculated in 5 mL of LB containing kanamycin (50 µg/mL), and left to grow overnight at 37 °C. The pCA-gp37 plasmids were extracted from the overnight cultures using the GeneJET Plasmid Miniprep Kit (Fisher Scientific, USA) according to the manufacturer's instructions. Plasmids were analysed by 1 % AGE, as previously described in Chapter 2, section 2.2.5.

3.2.2.5 Restriction enzyme analysis

The pCA-gp37 DNA obtained after extraction was subjected to restriction digestion using BamHI and XhoI as described in section 3.2.1.5. Subsequently, the plasmid samples were sent to Inqaba Biotechnical Industries (Pty) Ltd (South Africa) for sequencing using the plasmid sequencing primers, T7 forward and T7 terminator reverse, respectively.

3.2.2.6 Sanger sequencing

The chromatogram file generated from Sanger sequencing was processed using the ABI base recall programme (Elyazghi et al., 2017) to automatically correct ambiguities in both the forward and reverse sequences. Poor-quality sequence of regions at the beginning and end of

fragments were automatically identified and trimmed. The trimmed sequences were then analysed in Geneious R11. Using the ABI recall programme as reference, the trimmed sequences underwent manual trimming to further remove regions of poor-quality sequence that were not trimmed automatically by the ABI base recall programme. The reference sequence for pCA-gp37 was obtained from Benchling as a FASTA-formatted file and was processed using Geneious R11. A multiple alignment was performed for both the forward and reverse sequences against the pCA-gp37 as reference, using the Clustal W method in Geneious R11. This was done to confirm the reading frame of 6×His-SUMO-gp37 in pCA528.

3.3 Results

3.3.1 Cloning of *gp37* into pJET1.2/blunt

3.3.1.1 *In silico* pJET-gp37 plasmid map

Figure 3.1 illustrates the structure of the recombinant plasmid pJET1.2/blunt containing the *gp37* insert, generated using SnapGene® software. The plasmid map illustrates the position of *gp37* in the MCS, the incorporated restriction enzyme recognition sites (BamHI and XhoI), and the pJET sequencing primer positions. The size of the recombinant pJET-gp37 was 3752 bp.

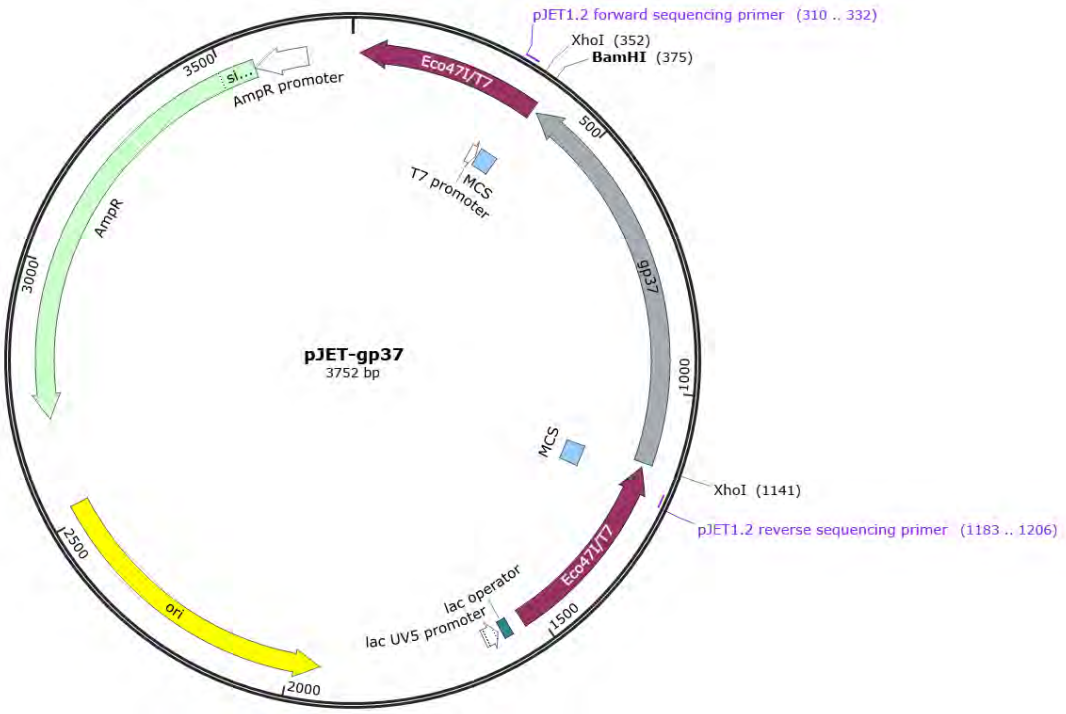


Figure 3.1. *In silico* construction of the pJET-gp37 map. Abbreviation used: AmpR- Ampicillin resistance gene, MCS- Multiple cloning site, *Eco47IR*- Lethal gene and Ori- origin of replication.

3.3.1.2 Confirmation of pJET-gp37

To confirm the presence of the insert in pJET2.1/blunt, colony PCR was performed, followed by plasmid extraction and restriction enzyme analysis. The results are presented in Figure 3.2 below. Figure 3.2A shows colony PCR of pJET-gp37 from four different colonies. The expected size of the PCR amplicon was 897 bp. A PCR amplicon band was observed running above 700 bp at approximately 900 bp in lanes 2-4, while no band was observed in the NTC as expected in lane 1. Figure 3.2B shows plasmid extraction of pJET-gp37 from the same four colonies mentioned above. The expected size of the pJET-gp37 was approximately 3752 bp. A bright supercoiled DNA band was observed running above 2000 bp in lanes 1-4. A linear band was expected above the supercoiled band at approximately 3752 bp, but it was not present. Figure 3.2C shows the results of the restriction enzyme analysis of the plasmid DNA extracted in Figure 3.2A lane 2. Lane 1 shows the undigested pJET-gp37, revealing a supercoiled DNA band running above 2000 bp. Lanes 2 and 3 show single digestions with BamHI and XhoI, respectively, which reveal clear, linearised bands at approximately 4000 bp. Lane 4 shows a double digestion with both BamHI and XhoI, which revealed two distinct bands, one running above 3000 bp and the other running above 700 bp at approximately 800 bp. Figure 3.2D shows the gel purified *gp37* fragment isolated from the double-digested plasmid DNA, detected running at approximately 800 bp in lane 1.

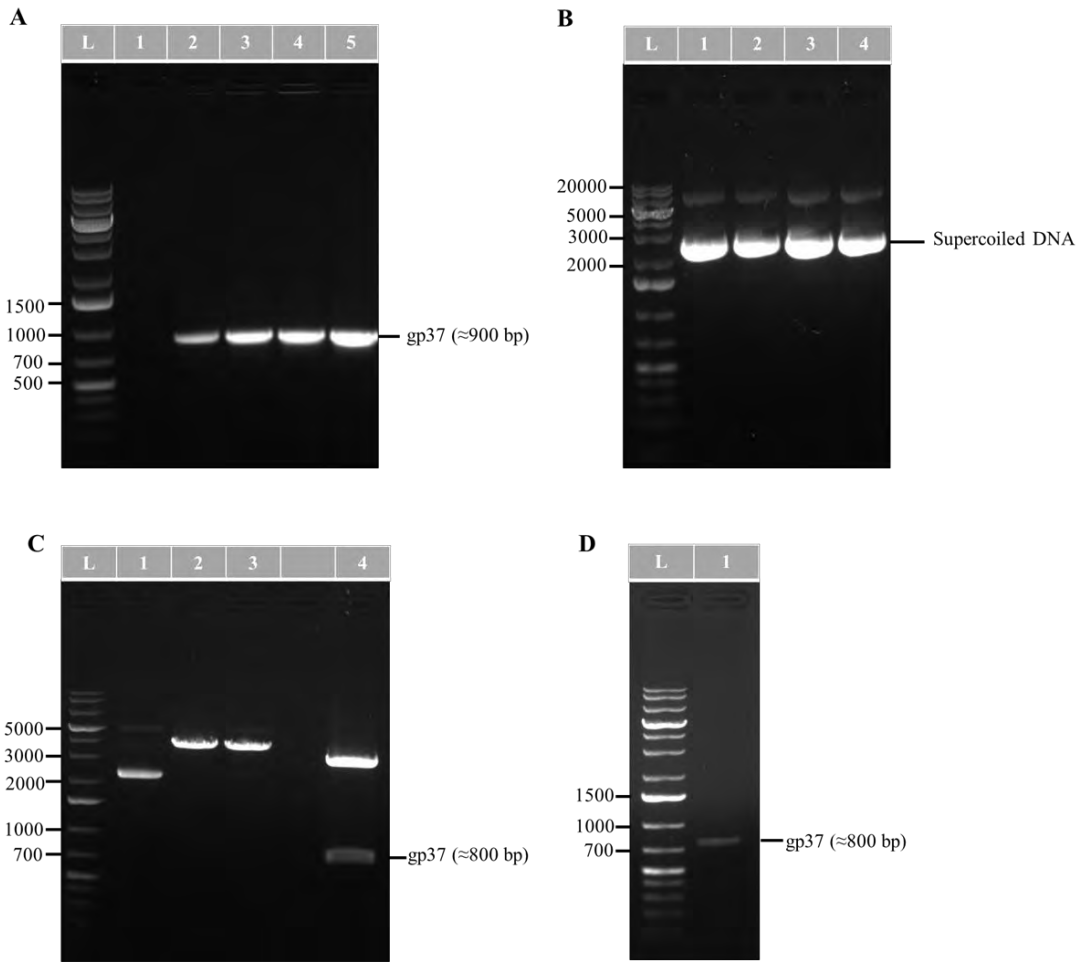


Figure 3.2. Agarose gel images showing colony PCR, plasmid extraction, restriction enzyme analysis of pJET-gp37 and gel purification of the *gp37* fragment. A: Colony PCR, lane 1: NTC, lane 2-5: *gp37* PCR product (colonies 1-4). B: Plasmid extraction, lane 1-4: Recombinant pJET-gp37. C: Restriction enzyme analysis of plasmid 2 (colony 2), lane 1: Undigested pJET-gp37, lane 2: Single digest with BamHI, lane 3: Single digest with XhoI, and lane:4 Double digest with XhoI and BamHI. D: Gel purification, lane 1: Purified *gp37* fragment. For all panels, lane L: GeneRuler 1 Kb plus DNA ladder.

3.3.2 Sub-cloning *gp37* into pCA528

3.3.2.1 Target vector preparation

To prepare the pCA528 backbone, plasmid extractions were performed, followed by restriction enzyme analysis and gel purification as shown in Figure 3.3 below. The pCA528 vector was extracted from four overnight cultures, with the resulting plasmid DNA shown in Figure 3.3A. The expected size of pCA528 was 5597 bp. Bright supercoiled DNA bands were observed running above 3000 bp in lanes 1-4, followed by faint linearised bands running above 5000 bp.

Figure 3.3B shows the results of the restriction enzyme analysis of pCA528. A bright supercoiled DNA band was observed running above 3000 bp in lane 1, representing the undigested DNA. Lane 2 shows the double digestion of the pCA528 backbone with BamHI and XhoI. The expected size of the digested backbone was approximately 5557 bp. A bright linear DNA band was observed running above 5000 bp. The linearised pCA528 backbone was purified from the agarose gel, as shown in Figure 3.3C. A purified pCA528 backbone was observed running at approximately 5000 bp, as expected.

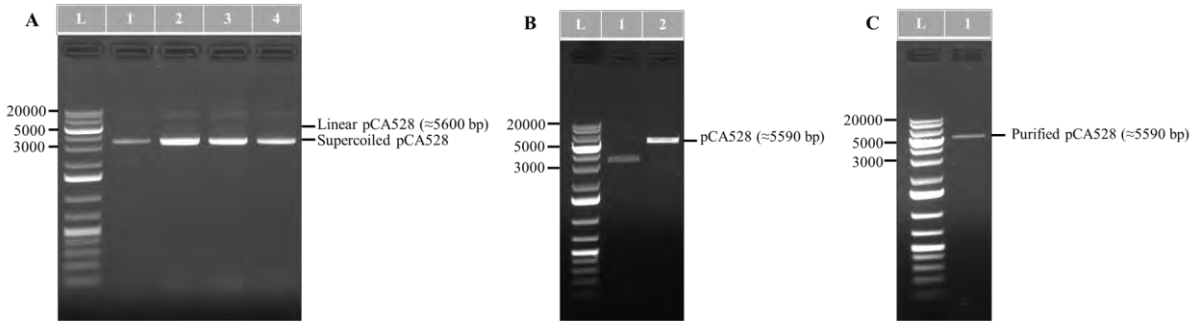


Figure 3.3. Agarose gel images showing plasmid extraction, restriction enzyme analysis and gel purification of pCA528. A: Plasmid extraction, lane 1-4: pCA528 DNA. B: Restriction enzyme analysis, lane 1: Undigested pCA528, and lane 2: Double digest with BamHI and XhoI. C: Gel purification, lane 1: Purified pCA528 backbone. For all panels, lane L: GeneRuler 1 Kb plus DNA ladder.

3.3.2.2 *In silico* pCA-gp37 map

Figure 3.4 illustrates the structure of the recombinant plasmid pCA528 containing the *gp37* insert, generated using SnapGene® software. The vector map shows the position of *gp37* in the vector and the recognition sites for BamHI and XhoI restriction enzymes, respectively. The *gp37* was strategically positioned to be in frame with the 6×His-SUMO tag, as illustrated below. The size of the pCA-gp37 was 6323 bp.

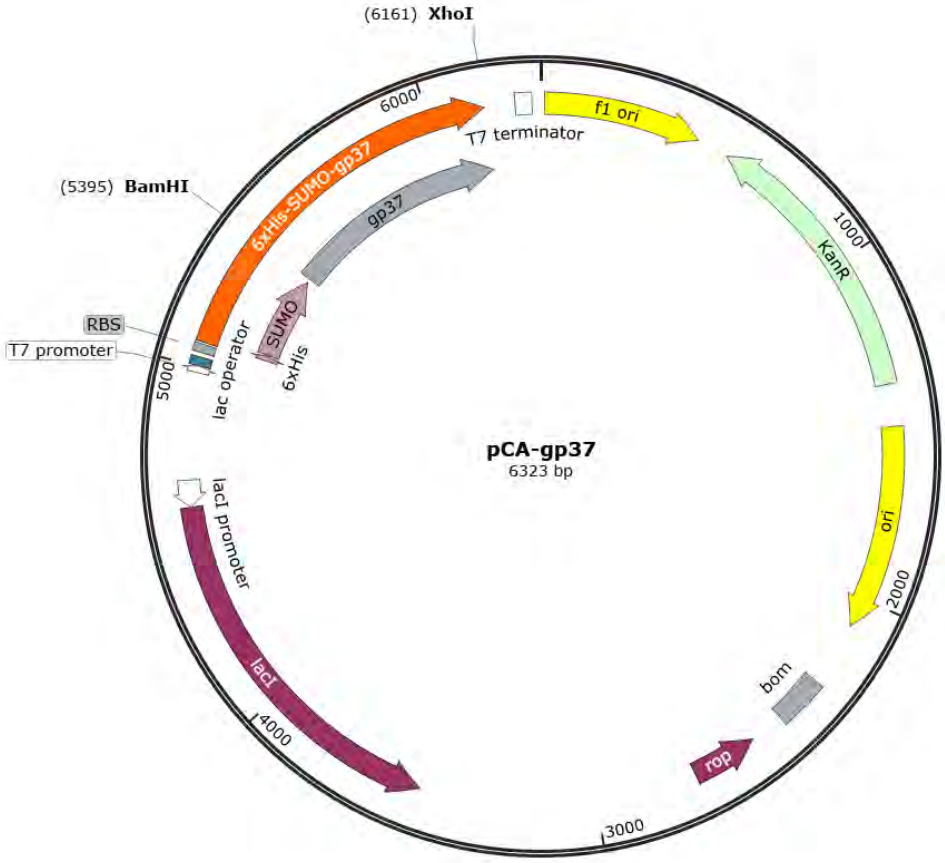


Figure 3.4. *In silico* construction of the pCA-gp37 map. Abbreviation used: RBS- Ribosome binding site, SUMO- Small ubiquitin-like modifier, 6×His- Polyhistidine tag, LacI- *Lac* repressor, *gp37*- The insert, 6×His-SUMO-gp37- Recombinant protein, KanR- Kanamycin resistance gene and Ori- Origin of replication.

3.3.2.3 Colony PCR and plasmid extraction

To confirm the presence of the insert in pCA528, colony PCR was performed, followed by plasmid extraction as shown below (Figure 3.5). Figure 3.5A shows colony PCR conducted using the oligonucleotides designed for cloning. The expected size of the *gp37* amplicon was approximately 778 bp. A bright band was observed running above 700 bp at approximately 800 bp. No band appeared in the NTC in lane 4 as expected. The pCA-gp37 DNA was extracted from 4 overnight cultures, with the resulting DNA shown in Figure 3.5B. Bright supercoiled DNA bands were observed running above 4000 bp in lanes 1-3, followed by a linear band running above 5000 bp at approximately 6000 bp.

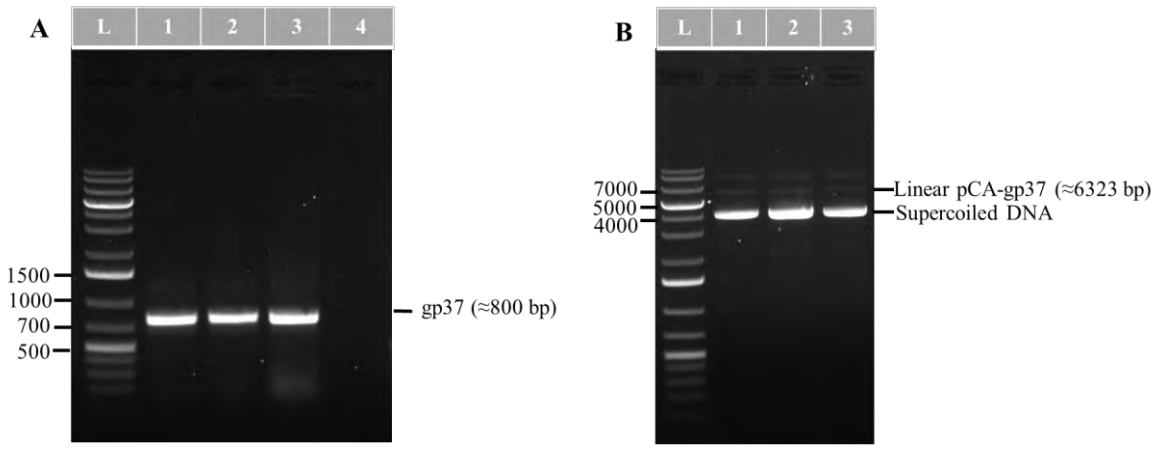


Figure 3.5. Agarose gel images showing colony PCR and plasmid extraction of the pCA-gp37. A: Colony PCR, lane 1-3: *gp37* PCR product (colonies 1-3), and lane 4: NTC. B: Plasmid extraction, lane 1-3: pCA-gp37 from colonies 1-3. For both panels, lane l: GeneRuler 1 Kb plus DNA ladder.

3.3.2.4 Restriction enzyme analysis

To confirm the integrity of the pCA-gp37, a restriction enzyme analysis was conducted, as shown below (Figure 3.6). Lane 1 shows the undigested pCA-gp37 DNA, exhibiting supercoiled DNA above 4000 bp. Single digests using BamHI and XhoI were observed in lanes 2 and 3, respectively, each revealing a clear, single linearised band above 5000 bp. In lane 4, the double digest with both BamHI and XhoI shows two distinct bands: one above 5000 bp and the other around 800 bp.

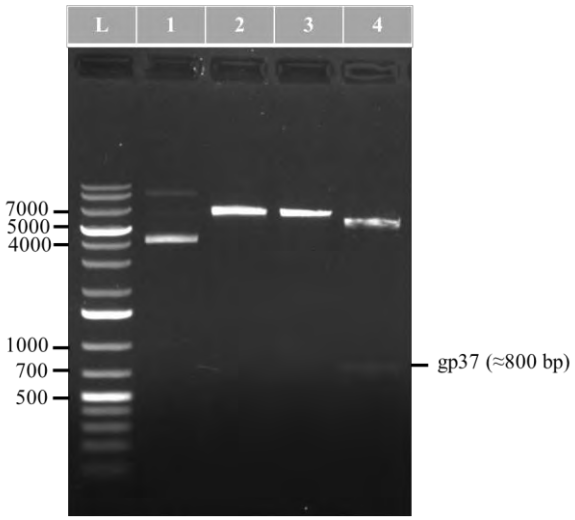


Figure 3.6. Agarose gel image showing the restriction enzyme analysis of the pCA-gp37. Lane L: GeneRuler 1 Kb plus DNA ladder, lane 1: Undigested pCA-gp37 DNA, lane 2: Single digest with BamHI, lane 3: Single digest with XhoI, and lane 4: Double digest with BamHI and XhoI.

