Investigating the expression of three small open reading frames encoded on Helicoverpa armigera stunt virus RNA 1

Dissertation submitted in fulfilment of the requirements for the degree of

Master of Science
(Microbiology)

in the

Department of Biochemistry and Microbiology
Rhodes University

Mart-Mari de Bruyn

January 2017

Abstract

The Helicoverpa armigera stunt virus (HaSV), belonging to the Family *Alphatetraviridae* (Genus: *Omegatetravirus*), is a non-enveloped insect virus encapsidating a bi-partite, positive-sense single-stranded RNA genome. RNA1 encodes the replicase, as well as three small open reading frames (ORFs) arranged in tandem, and overlapping with the 3' end of the replicase ORF. These ORFs, designated p11, p15 and p8, encode putative proteins of unknown function. The p11 and p15 ORFs are conserved in the genome of the related *Omegatetravirus*, Dendrolimus punctatus tetravirus. In HaSV, the stop codon of p11 is followed immediately by the start of p15, whereas the stop of p15 and start of p8 are separated by a glycine intercodon. Furthermore, only p11 is known to have a recognizable Kozak sequence. The aim of this study was to determine the expression and function of these three small proteins in the HaSV infectious lifecycle.

The authenticity of the viral cDNA sequence, encoding the three small ORFs, was validated by sequencing multiple cDNA clones of the relevant region in viral RNA (vRNA), purified from infectious HaSV particles. The sequence of all three ORFs was conserved in seven cDNA clones, while point mutations were observed in each of two remaining cDNA clones, suggesting that the ORFs were conserved in infectious virus. Polyclonal antisera were raised against a p11 peptide, and a recombinant p15-p8 fusion protein (p23) expressed and purified from *Escherichia coli*. The affinity of the anti-p23 antiserum was confirmed by western blot analysis, while that of the anti-p11 antiserum was confirmed using immunofluorescence microscopy, as attempted expression of recombinant p11 in *E. coli* appeared to be toxic. The antisera were used to detect expression of the small proteins in HaSV-infected *H. armigera* larvae by western blot analysis. A band migrating at approximately 34 kDa was detected by both antisera in infected larvae, absent in uninfected larvae, suggesting the expression of a p11-p15-p8 polyprotein. Protein bands of 11 kDa and 8 kDa were also detected by the anti-p11 and anti-p23 antisera, respectively. Bioinformatic analysis revealed that the polyprotein would be produced by a novel type of stop codon read-through, however the mechanism required for individual expression could not be definitively determined.

The mechanism by which these ORFs are translated was further investigated by expressing p11-p15, tagged with FLAG and enhanced green flourescent protein (EGFP) at its amino- and carboxyl-termini respectively (FLAG-p11-p15-EGFP), in *Spodoptera frugiperda* (Sf9) cells detected by flourescence microscopy. Punctate structures were observed throughout the cytoplasm that were also detected with anti-FLAG, anti-p11 and anti-p23 antisera, complementing results obtained in previous studies. Since p15 does not exhibit a strong recognizable Kozak like p11, the dependency of p15 expression on that of p11 was investigated by mutating this construct such that p15 occurred in a +1 frame to p11. Both EGFP and anti-p23 fluorescence was detected with the same cytoplasmic distribution as the unmutated construct, whereas nothing was detected by anti-FLAG and anti-p11. Preliminary results therefore suggested p15 may also be expressed as a discrete protein, independent of p11.

Acknowledgements

The financial assistance from the Rhodes University Prestigious Scholarship and the National Research Foundation (NRF), towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to Rhodes University or the donor.

I also wish to acknowledge the people, without who, none of this work would have been possible. Firstly, I would like to thank my supervisor, Prof. Dorrington for all the guidance and support you offered in the past three years. You are a fantastic scientist and incredible role model to not only myself, but all of your students. To Dr. Meesbah Jiwaji, an endless bucket of thanks! All of your patience, advice, and help over the years has been truly invaluable. The countless hours spent in confocal, pheriously, thank you. A special thanks goes out to Val Hodgson for maintaining tissue culture, and dealing with the never ending admin. Without you, I would have gone mad even sooner… To the rest of my colleagues in the Dorrington Research Group, you are all amazing. Thank you for all of the insightful conversations and discussions. Also, thank you for putting up, and especially joining in the lame joke parades. Thank you to Kudzai Mtambanengwe and Prof. Caroline Knox for supplying us with *H. armigera* eggs and larvae. We came to you, our neighbouring laboratory, to borrow some eggs time and time again, and you never hesitated to help.

Then to my family, without you, I would not be where I am today. Dankie dat julle my altyd ondersteun en aanspoor om te doen wat my hart gelukkig maak. Dat julle alles gedoen en gegee het om dit vir my moontlik te maak, maak nie saak die nagevolge daarvan. Julle is ongelooflik, en ek is oneindig lief vir julle. Laaste, maar definitief nie minste nie, wil ek alle dank en lof gee aan God. Dankie dat U my seën met talle geleenthede om ‘n beter persoon en wetenskaplike te word, en met soveel wonderlike mense in my lewe.

Table of contents

Chapter 2: Materials and methods

Chapter 3: Results

Chapter 4: Discussion

List of tables and figures

List of tables:

List of figures:

List of abbreviations and virus names

General:

(-) – negative sense

(+) – positive sense

6xHis - 6xhistidine/hexahistidine

aa – amino acid(s)

AGE – agarose gel electrophoresis

Amp – ampicillin

BSA - bovine serum albumin

BLAST - Basic Local Alignment Search Tool

bp – base pair

BV – baculovirus

cDNA - complementary DNA

CITE - cap-independent translation enhancer

cm^2 - squared centimetre

CO_2 – carbon dioxide

C-terminus – carboxyl terminus

DAPI - 4', 6-diamidino-2-phenylindole

DNA - deoxyribonucleic acid

dNTPs - deoxynucleoside triphosphates

ds - double-stranded

EDTA - ethylenediaminetetraacetic acid

EGFP - enhanced green fluorescent protein

eIF - eukaryotic initiation factor

ER – endoplasmic reticulum

FBS – foetal bovine serum

Gent - gentamicin

HEL - helicase

HRP – horse radish peroxidase

HSVtk - Herpes simplex virus thymidine kinase

IgG – immunoglobulin G

IPTG - isopropyl β-D-1-thiogalactopyranoside

IRES - internal ribosomal entry site

ITAF- IRES trans-activating factors

Kan - kanamycin

kb – kilobase(s)

kDa – kilodalton(s)

LB - Luria-Bertani

M – molar

^{m7}GpppN - 7-methylguanosine cap

Mg – milligram

mg/ml – milligram per millilitre

mM – millimolar

mRNA – messenger RNA

MT – methyl transferase

nm – nanometer

nt- nucleotide(s)