3.3.2.5 Sanger sequence analysis of pCA-gp37

To validate the pCA-gp37, Sanger sequencing was performed by Inqaba Biotechnical Industries (Pty) Ltd (South Africa), and the resulting sequences underwent analysis using Geneious R11 software, as detailed in Figure 3.7. The alignment presents a continuous sequence from the start codon (5072 bp) to the stop codon (6160 bp) of the pCA-gp37 across Figure 3.7A, B, and C. The pCA-gp37 served as the reference sequence, as indicated in Figure 3.7. Sequence 1 represents the forward sequenced-version, while sequence 2 represents the reverse-sequenced version. Regions of poor sequencing in both the forward and reverse sequences were trimmed at the ends, as depicted in Figure 3.7A and C. Both sequences were aligned to pCA-gp37, as shown in Figure 3.7. In Figure 3.7A, the depiction illustrates the N-terminus of the recombinant protein, containing a start codon, 6×His, and the initial 24 amino acids of the SUMO tag. Figure 3.7B provides an overview of the C-terminus of the SUMO tag, highlighting the BamHI restriction site and introducing the N-terminus of gp37. Finally, Figure 3.7C presents the C-terminus of the recombinant protein, displaying the tail region of gp37 terminated by an in-frame stop codon, along with the location of the XhoI restriction site. The alignment confirmed 100 % identity between the sequenced regions and the reference sequence. No SNPs or indels were detected between the start and stop codons in both the forward and reverse sequenced versions of pCA-gp37.

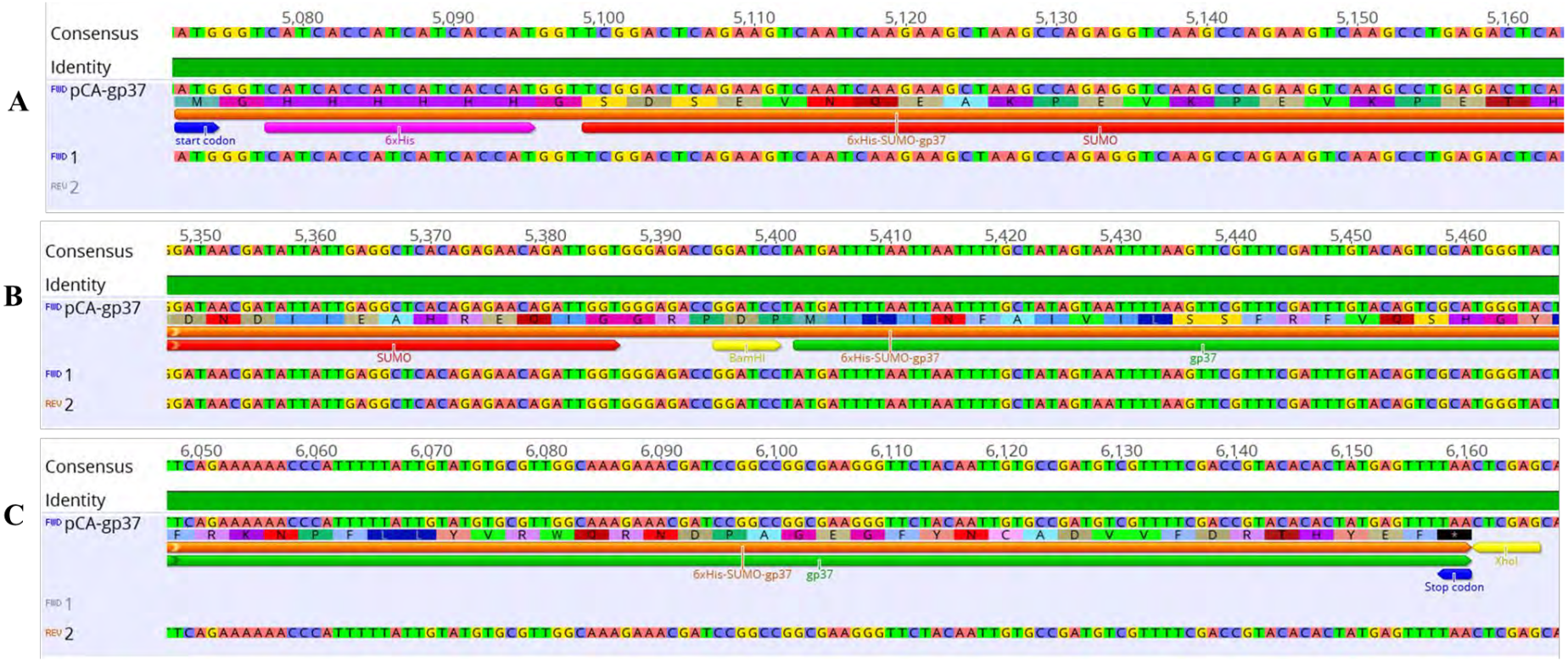


Figure 3.7. Sequence alignment of the regions flanking the *gp37* insert in pCA-gp37. The pCA-gp37 was used as a reference sequence, Sequence 1 and 2 are the forward and reverse sequences generated following Sanger sequencing of pCA-gp37. A) shows N-terminus of the recombinant protein, B) shows the central region of the recombinant protein, and C) shows the C-terminus of the recombinant protein. The translated amino acid sequence is shown from the start and to the stop codons in blue, with the 6×His tag in purple, the SUMO tag in red, the enzyme recognition sites of BamHI and XhoI in yellow, and the recombinant protein denoted as 6×His-SUMO-gp37 in orange.

3.4 Discussion

The aim of this chapter was to construct a recombinant plasmid containing the *gp37* gene in preparation for protein expression and purification. This was achieved by initially cloning the purified *gp37* amplicon into pJET1.2/blunt and verifying the recombinant plasmid. Cloning is a crucial step in thoroughly investigating genetic elements. This process combines a vector with an insert, facilitated by an enzyme such as DNA ligase. Subsequently, the recombinant plasmid is reproduced in a bacterial host cell, generating numerous identical copies of the DNA fragment for further analysis (Green & Sambrook, 2020).

In this study, pJET1.2/blunt was used as an intermediate vector as it offers a screening method for transformants containing the insert, via insertional inactivation lethal gene (*eco47IR*), which is not the case with direct cloning into the target vector. Additionally, restricting the recombinant plasmid yields the insert with the correct sticky ends for gel purification and ligation, making it a more efficient process than restricting a PCR amplicon for direct cloning. This is because restricting a PCR product may yield fragments cut once with either enzyme (incorrect sticky ends) or fragments cut with both enzymes (correct sticky ends), which makes it difficult to distinguish as they appear the same size on the gel. Another advantage of using an intermediate vector it that it serves as a stable source of the PCR product for downstream applications. This means that PCR doesn't have to be repeated since the product is already cloned into a vector and the recombinant plasmid can be stored for long periods of time as a glycerol stock. The *gp37* amplicon was ligated into pJET1.2/blunt, to generate the pJET-gp37. The presence of the insert was confirmed by colony PCR, and plasmid extraction, followed by restriction enzyme analysis. Restriction enzyme analysis using BamHI and XhoI successfully produced the *gp37* product at approximately 800 bp which was gel purified in preparation for sub-cloning into the pCA528 expression vector.

The pCA528 vector contains a T7 promoter that is inducible by IPTG, a kanamycin gene for the selection of bacterial cells carrying the plasmid, a 6×His tag consisting of six histidine residues that coordinate with transition metal ions such as Ni^{2+} immobilized on beads or resin for purification (Puckett, 2015), and a SUMO protein tag that has been demonstrated to significantly improve protein stability and solubility when used as an N-terminal fusion (Guerrero et al., 2015; Marblestone et al., 2006). The chosen expression vector, pCA528, is compatible with Rosetta (DE3) host cells. It features an inducible promoter regulated by IPTG for controlling gene expression levels. The plasmid is capable of accommodating the target

gene, and the chosen restriction enzymes are compatible with it. To prepare for sub-cloning the insert DNA, the pCA528 vector was restricted using BamHI and XhoI, and the resulting vector backbone was purified from an agarose gel. This was done to ensure that the backbone contains compatible cohesive ends with the *gp37* product. The purified *gp37* product was successfully ligated into the expression vector pCA528 and transformed into *E. coli* bacterial cells.

In this study, TOP10 *E. coli* competent cells were used to transform pJET-gp37 and pCA-gp37 after DNA ligations. This bacterial strain provides a transformation efficiency of 1×10^9 colony-forming units (CFUs) per microgram (μg) of supercoiled DNA, indicating that for every microgram of supercoiled DNA used, approximately 10 billion viable bacterial cells are expected to successfully take up the introduced DNA (Pope & Kent, 1996). TOP10 *E. coli* cells are well characterised, making it easier to troubleshoot and optimise cloning.

To ensure correct ligation, the pCA-gp37 was verified by colony PCR, plasmid extraction, and restriction enzyme analysis. Screening of colonies was conducted using oligonucleotides designed for cloning, revealing bands of the expected size. The recombinant plasmid was detected at the expected size, approximately 6323 bp. The recombinant plasmid was successfully digested using both BamHI and XhoI during restriction enzyme analysis, resulting in a band of approximately 800 bp. The recombinant plasmids sent for Sanger sequencing were analysed using Geneious R11 software. Initially, both forward and reverse sequences were trimmed to remove low-quality regions, typically found at the ends, which are the most ambiguous. This step increase's identity and coverage rates while reducing the occurrences of mismatches and gaps (Elyazghi et al., 2017). The purpose of this was to confirm in-frame cloning of the gene and verify whether the inserted DNA sequence is correct and matches the intended sequence. Sequence analysis of the recombinant plasmids confirmed that the *gp37* gene was in-frame with the 6×His-SUMO tag, displaying 100 % identity with the reference sequence. No SNPs or indels were detected between the start and stop codons.

In conclusion, the *gp37* purified amplicon was cloned into pJET1.2/blunt. The resulting pJET-gp37 was successfully verified by colony PCR, plasmid extraction and restriction enzyme analysis. The digested *gp37* product was gel-purified and sub-cloned into the expression vector, pCA528. Subsequently, the pCA-gp37 was successfully verified by colony PCR, plasmid extraction, restriction enzyme analysis, and sequencing. The next chapter involves expressing the recombinant protein in Rosetta (DE3) *E. coli* cells for downstream experiments, such as solubility analysis and protein purification.

Chapter 4

Expression and purification of 6×His-SUMO-gp37 in Rosetta (DE3) *Escherichia coli* competent cells

4.1 Introduction

In chapter 3, a recombinant plasmid was constructed for expression and purification of 6×His-SUMO-gp37 in a bacterial system for downstream bioassay evaluation. This chapter describes the expression of the recombinant protein in Rosetta (DE3) *E. coli* cells, followed by detection using Western analysis, solubility analysis, and purification under native and denaturing conditions. The Rosetta (DE3) *E. coli* host strain, a derivative of BL21, was chosen for expressing the recombinant protein (Baca & Hol, 2000; Kane, 1995). It is specifically designed to improve the expression of eukaryotic proteins with rarely used *E. coli* codons. This strain carries the chloramphenicol-resistant plasmid pRARE, which supplies tRNAs for AGG, AGA, CUA, CCC, and GGA codons (Novy et al., 2001). The DE3 designation indicates lysogeny for a λ prophage containing an IPTG-inducible T7 RNA polymerase, facilitating protein expression using T7 promoters in vectors (Rosano & Ceccarelli, 2009).

The expression of recombinant protein begins with a plasmid encoding the desired protein. The plasmid is then introduced into designated host cells, followed by the cultivation of these cells and induction of protein expression. The process ends with cell lysis and analysis to verify the presence of the protein (QIAexpressionist, 2002). In this study, expression of 6×His-SUMO-gp37 was conducted at different temperatures, and a time-course induction study was conducted to assess the level of protein expression. This involved collecting samples at various time points, including the control, and then adjusting the volume to account for cell growth (Vera et al., 2007). The empty vector was used as a control, to ensure that no other protein is being expressed other than the target protein. The recombinant protein was expressed using the pCA528 vector, which features a high-copy-number origin of replication and utilises the T7 promoter for gene transcription. Expression is controlled by the presence of the *lac* operon, to which the *lac* repressor (LacI) binds blocking gene transcription. Protein expression is induced by adding isopropyl β-D-1-thiogalactopyranoside (IPTG), which binds to the *lac* repressor protein, inactivating it (Rosano & Ceccarelli, 2014). Once the lac repressor is inactivated, the RNA polymerase of the host cells transcribes the sequence downstream from the promoter. Subsequently, the generated transcripts undergo translation to form the recombinant protein

(QIAexpressionist, 2002; Rosano & Ceccarelli, 2014). To visualise the protein being induced at different time points post induction, samples were analysed by Sodium dodecyl-sulphate polyacrylamide gel electrophoresis (SDS-PAGE) and Coomassie staining. Western blotting analysis was subsequently employed to identify the recombinant protein, using a monoclonal Histidine antibody that specifically recognises the Histidine tag attached to the recombinant protein. This monoclonal His-antibody provides high specificity, ensuring that it binds to the target protein with precision and lowers the chances of nonspecific binding (Nguyen et al., 2007; Shivachandra et al., 2012).

To purify the recombinant protein, it is necessary to perform a solubility analysis to confirm if the target protein is in a soluble or insoluble state. Soluble protein is required for downstream applications, such as protein purification and quantification. The host strain, vectors and growth conditions play a significant role in the solubility of the recombinant protein (Rosano & Ceccarelli, 2014). The temperature used to grow *E. coli* for protein expression can affect protein solubility and yield, with temperatures of 15 - 25 °C often proving optimal for poorly soluble proteins (Geng & Carstens, 2006; San-Miguel et al., 2013). In this study, 6×His-SUMO-gp37 was expressed at lower temperatures. High-temperature growth yields more protein synthesis but often results in a significant portion of recombinant protein being incorporated into inclusion bodies (Peti & Page, 2007). In contrast, lower temperatures facilitate better folding of the recombinant protein, maintaining its solubility (Geng & Carstens, 2006; Hwang et al., 2014; Peti & Page, 2007; Rosano & Ceccarelli, 2014). The formation of these inclusion bodies is also influenced by factors such as the nature of the protein, the host cell, and the expression level, which is determined by the choice of vector, growth conditions and induction (Bhat et al., 2018; Hwang et al., 2014). Therefore, it is important to optimise induction conditions prior to protein purification. The purification of 6×His-SUMO-gp37 was carried out using Nickel-Nitrilotriacetic acids (Ni-NTA) affinity chromatography. This technique consists of Ni-NTA attached to Sepharose® CL-6B, providing a substantial binding capacity and minimal occurrence of nonspecific binding (QIAexpressionist, 2002). Protein purification under both native and denaturing conditions can be conducted using this method (De Young et al., 1993; Franken et al., 2000; QIAexpressionist, 2002; Rashid et al., 2005). Most proteins in the inclusion bodies are solubilised with detergents or denaturants, such as 8 M Urea, before the purification process begins (Franken et al., 2000).

The aim of this chapter was to express the recombinant 6×His-SUMO-gp37 in Rosetta (DE3) *E. coli* competent cells in preparation for purification. The specific objectives were: i) to

transform pCA-gp37 into Rosetta (DE3) *E. coli* competent cells, ii) to express 6×His-SUMO-gp37 at different temperatures and analyse by SDS-PAGE, followed by Western blot detection, and iii) to purify the 6×His-SUMO-gp37 under both native and denaturing conditions.

4.2 Methods and materials

4.2.1 Competent cell preparation

A starter culture of Rosetta (DE3) *E. coli* cells was prepared by inoculating a single colony into 5 mL of fresh LB broth (10 g/L Tryptone, 5 g/L Yeast Extract, and 10 g/L NaCl) and allowing the cells to grow overnight in a shaking incubator at 37 °C. The cells were then diluted by transferring 5 mL of the overnight culture into 95 mL of fresh LB broth and incubated at 37 °C in a shaking incubator while measuring the optical density at OD_{600} until it reached the mid-log phase (0.4 to 0.6). The grown cell culture was incubated on ice for about 30 minutes. The culture was split into two 50 mL centrifuge tubes and centrifuged at 4000 ×g for 5 minutes at 4 °C. The supernatant was discarded, and the pellets were resuspended in 30 mL of ice-cold 100 mM $MgCl_2$ by gently vortexing, then incubated on ice for 20 minutes. The cells were centrifuged at 4000 ×g for 5 minutes at 4 °C, and the supernatants were discarded. The pellets were resuspended in 25 mL of ice-cold 100 mM $CaCl_2$ by gently pipetting and incubated on ice for 2 hours. The tubes were then centrifuged at 4000 ×g for 5 minutes at 4 °C, and the supernatants were discarded. The cells were resuspended in 5 ml of ice-cold 100 mM $CaCl_2$ and 5 mL of ice-cold 15 % glycerol. The cells were then transferred into sterile 1.5 mL tubes to make 200 µL aliquots and stored at -80 °C.

4.2.2 Transformation of Rosetta (DE3) *E. coli* competent cells

To transform Rosetta (DE3) *E. coli* competent cells, 2 µL (46.5 ng/µL) of pCA-gp37 and 2 µL of pCA528 (94.5 ng/µL) as a negative control were transferred into two separate tubes, each containing 50 µL of Rosetta (DE3) *E. coli* competent cells. The tubes were then placed on ice for 20 minutes and heat-shocked at 42 °C for 45 seconds. Subsequently, 950 µL of fresh LB broth was added, and the tubes were incubated in a 37 °C shaking water bath for 60 minutes. The cells were harvested by centrifugation at 5000 ×g for 1 minute, and 900 µL of the supernatant was discarded. The remaining 100 µL of cells were resuspended and spread onto the LB agar plate containing 50 µg/mL kanamycin. The plates were then placed in an incubator overnight at 37 °C.

4.2.3 The expression of 6×His-SUMO-gp37 in Rosetta (DE3) *E. coli* cells

The transformed Rosetta (DE3) *E. coli* competent cells, containing pCA-gp37 and pCA528 as a negative control, were inoculated into 5 mL of LB broth containing 50 µg/mL kanamycin and incubated overnight at 37 °C with shaking. The cells were then diluted by transferring 5 mL of the overnight culture into 95 mL of fresh LB broth containing 50 µg/mL kanamycin. The mixture was incubated at 37 °C in a shaking incubator while measuring the optical density at OD_{600} until it reached a range between 0.6 and 0.8. Expression was induced with 1 mM IPTG, and a time course induction study was performed at 18 °C and 25 °C, respectively, with samples collected at 0-, 3-, 5-, and 24- hours post-induction (hpi). The cells were harvested by centrifugation at 5000 ×g for 5 minutes, and the pellets were resuspended in phosphate-buffered saline (PBS) (137 mM NaCl, 27 mM KCL, 10 mM KH_2PO_4, 1.8 mM Na_2HPO_4, pH 7.4), determined by the absorbance reading to standardise the OD_{600} to 0.5 using the formula: Volume of PBS (µL) = (OD_{600} ÷ 0.5) × 150 µL. The samples were prepared for SDS-PAGE and Western blot analysis.

4.2.4 SDS-PAGE analysis

The harvested cells, following induction and resuspension in PBS, were treated by boiling with the addition of 4× SDS sample buffer (0.1 % Bromophenol blue; 10 % SDS; 30 % glycerol (v/v); 50 mM Tris; 0.5 % β-mercaptoethanol) in a ratio of 4:1 for 10 minutes at 100 °C. The samples were then resolved using a 4 % acrylamide stacking gel (SureCast™ Acrylamide Solution (40 %) (Thermo Fisher Scientific, USA), 0.5 M Tris pH 6.8, 10 % SDS, 10 % APS, 0.3 % TEMED, and ddH_2O) and a 12 % acrylamide resolving gel (SureCast™ Acrylamide Solution (40 %) (Thermo Fisher Scientific, USA), 1.5 M Tris pH 8.8, 10 % SDS, 10 % APS, 0.3 % TEMED, and ddH_2O). The gels were then transferred into the electrophoresis tank, and electrophoresis buffer (250 mM Tris-base, 1.92 M glycine, and 1 % (w/v) SDS) was added. The boiled samples and PageRuler™ prestained protein ladder (Thermo Fisher Scientific, USA) were loaded into the wells. The samples were electrophoresed for 60 minutes at 200 Volts (V) using the Bio-Rad Mini protein 3 electrophoresis system (Bio-Rad, USA) or Invitrogen™ Mini Gel Tank (Thermo Fisher Scientific, USA). The gels were stained overnight with Coomassie brilliant blue R250 stain (3 g/L Coomassie dye R250, 40 % methanol, and 7 % Glacial acetic acid) overnight. Then they were destained with destain buffer (40 % methanol and 7 % Glacial acetic acid), which was changed by adding a fresh solution. The gels were visualised using the ChemiDoc™ XRS+ (Bio-Rad, USA) with Image Lab™ (Bio-Rad, USA) software.

4.2.5 Western blot analysis

Samples were again resolved on SDS-PAGE gels as per Section 4.2.4 above without Coomassie staining for analysis by Western blot. The gel and nitrocellulose membrane (Bio-Rad, USA) were sandwiched between sheets of filter paper and then transferred into the electrophoresis tank. The tank was filled with Western transfer buffer (25 mM Tris; 192 mM glycine, 20 % methanol), and the protein transfer was initiated at 90 V for 60 minutes. Western blotting was conducted using the BM Chemiluminescence Western Blotting Kit (Mouse/Rabbit) (Roche, Germany), following the manufacturer's guidelines. The membrane was recovered and incubated overnight in 1 % blocking reagent prepared in Tris-buffered saline (TBS) (50 mM Tris-HCl, pH 7.5, 150 mM NaCl). The membrane was washed twice for 10 minutes each with 10 ml of TBS-T buffer (0.1 % Tween20 in TBS), followed by a 10-minute wash with 10 mL of TBS buffer. The membrane was then incubated with primary anti-His6(2) antibodies (Roche, Germany) diluted 1:3000 in 0.5 % blocking reagent for 1 hour with shaking, followed by two 10-minute washes with TBS-T and a 10-minute wash with TBS buffer. The membrane was further incubated with secondary POD-labelled mouse/rabbit antibodies (BM Western blotting kit) (Roche, Germany), diluted 1: 12000 in 0.5 % blocking reagent for 30 minutes, followed by two 10-minute washes with TBS-T, and then 10-minute washes with TBS buffer. The detection solution was pre-mixed and applied to the membrane. Proteins were visualised using the ChemiDoc™ XRS+ (Bio-Rad, USA) with Image Lab™ (Bio-Rad, USA) software.

4.2.6 Solubility analysis of 6×His-SUMO-gp37

To prepare the lysate, a 50 mL culture induced for 24 hours at 18 °C with 1 mM IPTG to express 6×His-SUMO-gp37 was harvested by centrifugation at 5000 ×g for 10 minutes at 4 °C. The pellets were resuspended in 2.5 mL of lysis buffer (100 mM Tris, 300 mM NaCl, and 10 mM imidazole, pH 8), which contained 100 µg/mL of lysozyme and 1 mM of the protease inhibitor Phenylmethylsulphonyl fluoride (PMSF). The lysates underwent three freeze-thaw cycles and were sonicated on ice (5 cycles with 15-second pulses). Subsequently, 1 mL of the sonicated whole lysate sample representing total protein was collected. The remaining sample was centrifuged at 5000 ×g for 20 minutes to obtain the supernatant (soluble fraction) and cell pellet (insoluble fraction). The pellets were resuspended in 2.5 mL of PBS and each sample prepared for analysis by SDS-PAGE and Western blot analysis as previously described in section 4.2.4 and 4.2.5, respectively.

4.2.7 Purification of 6×His-SUMO-gp37 and SUMO-GFP using Ni-NTA affinity chromatography

4.2.7.1 Green Fluorescent Protein (GFP)

A plate containing the recombinant pCri-11b-GFP transformed in XJb::λ (DE3) *E. coli* cell was provided by Dr G. Abrahams (Rhodes University, South Africa). The vector used to clone GFP contained both SUMO and His-Tag, which facilitate protein purification The expressed protein was denoted as SUMO-GFP and was expressed for 24 hours at 18 °C, as previously described in section 4.2.3 for protein purification preparation. The purification of 6×His-SUMO-gp37 and SUMO-GFP was conducted simultaneously under the same conditions.

4.2.7.2 Protein purification of 6×His-SUMO-gp37 and SUMO-GFP under native conditions

To purify the recombinant protein under native conditions (cells lysed with lysozyme), the samples were prepared as previously described in section 4.2.6. 0.2 mL of clarified supernatants (soluble fraction) were purified using the His Pur™ Ni-NTA Spin Purification kit, 0.2 mL (Thermo Fisher Scientific, USA), following the manufacturer's guidelines. The columns were washed five times with 0.4 mL washing buffer (200 mM sodium phosphate, 300 mM NaCl, and 25 mM imidazole, pH 7.4), and the proteins were eluted with 0.2 mL elution buffer (200 mM sodium phosphate, 300 mM NaCl, and 250 mM imidazole, pH 7.4) in triplicate. The proteins were analysed using SDS-PAGE and Western blot analysis as previously described in section 4.2.4 and 4.2.5, respectively.

4.2.7.3 Protein purification of 6×His-SUMO-gp37 and SUMO-GFP under denaturing conditions

To purify the recombinant proteins under denaturing conditions (cells lysed with Urea), the bacterial pellet obtained through centrifugation after cell culturing was resuspended in 2.5 mL of lysis buffer (8 M Urea, 100 mM Tris, 300 mM NaCl, and 10 mM imidazole, pH 8). The cells were lysed by sonication (5 cycles with 15-second pulses) on ice. 1 mL of the sonicated sample was collected and represented the whole lysate for SDS PAGE analysis, and the remaining cell lysates were centrifuged at 5000 ×g for 20 minutes to collect the supernatants (soluble fraction). The pellets were resuspended in 2.5 mL of PBS and prepared for SDS-PAGE analysis. The clarified supernatants, 0.2 mL in volume, were purified using the His Pur™ Ni-NTA Spin purification kit (Thermo Fisher Scientific, USA). The columns were washed twice with a series

of Urea concentrations, using 0.4 mL of each of the washing buffers as shown in Table 4.1. The proteins were then eluted with 0.2 mL of elution buffer (100 mM Tris, 300 mM NaCl, and 250 mM imidazole, pH 8). The proteins were analysed using SDS-PAGE and Western blot as previously described in section 4.2.4 and 4.2.5, respectively.

Table 4.1. Description of wash buffers used in denaturing purification experiments.

Washes	Compositions
Wash 1	8 M Urea, 100 mM Tris, 300 mM NaCl, and 25 mM imidazole, pH 8
Wash 2	6 M Urea, 100 mM Tris, 300 mM NaCl, and 25 mM imidazole, pH 8
Wash 3	4 M Urea, 100 mM Tris, 300 mM NaCl, and 25 mM imidazole, pH 8
Wash 4	2 M Urea, 100 mM Tris, 300 mM NaCl, and 25 mM imidazole, pH 8
Wash 5	100 mM Tris, 300 mM NaCl, and 25 mM imidazole, pH 8

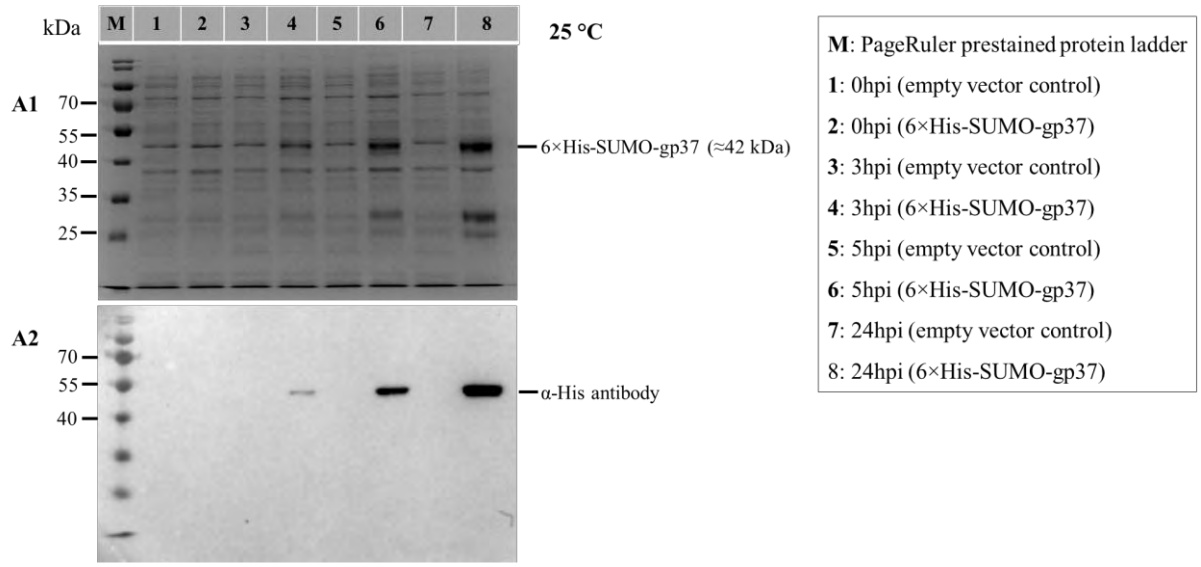
4.3 Results

4.3.1 Expression of 6×His-SUMO-gp37 at 25 °C and 18 °C

A time-course induction study of the total 6×His-SUMO-gp37 was conducted against the empty vector control at different temperatures, as depicted in Figure 4.1. The results show the total protein expressed in different time points (0-, 3-, 5- and 24-hours post-induction) at both 25 °C and 18 °C. The expected size of 6×His-SUMO-gp37 was 42 kDa. At 25 °C (Figure 4.1A1), a protein of approximately 42 kDa was visible at 3 hours post-induction (hpi) in lane 4, with increased intensity at 5 hpi (lane 6) and 24 hpi (lane 8). Western blot analysis using anti-His antibodies confirmed the expression of 6×His-SUMO-gp37 as a 42 kDa His-tag fusion protein (Figure 4.1A2). At 0 hpi (lane 2), no His-tagged protein was expressed, as confirmed by Western blot analysis showing the absence of 6×His-SUMO-gp37 (Figure 4.1A2). In lanes 1, 3, 5, and 7 of the empty vector, there was no protein expression, as confirmed by Western blot, indicating the absence of the His-tagged protein at the expected size (Figure 4.1A1). Additionally, nonspecific bands were detected with increasing intensity between 25 and 35 kDa

in lanes 6 and 8 (Figure 4.1A1), which were not detected by the Western blot analysis (Figure 4.1A2).

Similar outcomes were observed at 18 °C (Figure 4.1B1), where a protein of approximately 42 kDa was visible at 3 hpi (lane 4) and increased in intensity at 5 hpi (lane 6) and 24 hpi (lane 8). Western blot analysis using anti-His antibodies confirmed the expression of 6×His-SUMO-gp37 as a 42 kDa His-tag fusion protein (Figure 41B2). At 0 hpi (lane 2), no protein was observed, as confirmed by Western blot analysis showing no detection of 6×His-SUMO-gp37 (Figure 4.1B2). In lanes 1, 3, 5, and 7 of the empty vector, there was no protein expression, as confirmed by Western blotting, indicating the absence of the His-tagged protein at the expected size (Figure 4.1B1). Additionally, nonspecific bands were also detected with increasing intensity at approximately 25 kDa in the empty vector (Figure 4.1B1), and these bands were also detected by Western blot analysis (Figure 4.1B2)



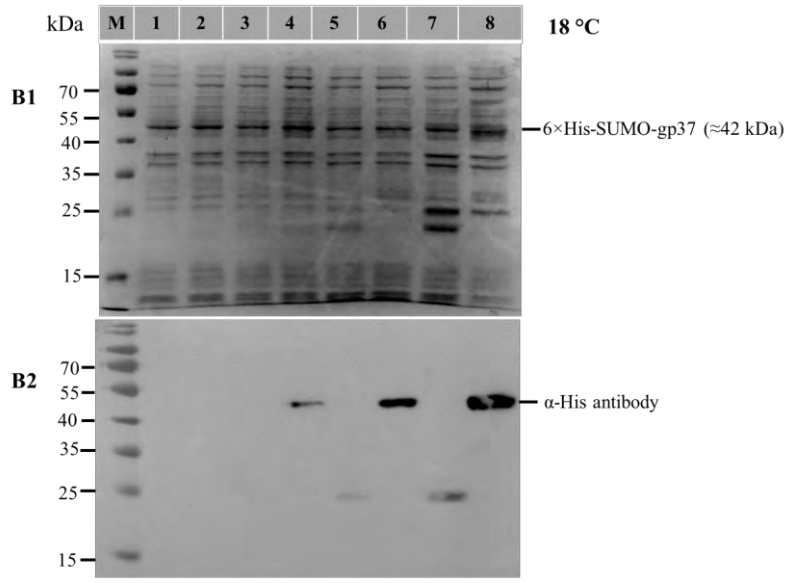


Figure 4.1. Expression of 6×His-SUMO-gp37 at two different temperatures. Panel A1 shows SDS-PAGE analysis at 25 °C, while Panel B1 shows SDS-PAGE analysis at 18 °C. Panel A2 and B2 show Western blot analyses. Lane M represents the protein ladder in kilodaltons (kDa), lanes 1,3,5, and 7 represent samples taken at 0-, 3-, 5- and 24- hpi (empty vector control), and lanes 2, 4, 6 and 8 represent samples taken at 0-, 3-, 5- and 24 hpi (6×His-SUMO-gp37), respectively.

4.3.2 Solubility analysis of 6×His-SUMO-gp37 under native conditions

The results show cells that were lysed with lysozyme under native conditions and centrifuged to separate the whole lysate (total protein), insoluble (pellet fraction), and soluble (supernatant fraction). 6×His-SUMO-gp37 was harvested 24 hours post-induction at both 25 °C and 18 °C, as shown in Figure 4.2 below. At 25 °C (Figure 4.2A1), both the whole lysate and pellet fractions show the presence of 6×His-SUMO-gp37 with high band intensity, visible in lanes W and P, respectively. This was confirmed by Western blot analysis using the anti-His antibodies which showed the detection of bands between 55 and 40 kDa for both the whole lysate and pellet fractions (Figure 4.2A2). No 6×His-SUMO-gp37 was detected in the soluble fraction, as shown in lane S (Figure 4.2A1), and this absence was further confirmed by Western blot analysis using the anti-His antibodies (Figure 4.2A2).

At 18 °C (Figure 4.2B1), similar outcomes were observed in both the whole lysate and pellet fractions in lanes W and P, respectively. However, a band of 6×His-SUMO-gp37 was detected in the soluble fraction, as shown in lane S (Figure 4.2B1), and this observation was further

confirmed by Western blot analysis using the anti-His antibodies, which detected the respective target bands (Figure 4.2B2).

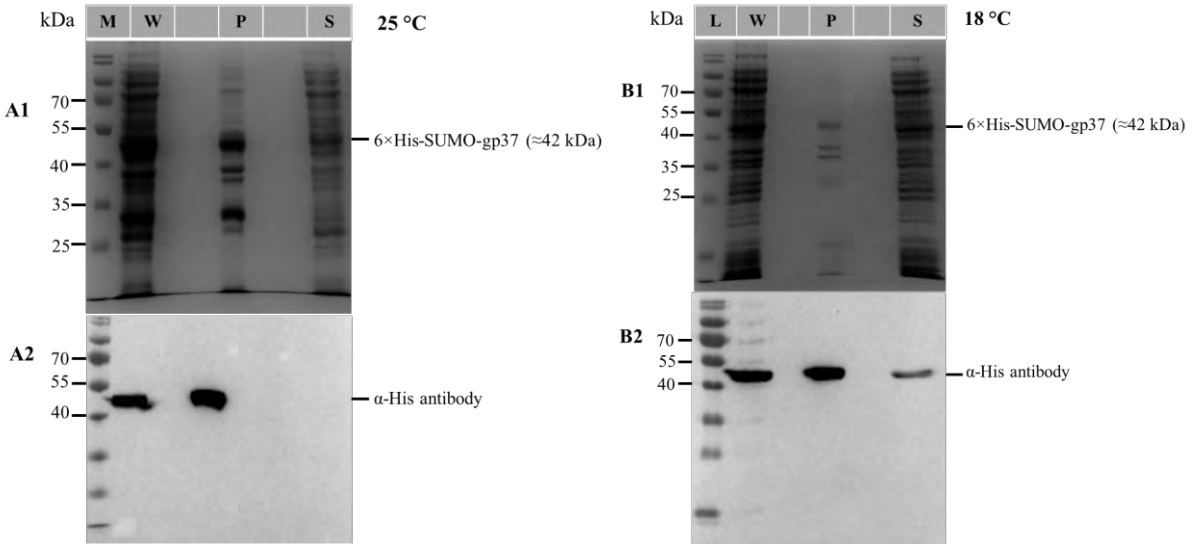


Figure 4.2. Solubility analysis of 6×His-SUMO-gp37 under native conditions (cells lysed with lysozyme) at two different temperatures. Panel A1 shows SDS-PAGE analysis of proteins expressed at 25 °C, while Panel B1 shows SDS-PAGE analysis of proteins expressed at 18 °C. Panel A2 and B2 show Western blot analyses of proteins expressed at 25 and 18 °C, respectively. Lane M represents the protein marker in kilodaltons (kDa), lane W represents the whole lysate (total protein), lane P represents the pellet fraction (insoluble after centrifugation), and lane S represents the supernatant fraction (soluble after centrifugation).

4.3.3 Purification of 6×His-SUMO-gp37 and SUMO-GFP using affinity chromatography

6×His-SUMO-gp37 and SUMO-GFP were expressed at 18 °C, purified under native conditions (cells lysed with lysozyme) and denaturing conditions (cells lysed with Urea) (Figure 4.3). This section refers to bacterial cells that were lysed with lysozyme. Under native conditions (Figure 4.3A and B), 6×His-SUMO-gp37 was detected in the flow-through, as shown in lane FT, however, no protein was eluted, as indicated in lanes E1-E3 (Figure 4.3A). No band running at the size of 6×His-SUMO-gp37 was detected in the wash fraction, as expected in lane W1 (Figure 4.3A). Bands of approximately 42 kDa representing SUMO-GFP, were observed in the whole lysate (lane W) and Supernatant (lane S), while no band was observed in the pellet (lane P) fraction. SUMO-GFP was lost in the flow-through, as shown in lane FT, but was subsequently eluted with a clear band observed in each lane, E1-E3 (Figure 4.3B).

This section refers to bacterial cells that were lysed with Urea. Under denaturing conditions (Figure 4.3C1 and D), a 42 kDa band representing 6×His-SUMO-gp37 was observed in the whole lysate (lane W), supernatant (lane S), and pellet (lane P) fractions. A band of lower intensity representing 6×His-SUMO-gp37 was observed in the flow through, as shown in lane FT, and in lanes E1-E3, with nonspecific bands detected between 25 and 35 kDa (Figure 4.3C1). Only a little protein was detected in the wash fraction, as shown in lane W1 (Figure 4.3C1). This was confirmed by Western blot analysis using anti-His antibodies, which showed the expected bands at approximately 42 kDa, along with nonspecific bands detected at a higher molecular weight. SUMO-GFP also exhibited bands of high intensity in the whole lysate (lane W) and Supernatant (lane S), while a low intensity band was observed in the pellet (lane P). No protein was lost in flow-through, as shown in lane FT, with SUMO-GFP subsequently eluted, producing clear high intensity bands in lanes E1-E3 (Figure 4.3D).