N-terminus – amino terminus

NTPs - nucleoside triphosphates

OD_{600} - optical density at 600 nm

ORF - open reading frame

P1, 2 – passage1, 2

PBS - phosphate buffered saline

PCR - polymerase chain reaction

pfu - plaque-forming units

RBS - ribosome binding site

RdRp - RNA dependent RNA polymerase

REP – replicase

RNA - ribonucleic acid

RT - reverse transcriptase

RT-PCR - reverse transcription polymerase chain reaction

SDS - sodium dodecyl sulphate

SDS-PAGE - sodium dodecyl sulphate polyacrylamide gel electrophoresis

ss – single-stranded

SV - Simian virus

TBS – tris buffered saline

Tet - tetracycline

Tris - tris-2-amino-2-(hydroxymethyl)-1,3-propandiol

tRNA – transfer RNA

$tRNA_i$ – initiator tRNA

TURBS - termination upstream ribosomal binding site

UTR - untranslated region

VCAP - viral capsid protein precursor

vp/mL - virus particles per millilitre

VPg - viral genome-linked protein

vRNA – viral RNA

X-gal - 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

µg - microgram

µg/ml - microgram per millilitre

µl - microlitre

Cells:

E. coli - Escherichia coli

FB33 – *Helicoverpa zea* fat body cell line

HeLa - HeLa S3 immortalised cervical cancer cell line

MG8 – *Helicoverpa zea* midgut cell line

Se-1 – Spodoptera exigua embryonic cell line

Sf9 - *Spodoptera frugiperda* cell line 9

Viruses:

AcMNPV - Autographa californica multiple nuclearpolyhedrovirus

CaMV - Cauliflower mosaic virus

CPV- Cricket paralysis virus

DpTV - Dendrolimus punctatus tetravirus

ECMV- Encephalomyocarditis virus

EeV - Euprosterna elaeasa virus

HaSV - Helicoverpa armigera stunt virus

HCV- Hepatitis C virus

HIV- Human Immunodeficiency virus

NβV- Nudaurelia capensis β virus

NωV - Nudaurelia capensis omega virus

PCRPV - Pelargonium chlorotic ring pattern virus

PrV – Providence virus

TaV - Thosea asigna virus

WNV- West Nile virus

Nucleotides:

A - adenine

C - cytosine

G - guanine

T - thymine

U - uracil

Amino acids:

A	- Ala	-	Alanine
C	- Cys	-	Cysteine
D	- Asp	-	Aspartic acid
E	- Glu	-	Glutamic Acid
F	- Phe	-	Phenylalanine
G	- Gly	-	Glycine
H	- His	-	Histidine
K	- Lys	-	Lysine
L	- Leu	-	Leucine
M	- Met	-	Methionine
N	- Asn	-	Asparagine
P	- Pro	-	Proline
R	- Arg	-	Arginine
S	- Ser	-	Serine
T	- Thr	-	Threonine
V	- Val	-	Valine
Y	- Tyr	-	Tyrosine

Chapter 1:
Literature review

1.1 Introduction

Tetraviruses are small (~40 nm in diameter), non-enveloped positive-sense (+) single-stranded (ss) RNA viruses, characterized by their *T*=4 icosahedral capsid symmetry (Moore *et al.*, 1985). Originally isolated from respective Lepidopteran larvae or Lepidopteran-derived cell lines, these viruses are generally renowned for their unusual narrow host range. Tetraviruses further exhibit high tissue specificity, infecting the larval midgut cells. Infection results in shedding of these cells into the lumen of the midgut, consequently affecting larval feeding resulting in stunting and mortality (Brooks *et al.*, 2002). This has raised interest in the use of tetraviruses as bio-pesticides against respective lepidopteran agricultural pests. The Darna trima virus (DtV) has been successfully used to control field populations of *D. trima*, a pest of oil palms in Malaysia (Moore, 2001). The Helicoverpa armigera stunt virus (HaSV) is the only other tetravirus demonstrated to effectively control field populations of their host, the cotton boll worm (*H. armigera*). This is an economically detrimental pest of various crop hosts, including cotton, tomato and rice in Africa, Asia and Australia (Christian *et al.,* 2005).

1.2 Tetravirus capsid structure

The structure and assembly of tetravirus capsids, in particular Nudaurelia capensis omega virus (NωV), have been extensively studied since their discovery. A study by Olson *et al.* (1990) revealed that the NωV capsid contains 240 copies of a 71 kDa capsid precursor protein (VCAP), with each subunit exhibiting one of four closely related conformations (Fig. 1.1). These four conformations arrange to form the characteristic *T*=4 icosahedral symmetry of tetraviruses (Olson *et al.,* 1990). Each capsid subunit (A, B, C and D) consists of an internal helical domain, an eight-stranded β-barrel and an exterior immunoglobulin-like domain (Fig. 1.1) (Munshi *et al.*, 1996; Speir and Johnson, 2008).

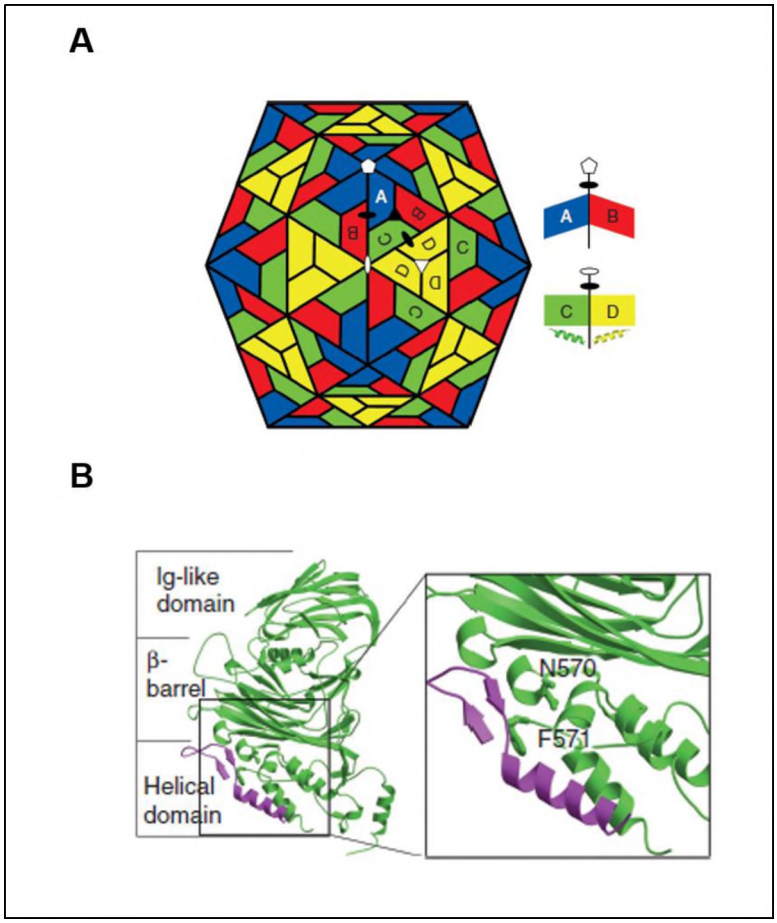

Figure 1.1: Schematic representation of NωV capsid structure. A: Subunit organization into T=4 icosahedral virus particle, with icosahedral symmetry axes (⬠, ⬭ and ▵) and quasi-symmetry indicated (⬬ and ▲). **B:** Ribbon diagram of the C subunit, consisting of the three domains. The top of the diagram represents the outer surface of the capsid subunit, exhibiting the immunoglobulin-like domain (Ig-like). Residues cleaved during maturation represented as N570 (asparagine-570) and F571 (phenylalanine-571). (Taken from Speir and Johnson, 2008).