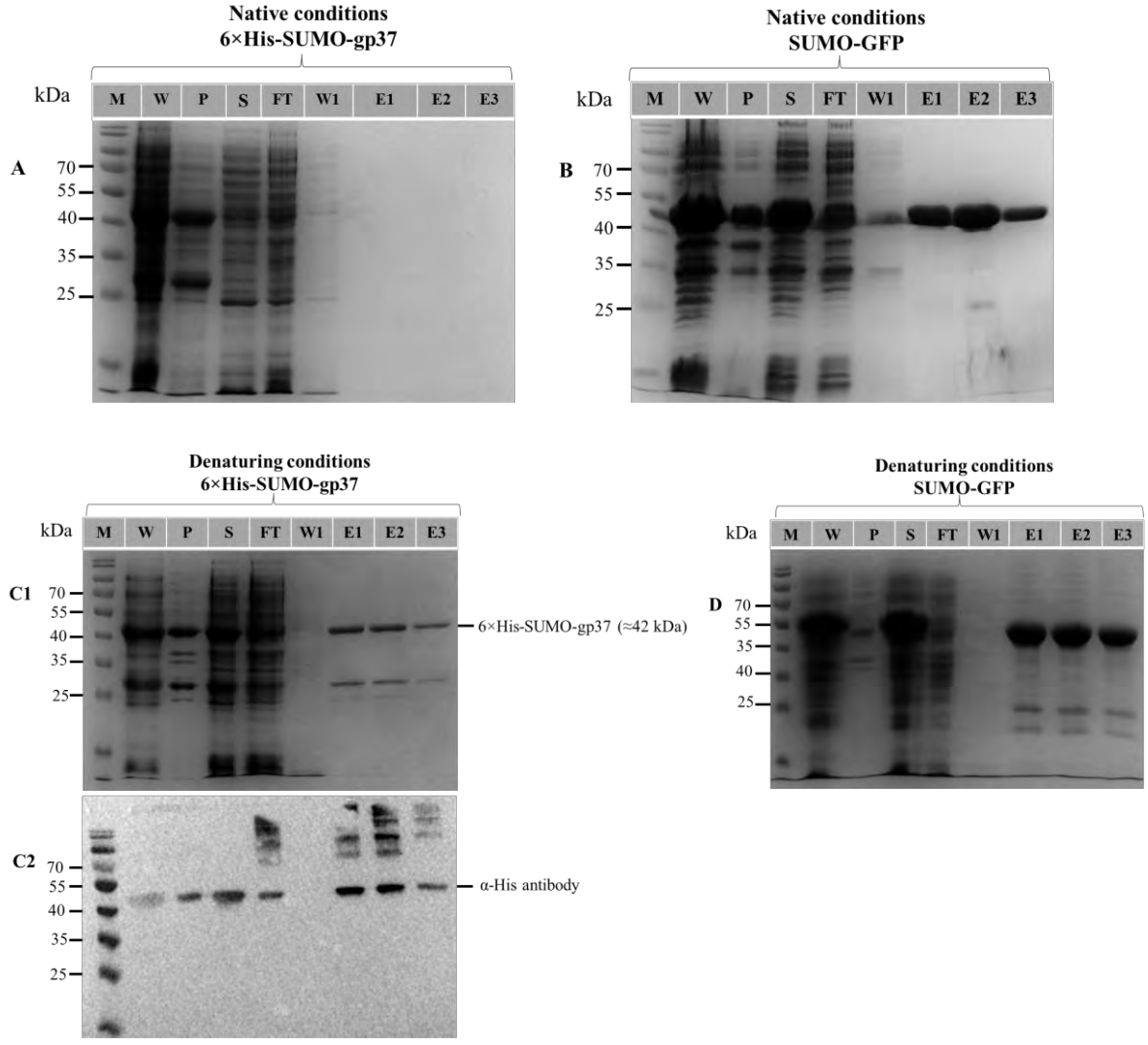


Figure 4.3. Purification of 6×His-SUMO-gp37 and SUMO-GFP under native (cells lysed with lysozyme) and denaturing conditions (cells lysed with Urea). SDS-PAGE analysis under native conditions of A) 6×His-SUMO-gp37 and B) SUMO-GFP. SDS-PAGE analysis under denaturing condition of C1) 6×His-SUMO-gp37 and D) SUMO-GFP. Panel C2 shows Western blot analysis of 6×His-SUMO-gp37 under denaturing conditions. Lane M represents the Page Ruler pre-stained protein marker in kilodalton (kDa), lane W represents the whole lysate (total protein), lane P represents the pellet fraction (insoluble after centrifugation), lane S represents the supernatant fraction (soluble after centrifugation). Lane FT represents the flow-through fraction, lane W1 represents the wash fraction, and lanes E1-E3 represent the eluates, respectively.

4.4 Discussion

The overall aim of this chapter was to express 6×His-SUMO-gp37 in a bacterial system at different temperatures and then analyse the protein through SDS-PAGE with Coomassie staining and Western blot analysis using anti-His antibodies.

The host cell plays a significant role in the production of the target protein. In this study, *E. coli* was chosen as a host organism due to the utilisation of a bacterial expression vector, its fast growth rate (Sezonov et al., 2007), well characterised biology, ability to achieve high density using inexpensive culture reagents (Shiloach & Fass, 2005), and efficient transformation with exogenous DNA (Pope & Kent, 1996). However, the heterogeneous expression of recombinant protein in *E. coli* often leads to protein aggregation into inclusion bodies. This presents a significant challenge in producing soluble recombinant proteins with proper biological functions (Bhatwa et al., 2021).

A time course induction study was performed in which transformed cells were induced at 25 °C and 18 °C for 0-, 3-, 5- and 24 hours. The results of SDS-PAGE of samples collected at various time points showed that a protein of approximately 42 kDa was visible at 3 hpi, with maximal expression at 24 hpi. Western blot analysis of the samples using anti-His antibodies confirmed that 6×His-SUMO was expressed as a 42 kDa His-SUMO-tag fusion protein. No leaky expression was observed before IPTG induction, as confirmed by Western blot analysis, indicating tight regulation of the gene in the plasmid. Additionally, a preliminary time course induction study of 6×His-SUMO-gp37 was performed at 37 °C, yielding similar results (results not shown). Inducing at 37 °C has disadvantages despite the high protein synthesis rate, it often leads to the incorporation of recombinant protein into inclusion bodies (Geng & Carstens,

2006). To optimise solubility of 6×His-SUMO-gp37, the recombinant protein was expressed at lower temperatures. At lower temperatures, aggregation is reduced due to the temperature-dependent nature of hydrophobic interactions (Kataeva et al., 2005; Pinsach et al., 2008; Sahdev et al., 2008; Volontè et al., 2008). Additionally, decreased temperature slows protein production rates, allowing newly transcribed recombinant proteins with more time for proper correct folding (Chou, 2007; QIAexpressionist, 2002; Rosano & Ceccarelli, 2014). The next step involved conducting a solubility analysis at both 18 °C and 25 °C. Solubility analysis is a critical step in protein purification as it ensures that the target protein is in a soluble form, essential for downstream applications such as protein purification and quantification (Rosano & Ceccarelli, 2014; Vera et al., 2007). The solubility analysis revealed that some 6×His-SUMO-gp37 was soluble at 18 °C but predominantly present in the insoluble fraction at both 18 °C and 25 °C. Additionally, a preliminary solubility analysis of 6×His-SUMO-gp37 harvested at 5- and 24 hpi at 37 °C showed that the protein was predominantly present in the insoluble fraction rather than soluble fraction (results not shown). This suggests that the protein might be aggregating, resulting in the formation of inclusion bodies. The formation of protein inclusion bodies in *E. coli* is a result of an unbalance equilibrium involving proper folding, aggregation, and degradation (Donovan et al., 1996). Various factors contribute to this process, including host cell metabolism, protein synthesis, modification machinery, properties of the target protein such as signal peptides or transmembrane domains, and environmental conditions (Bhatwa et al., 2021).

The Rosetta (DE3) cells were used as a host strain for expressing the recombinant protein. These cells carry the chloramphenicol-resistance plasmid, pRARE, which supplies tRNAs for AGG, AGA, CUA, CCC, and GGA codons under the control of their native promoter (Novy et al., 2001). *In silico* analysis revealed that 6×His-SUMO-gp37 contains 12 AGG, 27 AGA, 6 CCC, 14 CUA and 18 GGA codon, with expression likely benefiting from the presence of the pRARE plasmid. These rare codons are capable of causing translational frame shifting, premature translation termination, and misincorporation of amino acids that reduce the quality of the product or prevent a good expression yield (Calderone et al., 1996; Gerchman et al., 1994; Kane, 1995; Spanjaard & Van Duin, 1988). Rosetta (DE3) cells were initially cultured on chloramphenicol LB agar plates to verify the presence of the plasmid. The cells were able to grow abundantly, confirming the presence of the chloramphenicol resistance plasmid, pRARE. Even though Rosetta (DE3) cells improve expression yield, it has been reported that using bacterial strains that overcome codon bias can also lead to protein misfolding and

insolubility (Rosano & Ceccarelli, 2009). This suggests that such may be the case with 6×His-SUMO-gp37, which was predominantly found in the insoluble fraction. Eukaryotic proteins expressed in *E. coli* cells often form insoluble inclusion bodies (Rosano & Ceccarelli, 2014). This is due to the inability of bacterial cells to carry out post-translational modifications such as, glycosylation, phosphorylation, acetylation, and proteolytic processing, which are critical for the formation of folded, active protein (Wingfield, 2015; Zhang et al., 2004). The SUMO tag has been demonstrated to significantly improve protein stability and solubility when used as an N-terminal fusion (Guerrero et al., 2015; Marblestone et al., 2006). However, the mechanism behind the solubility enhancement by the SUMO tag is unknown. Surprisingly, the presence of the SUMO tag in pCA528 did not enhance the solubility of 6×His-SUMO-gp37, as the protein remained predominantly in the insoluble fraction at both temperatures. The solubility of the protein can be improved by changing expression conditions, such as lowering temperatures (25 - 15 °C), media, host strains, expression vector, and concentrations of the inducer. All these options were implemented, except for lowering the concentration of IPTG. In this study, a high concentration of 1 mM IPTG was used to induce expression of 6×His-SUMO-gp37. It has been observed that reducing the concentration of IPTG improves solubility of the protein. For example, the solubility of recombinant cyclomaltodextrinase (CDase) is notably influenced by the inducer concentration. Specifically, when induced with 0.05 mM IPTG, the protein remains soluble and active. However, doubling the inducer concentration to 0.1 mM results in the expressed protein becoming insoluble and inactive (Turner et al., 2005). This can be explored in future experiments.

Proteins that are soluble in the cytoplasm can be purified under native conditions (Bornhorst & Falke, 2000). In this study, affinity purification under native conditions revealed that 6×His-SUMO-gp37 was mainly in the flow-through, indicating poor binding to the purification column. Consequently, a soluble SUMO-GFP was utilised to confirm the integrity of the purification kit. The SUMO-GFP was successfully purified under both native conditions (cells lysed with lysozyme) and denaturing conditions (cells lysed with Urea), showing high levels of elution of a GFP sized protein, indicating the efficiency of the purification kit. However, even after thorough testing of the purification kit, 6×His-SUMO-gp37 under native conditions could still not be purified. There are several reasons reported in the literature as to why the recombinant protein may not bind to the column, such as the His-tag not being accessible, protein precipitation in the column, and protein aggregation or existing in a multimeric form (Bornhorst & Falke, 2000; QIAexpressionist, 2002; Rosano & Ceccarelli, 2014). Purification

under native conditions may face hindrances if the target protein is insoluble, forms aggregates in inclusion bodies, or possesses a tertiary structure that occludes the polyhistidine affinity tag (Bornhorst & Falke, 2000; Francis & Page, 2010). Consequently, protein purification may require the use of denaturing detergents such as 8 M urea during the process.

Affinity purification under denaturing conditions successfully resulted in the elution of 6×His-SUMO-gp37 with contaminants at a lower molecular weight. The contaminants may be attributed to binding and washing conditions not being stringent enough, contaminations that are associated with tagged proteins, or contamination in the form of truncated protein (Bornhorst & Falke, 2000). Western blot analysis was used to confirm the denatured target protein, and nonspecific bands were observed at higher molecular weights in both the flow-through and the eluates, but not at lower molecular weights. This suggests that nonspecific bands at high molecular weights may be caused by high concentrations of antibodies or may have resulted from denaturing conditions. The use of 8 M urea increases solubility of the protein by disrupting the non-covalent interactions that maintain its native, folded structure, causing the protein to lose its three-dimensional structure and unfold (De Young et al., 1993; Rashid et al., 2005).

Proteins purified under denaturing conditions need to be refolded into their active state by dialysing away the denaturants (Bornhorst & Falke, 2000). In some cases, protein refolding can occur while the protein is bound to the resin (Sinha et al., 1994). At high urea concentration (8 M), proteins are denatured due to the chaotropic effect. In contrast, at low urea concentrations, they tend to stabilise the structure of the target protein and facilitate protein refold (Hevehan & De Bernardez Clark, 1997; Singh & Panda, 2005). In this study, a series of urea concentration ranging from 6 M to 1 M, was used to refold the denatured protein. The wash buffers during purification contained urea (6- 1 M) to wash away nonspecific proteins and simultaneously refold the protein, yielding pure refolded protein. However, the gradual removal of urea from the solubilised protein is required to achieve high refolding efficiency (Okada et al., 2009), which can be time consuming. No protein degradation was observed in the SDS-PAGE analysis in this study, suggesting that the protein was stable. However, due to time constraint, no experiments were conducted to validate the refolded recombinant protein, and no downstream applications were performed, such as dialysis and quantification. Subsequently, the pelleted fraction of 6×His-SUMO-gp37 obtained under native condition was used to conduct a preliminary biological assay.

In conclusion, the 6×His-SUMO-gp37 was efficiently expressed at both 25 °C and 18 °C, and its purification under denaturing conditions was achieved. However, the native form of 6×His-SUMO-gp37 proved to be insoluble, hindering its purification. The next chapter will focus on utilising the insoluble protein pellet to assess the biological activity of 6×His-SUMO-gp37 on *T. leucotreta* larval mortality when combined with CrleGV-SA in a laboratory biological assay.

Chapter 5

Evaluation of 6×His-SUMO-gp37 on *Thaumatotibia leucotreta* larval mortality when combined with CrleGV-SA in laboratory biological assay

5.1 Introduction

In chapter 4, the 6×His-SUMO-gp37 was expressed at different temperatures. However, it was predominantly found in the insoluble fraction. Consequently, the 6×His-SUMO-gp37 in the pellet fraction (cells lysed with lysozyme) was used to conduct the biological assay. The Biological assays are employed to determine the virulence of a virus against a specific host. The aim of this chapter was to evaluate the effect of pelleted 6×His-SUMO-gp37 on *T. leucotreta* larval mortality when combined with CrleGV-SA in laboratory biological assays. Biological assays are employed to determine the virulence of a virus against a specific host. The virulence of a pathogen can be assessed through experimental means, using biological assays to collect data, which can then be used to construct mortality curves (Cory & Bishop, 1997). Comparing these mortality curves allows evaluation of the relative virulence of different treatments and determination of the doses required to cause a specific percentage of the population's mortality. These doses, or concentrations, can vary depending on the method of virus administration; 'dose' refers to the quantity of OBs ingested, whereas 'concentration' signifies the density of OBs administered. The evaluation of biological activity can be quantified differently depending on the chosen method, either as the lethal concentration (LC), lethal dose (LD), or lethal time (LT) (Grzywacz et al., 2004).

In this study, two concentration doses, corresponding to LC_{50} and LC_{90}, were prepared for the biological assays. The LC_{50} and LC_{90} values for CrleGV-SA against *T. leucotreta* have been reported as 4.09×10^3 and 1.18×10^5 OBs/mL, respectively (Moore et al., 2011). Additionally, in a recent study by Taylor (2021), the LC_{50} and LC_{90} values for CrleGV-SA were reported as 1.54×10^4 and 4.10×10^5 OBs/mL, respectively. The LC_{90} represents the concentration level required to kill 90 % of larvae in a given sample, while the LC_{50} indicates the concentration of the virus required to kill 50 % of the larvae in a sample (Moore et al., 2011; Thomas & Elkinton, 2004).

Several methods have been developed to evaluate the biological efficacy of baculoviruses. These methods include droplet feeding, diet incorporation, and surface dose assays, each of which comes with its own set of strengths and weaknesses (Grzywacz et al., 2004). This study used the surface dose method to evaluate the biological activity of 6×His-SUMO-gp37 on *T. leucotreta* larval mortality when combined with CrleGV-SA. This method involves applying the virus to the surface of the diet, and then allowing larvae to feed. It is effective for larvae that burrow into their food source, as they ingest a single dose while tunnelling through the surface. The precise number of OBs consumed by each larva cannot be determined, thus the dose-response must be determined based on concentration rather than the number of OBs ingested. Moreover, this method has been extensively used to test baculoviruses against *T. leucotreta* neonates, establishing a strong foundation for the current study (Jukes, 2018; Moore et al., 2011; Opoku-Debrah et al., 2016; Taylor, 2021). The pelleted fraction transformed with pCA528 (empty vector) was used as a control in combination with CrleGV-SA to ensure that no other factors were affecting the mortality of larvae. To analyse the results, an analysis of variance (ANOVA) was used to study differences between two or more independent group means. This provides evidence of the equality of means between the treatment groups (Steel & Torrie, 1986; Quinn & Keough, 2002).

The aim of this study was to use a surface dose biological assay to evaluate the mortality of *T. leucotreta* neonate larvae exposed to lethal concentrations of CrleGV-SA, both independently and in combination with either bacterial extracts expressing 6×His-SUMO-gp37 or bacterial extracts transformed with pCA528 (empty vector). The specific objectives were: i) to determine the working solutions of the concentration doses of CrleGV-SA, ii) to prepare the virus treatments containing either 6×His-SUMO-gp37 or pCA528 and ddH_2O (control), and iii) to conduct a surface dose bioassay and statistically analyse the results using a one-way ANOVA in R.

5.2 Methods and materials

5.2.1 6×His-SUMO-gp37 and pCA528 (empty vector) preparations

A new induction of 6×His-SUMO-gp37 and pCA528 (empty vector) was conducted at 18 °C for 24 hours under native conditions (cells lysed with lysozyme), following the protocol described in Chapter 4, section 4.2.3. The target protein, 6×His-SUMO-gp37, and the control, pCA528, were obtained by lysing the cells, centrifuging the samples, and collecting the insoluble fractions in pellet form. These pellets were then resuspended in PBS and used as

treatments. These treatments were denoted as bacterial extracts expressing 6×His-SUMO-gp37 and bacterial extracts transformed with pCA528, respectively. The samples were then analysed by SDS-PAGE and Western blotting, as previously described in Chapter 4, sections 4.2.4 and 4.2.5, respectively.

The concentrations of the pellet fractions were determined using a protocol outlined by Desjardins et al. (2009). Measurement of the pellets' concentration was carried out using a Nanodrop® spectrophotometer (Thermo Fisher Scientific, USA), by assessing the absorbance at 280 nm (1 Abs = 1 mg/mL). Initially, 2 µL of PBS served as a blanking solution. Following blanking, the solution was wiped off from both the lower and upper pedestals. Subsequently, 2 µL of either bacterial extracts of 6×His-SUMO-gp37 or bacterial extracts of pCA528 was measured in triplicate, using a fresh aliquot of the sample for each measurement. The estimated concentration of bacterial extracts expressing 6×His-SUMO-gp37 was 7.707 mg/mL, while that of bacterial extracts transformed with pCA528 was 2.373 mg/mL. A final concentration of 100 µg/mL of the pellet fraction was utilised as treatments in biological assays against *T. leucotreta* neonates.

5.2.2 Acquisition of neonates

The *T. leucotreta* eggs were laid on wax paper, cut to fit into the Petri dishes, and washed for 15 seconds in a 1 % (w/v) sodium hypochlorite solution, followed by a rinse in sterile water. The sheets were allowed to dry in a laminar flow cabinet. Afterward, they were placed into petri dishes, sealed with Parafilm, and incubated in the controlled environment (CE) room at 25 °C until the eggs hatched.

5.2.3 Surface-dose biological assay

Dosages with concentrations of 2.96×10^4 and 2.96×10^5 OBs/mL were prepared and used to conduct a surface biological assay, each in triplicate. Each assay was carried out using neonate *T. leucotreta* larvae that were reared on an artificial diet. The artificial diet was prepared by baking 250 g of diet, as described by Moore et al. (2014), mixed with 250 mL of ddH_2O at 200 °C for 30 minutes in a flat baking dish. Plugs of the diet were transferred into 24-well plates and compressed with a syringe plunger. Each experiment consisted of four 24 well plates, with each plate inoculated with one of the four treatments: CrleGV-SA alone, CrleGV-SA in combination with the bacterial extracts expressing 6×His-SUMO-gp37 (100 µg/mL), CrleGV-SA in combination with the pelleted bacteria transformed with pCA528 (100 µg/mL), and ddH_2O as a control. Experiments were performed in triplicate with two different CrleGV-SA

concentrations tested: a low dose of 2.96×10^4 OBs/mL and a high dose of 2.96×10^5 OBs/mL, as described in Table 5.1. Each well received 100 µL of the respective treatment per plate, and plates were left with the lids off to allow the wells to dry before transferring a single neonate larva to each inoculated well. Once all four plates had been prepared, neonate larvae were left to feed on the treated diet for 7 days in a CE room maintained at 25 °C. After 7 days, diet plugs in each well were inspected for surviving larvae or cadavers, and the number of dead larvae was recorded for each treatment.

Table 5.1. List of the four treatments used for biological assays at two concentrations of CrleGV-SA OBs. The descriptions of the treatments were abbreviated as shown under the two concentration doses of CrleGV-SA.

Descriptions	**Treatments**	
	CrleGV-SA Low Dose (2.96×10^4 OBs/mL)	**CrleGV-SA High Dose (2.96×10^5 OBs/mL)**
Bacterial extracts expressing 6×His-SUMO-gp37 (100 µg/mL) + CrleGV-SA	CrleGV-SA L + gp37	CrleGV-SA H + gp37
Bacterial extracts transformed with pCA528 (100 µg/mL) + CrleGV-SA	CrleGV-SA L + pCA528	CrleGV-SA H + pCA528
CrleGV-SA alone	CrleGV-SA L	CrleGV-SA H
ddH_2O (Control)	Water	Water

5.2.4 Statistical analysis

The data obtained from each bioassay were analysed using RStudio version 1.2.1335 © 2009-2019, and percentage mortality was compared using a one-way ANOVA, following the example provided (https://www.scribbr.com/statistics/anova-in-r/). Conducting ANOVA in R involves loading the data, performing the ANOVA test, conducting a post-hoc test, plotting the results on a graph, and reporting the outcomes.