During maturation, the NωV capsid undergoes autocatalytic cleavage within the internal helical domain, between the asparagine (N) residue at position 570 and the phenylalanine (F) at position 571 to yield 64 kDa (β) and 7 kDa (γ) proteins (Agrawal and Johnson, 1992; Hanzlik *et al.,* 1995; Canady *et al.*, 2000). Cleavage *in vitro* is triggered by a drop in pH (pH 5.0), during which the procapsid condenses and rearranges (Canady *et al.*, 2001). A drop in cytoplasmic pH, that accompanies apoptosis in yeast cells, has been shown to trigger virus-like particle maturation. It is therefore hypothesized that apoptosis would be necessary for the successful spread of infection (Tomasicchio *et al.*, 2007).

1.3 Taxonomic classification

Tetraviruses were previously classified under a single family, the *Tetraviridae*, and further divided into two genera based on the number of viral RNA (vRNA) molecules that they encapsidated. Tetraviruses with one or two vRNAs were therefore classified as beta- and omegatetraviruses, respectively (Dorrington and Short, 2010). However, phylogenetic analysis of the RNA dependent RNA polymerase (RdRp) domains of respective tetravirus replicases (REP) (Walter *et al.*, 2010; Zeddam *et al.,* 2010) have led to the reclassification of tetraviruses into three distinct families (Table 1.1). These families are the *Permutotetraviridae, Carmotetraviridae* and *Alphatetraviridae*, as their RdRp domains each resembled one of the three superfamilies of viral RdRps (Koonin, 1991; ICTV Taxonomy Release, 2011).

All viral RdRp domains exhibit palm, thumb and finger subdomains. Of the three subdomains, motifs found in the palm subdomain are predominantly conserved throughout all RdRp superfamilies (Poch *et al.,* 1989; Koonin, 1991). Motif A contains two conserved aspartic acid (D) residues, separated by four or five variable residues, whilst motif C contains a conserved D-D dipeptide generally preceded by a glycine residue (G) (Poch *et al.,* 1989). The negative charge associated with these residues aids in the binding of divalent cations (Mg^{2+} or Mn^{2+}) involved in catalysis (Gorbalenya *et al.,* 2002). Motif B of RdRps, on the other hand, contains N, which is responsible for discriminating between nucleoside triphosphates (NTPs) and deoxynucleoside triphosphates (dNTPs). This allows the REP to selectively produce vRNA from NTPs (Gohara *et al.,* 2000; Butcher *et al.,* 2001).

The canonical arrangement of these motifs of the palm subdomain is A-B-C (Poch *et al.,* 1989; Koonin, 1991). A characteristic permutated arrangement of these motifs (C-A-B) are observed within the RdRp domain of the permutotetraviruses (Gorbalenya *et al.,* 2002). Phylogenetic analysis of the RdRp domains have shown that members of this family, the Thosea asigna virus (TaV) and Euprosterna elaeasa virus (EeV), cluster with Infectious pancreatic necrosis virus (IPNV) and Infectious bursal disease virus (IBDV), both known double-stranded (ds) birnaviruses exhibiting picornavirus-like RdRp domains (Gorbalenya *et al.,* 2002; Zeddam *et al.,* 2010).

Table 1.1: Classification of known tetraviruses.

Virus species	Acronym	Host family	Geographic origin	GenBank accession no.	Reference
Family: *Permutotetraviridae* Genus: *Alphapermutotetravirus*					
Euprosterna elaeasa virus*	EeV	*Limacodidae*	South America	AF461742	Genty, 1976
Thosea asigna virus*	TaV	*Limacodidae*	Malaysia and USA	AF282930, AF062037	Reinganum *et al*., 1978
Family: *Carmotetraviridae* Genus: *Alphacarmotetravirus*					
Providence virus*	PrV	*Noctuidae*	USA	GU991616	Pringle *et al*., 2003
Family: *Alphatetraviridae* Genus: *Betatetravirus*					
Antheraea eucalypti virus	AeV	*Saturniidae*	Australia		Grace and Mercer, 1965
Darna trima virus	DtV	*Limacodidae*	Malaysia		Reinganum *et al*., 1978
Calliteara pudibunda virus	CpV	*Lymantriidae*	UK		Greenwood and Moore, 1984
Nudaurelia capensis beta virus*	NβV	*Saturniidae*	South Africa	AF102884	Hendry *et al*., 1968
Philosamia cynthia x ricini virus	PxV	*Saturniidae*	UK		Reinganum *et al*., 1978
Pseudoplusia includens virus	PiV	*Noctuidae*	USA		Chao *et al*., 1983
Trichoplusia ni virus	TnV	*Noctuidae*	USA		Morris *et al*., 1979
Genus: *Omegatetravirus*					
Dendrolimus punctatus virus*	DpTV	*Lasiocampidae*	China	RNA1–AY594352 RNA2–AY594353	Yi *et al.,* 2005
Helicoverpa armigera stunt virus*	HaSV	*Noctuidae*	Australia	RNA1–U18246, EU345431, KX423453 RNA2–AY594353	Hanzlik *et al.,* 1993
Nudaurelia capensis omega virus**	NωV	*Saturniidae*	South Africa	RNA2–DQ054400	Hendry *et al.,* 1985

*- Complete vRNA sequence has been determined; **- partial vRNA sequence has been determined.

Providence virus (PrV), however, is currently the only member of the *Carmotetraviridae* (ICTV Taxonomy Release, 2011). Originally isolated from an insect midgut cell culture line (*Helicoverpa zea* MG8), it was classified as a tetravirus due to its characteristic *T*=4 capsid symmetry (Pringle *et al.*, 2003). Its RdRp domain, however, exhibits a closer relationship to plant carmo-like viruses, namely tombus- and umbraviruses (Walter *et al.,* 2010). It is therefore hypothesized that PrV originated from a recombination event between a plant and insect virus (Dorrington *et al.*, 2011).

The REP of alphatetraviruses exhibit three conserved functional domains that are present in the alpha-like superfamily (Koonin, 1991); namely a methyltransferase (MT) responsible for attaching 5' caps to the vRNA, the helicase (HEL) that causes the unwinding of dsRNA intermediates during replication and the RdRp domain (Gordon *et al.*, 1995; Dorrington *et al.,* 2011). The *Alphatetraviridae* family is divided into one of two genera namely, *Betatetravirus* and *Omegatetravirus,* based upon the number of vRNAs encapsidated. The Nudaurelia capensis β virus (NβV) is currently the only betatetravirus for which the complete genome sequence is known, and like other pemuto- and carmotetraviruses, it has a monopartite genome (Dorrington *et al.*, 2011).