5.3 Results

5.3.1 Solubility analysis of 6×His-SUMO-gp37 and pCA528 (empty vector)

Fresh solubility analyses of both 6×His-SUMO-gp37 and pCA528 (empty vector) were conducted under native conditions (cells lysed with lysozyme) in preparation for the biological assay (Figure 5.1). The results distinguished the presence of 6×His-SUMO-gp37 in the recombinant vector from that in the empty vector. In solubility analysis (S1), 6×His-SUMO-gp37 was predominantly in the pellet fraction (lane P) compared to the supernatant fraction (lane S) at approximately 42 kDa. Lane W shows the whole lysate (total protein) of 6×His-SUMO-gp37. Western blotting analysis using anti-His antibodies showed that bands detected in lane W and P are 6×His-SUMO-gp37. In solubility analysis (S2), no His-tag protein was detected in the pCA528. This was confirmed by Western blot analysis using anti-His antibodies, which showed the absence of the His-tagged protein. The concentration of the pellet fraction of 6×His-SUMO-gp37 was 7.707 mg/mL with a standard deviation of 0.348 mg/mL, while the concentration of the pellet fraction of pCA528 (empty vector) was 2.373 mg/mL with a standard deviation of 0.027 mg/mL.

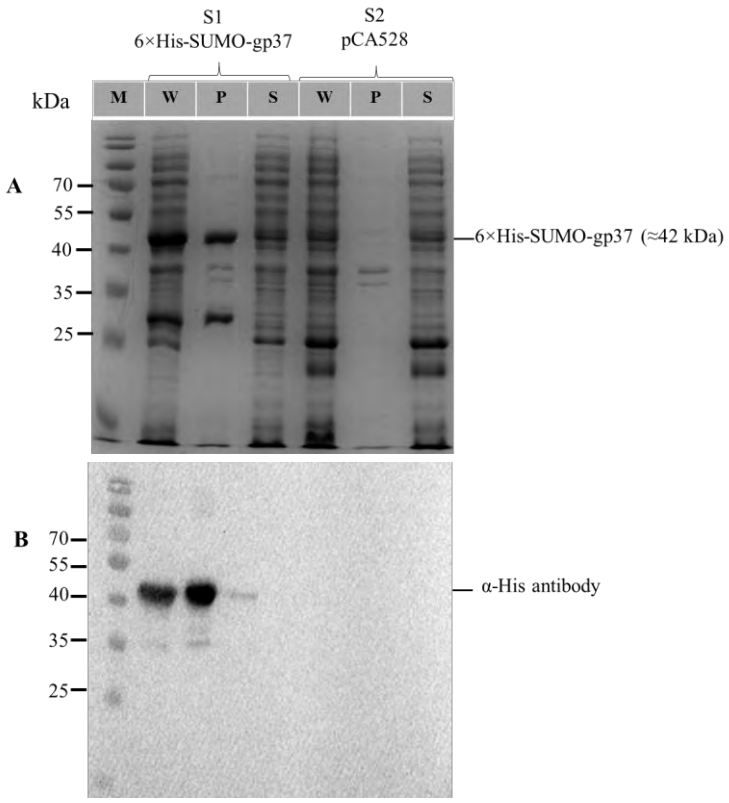


Figure 5.1. Solubility analysis of 6×His-SUMO-gp37 and pCA528. Panel A shows SDS-PAGE analysis, and Panel B shows Western blot analysis, respectively. S1 indicates the solubility

analysis of 6×His-SUMO-gp37, while S2 indicates the solubility analysis of pCA528 (empty vector). Lane M represents the protein marker in kilodaltons (kDa), lane W represents the whole lysate (total protein), lane P represents the pellet fraction (insoluble after centrifugation), and lane S represents the supernatant fraction (soluble after centrifugation).

5.3.2 Biological activity of CrleGV-SA in combination with 6×His-SUMO-gp37 against *T. leucotreta* neonates

The bacterial extracts expressing 6×His-SUMO-gp37 was used as an additive to CrleGV-SA to evaluate its biological effect on larval mortality in laboratory bioassays. Three replicates were performed for the two dosages of CrleGV-SA, each in combination with bacterial extracts expressing 6×His-SUMO-gp37 and bacterial extracts transformed with pCA528. Additionally, ddH_2O was used separately as a control. The Table below shows the average percentage of larval mortality for each treatment against *T. leucotreta* (Table 5.2). A 4.5 % difference was observed between CrleGV-SA L + gp37 and CrleGV-SA L. A 3.9 % difference was also observed between CrleGV-SA L + pCA528 and CrleGV-SA L. A 5.6 % difference was observed between CrleGV-SA H + gp37 and CrleGV-SA H. The same outcomes were observed for CrleGV-SA H + pCA528. No difference was observed between CrleGV-SA H + gp37 and CrleGV-SA H + pCA528. The average percentage mortality of all three replicates is presented in Table 5.2.

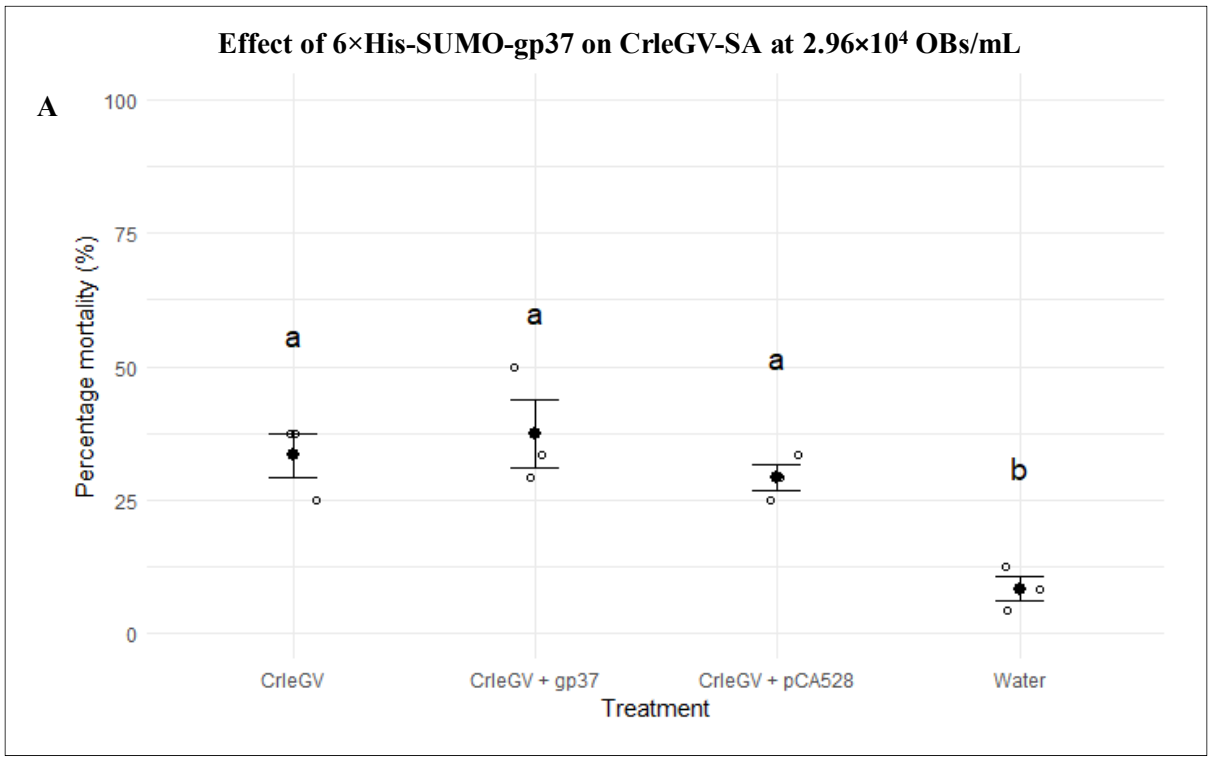
Table 5.2. The average percentage mortality and standard deviation (SD) for each dose-response biological assay replicate.

	Low Dose (2.96×10^4 OBs/mL)		High Dose (2.96×10^5 OBs/mL)	
Treatments	Mortality (%)	SD	Mortality (%)	SD
Bacterial extracts expressing 6×His-SUMO-gp37 + CrleGV-SA	37.5 %	11.03	88.9 %	2.40
Bacterial extracts transformed with pCA528 + CrleGV-SA	29.1 %	4.17	88.9 %	4.81

CrleGV-SA alone	33 %	7.22	83.3 %	7.22
ddH_2O	8.3 %	4.17	9.2 %	2.41

5.3.3 One-way ANOVA statistical analysis

A one-way ANOVA statistical analysis was conducted using R to compare the mortality observed between treatments for each dose against *T. leucotreta* neonates (Figure 5.2). A statistical difference was observed between treatments in experiments conducted with both low ($F(3) = 9.667$, $p = 0.00489$) and high ($F(3) = 210.3$, $p = 6.02^{08}$) concentrations of CrleGV-SA OBs. A Tukey HSD post-hoc test was conducted to evaluate differences between treatments for each tested concentration. No significant differences ($p > 0.05$) were observed between treatments with CrleGV-SA combined with bacterial extracts expressing 6×His-SUMO-gp37, CrleGV-SA combined with bacterial extracts transformed with pCA528, and CrleGV-SA alone at both lower and higher CrleGV-SA concentrations (Figure 5.2A and B). A significant difference ($p < 0.05$) was observed between virus treatments and the control (water only) treatment at both CrleGV-SA concentrations (Figure 5.2A and B).



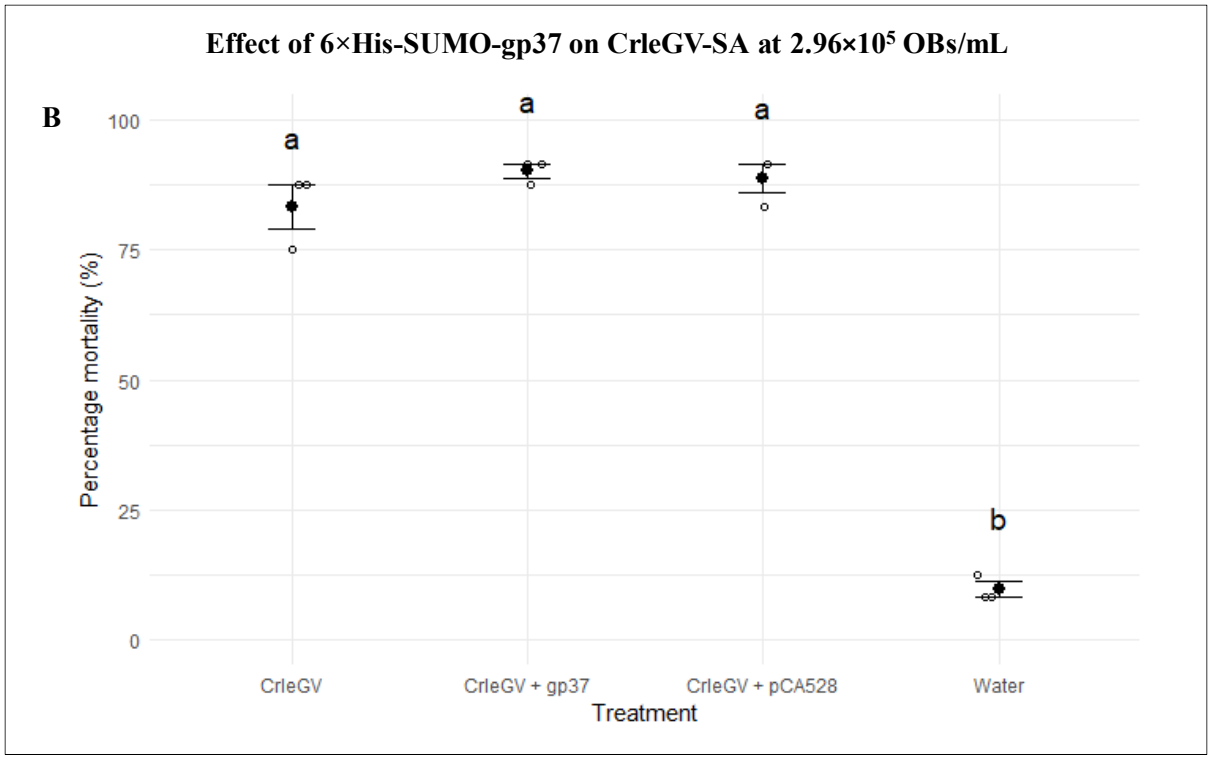


Figure 5.2. Evaluation of the effect of 6×His-SUMO-gp37 on CrleGV-SA against *T. leucotreta* A) low (2.96×10^4 OBs/mL) and B) High (2.96×10^5 OBs/mL) concentrations. Treatments include CrleGV-SA alone (CrleGV), CrleGV-SA combined with bacterial extracts expressing 6×His-SUMO-gp37 (CrleGV + gp37), CrleGV-SA combined with bacterial extracts transformed with the pCA528 vector (CrleGV + pCA528), and ddH_2O control (water). Different letters (a and b) indicate significant differences between treatments.

5.4 Discussion

This chapter focused on the relative virulence of a mixed infection of CrleGV-SA and pelleted bacteria expressing 6×His-SUMO-gp37 on *T. leucotreta* larvae in laboratory bioassays. The working solutions of CrleGV-SA used for the bioassay were 2.96×10^4 OBs/mL and 2.96×10^5 OBs/mL. A preliminary experiment was conducted to validate the CrleGV-SA concentration doses. The results showed that the low and high doses killed approximately 40 % and 90 % of the *T. leucotreta* larval population, respectively (results not shown). Consequently, when a combination of treatments is applied, an increase or decrease in mortality can be observed.

The target protein was obtained by lysing cells with lysozyme, centrifuging the sample, and collecting the pelleted insoluble fraction. The bacterial extracts expressing 6×His-SUMO-gp37 were subsequently used in combination with CrleGV-SA. A final concentration of 100 µg/mL of 6×His-SUMO-gp37 was used to conduct bioassays. In another study, bioassays revealed that

the inclusion of 30.0 µg/mL of purified gp37 in the inoculum significantly enhance the infectivity of SeMNPV and AcMNPV (Liu et al., 2011). Since the cell lysate of 6×His-SUMO-gp37 was not purified, and no accurate quantification of the recombinant protein was conducted, the final concentration of 6×His-SUMO-gp37 was increased to 100 µg/mL compared to the study conducted by Liu et al. (2011). To obtain the control, cells were lysed with lysozyme, centrifuged, and the insoluble fractions in the pelleted form were collected. The bacterial extracts transformed with pCA528 (empty vector) were used as a control, and a concentration of 100 µg/mL was applied to ensure that no other factors were affecting the function of the recombinant protein. Additionally, a ddH_2O control was used to ensure that no other external factors were influencing the mortality of *T. leucotreta* neonates.

The final objective was to conduct surface bioassays using the prepared virus treatments on *T. leucotreta* and statistically analyse the results using R. A biological assay was conducted, wherein the virus was administered using a surface dose method. This method was chosen because it resembles the way the virus would be administered in the field and how the larvae likely feed on the viruses in the field (Moore et al., 2011). The larvae burrow into the fruit without causing visible damage, feeding inside throughout their larval phase. To introduce viruses to the larvae, applying OBs to the fruit before larval entry is the most practical method. This allows larvae to inadvertently consume the OBs while burrowing, making the surface dose protocol the most comparable field method for virus contact. In South Africa's citrus industry, where the *T. leucotreta* pest is a concealed threat, neonate larvae serve as the most accessible stage for field targeting. Therefore, they are frequently used to evaluate the biological activity of CrleGV in controlling the pest (Moore et al., 2011; Opoku-Debrah et al., 2016).

To analyse larval mortality in each treatment, the average larval mortalities were subjected to a one-way ANOVA test. This statistical technique determines if treatment means significantly differ from the overall treatment by comparing the variance within each treatment to the overall variance of the entire dataset (Bevans, 2023). This study revealed a statistical difference between treatments in experiments conducted with both low and high concentrations of CrleGV-SA. A p-value below 0.05 indicates differences among the treatments, while a p-value above 0.05 indicates no difference among the treatments. A Turkey multiple comparison of the means revealed that no significant differences were observed between CrleGV-SA in combination with either bacterial extracts expressing 6×His-SUMO-gp37, or bacterial extracts transformed with pCA528 at both low and high concentrations. However, there was a significant difference between the virus treatments and ddH_2O control at both concentrations.

This difference was expected, as no virus was present in the ddH_2O treatment, allowing the neonates to infest the diet without restriction. The mortality observed in the no virus treatment may have resulted from factors such as colony health and the way the neonates were handled. The result obtained indicate that the bacterial extracts expressing 6×His-SUMO-gp37 as an additive to CrleGV-SA do not affect the mortality of *T. leucotreta* neonate larvae when compared to the virus alone.

The function of proteins can be influenced by various factors, including bacterial inclusion bodies, incomplete folding, protein tags (Rosano & Ceccarelli, 2014) and protein concentration. Inclusion bodies, which are accumulations of protein aggregates observed in recombinant *E. coli,* results from an imbalance between protein aggregation and solubilisation (García-Fruitós et al., 2012). While inclusion bodies are functional; their biological activity relies on proper protein folding and a native or native-like secondary structure of the embedded protein (García-Fruitós et al., 2005, 2012). In this study, pelleted bacteria expressing the recombinant protein were used to conduct biological assays. 6×His-SUMO-gp37 was contaminated with bacterial cell debris, which can compromise the biological activity of the protein (García-Fruitós et al., 2005) and make it difficult to determine the total material of 6×His-SUMO-gp37. However, it is unknown if the bacterial cell debris had an effect on 6×His-SUMO-gp37. Incomplete protein folding leads to protein inactivity (Oliveira & Domingues, 2018). In such cases, the protein may attain a stable conformation, but the specific structure of the active site remains unsuitable for proper functionality (Gonzalez-Montalban et al., 2007; Martínez-Alonso et al., 2008). For structural or biochemical studies on a recombinant protein, it is necessary to remove the peptide tags (Oliveira & Domingues, 2018; Perron-Savard et al., 2005). This elimination is crucial as peptide tags can potentially interfere with both the activity and structure of the protein (Guerrero et al., 2015; Perron-Savard et al., 2005; Wu & Filutowicz, 1999). The activity of 6×His-SUMO-gp37 might have been affected by the concentration used, which could have been either too high or too low. The estimated concentration might have been invalid since no accurate quantification was conducted. It's important to note that the Nanodrop spectrophotometer was used to estimate the concentration of 6×His-SUMO-gp37. This device measures absorbance primarily from tryptophan, tyrosine, phenylalanine residues, or cysteine-cysteine disulphide bonds and makes estimations based on the average frequency of these amino acids in the mixtures. However, estimation of protein concentrations using Nanodrop spectrophotometer is recommended for purified proteins, not complex mixtures of macromolecules (Desjardins et al., 2009). Ongoing efforts are needed to develop and optimise

the expression system to enhance protein solubility. Solubility is considered a key indicator of successful protein conformation and functional quality at a molecular level (Gonzalez-Montalban et al., 2007), as soluble protein can be purified and dialysed to remove any external factors that might contaminate the protein. This makes it easier to safely conclude on the effect caused by the protein and not by any *E. coli* debris or expression components. Moreover, soluble purified protein is necessary for accurate quantification, enabling comprehensive bioassays with different protein: virus ratios to be conducted effectively.

In conclusion, this chapter reported on the biological activity of CrleGV-SA when applied in combination with the bacterial extracts expressing 6×His-SUMO-gp37 on *T. leucotreta* neonates. No improvement in virulence was observed when CrleGV-SA, in combination with the bacterial extracts expressing 6×His-SUMO-gp37, was applied compared to CrleGV-SA alone in both concentrations. More research is required to effectively determine whether there is a synergistic, antagonistic or any other effect.

Chapter 6

General discussion

6.1 Thesis overview

Thaumatotibia leucotreta is a serious pest endemic to Africa, posing a significant threat to the export of fresh citrus in South Africa, while also causing extensive damage to citrus crops (Moore, 2021). Classified as a phytosanitary risk by several South African export markets (Hattingh et al., 2020), this pest demands effective control measures. Two baculoviruses, CrleGV-SA and CrpeNPV, infect *T. leucotreta*, leading to larval mortality (Moore & Jukes, 2023). Both viruses are used in integrated pest management programmes to reduce fruit damage in agricultural fields. Notably, CrleGV-SA has been used against *T. leucotreta* for nearly 20 years in South Africa, demonstrating its importance in pest control strategies (Moore et al., 2015; van der Merwe et al., 2017). However, the efficacy of these control options is hindered by factors such as virulence and the slow speed of kill. To address these limitations, there is potential in exploring synergistic relationships between baculoviruses that infect the same host. Several infectivity proteins, including enhancin and gp37, have shown promise as additives to enhance the infection of baculoviruses, paving the way for more effective pest management strategies. For example, TnGV enhancin was effectively expressed in a baculovirus system, purified using gel filtration and ion exchange chromatography. Bioassay results revealed that enhancin facilitated NPV infections by digesting the peritrophic membrane (PM) and increased the amount of the virus's contacts with susceptible tissue (Lepore et al., 1996). In a recent study by Ricarte-Bermejo et al. (2021), TnGV enhancin was expressed in a baculovirus system, and the soluble fraction of infected cells containing enhancin was used for bioassay, without further purification. Combining AcMNPV OBs with the expressed protein resulted in a significant improvement in lethal concentrations against *Spodoptera exigua* (Ricarte-Bermejo et al., 2021). In a study by Liu et al. (2000), HearGV *enhancin* was cloned and expressed in bacterial cells. The expressed enhancin product exhibited promising results in a preliminary bioassay, enhancing larval mortality infected with HaNPV by 31.7 % to 34.1 % over 7 days post-infection. Moreover, the LT_{50} decreased by at least 1.5 to 2.1 days, indicating the potential insecticidal activity of the expressed enhancin (Liu et al., 2000). Furthermore, the expressed and purified truncated CpGV gp37 was shown to enhance the infectivity of NPVs on *S. exigua* larvae using droplet feeding bioassays (Liu et al., 2011). The

presence of 30.0 µg/mL of gp37 significantly enhanced the infectivity of SeMNPV, as evidenced by a remarkable reduction in the LC_{50} value. Additionally, 30.0 µg/mL of gp37 significantly enhanced the infectivity of AcMNPV against third-instar larvae of *S. exigua* (Liu et al., 2011). Further studies revealed that CpGV gp37 was capable of altering the PM structure of *S. exigua* (Liu et al., 2019). The above observations prompted the current study, which investigated the expression and evaluation of CrpeNPV gp37 as a formulation additive for enhanced infectivity with CrleGV-SA and improved *T. leucotreta* control. The genomic DNA of CrpeNPV was extracted from OBs, and oligonucleotides targeting the CrpeNPV *gp37* gene were designed for PCR amplification in preparation for cloning (Chapter 2), followed by cloning *gp37* amplicon into the intermediate vector, pJET1.2/blunt, and subcloning the *gp37* product into the bacterial expression vector, pCA528. Verification of the recombinant constructs was conducted through colony PCR, plasmid extraction, restriction enzyme analysis, and Sanger sequencing (Chapter 3). The recombinant protein was expressed and purified using Nickel column affinity chromatography and analysed using SDS-PAGE and Western blotting (Chapter 4). The expressed recombinant protein was used as an additive to determine its effect on the biological activity of CrleGV-SA against neonate *T. leucotreta* larvae in biological assays.

6.2 Genome analysis and cloning of CrpeNPV *gp37*

South African baculoviruses provide a rich bioresource library, with the CrpeNPV isolate containing the *gp37* gene. The genome of CrpeNPV has undergone genetic characterisation, and the full genome sequence is available from GenBank (Accession no: NC_055500) (Marsberg et al., 2018). In this study, CrpeNPV DNA was extracted from OBs using a Quick-DNA Mini Prep Plus kit (Zymo Research, CA, USA), with a protocol described by Mela (2022). This protocol effectively dissolves the polyhedron matrix and neutralises the pH before using the kit. The kit allows for the easy extraction of high-yield total DNA and has been used to extract gDNA from different baculoviruses (Bennett, 2022; Mela, 2022; Taylor, 2021). In this study, unique oligonucleotides were designed to specifically target the CrpeNPV *gp37* gene in the genome. The *gp37* gene sequence was then amplified using PCR, and the resulting amplicon was gel purified in preparation for cloning. To clone the purified *gp37* product, the intermediate vector, pJET1.2/blunt was used, followed by subcloning into the expression vector pCA528. The pCA528 expression vector was selected and transformed into Rosetta (DE3) *E. coli* cells as the host organism. The vector was chosen for its strength, cell specificity, inducibility, compatibility, regulation, and size. It contains a T7 promoter, a multiple cloning

site, an N-terminal 6×His-SUMO tag, and a kanamycin resistance gene, all essential for expressing the recombinant protein (Kuo et al., 2014; Marblestone et al., 2006). The plasmid is compatible with Rosetta (DE3) host cells. It contains an inducible promoter that can be activated by IPTG to control the level of gene expression (Rosano & Ceccarelli, 2014). The plasmid can incorporate the target gene, and the selected restriction enzymes are compatible with the plasmid. After cloning, the recombinant plasmid was verified through colony PCR, plasmid extraction, and restriction enzyme analysis before being sent for sequencing. Analysis of the sequenced recombinant plasmid revealed that no SNPs or indels were detected between the start and stop codons, confirming that the *gp37* gene was in frame with the 6×His and SUMO tag, displaying 100 % identity with the reference sequence of pCA-gp37. These verifications of the recombinant plasmid were conducted to enable protein expression.

6.3 The expression and purification of the recombinant protein, CrpeNPV gp37

Bacterial expression systems are commonly used for production of heterologous gene products from both eukaryotic and prokaryotic cells (Glick & Whitney, 1987). The termed 'heterologous protein production' refers to the expression of recombinant proteins in foreign cells. *Escherichia coli* is recognised as a leading host for producing recombinant proteins (Rosano & Ceccarelli, 2014). In this study, the recombinant protein of CrpeNPV gp37 was expressed in Rosetta (DE3) *E. coli* cells, building upon research conducted by Liu et al. (2011), who expressed a truncated CpGV gp37 in BL21 *E. coli* cells. Rosetta host strains, BL21 derivatives, are specifically engineered to facilitate the expression of eukaryotic proteins containing rarely used codons within *E. coli* (Baca & Hol, 2000; Kane, 1995). The sequence analysis of the CrpeNPV *gp37* gene revealed the presence of these rare codons namely, AGG, AGA, CUA, CCC, and GGA. Since CrpeNPV gp37 is a eukaryotic protein expressed by a prokaryotic cell, the presence of numerous rare codons in the foreign mRNA within *E. coli* can lead to the depletion of low abundance tRNAs. Consequently, this depletion may result in amino acid misincorporation, impacting both the level and quality of heterologous protein expression (Rosano & Ceccarelli, 2009). The Rosetta (DE3) *E. coli* strain contains a plasmid that supplies these rare codons. Using *E. coli* host strains as an expression system, come with several advantages, such as unparalleled fast-growing kinetics (Sezonov et al., 2007). However, it's important to note that the expression of a recombinant protein may impose a metabolic burden on the microorganism, potentially leading to a significant decrease in generation time (Bentley et al., 1990). In this study, the expression of 6×His-SUMO-gp37 was abundant at different time

points at both 25 °C and 18 °C, and no challenges were encountered in its expression. Nevertheless, along with advantages comes drawbacks, as the expression of heterologous proteins in *E. coli* often results in the accumulation of insoluble proteins known as inclusion bodies (Wingfield, 2015). Solubility analysis of 6×His-SUMO-gp37 revealed that the protein is soluble at 18 °C but predominantly present in the insoluble fraction at both 25 °C and 18 °C. The solubility of the protein can be enhanced by modifying expression conditions, including temperature (15 - 25 °C), IPTG concentrations (1 - 0.05 mM) (Turner et al., 2005), media, host strains, and the use of fusion tags such as SUMO (Kuo et al., 2014; QIAexpressionist, 2002). All these options were explored in this study, except for reducing the concentration of IPTG. A concentration of 1 mM of IPTG was used, following the recommended guidelines (QIAexpressionist, 2002). In a study by Liu et al. (2011), CpGV gp37 was expressed at 28 °C, and no solubility analysis was reported. Low solubility or the formation of aggregates can result from the non-covalent association of two or more polypeptide chains, which may or not retain their native fold (Murphy & Roberts, 2013). Another reason for the insolubility of 6×His-SUMO-gp37 could be the bacterial cells' inability to carry out post-translational modifications, such as glycosylation, phosphorylation, acetylation, and proteolytic processing, which are critical for the formation of folded, active protein (Wingfield, 2015; Zhang et al., 2004), or the limited ability to carry out extensive disulfide bond formation (Fakruddin et al., 2013). Additionally, gp37 is a glycoprotein (Gross et al., 1993). Sequence analysis reveals that it has a N-terminal signal peptide and a transmembrane domain. In the study conducted by Liu et al. (2011), CpGV gp37 was truncated to remove the signal sequences and membrane helices at the N-terminus before cloning. In this study, the full-length CrpeNPV *gp37* was cloned in preparation for protein expression. Perhaps the presence of the membrane domain and signalling sequence played a role in the formation of insoluble 6×His-SUMO-gp37, suggesting their removal in future experiments. The solubility of the protein is crucial for purifying recombinant proteins, as it profoundly impacts protein function and downstream applications, including quantification, stability, and storage (Jamrichová et al., 2017; Olson, 2016; Senisterra et al., 2012). These factors bear significant importance in bioassays, where the precise quantification of recombinant protein is essential to determine the concentration required for the bioassays. Numerous studies have demonstrated the necessity of using purified soluble proteins for downstream experiments, particularly bioassays (Lepore et al., 1996; Liu et al., 2011). In this study, efforts were made to purify 6×His-SUMO-gp37 under native conditions. However, attempts to purify 6×His-SUMO-gp37 under native conditions were unsuccessful. This failure could be attributed to the observation of less soluble protein in the soluble fraction,

prompting purification under native conditions. The study by Liu et al. (2011) indicated the purification of CpGV gp37, though it remains unclear whether it was achieved under native or denaturing conditions.

Purification of 6×His-SUMO-gp37 under denaturing conditions proved successful, and the recombinant protein was confirmed via Western blot analysis. However, denatured proteins must be refolded into their native state (Bornhorst & Falke, 2000). Refolding attempts were made using a series of urea concentrations ranging from 6 M to 1 M. Urea, acting as a chaotropic agent, can denature proteins at high concentrations (8 M urea) (Rudolph & Lilie, 1996) and refold proteins at lower concentrations (Singh & Panda, 2005). The wash buffers during purification contained urea (6 - 1 M) to eliminate nonspecific proteins and concurrently refold the protein, yielding pure refolded protein. However, achieving high refolding efficiency necessitates the gradual removal of urea from the solubilised protein, which can be time consuming (Okada et al., 2009). Contaminants were observed in the lower molecular weight range of the eluates, undetected by Western blot analysis, indicating nonspecific bands that also adhere to purification columns. Impurities may persist after His-tag purification due to proteins' natural affinity for nickel used in the purification column, leading to nonspecific binding. Additionally, overexpression of the target protein can sometimes result in the co-purification of other proteins that form complexes with the target (Andersen et al., 2013). No further experiments were conducted to validate the refolded recombinant protein. Numerous reviews have focused on protein purification and structural characterisation (Manta et al., 2011; Saraswat et al., 2013; Yadav et al., 2016), proposing workflows for quality control (Lebendiker et al., 2014; Raynal et al., 2014) applicable to CrpeNPV gp37 expression, offering guidance in the production of soluble and reliable recombinant proteins. Continual optimisation of solubility analysis of the recombinant protein is crucial as soluble protein can be purified, dialysed and quantified, making it easier to safely conclude on the effect caused by the protein and not by any *E. coli* debris or expression components. Additionally, accurate quantification is essential for formulating the protein as an additive.

6.4 Biological assay using 6×His-SUMO-gp37

In this study, a surface dose bioassay was conducted using bacterial extracts expressing 6×His-SUMO-gp37. Due to time constraints, the biological assays were carried out as a small-scale experiment, using both low and high concentrations of CrleGV-SA in combination with 6×His-SUMO-gp37. A preliminary experiment was conducted to validate the CrleGV-SA

concentration doses. The results showed that the low and high doses killed 40 % and 90 % of the *T. leucotreta* population larvae over 7 days, respectively. Consequently, when a combination of virus treatments is applied, an increase or decrease in mortality can be observed. These two doses facilitated the straightforward evaluation of 6×His-SUMO-gp37 on the biological activity of CrleGV-SA. The bioassay results showed that the combination of 100 µg/mL of bacterial extracts expressing 6×His-SUMO-gp37 with CrleGV-SA had no effect on *T. leucotreta* larval mortality compared to CrleGV-SA alone. It is important to note that the protein used as an additive is from the bacterial extracts expressing 6×His-SUMO-gp37, not the purified recombinant protein of CrpeNPV gp37. This was due to the inability to purify the recombinant protein under native conditions. Therefore, CrpeNPV gp37 was contaminated with pelleted bacterial debris and consisted of both 6×His and SUMO tag, all of which can interfere with the structure and function of the protein (García-Fruitós et al., 2005; Yadav et al., 2016). A control was also conducted in which CrleGV-SA was used in combination with bacterial extracts transformed with the empty vector, which also didn't show any significant effect on *T. leucotreta* larval mortality compared to CrleGV-SA alone. This control only validates that the bacterial contaminants and the 6×His and SUMO tag do not have any effect on *T. leucotreta* larval mortality. However, it does not verify their effect on the biological activity of CrpeNPV gp37. It has been reported that large tags can interfere with the structure and function of the protein (Oliveira & Domingues, 2018; Perron-Savard et al., 2005). The recombinant protein contains a SUMO tag (11.1 kDa), suggesting its removal in future experiments. It is important to note that, due to time constraints, this was a preliminary study. Despite having a solid design, adequate controls, and analysing the data as thoroughly as possible, a more comprehensive bioassay is necessary to generate a dose and time response curve. This will help determine whether there is a synergistic, antagonistic, or any other effect.

6.5 Potential future research

Although CrpeNPV gp37 was successfully expressed in a bacterial system, future work is required in order to investigate the effects of the protein when used as an additive to CrleGV in biological assays against *T. leucotreta* larvae. For example, continuous development and optimisation of the expression system is necessary to improve protein solubility, enabling downstream experiments such as protein purification, quantification, and the evaluation of its effect on the biological activity of baculovirus-based biopesticides. This can also be achieved by optimising the expression system and purifying the recombinant protein under denaturing conditions. In this study, the recombinant protein was successfully purified under denaturing

conditions as shown in chapter 4, section 4.3.3. Attempts were made to refold the denatured recombinant protein; however, no further experiments, such as dialysis, quantification and testing the biological activity of the protein in biological assays, were conducted due to time constraints. Future work could include exploring the purification of CrpeNPV gp37 under denaturing conditions with an appropriate protocol and a refolding kit.

Soluble protein can also be achieved through codon optimisation of the gene before cloning and expression. This strategy involves modifying the DNA sequence in the codon region, while keeping the encoded protein amino acid sequence unchanged (Itkonen et al., 2014). Codon optimisation has been used to enhance protein expression and improve solubility by increasing the translational efficiency of a gene of interest, accommodating the codon bias of the host organism (Fei et al., 2015; Gao et al., 2013; Liu et al., 2013). In this study, this approach was not performed due to the expression vector and host strains used, which contain essential components such as the SUMO tag and a plasmid (pRARE) that supplies tRNA rare codons, crucial for improving protein expression and solubility (Kuo et al., 2014; Novy et al., 2001). However, for future experiments, codon optimisation can be a valuable strategy to enhance protein solubility during recombinant expression.

Other expression systems, such as baculovirus vectors and yeast-based systems, can be used to express and purify CrpeNPV gp37. A baculovirus expression vector is well-documented for heterologous expression of proteins in insect cells (Luckow, 1990; Miller, 1988; O'Reilly et al., 1994), and can potentially be used in the future to express CrpeNPV gp37. In a recombinant AcMNPV isolate expressing the TnGV *enhancin 3* gene, a combination of AcMNPV OBs with the expressed protein resulted in ~3- and ~4.7-fold improvements in lethal concentrations against second and fourth instar *S. exigua* (Ricarte-Bermejo et al., 2021), proving baculovirus vectors to be effective tools in producing soluble proteins.

The yeast-based system used to express recombinant proteins is a valuable tool, providing benefits such as essential post-translational modifications (glycosylation, phosphorylation, acetylation, proteolytic processing), crucial for ensuring proper protein folding, stability, and functionality (Vieira Gomes et al., 2018). This has never been tested on baculovirus proteins; however, is has been shown to be effective in expressing and purifying eukaryotic proteins. The expression of recombinant S and preS2-S of ayw2 HBV subtype was successful in both *S. cerevisiae* and *P. pastoris*, but higher yields were achieved using the inducible Gal system in *S. cerevisiae*. This suggests that co-expression of *S* or *preS2-S* genes under inducible and

constitutive promoters in yeast cells leads to significant increase in the expression levels of the corresponding proteins (Hadiji-Abbes et al., 2009). In the purification process, the protein can be obtained by incorporating an export signal along with the 6×His tag. This modification directs the proteins to the medium, facilitating their purification through Ni-NTA affinity chromatography (Abelson et al., 2004; QIAexpressionist, 2002). The yeast-based system is a promising system capable of expressing eukaryotic proteins. These expression systems can be used not only for CrpeNPV gp37 but also for other viral enhancing proteins, such as enhancin, ODV-E66 or other gp37 encoded by other baculoviruses such as HearNPV and CpGV.

6.6 Conclusion

The primary aim of this study was to express CrpeNPV gp37 in a bacterial system and then evaluate its effect on *T. leucotreta* larval mortality when combined with CrleGV-SA in laboratory bioassays. A surface dose biological assay was conducted, which showed that the bacterial extracts expressing 6×His-SUMO-gp37, when combined with CrleGV-SA, had no effect on *T. leucotreta* larval mortality compared to CrleGV-SA alone. Even though no effect was observed, *gp37* was successfully cloned, and the full-length protein was expressed abundantly on a small scale. This marks an important first step in evaluating it as an additive. The solubility of the protein was further analysed, a task never undertaken before for CrpeNPV gp37. This provides new knowledge about CrpeNPV gp37 and establishes a platform for further experiments aimed at solubilising it or expressing it in a different system. The main objectives of this study were achieved; however, further work may be needed, particularly in protein expression and purification, to enable effective downstream experiments such as quantification and stability, which are important when conducting biological assays. This study provides insightful information on the production of CrpeNPV gp37, useful for investigating the characteristics of gp37 and its interaction with CrleGV-SA. The ultimate goal is to develop an effective and low-cost formulation for controlling citrus pests.

References

Abelson, J. N., Simon, M. I., Guthrie, C., & Fink, G. R. (2004). *Guide to yeast genetics and molecular biology*. Gulf Professional Publishing.

Andersen, K. R., Leksa, N. C., & Schwartz, T. U. (2013). Optimized E. coli expression strain LOBSTR eliminates common contaminants from His-tag purification. *Proteins: Structure, Function, and Bioinformatics*, *81*(11), 1857–1861.

Angélini, A., Amargier, A., Vandamme, P., & Duthoit, J. L. (1965). Une virose á granules chez le lepidoptére *Argyroploce leucotreta*. Coton Fibres Trop. 20, 277-282.

Asser-Kaiser, S., Fritsch, E., Undorf-Spahn, K., Kienzle, J., Eberle, K. E., Gund, N. A., Reineke, A., Zebitz, C. P. W., Heckel, D. G., & Huber, J. (2007). Rapid emergence of baculovirus resistance in codling moth due to dominant, sex-linked inheritance. *Science*, *317*(5846), 1916–1918.

Baca, A. M., & Hol, W. G. J. (2000). Overcoming codon bias: a method for high-level overexpression of Plasmodium and other AT-rich parasite genes in Escherichia coli. *International Journal for Parasitology*, *30*(2), 113–118.

Barrera, G. P., Villamizar, L. F., Araque, G. A., Gómez, J. A., Guevara, E. J., Cerrudo, C. S., & Belaich, M. N. (2021). Natural coinfection between novel species of baculoviruses in spodoptera ornithogalli larvae. *Viruses*, *13*(12), 2520.

Beas-Catena, A., Sánchez-Mirón, A., García-Camacho, F., Contreras-Gómez, A., & Molina-Grima, E. (2014). Baculovirus biopesticides: an overview. *JAPS: Journal of Animal & Plant Sciences*, *24*(2), 362–373.

Behjati, S., & Tarpey, P. S. (2013). What is next generation sequencing? *Archives of Disease in Childhood-Education and Practice*, *98*(6), 236–238.

Benchling, I. B. (2023). *Biology software*. Retrieved from https://benchling.com.

Bennett, P. M. (2008). Plasmid encoded antibiotic resistance: acquisition and transfer of antibiotic resistance genes in bacteria. *British Journal of Pharmacology*, *153*(S1), S347–S357.

Bennett, T. T. (2022). Genetic analysis and field application of a UV-tolerant strain of CrleGV for improved control of Thaumatotibia leucotreta. Master's Thesis, Rhodes University, Makhanda, South Africa.

Bentley, W. E., Mirjalili, N., Andersen, D. C., Davis, R. H., & Kompala, D. S. (1990). Plasmid-encoded protein: the principal factor in the "metabolic burden" associated with recombinant bacteria. *Biotechnology and Bioengineering*, *35*(7), 668–681.

Bevans, R. (2023, June 22). *ANOVA in R | A Complete Step-by-Step Guide with Examples*. Scribbr. https://www.scribbr.com/statistics/anova-in-r/

Bhat, E., Abdalla, M., & Rather, I. (2018). Key factors for successful protein purification and crystallization. *Glob. J. Biotechnol. Biomater. Sci*, *4*(1), 1–7.