Omegatetraviruses, including the Helicoverpa armigera stunt virus (HaSV), Dendrolimus punctatus tetravirus (DpTV) and Nudaurelia capensis omega virus (NωV), are the only tetraviruses encapsidating a bipartite genome. Unlike monopartite tetraviruses, which encode both structural and non-structural proteins on a single vRNA, the omegatetraviruses encode these on respective vRNA molecules (Dorrington *et al.*, 2011). The segmented genome of omegatetraviruses has led researchers to propose an origin of a recombinatory nature. Dorrington *et al.* (2011) suggested that recombination could have occurred between two monopartite parent tetraviruses, with vRNA encoding REP from NβV (or NβV-like) and subgenomic RNA (sgRNA) encoding VCAP from PrV (or PrV-like). The sgRNA could then have evolved into a second molecule by means of mutation and/or autonomization (Dorrington *et al.*, 2011).

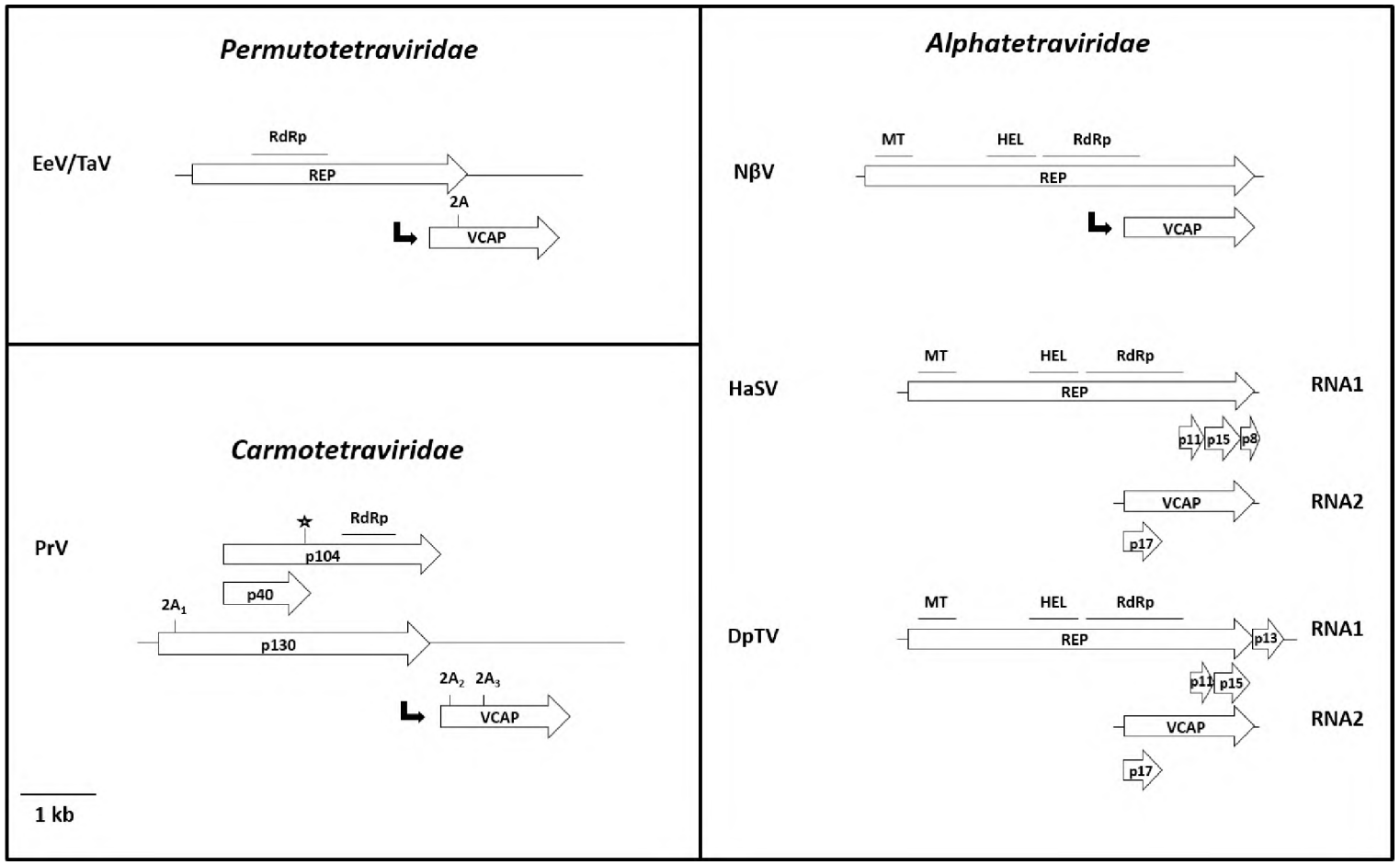

Figure 1.2: Genome organisation of the *Alphatetraviridae*, *Permutotetraviridae* and *Carmotetraviridae*. Thosea asigna virus (TaV); Euprosterna elaeasa virus (EeV); Providence virus (PrV); NβV- Nudaurelia capensis β virus; Helicoverpa armigera stunt virus (HaSV); Dendrolimus punctatus tetravirus (DpTV); VCAP- viral capsid protein precursor; MT- methyltransferase; HEL- helicase; RdRp- RNA dependent RNA polymerase; 2A- 2A-like processing sequences; ↳ - subgenomic RNA; ☆ - read-through stop codon. The scale bar represents 1 kb nucleotides. (Adapted from Walter *et al.*, 2010).

1.4 Genome organisation

1.4.1 *Permutotetraviridae*

The monopartite genome of TaV is 6.5 kb in length, whereas that of EeV is 5.7 kb. Both permutotetraviruses encode the VCAP open reading frame (ORF) overlapping with the carboxyl terminus (C-terminus) of the REP. The VCAP (82 kDa) is translated from sgRNA, and co-translationally processed by 2A-like cleavage sites present at the amino-terminus (N-teminus) of the sequence (Fig. 1.2) (Pringle *et al.,* 1999). This produces a second protein of 17 kDa in addition to VCAP, the mechanism of translation will be discussed later on in this review. The 17 kDa protein of TaV has been proposed to function non-structurally, as it has not been detected in virus particles and since recombinant baculoviruses were capable of producing virus-like particles in its absence (Pringle *et al.*, 2001). Furthermore, bioinformatic analysis of TaV and EeV vRNA indicated the presence of a 3' pseudoknot proposed to increase vRNA stability (Zeddam *et al.,* 2010).

1.4.2 *Carmotetraviridae*

PrV has the most complex genome among the tetraviruses, maximizing the genetic potential of its small genome (6.4 kb). Like TaV and EeV, PrV produces the VCAP (81 kDa) from sgRNA, however has two 2A-like processing sites ($2A_2$ and $2A_3$) present at its N-terminus (Fig. 1.2) (Pringle *et al.*, 2003). *In vitro* studies have demonstrated that approximately 99% and 94% cleavage activity for the $2A_2$ and $2A_3$ sites, respectively (Luke *et al.,* 2008). This activity was not however in the wild-type surrounding of the PrV sequence. Results obtained from coupled *in vitro* transcription-translation experiments, with these sites occurring in the native context, showed only $2A_3$ to be funtional, since a 15 kDa protein was detected in addition to the VCAP. Had $2A_2$ been active, two additional proteins of 7 kDa and 8 kDa would have been detected, as opposed to the single 15 kDa protein (Walter *et al.*, 2010). The function of this additional protein is not yet known.