Bhatwa, A., Wang, W., Hassan, Y. I., Abraham, N., Li, X.-Z., & Zhou, T. (2021). Challenges associated with the formation of recombinant protein inclusion bodies in Escherichia coli and strategies to address them for industrial applications. *Frontiers in Bioengineering and Biotechnology*, *9*, 630551.

Biedma, M. E., Salvador, R., Ferrelli, M. L., Sciocco-Cap, A., & Romanowski, V. (2015). Effect of the interaction between Anticarsia gemmatalis multiple nucleopolyhedrovirus and Epinotia aporema granulovirus, on A. gemmatalis (Lepidoptera: Noctuidae) larvae. *Biological Control*, *91*, 17–21.

Boogaard, B., Van Oers, M. M., & Van Lent, J. W. M. (2018). An advanced view on baculovirus per os infectivity factors. *Insects*, *9*(3), 84.

Bornhorst, J. A., & Falke, J. J. (2000). [16] Purification of proteins using polyhistidine affinity tags. In *Methods in enzymology* (Vol. 326, pp. 245–254). Academic Press.

Briese, D. T., & Mende, H. A. (1983). Selection for increased resistance to a granulosis virus in the potato moth, Phthorimaea operculella (Zeller)(Lepidoptera: Gelechiidae). *Bulletin of Entomological Research*, *73*(1), 1–9.

Calderone, T. L., Stevens, R. D., & Oas, T. G. (1996). High-level Misincorporation of Lysine for Arginine at AGA Codons in a Fusion Protein Expressed inEscherichia coli. *Journal of Molecular Biology*, *262*(4), 407–412.

Casadaban, M. J., & Cohen, S. N. (1980). Analysis of gene control signals by DNA fusion and cloning in Escherichia coli. *Journal of Molecular Biology*, *138*(2), 179–207.

Casali, N. (2003). Escherichia coli host strains. *E. Coli Plasmid Vectors: Methods and Applications*, 27–48.

Cheng, X.-W., Krell, P. J., & Arif, B. M. (2001). P34. 8 (GP37) is not essential for baculovirus replication. *Journal of General Virology*, *82*(2), 299–305.

Chisoro-Dube, S., & Roberts, S. (2021). Innovation and inclusion in South Africa's citrus industry. Innovation and Inclusion in Agro-processing Working Paper v1.

Chou, C. P. (2007). Engineering cell physiology to enhance recombinant protein production in Escherichia coli. *Applied Microbiology and Biotechnology*, *76*, 521–532.

Coombes, C. A., Hill, M. P., Moore, S. D., & Dames, J. F. (2016). Entomopathogenic fungi as control agents of Thaumatotibia leucotreta in citrus orchards: field efficacy and persistence. *BioControl*, *61*, 729–739.

Copping, L. G. (2004). *The manual of biocontrol agents.* British Crop Protection Council.

Cory, J. S., & Bishop, D. H. L. (1997). Use of baculoviruses as biological insecticides. *Molecular Biotechnology*, *7*, 303–313.

d Steel, R. G., & Torrie, J. H. (1986). *Principles and procedures of statistics: a biometrical approach*. New York, NY, USA: McGraw-Hill.

Daryaei, F. A., Langroudi, R. P., & Golestani Eimani, B. (2017). Production and evaluation of a new recombinant cloning vector for the. *Clostridium Novyi*, 106–111.

De Young, L. R., Dill, K. A., & Fink, A. L. (1993). Aggregation and denaturation of apomyoglobin in aqueous urea solutions. *Biochemistry*, *32*(15), 3877–3886.

Department of Agriculture, Land Reform and Rural Development (DALRRD). (2022). Annual report 2022/2023. https://www.dalrrd.gov.za/index.php/publications/15-annual-report.

Desjardins, P., Hansen, J. B., & Allen, M. (2009). Microvolume protein concentration determination using the NanoDrop 2000c spectrophotometer. *JoVE (Journal of Visualized Experiments)*, *33*, e1610.

Domingues, L. (Ed.). (2017). *PCR.* Springer New York.

Donovan, R. S., Robinson, C. W., & Glick, B. R. (1996). Optimizing inducer and culture conditions for expression of foreign proteins under the control of the lac promoter. *Journal of Industrial Microbiology*, *16*, 145–154.

Durfee, T., Nelson, R., Baldwin, S., Plunkett III, G., Burland, V., Mau, B., Petrosino, J. F., Qin, X., Muzny, D. M., & Ayele, M. (2008). The complete genome sequence of Escherichia coli DH10B: insights into the biology of a laboratory workhorse. *Journal of Bacteriology*, *190*(7), 2597–2606.

Eberle, K. E., & Jehle, J. A. (2006). Field resistance of codling moth against Cydia pomonella granulovirus (CpGV) is autosomal and incompletely dominant inherited. *Journal of Invertebrate Pathology*, *93*(3), 201–206.

Elyazghi, Z., El Yazouli, L., Sadki, K., & Radouani, F. (2017). ABI base recall: Automatic correction and ends trimming of DNA sequences. *IEEE Transactions on NanoBioscience*, *16*(8), 682–686.

Fakruddin, M., Mohammad Mazumdar, R., Bin Mannan, K. S., Chowdhury, A., & Hossain, M. N. (2013). Critical factors affecting the success of cloning, expression, and mass production of enzymes by recombinant E. coli. *International Scholarly Research Notices*, *2013*.

Fei, D., Zhang, H., Diao, Q., Jiang, L., Wang, Q., Zhong, Y., Fan, Z., & Ma, M. (2015). Codon optimization, expression in Escherichia coli, and immunogenicity of recombinant Chinese Sacbrood Virus (CSBV) structural proteins VP1, VP2, and VP3. *PLoS One*, *10*(6), e0128486.

Ferrelli, M. L., & Salvador, R. (2023). Effects of Mixed Baculovirus Infections in Biological Control: A Comprehensive Historical and Technical Analysis. *Viruses*, *15*(9), 1838.

Francis, D. M., & Page, R. (2010). Strategies to optimize protein expression in E. coli. *Current Protocols in Protein Science*, *61*(1), 5–24.

Franken, K. L. M. C., Hiemstra, H. S., van Meijgaarden, K. E., Subronto, Y., Den Hartigh, J., Ottenhoff, T. H. M., & Drijfhout, J. W. (2000). Purification of his-tagged proteins by immobilized chelate affinity chromatography: the benefits from the use of organic solvent. *Protein Expression and Purification*, *18*(1), 95–99.

Furlong, M. J., Wright, D. J., & Dosdall, L. M. (2013). Diamondback moth ecology and management: problems, progress, and prospects. *Annual Review of Entomology*, *58*, 517–541.

Gao, J., Meng, C., Chen, Z., Li, C., & Liu, G. (2013). Codon optimization of the rabbit hemorrhagic disease virus (RHDV) capsid gene leads to increased gene expression in Spodoptera frugiperda 9 (Sf9) cells. *Journal of Veterinary Science*, *14*(4), 441–447.

Garavaglia, M. J., Miele, S. A. B., Iserte, J. A., Belaich, M. N., & Ghiringhelli, P. D. (2012). The ac53, ac78, ac101, and ac103 genes are newly discovered core genes in the family Baculoviridae. *Journal of Virology*, *86*(22), 12069–12079.

García-Fruitós, E., González-Montalbán, N., Morell, M., Vera, A., Ferraz, R. M., Arís, A., Ventura, S., & Villaverde, A. (2005). Aggregation as bacterial inclusion bodies does not imply inactivation of enzymes and fluorescent proteins. *Microbial Cell Factories*, *4*(1), 1–6.

García-Fruitós, E., Vázquez, E., Díez-Gil, C., Corchero, J. L., Seras-Franzoso, J., Ratera, I., Veciana, J., & Villaverde, A. (2012). Bacterial inclusion bodies: making gold from waste. *Trends in Biotechnology*, *30*(2), 65–70.

Gebhardt, M. M., Eberle, K. E., Radtke, P., & Jehle, J. A. (2014). Baculovirus resistance in codling moth is virus isolate-dependent and the consequence of a mutation in viral gene pe38. *Proceedings of the National Academy of Sciences*, *111*(44), 15711–15716.

Geng, J., & Carstens, R. P. (2006). Two methods for improved purification of full-length mammalian proteins that have poor expression and/or solubility using standard Escherichia coli procedures. *Protein Expression and Purification*, *48*(1), 142–150.

Gerchman, S. E., Graziano, V., & Ramakrishnan, V. (1994). Expression of chicken linker histones in E. coli: sources of problems and methods for overcoming some of the difficulties. *Protein Expression and Purification*, *5*(3), 242–251.

Glare, T., Caradus, J., Gelernter, W., Jackson, T., Keyhani, N., Köhl, J., Marrone, P., Morin, L., & Stewart, A. (2012). Have biopesticides come of age? *Trends in Biotechnology*, *30*(5), 250–258.

Glick, B. R., & Whitney, G. K. (1987). Factors affecting the expression of foreign proteins in Escherichia coli. *Journal of Industrial Microbiology and Biotechnology*, *1*(5), 277–282.

Gonzalez-Montalban, N., Garcia-Fruitos, E., & Villaverde, A. (2007). Recombinant protein solubility—does more mean better? *Nature Biotechnology*, *25*(7), 718–720.

Granados, R. R., Fu, Y., Corsaro, B., & Hughes, P. R. (2001). Enhancement of Bacillus thuringiensis toxicity to lepidopterous species with the enhancin from Trichoplusia ni granulovirus. *Biological Control*, *20*(2), 153–159.

Green, M. R., & Sambrook, J. (2020). Cloning in plasmid vectors: blunt-end cloning. *Cold Spring Harbor Protocols*, *2020*(11), pdb-prot101246.

Gross, C. H., Wolgamot, G. M., Russell, R. L., Pearson, M. N., & Rohrmann, G. F. (1993). A 37-kilodalton glycoprotein from a baculovirus of Orgyia pseudotsugata is localized to cytoplasmic inclusion bodies. *Journal of Virology*, *67*(1), 469–475.

Grzywacz, D., & Moore, S. (2017). Production, formulation, and bioassay of baculoviruses for pest control. In *Microbial control of insect and mite pests* (pp. 109–124). Academic Press.

Grzywacz, D., Rabindra, R. J., Brown, M., Jones, K. A., & Parnell, M. (2004). The Helicoverpa armigera NPV production manual. *Natural Resources Institute, Chatham*, *107*.

Guerrero, F., Ciragan, A., & Iwaï, H. (2015). Tandem SUMO fusion vectors for improving soluble protein expression and purification. *Protein Expression and Purification*, *116*, 42–49.

Hadiji-Abbes, N., Borchani-Chabchoub, I., Triki, H., Ellouz, R., Gargouri, A., & Mokdad-Gargouri, R. (2009). Expression of HBsAg and preS2-S protein in different yeast based system: a comparative analysis. *Protein Expression and Purification*, *66*(2), 131–137.

Hara, S., Tanada, Y., & Omi, E. M. (1976). Isolation and characterization of a synergistic enzyme from the capsule of a granulosis virus of the armyworm, Pseudaletia unipuncta. *Journal of Invertebrate Pathology*, *27*(1), 115–124.

Harrison, R., & Hoover, K. (2012). Baculoviruses and other occluded insect viruses. *Insect Pathology*, 73–131.

Harrison, R. L., Herniou, E. A., Jehle, J. A., Theilmann, D. A., Burand, J. P., Becnel, J. J., Krell, P. J., van Oers, M. M., Mowery, J. D., & Bauchan, G. R. (2018). ICTV virus taxonomy profile: Baculoviridae. *Journal of General Virology*, *99*(9), 1185–1186.

Harrison, S. C. (2015). Viral membrane fusion. *Virology*, *479*, 498–507.

Hatting, J. L., Moore, S. D., & Malan, A. P. (2019). Microbial control of phytophagous invertebrate pests in South Africa: Current status and future prospects. *Journal of Invertebrate Pathology*, *165*, 54–66.

Hattingh, V., Moore, S., Kirkman, W., Goddard, M., Thackeray, S., Peyper, M., Sharp, G., Cronjé, P., & Pringle, K. (2020). An improved systems approach as a phytosanitary measure for Thaumatotibia leucotreta (Lepidoptera: Tortricidae) in export citrus fruit from South Africa. *Journal of Economic Entomology*, *113*(2), 700–711.

Hawtin, R. E., Arnold, K., Ayres, M. D., Zanotto, P. M. de A., Howard, S. C., Gooday, G. W., Chappell, L. H., Kitts, P. A., King, L. A., & Possee, R. D. (1995). Identification and preliminary characterization of a chitinase gene in the Autographa californica nuclear polyhedrosis virus genome. *Virology*, *212*(2), 673–685.

Hegedus, D., Erlandson, M., Gillott, C., & Toprak, U. (2009). New insights into peritrophic matrix synthesis, architecture, and function. *Annual Review of Entomology*, *54*, 285–302.

Herniou, E. A., Olszewski, J. A., Cory, J. S., & O'Reilly, D. R. (2003). The genome sequence and evolution of baculoviruses. *Annual Review of Entomology*, *48*(1), 211–234.

Hevehan, D. L., & De Bernardez Clark, E. (1997). Oxidative renaturation of lysozyme at high concentrations. *Biotechnology and Bioengineering*, *54*(3), 221–230.

Hou, D., Kuang, W., Luo, S., Zhang, F., Zhou, F., Chen, T., Zhang, Y., Wang, H., Hu, Z., & Deng, F. (2019). Baculovirus ODV-E66 degrades larval peritrophic membrane to facilitate baculovirus oral infection. *Virology*, *537*, 157–164.

Hukuhara, T., & Zhu, Y. (1989). Enhancement of the in vitro infectivity of a nuclear polyhedrosis virus by a factor in the capsule of a granulosis virus. *Journal of Invertebrate Pathology*, *54*(1), 71–78.

Hunter-Fujita, F. R., Entwistle, P. F., Evans, H. F., & Crook, N. E. (1998). *Insect viruses and pest management.* John Wiley & Sons Ltd.

Hwang, P. M., Pan, J. S., & Sykes, B. D. (2014). Targeted expression, purification, and cleavage of fusion proteins from inclusion bodies in Escherichia coli. *FEBS Letters*, *588*(2), 247–252.

Idamokoro, E. M., Hosu, Y. S., Oyedeji, O. O., Miya, G. M., Kuria, S. K., & Oyedeji, A. O. (2022). A comparative analysis of the proximate and mineral composition of whole Citrus

limon and Citrus clementina as a prospective alternative feed resource for livestock farming in South Africa. *Frontiers in Sustainable Food Systems*, *6*, 1021175.

Ishimwe, E., Hodgson, J. J., Clem, R. J., & Passarelli, A. L. (2015). Reaching the melting point: Degradative enzymes and protease inhibitors involved in baculovirus infection and dissemination. *Virology*, *479*, 637–649.

Itkonen, J. M., Urtti, A., Bird, L. E., & Sarkhel, S. (2014). Codon optimization and factorial screening for enhanced soluble expression of human ciliary neurotrophic factor in Escherichia coli. *BMC Biotechnology*, *14*(1), 1–11.

Jamrichová, D., Tišáková, L., Jarábková, V., & Godány, A. (2017). How to approach heterogeneous protein expression for biotechnological use: An overview. *Nova Biotechnologica et Chimica*, *16*(1).

Javed, M. A., Biswas, S., Willis, L. G., Harris, S., Pritchard, C., van Oers, M. M., Donly, B. C., Erlandson, M. A., Hegedus, D. D., & Theilmann, D. A. (2017). Autographa californica multiple nucleopolyhedrovirus AC83 is a per os infectivity factor (PIF) protein required for occlusion-derived virus (ODV) and budded virus nucleocapsid assembly as well as assembly of the PIF complex in ODV envelopes. *Journal of Virology*, *91*(5), 10–1128.

Jehle, J. A., Schulze-Bopp, S., Undorf-Spahn, K., & Fritsch, E. (2017). Evidence for a second type of resistance against Cydia pomonella granulovirus in field populations of codling moths. *Applied and Environmental Microbiology*, *83*(2), e02330-16.

Jukes, M. D. (2018). Baculovirus synergism: investigating mixed alphabaculovirus and betabaculovirus infections in the false codling moth, Thaumatotibia leucotreta, for improved pest control. PhD Thesis, Rhodes University, Makhanda, South Africa.

Kalendar, R. (2021). A guide to using FASTPCR software for PCR, in silico PCR, and oligonucleotide analysis. *PCR primer design*, 223-243.

Kane, J. F. (1995). Effects of rare codon clusters on high-level expression of heterologous proteins in Escherichia coli. *Current Opinion in Biotechnology*, *6*(5), 494–500.

Kataeva, I., Chang, J., Xu, H., Luan, C.-H., Zhou, J., Uversky, V. N., Lin, D., Horanyi, P., Liu, Z. J., & Ljungdahl, L. G. (2005). Improving Solubility of Shewanella o neidensis MR-1 and Clostridium thermocellum JW-20 Proteins Expressed into Esherichia coli. *Journal of Proteome Research*, *4*(6), 1942–1951.

Kessler, P., & Zingg, D. (2008, July). New baculovirus products offer solutions for the biological control of Cydia pomonella and Cryptophlebia leucotreta. In *Proceedings of the 23rd International Congress of Entomology, Durban, South Africa* (Vol. 612).

Knox, C., Moore, S. D., Luke, G. A., & Hill, M. P. (2015). Baculovirus-based strategies for the management of insect pests: a focus on development and application in South Africa. *Biocontrol Science and Technology*, *25*(1), 1-20.

Kuo, D., Nie, M., & Courey, A. J. (2014). SUMO as a solubility tag and in vivo cleavage of SUMO fusion proteins with Ulp1. *Protein Affinity Tags: Methods and Protocols*, 71–80.

Lacey, L. A., Grzywacz, D., Shapiro-Ilan, D. I., Frutos, R., Brownbridge, M., & Goettel, M. S. (2015). Insect pathogens as biological control agents: Back to the future. *Journal of Invertebrate Pathology*, *132*, 1–41.

Lebendiker, M., Danieli, T., & de Marco, A. (2014). The Trip Adviser guide to the protein science world: a proposal to improve the awareness concerning the quality of recombinant proteins. *BMC Research Notes*, *7*(1), 1–2.

Lei, C., Yang, S., Lei, W., Nyamwasa, I., Hu, J., & Sun, X. (2020). Displaying enhancing factors on the surface of occlusion bodies improves the insecticidal efficacy of a baculovirus. *Pest Management Science*, *76*(4), 1363–1370.

Lepore, L. S., Roelvink, P. R., & Granados, R. R. (1996). Enhancin, the granulosis virus protein that facilitates nucleopolyhedrovirus (NPV) infections, is a metalloprotease. *Journal of Invertebrate Pathology*, *68*(2), 131–140.

Li, Z., Li, C., Yang, K., Wang, L., Yin, C., Gong, Y., & Pang, Y. (2003). Characterization of a chitin-binding protein GP37 of Spodoptera litura multicapsid nucleopolyhedrovirus. *Virus Research*, *96*(1–2), 113–122.

Liu, X., Fang, W., Fan, R., Zhang, L., Lei, C., Zhang, J., Nian, W., Dou, T., An, S., & Zhou, L. (2019). Granulovirus GP37 facilitated ODVs cross insect peritrophic membranes and fuse with epithelia. *Toxins*, *11*(3), 145.

Liu, X., Ma, X., Lei, C., Xiao, Y., Zhang, Z., & Sun, X. (2011). Synergistic effects of Cydia pomonella granulovirus GP37 on the infectivity of nucleopolyhedroviruses and the lethality of Bacillus thuringiensis. *Archives of Virology*, *156*, 1707–1715.

Liu, X., Yang, G., Qiu, B., & Tian, P. (2000). Molecular cloning of enhancin gene from Helicoverpa armigera Granulosis virus and its expression in E. coli. *Wei Sheng Wu Xue Bao= Acta Microbiologica Sinica*, *40*(4), 379–383.

Liu, Y., Wu, C., Wang, J., Mo, W., & Yu, M. (2013). Codon optimization, expression, purification, and functional characterization of recombinant human IL-25 in Pichia pastoris. *Applied Microbiology and Biotechnology*, *97*, 10349–10358.

Love, C. N., Hill, M. P., & Moore, S. D. (2014). T haumatotibia leucotreta and the N avel orange: ovipositional preferences and host susceptibility. *Journal of Applied Entomology*, *138*(8), 600–611.

Luckow, V. A. (1990). Cloning and expression in heterologous genes in insect cells with baculovirus vectors. *Baculovirus Expression System and Biopesticides.*, 97–149.

Malan, A. P., Von Diest, J. I., Moore, S. D., & Addison, P. (2018). Control options for false codling moth, Thaumatotibia leucotreta (Lepidoptera: Tortricidae), in South Africa, with emphasis on the potential use of entomopathogenic nematodes and fungi. *African Entomology*, *26*(1), 14–29.

Manrakhan, A., Abeeluck, D., & Gokool, A. (2008). Assessment of damage by Cryptophlebia peltastica (Meyrick)(Lepidoptera: Tortricidae) in litchi orchards in Mauritius. *African Entomology*, *16*(2), 203–208.

Manta, B., Obal, G., Ricciardi, A., Pritsch, O., & Denicola, A. (2011). Tools to evaluate the conformation of protein products. *Biotechnology Journal*, *6*(6), 731–741.

Marblestone, J. G., Edavettal, S. C., Lim, Y., Lim, P., Zuo, X. U. N., & Butt, T. R. (2006). Comparison of SUMO fusion technology with traditional gene fusion systems: enhanced expression and solubility with SUMO. *Protein Science*, *15*(1), 182–189.

Marsberg, T. (2016). The isolation and genetic characterisation of a novel alphabaculovirus for the microbial control of Cryptophlebia peltastica and closely related tortricid pests. *Ph. D. Thesis*. Rhodes University, Makhanda, South Africa.

Marsberg, T., Jukes, M., Chambers, C., Hendriks, C., Knox, C., Hill, M., & Moore, S. (2017). The isolation of a novel alphabaculovirus and its potential for microbial control of key tortricid moth pests. *Microbial and Nematode Control of Invertebrate Pests IOBC-WPRS Bulletin., 129*, 175-178.

Marsberg, T., Jukes, M. D., Krejmer-Rabalska, M., Rabalski, L., Knox, C. M., Moore, S. D., ... & Szewczyk, B. (2018). Morphological, genetic and biological characterisation of a novel alphabaculovirus isolated from Cryptophlebia peltastica (Lepidoptera: Tortricidae). *Journal of invertebrate pathology*, *157*, 90-99.

Martínez-Alonso, M., González-Montalbán, N., García-Fruitós, E., & Villaverde, A. (2008). The functional quality of soluble recombinant polypeptides produced in Escherichia coli is defined by a wide conformational spectrum. *Applied and Environmental Microbiology*, *74*(23), 7431–7433.

Masson, T., Fabre, M. L., Ferrelli, M. L., Pidre, M. L., & Romanowski, V. (2019). Protein composition of the occlusion bodies of Epinotia aporema granulovirus. *Plos One*, *14*(2), e0207735.

Mela, T. (2022). Development and optimisation of a qPCR assay for the enumeration of Cryptophlebia leucotreta granulovirus (CrleGV) used for commercial applications. Master's Thesis, Rhodes University, Makhanda, South Africa.

Men, A. E., Wilson, P., Siemering, K., & Forrest, S. (2008). Sanger DNA sequencing. *Next Generation Genome Sequencing: Towards Personalized Medicine*, 1–11.

Miele, S. A. B., Garavaglia, M. J., Belaich, M. N., & Ghiringhelli, P. D. (2011). Baculovirus: molecular insights on their diversity and conservation. *International Journal of Evolutionary Biology*, *2011*.

Miller, L. K. (1988). Baculoviruses as gene expression vectors. *Annual Reviews in Microbiology*, *42*(1), 177–199.

Moore, S. D. (2021). Biological control of a phytosanitary pest (Thaumatotibia leucotreta): A case study. *International Journal of Environmental Research and Public Health*, *18*(3), 1198.

Moore, S. D., Hendry, D. A., & Richards, G. I. (2011). Virulence of a South African isolate of the Cryptophlebia leucotreta granulovirus to Thaumatotibia leucotreta neonate larvae. *BioControl*, *56*(3), 341–352.

Moore, S. D., Kirkman, W., Peyper, M., Thackeray, S. R., Marsberg, T., Albertyn, S., & Hill, M. P. (2018). Development of a Postharvest Cold Treatment for Cryptophlebia peltastica

(Lepidoptera: Tortricidae) for Export of Litchis from South Africa. *Journal of Economic Entomology*, *111*(6).

Moore, S. D., Kirkman, W., Richards, G. I., & Stephen, P. R. (2015). The Cryptophlebia leucotreta granulovirus—10 years of commercial field use. *Viruses*, *7*(3), 1284–1312.

Moore, S., & Jukes, M. (2019). Advances in microbial control in IPM: Entomopathogenic Viruses. In *Integrated management of insect pests* (pp. 593–648). Burleigh Dodds Science Publishing.

Moore, S., & Jukes, M. (2023). The History of Baculovirology in Africa. *Viruses*, *15*(7), 1519.

Moore, S., Kirkman, W., & Hattingh, V. (2015). The host status of lemons for the false codling moth, Thaumatotibia leucotreta (Meyrick)(Lepidoptera: Tortricidae) with particular reference to export protocols. *African Entomology*, *23*(2), 519–525.

Moore, S., Kirkman, W., & Stephen, P. (2004). Cryptogran. A virus for the biological control of false codling moth. *SA Fruit Journal (South Africa)*.

Motsoeneng, B., Jukes, M. D., Knox, C. M., Hill, M. P., & Moore, S. D. (2019). Genome analysis of a novel south African Cydia pomonella granulovirus (CpGV-SA) with resistance-breaking potential. *Viruses*, *11*(7), 658.

Mudgal, S., De Toni, A., Tostivint, C., Hokkanen, H., & Chandler, D. (2013). Scientific support, literature review and data collection and analysis for risk assessment on microbial organisms used as active substance in plant protection products–Lot 1 Environmental Risk characterisation. *EFSA Supporting Publications*, *10*(12), 518E.

Mullis, K. B. (1990). The unusual origin of the polymerase chain reaction. *Scientific American*, *262*(4), 56–65.

Murphy, R. M., & Roberts, C. J. (2013). Protein misfolding and aggregation research: some thoughts on improving quality and utility. *Biotechnology Progress*, *29*(5), 1109–1115.

Nakai, M., Takahashi, K., Iwata, K., Tanaka, K., Koyanagi, J., Ookuma, A., Takatsuka, J., Okuno, S., & Kunimi, Y. (2017). Acquired resistance to a nucleopolyhedrovirus in the smaller tea tortrix Adoxophyes honmai (Lepidoptera: Tortricidae) after selection by serial viral administration. *Journal of Invertebrate Pathology*, *145*, 23–30.

Nawawi, O., Abdullah, M. P., & Yusuf, C. Y. L. (2022). A strategy for in-house production of a positive selection cloning vector from the commercial pJET1. 2/blunt cloning vector at minimal cost. *3 Biotech*, *12*(9), 216.

Nguyen, H. D., Phan, T. T. P., & Schumann, W. (2007). Expression vectors for the rapid purification of recombinant proteins in Bacillus subtilis. *Current Microbiology*, *55*, 89–93.

Nicolopoulou-Stamati, P., Maipas, S., Kotampasi, C., Stamatis, P., & Hens, L. (2016). Chemical pesticides and human health: the urgent need for a new concept in agriculture. *Frontiers in Public Health*, *4*, 148.

Novy, R., Drott, D., Yaeger, K., & Mierendorf, R. (2001). Overcoming the codon bais of E. coli for enhanced protein expression. *inNovations*, *12*.

Okada, J., Maruyama, T., Motomura, K., Kuroki, K., Maenaka, K., Sakono, M., & Goto, M. (2009). Enzyme-mediated protein refolding. *Chemical Communications*, *46*, 7197–7199.

Olins, P. O., & Rangwala, S. H. (1989). A novel sequence element derived from bacteriophage T7 mRNA acts as an enhancer of translation of the lacZ gene in Escherichia coli. *Journal of Biological Chemistry*, *264*(29), 16973–16976.

Oliveira, C., & Domingues, L. (2018). Guidelines to reach high-quality purified recombinant proteins. *Applied Microbiology and Biotechnology*, *102*, 81–92.

Olson, B. J. S. C. (2016). Assays for determination of protein concentration. *Current Protocols in Pharmacology*, *73*(1), A-3A.

Opoku-Debrah, J. K., Hill, M. P., Knox, C., & Moore, S. D. (2013). Overcrowding of false codling moth, Thaumatotibia leucotreta (Meyrick) leads to the isolation of five new Cryptophlebia leucotreta granulovirus (CrleGV-SA) isolates. *Journal of Invertebrate Pathology*, *112*(3), 219–228.

Opoku-Debrah, J. K., Hill, M. P., Knox, C., & Moore, S. D. (2016). Heterogeneity in virulence relationships between Cryptophlebia leucotreta granulovirus isolates and geographically distinct host populations: Lessons from codling moth resistance to CpGV-M. *BioControl*, *61*, 449–459.

O'Reilly, D. R., Miller, L. K., & Luckow, V. A. (1994). *Baculovirus expression vectors: a laboratory manual*. Oxford University Press.

Peng, J., Zhong, J., & Granados, R. R. (1999). A baculovirus enhancin alters the permeability of a mucosal midgut peritrophic matrix from lepidopteran larvae. *Journal of Insect Physiology*, *45*(2), 159–166.

Peng, K., van Lent, J. W. M., Boeren, S., Fang, M., Theilmann, D. A., Erlandson, M. A., Vlak, J. M., & van Oers, M. M. (2012). Characterization of novel components of the baculovirus per os infectivity factor complex. *Journal of Virology*, *86*(9), 4981–4988.

Peng, K., van Oers, M. M., Hu, Z., van Lent, J. W. M., & Vlak, J. M. (2010). Baculovirus per os infectivity factors form a complex on the surface of occlusion-derived virus. *Journal of Virology*, *84*(18), 9497–9504.

Perron-Savard, P., De Crescenzo, G., & Moual, H. Le. (2005). Dimerization and DNA binding of the Salmonella enterica PhoP response regulator are phosphorylation independent. *Microbiology*, *151*(12), 3979–3987.

Peti, W., & Page, R. (2007). Strategies to maximize heterologous protein expression in Escherichia coli with minimal cost. *Protein Expression and Purification*, *51*(1), 1–10.

Pinsach, J., de Mas, C., López-Santín, J., Striedner, G., & Bayer, K. (2008). Influence of process temperature on recombinant enzyme activity in Escherichia coli fed-batch cultures. *Enzyme and Microbial Technology*, *43*(7), 507–512.

Pope, B., & Kent, H. M. (1996). High efficiency 5 min transformation of Escherichia coli. *Nucleic Acids Research*, *24*(3), 536–537.

Preston, A. (2003). Choosing a cloning vector. *E. Coli Plasmid Vectors: Methods and Applications*, 19–26.

Puckett, M. C. (2015). Hexahistidine (6xHis) fusion-based assays for protein-protein interactions. *Protein-Protein Interactions: Methods and Applications*, 365–370.

QIAexpressionist, A. (2002). A handbook for high-level expression and purification of 6xhis-tagged proteins. *Qiagen. P1-125*.

Quinn, G. P., & Keough, M. J. (2002). *Experimental design and data analysis for biologists*. Cambridge university press.

Rashid, F., Sharma, S., & Bano, B. (2005). Comparison of guanidine hydrochloride (GdnHCl) and urea denaturation on inactivation and unfolding of human placental cystatin (HPC). *The Protein Journal*, *24*, 283–292.

Raynal, B., Lenormand, P., Baron, B., Hoos, S., & England, P. (2014). Quality assessment and optimization of purified protein samples: why and how? *Microbial Cell Factories*, *13*(1), 1–10.

Reece-Hoyes, J. S., & Walhout, A. J. M. (2018). Gateway recombinational cloning. *Cold Spring Harbor Protocols*, *2018*(1), pdb-top094912.

Ricarte-Bermejo, A., Simón, O., Fernández, A. B., Williams, T., & Caballero, P. (2021). Bacmid Expression of Granulovirus Enhancin En3 Accumulates in Cell Soluble Fraction to Potentiate Nucleopolyhedrovirus Infection. *Viruses*, *13*(7), 1233.

Rohrmann, G. (2019). Baculovirus Molecular Biology, 4th Edition. In *Bethesda (MD): National Center for Biotechnology Information (US)* (Issue Md).

Rosano, G. L., & Ceccarelli, E. A. (2009). Rare codon content affects the solubility of recombinant proteins in a codon bias-adjusted Escherichia coli strain. *Microbial Cell Factories*, *8*(1), 1–9.

Rosano, G. L., & Ceccarelli, E. A. (2014). Recombinant protein expression in Escherichia coli: advances and challenges. *Frontiers in Microbiology*, *5*, 172.

Rudolph, R., & Lilie, H. (1996). In vitro folding of inclusion body proteins. *The FASEB Journal*, *10*(1), 49–56.

Sahdev, S., Khattar, S. K., & Saini, K. S. (2008). Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. *Molecular and Cellular Biochemistry*, *307*, 249–264.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., & Erlich, H. A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, *239*(4839), 487–491.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, *74*(12), 5463–5467.

San-Miguel, T., Pérez-Bermúdez, P., & Gavidia, I. (2013). Production of soluble eukaryotic recombinant proteins in E. coli is favoured in early log-phase cultures induced at low temperature. *Springerplus*, *2*, 1–4.

Saraswat, M., Musante, L., Ravidá, A., Shortt, B., Byrne, B., & Holthofer, H. (2013). Preparative purification of recombinant proteins: current status and future trends. *BioMed Research International*, *2013*.

Seiler, C. Y., Park, J. G., Sharma, A., Hunter, P., Surapaneni, P., Sedillo, C., Field, J., Algar, R., Price, A., & Steel, J. (2014). DNASU plasmid and PSI: Biology-Materials repositories: resources to accelerate biological research. *Nucleic Acids Research*, *42*(D1), D1253–D1260.

Senisterra, G., Chau, I., & Vedadi, M. (2012). Thermal denaturation assays in chemical biology. *Assay and Drug Development Technologies*, *10*(2), 128–136.

Sezonov, G., Joseleau-Petit, D., & d'Ari, R. (2007). Escherichia coli physiology in Luria-Bertani broth. *Journal of Bacteriology*, *189*(23), 8746–8749.

Shiloach, J., & Fass, R. (2005). Growing E. coli to high cell density—a historical perspective on method development. *Biotechnology Advances*, *23*(5), 345–357.

Shivachandra, S. B., Yogisharadhya, R., Ahuja, A., & Bhanuprakash, V. (2012). Expression and purification of recombinant type IV fimbrial subunit protein of Pasteurella multocida serogroup B: 2 in Escherichia coli. *Research in Veterinary Science*, *93*(3), 1128–1131.

Singh, S. M., & Panda, A. K. (2005). Solubilization and refolding of bacterial inclusion body proteins. *Journal of Bioscience and Bioengineering*, *99*(4), 303–310.

Sinha, D., Bakhshi, M., & Vora, R. (1994). Ligand binding assays with recombinant proteins refolded on an affinity matrix. *Biotechniques*, *17*(3), 509–512.

Slavicek, J. M. (2012). Baculovirus enhancins and their role in viral pathogenicity. *Molecular Virology*, 147–168.

Slavicek, J. M., & Popham, H. J. R. (2005). The Lymantria dispar nucleopolyhedrovirus enhancins are components of occlusion-derived virus. *Journal of Virology*, *79*(16), 10578–10588.

Spanjaard, R. A., & Van Duin, J. (1988). Translation of the sequence AGG-AGG yields 50% ribosomal frameshift. *Proceedings of the National Academy of Sciences*, *85*(21), 7967–7971.

Sugiura, N., Setoyama, Y., Chiba, M., Kimata, K., & Watanabe, H. (2011). Baculovirus envelope protein ODV-E66 is a novel chondroitinase with distinct substrate specificity. *Journal of Biological Chemistry*, *286*(33), 29026–29034.

Tanada, Y. (1985). A synopsis of studies on the synergistic property of an insect baculovirus: a tribute to Edward A. Steinhaus. *Journal of Invertebrate Pathology*, *45*(2), 125–138.

Tanada, Y., Himeno, M., & Omi, E. M. (1973). Isolation of a factor, from the capsule of a granulosis virus, synergistic for a nuclear-polyhedrosis virus of the armyworm. *Journal of Invertebrate Pathology*, *21*(1), 31–40.

Taylor, D. G. (2021). Baculovirus synergism for improved management of false codling moth Thaumatotibia leucotreta Meyr. (Lepidoptera: Tortricidae). Master's Thesis, Rhodes University, Makhanda, South Africa.

Thomas, S. R., & Elkinton, J. S. (2004). Pathogenicity and virulence. *Journal of Invertebrate Pathology*, *85*(3), 146–151.

Turner, P., Holst, O., & Karlsson, E. N. (2005). Optimized expression of soluble cyclomaltodextrinase of thermophilic origin in Escherichia coli by using a soluble fusion-tag and by tuning of inducer concentration. *Protein Expression and Purification*, *39*(1), 54–60.

van der Merwe, M., D. Jukes, M., Rabalski, L., Knox, C., K. Opoku-Debrah, J., D. Moore, S., Krejmer-Rabalska, M., Szewczyk, B., & P. Hill, M. (2017). Genome analysis and genetic stability of the Cryptophlebia leucotreta granulovirus (CrleGV-SA) after 15 years of commercial use as a biopesticide. *International Journal of Molecular Sciences*, *18*(11), 2327.

van Oers, M. M., Herniou, E. A., Jehle, J. A., Krell, P. J., Abd-Alla, A. M. M., Ribeiro, B. M., Theilmann, D. A., Hu, Z., & Harrison, R. L. (2023). Developments in the classification and nomenclature of arthropod-infecting large DNA viruses that contain pif genes. *Archives of Virology*, *168*(7), 182.

van Oers, M. M., & Vlak, J. M. (2007). Baculovirus genomics. *Current Drug Targets*, *8*(10), 1051–1068.

Venette, R. C., Davis, E. E., DaCosta, M., Heisler, H., & Larson, M. (2003). Mini risk assessment: false codling moth, Thaumatotibia (= Cryptophlebia) leucotreta (Meyrick)[Lepidoptera: Tortricidae]. *University of Minnesota, Department of Entomology, CAPS PRA, St. Paul, MN*, 1–30.

Vera, A., González-Montalbán, N., Arís, A., & Villaverde, A. (2007). The conformational quality of insoluble recombinant proteins is enhanced at low growth temperatures. *Biotechnology and Bioengineering*, *96*(6), 1101–1106.

Vieira Gomes, A. M., Souza Carmo, T., Silva Carvalho, L., Mendonça Bahia, F., & Parachin, N. S. (2018). Comparison of yeasts as hosts for recombinant protein production. *Microorganisms*, *6*(2), 38.

Volontè, F., Marinelli, F., Gastaldo, L., Sacchi, S., Pilone, M. S., Pollegioni, L., & Molla, G. (2008). Optimization of glutaryl-7-aminocephalosporanic acid acylase expression in E. coli. *Protein Expression and Purification*, *61*(2), 131–137.

Wang, M., & Hu, Z. (2019). Cross-talking between baculoviruses and host insects towards a successful infection. *Philosophical Transactions of the Royal Society B*, *374*(1767), 20180324.

Wang, P., & Granados, R. R. (1997). An intestinal mucin is the target substrate for a baculovirus enhancin. *Proceedings of the National Academy of Sciences*, *94*(13), 6977–6982.

Wang, X., Liu, X., Makalliwa, G. A., Li, J., Wang, H., Hu, Z., & Wang, M. (2017). Per os infectivity factors: a complicated and evolutionarily conserved entry machinery of baculovirus. *Science China Life Sciences*, *60*, 806–815.

Wennmann, J. T., Köhler, T., Gueli Alletti, G., & Jehle, J. A. (2015). Mortality of cutworm larvae is not enhanced by Agrotis segetum granulovirus and Agrotis segetum nucleopolyhedrovirus B coinfection relative to single infection by either virus. *Applied and Environmental Microbiology*, *81*(8), 2893–2899.

Williams, T., Virto, C., Murillo, R., & Caballero, P. (2017). Covert infection of insects by baculoviruses. *Frontiers in Microbiology*, *8*, 1337.

Wingfield, P. T. (2015). Overview of the purification of recombinant proteins. *Current Protocols in Protein Science*, *80*(1), 1–6.

Wu, J., & Filutowicz, M. (1999). Hexahistidine (His6)-tag dependent protein dimerization: a cautionary tale. *Acta Biochimica Polonica*, *46*(3), 591–599.

Yadav, D. K., Yadav, N., Yadav, S., Haque, S., & Tuteja, N. (2016). An insight into fusion technology aiding efficient recombinant protein production for functional proteomics. *Archives of Biochemistry and Biophysics*, *612*, 57–77.

Yang, S., Zhao, L., Ma, R., Fang, W., Hu, J., Lei, C., & Sun, X. (2017). Improving baculovirus infectivity by efficiently embedding enhancing factors into occlusion bodies. *Applied and Environmental Microbiology*, *83*(14), e00595-17.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T. L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, *13*, 1–11.

Zhang, Z., Gildersleeve, J., Yang, Y.-Y., Xu, R., Loo, J. A., Uryu, S., Wong, C.-H., & Schultz, P. G. (2004). A new strategy for the synthesis of glycoproteins. *Science*, *303*(5656), 371–373.

Zhu, S., Wang, W., Wang, Y., Yuan, M., & Yang, K. (2013). The baculovirus core gene ac83 is required for nucleocapsid assembly and per os infectivity of Autographa californica nucleopolyhedrovirus. *Journal of Virology*, *87*(19), 10573–10586.