The replicase ORF is expressed as two proteins due to a read-through stop codon; a 40 kDa accessory protein (p40) and the 104 kDa full length REP protein (p104) encoding the RdRp domain (Walter *et al.,* 2010). The mechanism by which read-through occurs will be described later on in this review. The p40 and p104 ORFs are located within one encoding a 130 kDa protein (p130) (Fig. 1.2), the function of which

is not known. A third 2A-like site ($2A_1$) occurs at the N-terminus of the p130 ORF (Fig. 1.2) which is predicted to produce two proteins of 17 kDa and 113 kDa. This site has been shown to exhibit 99% cleavage activity in *in vitro* studies, however, it is not known whether it does so in the native context (Luke *et al.,* 2008).

1.4.3 *Alphatetraviridae*

Unlike other tetraviruses, alphatetravirus vRNAs exhibit 5' 7-methylguanosine caps (^{m7}GpppN), and are not polyadenylated at the 3' terminus (Gordon *et al.,* 1995; Yi *et al.*, 2005). Furthermore, the vRNAs form a transfer RNA (tRNA)-like secondary structure at the 3' terminus; these structure are often found in plant virus RNAs that are not polyadenylated (Gordon *et al.*, 1999; Gordon *et al.,* 1995; Yi *et al.*, 2005). In contrast to the plant tRNA-like structures, the tRNA-like structures of alphatetraviruses that encode a valine (Val) anti-codon without a pseudoknot. This makes these structures more similar to those of cellular $tRNA^{Val}$ than those that are found in plant viruses (Gordon *et al.*, 1995). It is not known whether these structures are molecular fossils, or have a function in alphatetraviruses replication.

NβV, the only sequenced betatetravirus, has the simplest genome among the tetraviruses, encoding only the REP and VCAP on the 6.6 kb monopartite genome. It produces VCAP from sgRNA like other monopartite tetraviruses, however does not encode any 2A-like cleavage sites (Fig. 1.2) (Dorrington *et al.*, 2011). The genomes of omegatetraviruses are more complex, with characteristic bipartite genomes consisting of the larger vRNA1 (~5.3 kb) and the smaller vRNA2 (~2.5 kb) encoding the non-structural and structural proteins, respectively (Fig. 1.2).

HaSV and DpTV are currently the only classified members of the omegatetravirus genus for which the complete vRNA1 sequence is known (Hanzlik *et al.,* 1995; Yi *et al.*, 2005). While the sequence for vRNA2 of NωV is known, only a partial sequence for RNA1 is available (Agrawal and Johnson, 1992). Omegatetraviruses produce VCAP from vRNA2, which further encodes a 17 kDa protein partially overlapping with the VCAP ORF (Hanzlik *et al.,* 1995; Yi *et al.,* 2005; du Plessis *et al.,* 2005). This protein is not essential for the assembly of virus-like particles, though it has been shown to have a role in the selective packaging of vRNA into the virus particle (Hanzlik *et al.*, 1995; Mendes *et al.,* 2015).

The larger genomic vRNA1 encodes the 186 kDa REP (Gordon *et al.*, 1995; Yi *et al.*, 2005). A study by Zhou *et al.* (2006) demonstrated that the DpTV RdRp domain heterologously expressed in *Escherichia coli* BL21 (DE3), was capable of initiating primer-independent synthesis of vRNA *in vitro*. Overlapping with the DpTV and HaSV REP ORF, at the 3' end of vRNA1, are two small putative ORFs encoding proteins of 11 kDa (p11) and 15 kDa (p15) (Fig. 1.2) arranged in tandem (Gordon *et al.,* 1995; Yi *et al.*, 2005). It is tempting to speculate the presence of the p11 and p15 ORFs in NωV as well. Furthermore, both DpTV and HaSV encode a third overlapping ORF downstream of the aforementioned. DpTV encodes a 13 kDa protein (p13) in a +1 frame relative to p15, whereas HaSV encodes an 8 kDa protein (p8) in frame with p15 (Fig. 1.2). The small ORFs will be discussed in further detail later on in the review.

1.5 Tetravirus replication biology

1.5.1 Providence virus

Until recently, very little was known about the infectious life cycle of the tetraviruses. This was due to practical difficulties associated with conducting virus replication studies *in vivo*, and the lack of tissue culture cell lines susceptible to tetravirus infection (Dorrington and Short, 2010). Until recently, only PrV has been known to infect and replicate in cell culture (Pringle *et al.,* 2003). PrV was first isolated from a persistently infected the *H. zea* (MG8) cell line, and was shown to infect and replicate in *H. zea* fat body (FB33), *Spodoptera exigua* (Se-1) and human cervical cancer (HeLa) cells (Pringle *et al.*, 2003; Jiwaji *et al.,* 2016). *In vitro* data has further shown PrV to be capable of replication in plant (wheatgerm), insect and mammalian (rabbit reticulocyte) cell-free systems (Jiwaji *et al.,* 2016).

The PrV REP ORF is expressed as two proteins due to a read-through stop codon; an accessory protein (p40) and the full length REP protein (p104) encoding the RdRp domain (Walter *et al.,* 2010). The read-through stop sequence (**UAG**CAACUA) is classified as a group 1 read-through stop codon which has the consensus sequence (**UAG**CARYYA) (Harrell *et al.,* 2002). In *H. zea* MG8 cells, the p40 and the REP of PrV associates with detergent resistant membranes (DRMs) (Short and Dorrington, 2012). Immunofluorescence studies using anti-p40 antibodies demonstrated that the p40/p104 proteins co-localizes with dsRNA, leading to the conclusion that they form part of the replication complex (Short *et al.,* 2013). Using p104 fused to the enhanced

green fluorescent protein (EGFP) at the N-terminus, Short *et al.* (2013) showed that the p104 protein formed punctate structures in the cytoplasm, and localized to the Golgi apparatus and secretory vesicles in *Spodoptera frugiperda* (Sf9) cells. PrV replication is therefore proposed to occur in these organelles.

1.5.2 Helicoverpa armigera stunt virus

The HaSV REP has been shown to associate with DRMs, localizing to punctate structures throughout the cytoplasm of *S. frugiperda* (Sf9) cells (Short *et al.,* 2010; Short *et al.*, 2012). By expressing REP with EGFP fused at its C-terminus Short *et al.* (2010) showed that the HaSV REP co-localized with a marker for the endocytic pathway. Subsequent experiments have shown that the HaSV REP also localizes to the Golgi apparatus and secretory vesicles (James Short, unpublished data). This data suggests that PrV and HaSV RNA replication, and potentially other tetraviral replication, takes place in association with vesicles of the secretory pathway in the cytosol of the cells. The exact mechanism of replication and virus-host interactions have yet to be elucidated.

A recent study by Penkler *et al.* (2016) demonstrated that HaSV infection of *H. armigera* larvae is pH dependent, with structural changes triggered under alkaline conditions (pH9) required for binding of virus particles to host cells. Importantly, Penkler *et al.* (2016) were able to demonstrate that HaSV particles pre-incubated at pH 9.0 were able to bind and initiate a productive infection in *S. frugiperda* (Sf9) tissue culture cells. This new discovery creates an ideal opportunity for future transfection-independent studies on the replication biology of alphatetraviruses.

1.6 Small ORFs of omegatetraviruses

In DpTV, the ORFs for p11 and p15 encode proteins of 98 amino acids (aa) and 135 aa, respectively. When compared to HaSV, the small ORFs of p11 and p15 encode slightly longer proteins of 99 aa and 140 aa respectively (Fig. 1.4). Apart from their similar sizes, the position of the p11 and p15 ORFs are conserved in both HaSV and DpTV (Fig. 1.4) (Dorrington and Short, 2010). Furthermore, between HaSV and DpTV, p11 and p15 have a 68.7% and 53.4% amino acid identity. Overall, the p11-p15 region are 59.3% identical between these two viruses (Hughes, 2013). The conservation of the position and amino acid sequence of the p11 and p15 ORFs suggest that these

proteins may function in the infectious virus lifecycles of HaSV and DpTV (Dorrington and Short, 2010; Hughes, 2013). However, it is not known whether these proteins are indeed expressed in virus-infected cells and if so, by what mechanism they might be translated.

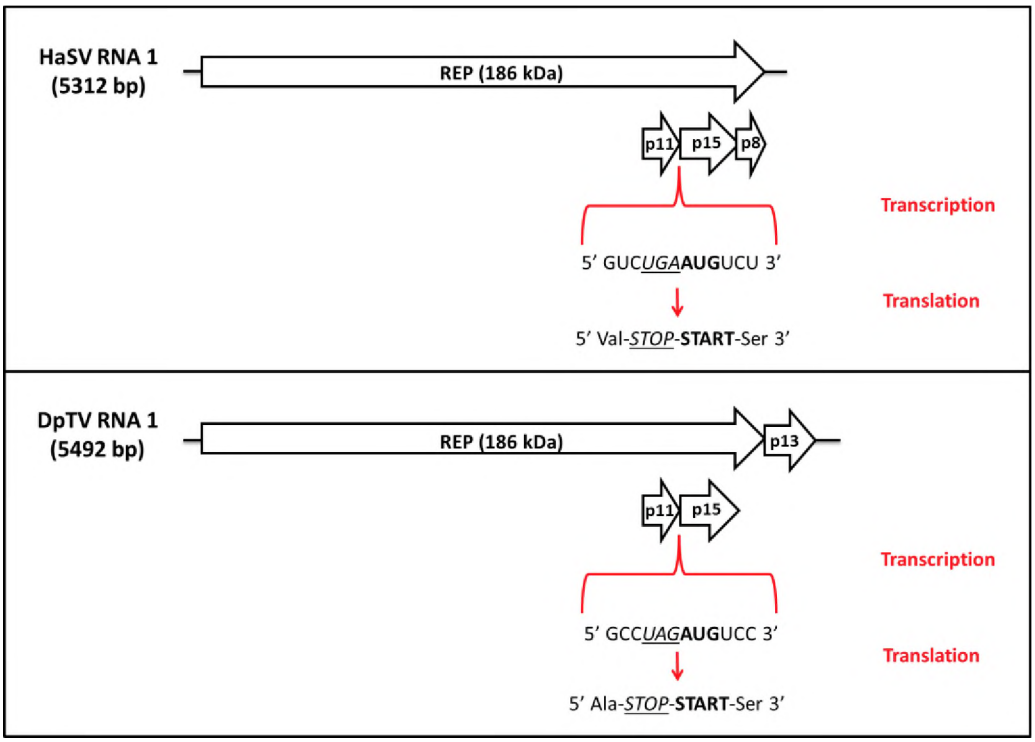

Figure 1.3: Schematic representation showing the organization of vRNA1 of Omegatetraviruses. Sequences in **bold** - start codons for initiation of translation; Sequences *underlined + italicized* – stop codons for termination of translation; Val – valine; Ser – serine; Ala – alanine; HaSV - Helicoverpa armigera stunt virus; DpTV - Dendrolimus punctatus tetravirus. (Adapted from Yi *et al.*, 2005).

Another conserved feature of p11 in both HaSV and DpTV is a predicted transmembrane domain. In HaSV, this region spans amino acid residues 52-69 whereas in DpTV it spans residues 54-72. While the predicted transmembrane domains occurs in a similar position in both HaSV and DpTV, the sequences share only an 18% identity and 36% similarity (Hughes, 2013). While there are many similarities in the p11 and p15 ORFs in DpTV and HaSV, there is a difference in the stop codon that separates the p11 and p15 ORFs. The stop codon in DpTV is UAG and in HaSV it is UGA (Fig. 1.3). The effect of this change on the expression of these small proteins during infection is currently not known.

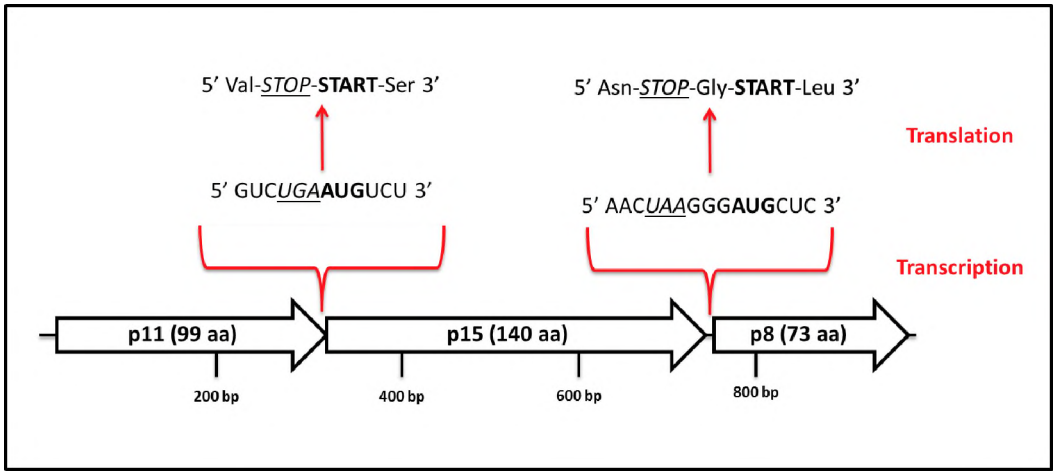

Figure 1.4: Schematic representation showing the organization of the three putative ORFs on vRNA1 of HaSV. Sequences in **bold** - start codons for initiation of translation; Sequences *underlined + italicized* – stop codons for termination of translation; Val – valine; Ser – serine; Asn –asparagine; Gly – glycine; Leu – leucine. (Adapted from: Gordon *et al.*, 1995).

The expression of the small ORFs of HaSV are, however, the focus of this study. If these are in fact expressed during infection, a non-canonical mechanism of translation would likely be required as the stop codon of p11 and start codon p15 occur immediately adjacent to each other, and that of p15 and p8 separated by a glycine inter codon (GGG) (Fig. 1.4). The mechanisms by which the small ORFs may be translated are therefore discussed below.

1.7 Translation initiation mechanisms employed by viruses

All viruses are known for their obligate intracellular parasitic nature as well as their reliance on host translation machinery for replication. Viruses have consequently evolved mechanisms to compete with the host cell for the resources that are required for translation. This section of the literature review focuses on mechanisms employed by RNA viruses to evade the host immune system and to exploit host cell initiation, elongation and termination to enhance the expression of viral proteins.

1.7.1 Initiation of translation

During canonical cap-dependent eukaryotic translation, proteins of the eukaryotic initiation factor (eIF4F) complex are the first to interact with nascent messenger RNA (mRNA) (Fig. 1.5A; reviewed in Jackson *et al.*, 2010). This complex consists of eIF4G, a cap-binding protein (eIF4E) that associates with the N-terminus of eIF4G, and an

ATP-dependent RNA HEL (eIF4A) that associates with an internal region of eIF4G (Fig. 1.5A). The poly(A)-binding protein (PABP) is responsible for binding the poly(A) tail of mRNA, and further interacts with eIF4G to form a closed loop translation complex. This complex serves to assess mRNA integrity and promotes cap and poly(A)-dependent translation (Imataka *et al.,* 2000). The circularized mRNA complex then associates to the 43S pre-initiation complex via eIF3, which binds directly to eIF4G (Schneider and Mohr, 2003).

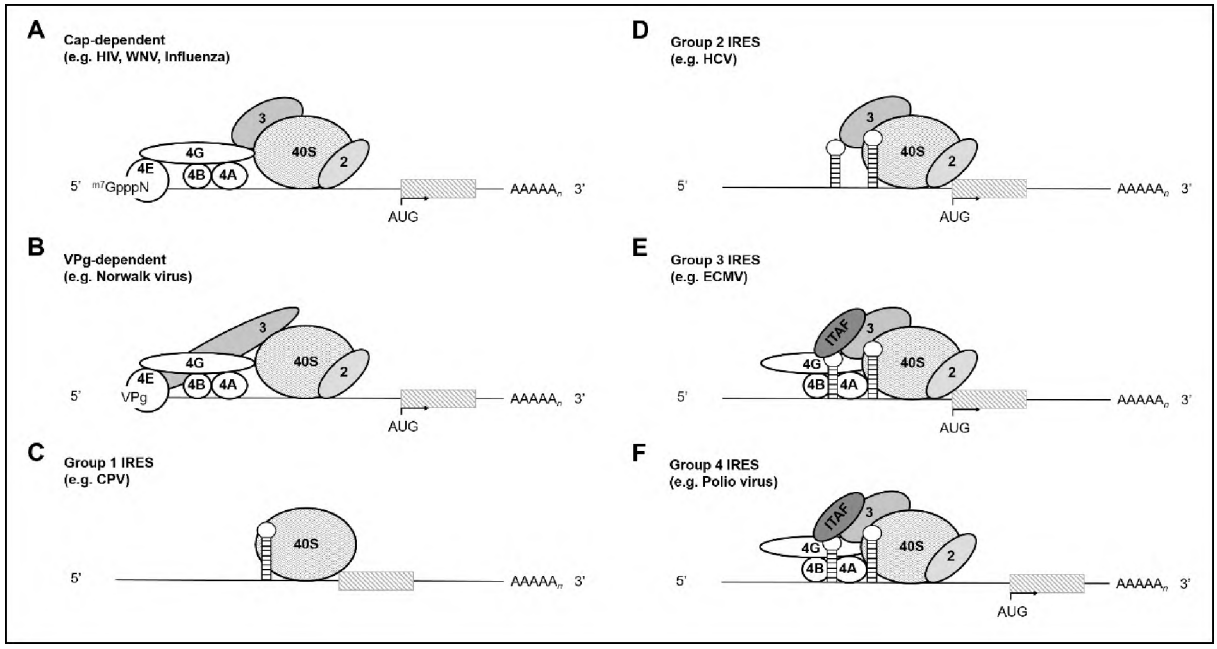

Figure 1.5: Comparison of cap-dependent, VPg-dependent and IRES initiation of translation. The 7-methylguanosine cap (^{m7}GpppN) is recognized by eIF4E in cap-dependent initiation of translation. The viral genome-linked protein (VPg) is used by viruses to recruit eIF4E to initiate cap-independent translation. Internal ribosome entry sites (IRESs) are also employed by viruses to initiate cap-independent translation, within a RNA molecule. There are four classes of IRESs, each able to recruit the 40S ribosome subunit, and different initiation factors where needed. ITAF- IRES trans-activating factors; HIV- Human Immunodeficiency virus; WNV- West Nile virus; ECMV- Encephalomyocarditis virus; HCV- Hepatitis C virus and CPV- Cricket paralysis virus. Patterned boxes represent ORFs translated (Adapted from Picard-Jean *et al.*, 2013).

Viruses have evolved strategies to mimic the structure of cellular mRNA, with a 5' cap (^{m7}GpppN) and 3' poly(A) tail, using the 5' cap and 3' tail to aid in the evasion of the host immune system. These features further enable and enhance translation of vRNA by the host cell machinery, however viruses rarely exhibit both. The (+)-ssRNA viruses that have both mechanisms include the alphaviruses and the hepeviruses (Decroly *et*

al., 2012). Other viruses have either the 5' cap with no poly(A) tail (e.g. flaviviruses and tetraviruses) or a poly(A) tail but no 5' cap (e.g. picornaviruses) (Picard-Jean *et al.*, 2013).

While some viruses exploit cap-dependent translation, they do not rely solely on this mechanism. Cellular mRNAs are monocistronic, so in general, one mRNA expresses one protein (Schneider and Mohr, 2003). To optimize their coding capacity, as they have limited genome sizes, viruses have evolved mechanisms to express multiple protein from a single vRNA (Fig. 1.6). To initiate translation in a cap-independent manner, viruses have evolved cis-acting RNA elements to circumvent the need for a 5' cap. This includes the use of covalently linked 5' viral genome-linked protein (VPg), 3'-cap-independent translation enhancer (3' CITE) or internal ribosomal entry sites (IRESs) to initiate viral protein translation (Walsh *et al.,* 2013; Picard-Jean *et al.*, 2013).

Members of the *Caliciviridae* and *Astroviridae* possess VPgs at the 5' terminus of their vRNA. The VPg of these viruses, acting as a cap substitute, interacts with eIF4E and eIF3 allowing initiation of vRNA translation (Fig. 1.5B) (Firth and Brierley, 2012). Like VPgs, 3' CITEs (present in Turnip crinkle virus) also bind initiation factors, and further interact with the 5' untranslated region (UTR), to place these in close proximity (Walsh *et al.*, 2013). IRESs, on the other hand, are complex RNA structures exhibiting characteristic multiple stem-loop structures that are involved in recruiting ribosomes to regions within the mRNA (reviewed in Picard-Jean *et al.*, 2013). IRESs are not unique to viruses as they have been found on some eukaryotic cellular mRNAs (Komar and Hatzoglou, 2011).

1.7.1.1 Internal ribosome entry sites

Viruses employing IRESs often have a long 5' UTR and lack a 5' cap (Fig. 1.5C-F) rendering them incompatible with cap-dependent translation (Kieft, 2008). As the 5' UTRs of vRNA often encode signals for replication and packaging, it is advantageous for the virus to use an IRES mechanism as it does not mask the signals at the 5' terminus (Schneider and Mohr, 2003). IRESs are mainly classified into four groups based on their structure, length and mechanism. Group 1 IRESs (found in Cricket paralysis virus) forms a structure that is 180 nt in length, which recruits the 40S ribosome subunit to the translation start site (Kieft, 2008). The group 1 IRESs are

considered the simplest IRES as no initiation factors or initiator methionyl-tRNAs (Met-tRNA$_i$) are required to initiate translation (Fig. 1.5C). Group 2 IRESs (that are present in the flaviviruses) are larger at 330 nt and are also capable of binding the 40S subunit directly (Jackson, 2005). These do, however, require eIF2, eIF3 and Met-tRNA$_i$ in order to initiate translation (Fig. 1.5D).

Group 3 and 4 IRESs represent some of the most complex IRESs, at 450 nt in length. Viruses exhibiting Group 3 IRESs include cardioviruses (e.g. Encephalomyocarditis virus), whereas enteroviruses (e.g. Polio virus) are known to exhibit Group 4 IRESs (Picard-Jean *et al.*, 2013). These IRESs cannot bind the 40S subunit directly, unlike the group 1 and 2 IRESs, and requires all eIFs (except eIF4E) and a Met-tRNA$_i$ (that is used in cap-dependent initiation of translation). In addition, the group 3 and group 4 IRESs require IRES trans-activating factors (ITAFs) (Komar and Hatzoglou, 2010). If the IRES recruits the 40S subunit directly to the AUG-start site, it is classified as a group 3 IRES, and if it does so upstream of the AUG-start site, it is classified as a group 4 IRES (Fig. 1.5E and F). The latter requires scanning or shunting of the ribosome to the AUG-start site for the initiation of translation to occur (Picard-Jean *et al.*, 2013).

1.7.1.2 Leaky scanning

Leaky scanning is commonly used by RNA viruses to translate multi-cistronic vRNAs, including the plant tombus- and umbraviruses, and the animal calici- and hepeviruses (Firth and Brierley, 2012). During leaky scanning, the ribosome does not recognize the first AUG as a start codon and an alternative start codon is recognized further downstream (Ryabova *et al.,* 2006). As a result of this process, many truncated and/or distinct proteins are produced from overlapping and non-overlapping ORFs (Fig. 1.6).

The ribosome is capable of scanning a large number of nucleotides to locate a start codon. The largest distance that has been demonstrated thus far is a distance of 895 nt in Rice tungro bacilliform tungrovirus (Fütterer *et al.,* 1997). Although the initiation of translation often uses an AUG start codon, other codons can also be recognized as initiation codons. Codons that are known to serve as start sites include CUG, GUG, ACG, AUU, AUA, AUC and UUG, with CUG exhibiting the greatest efficiency (Touriol *et al.,* 2003).

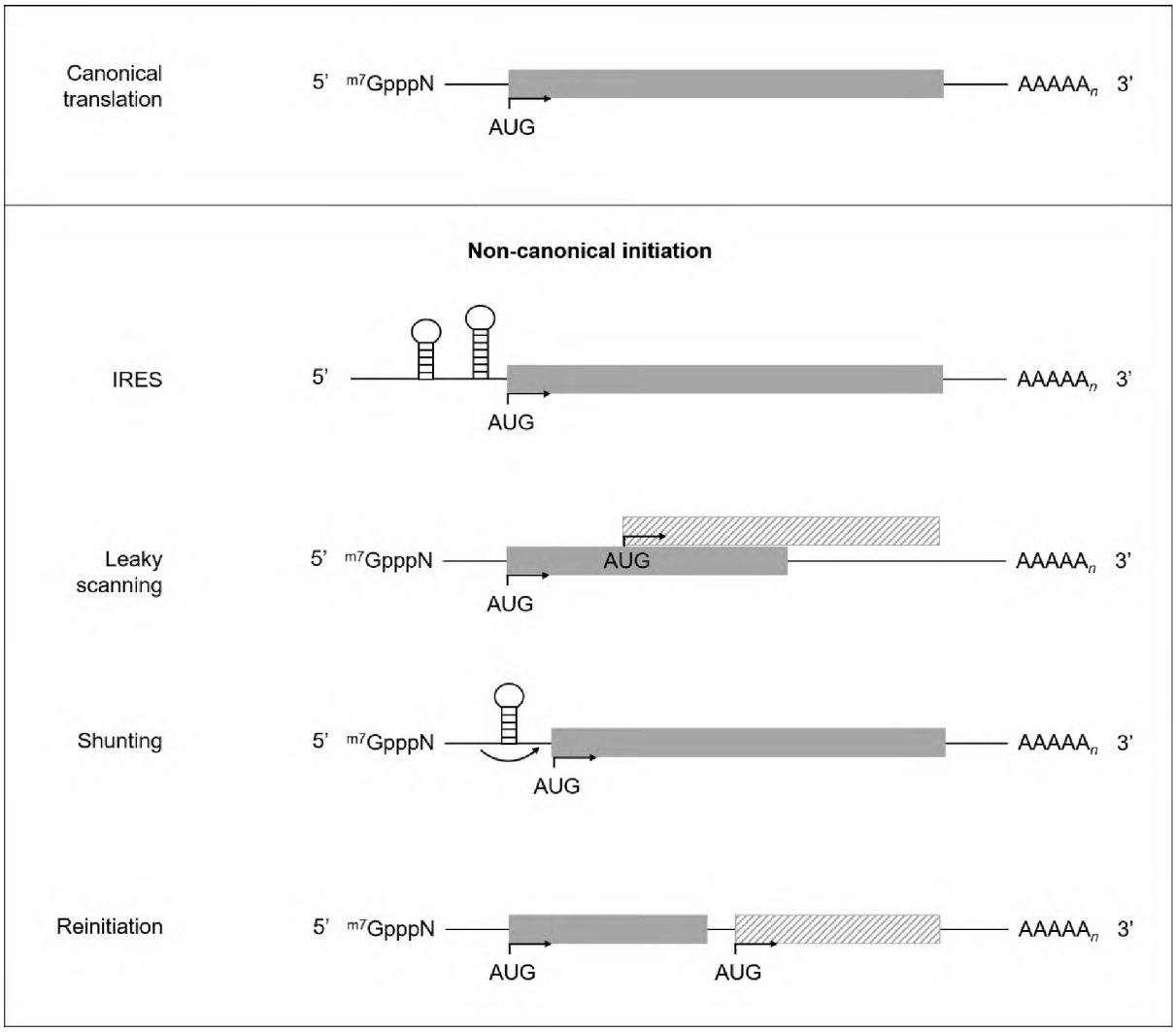

Figure 1.6: Mechanisms of non-canonical initiation of translation employed by RNA viruses. Arrows depict the initiation of translation and direction of ribosome movement. Solid blocks depict the predominantly translated ORF, whereas patterned blocks depict either additional ORFs translated or carboxyl-terminal extensions of the predominant ORF. (Adapted from Firth and Brierley, 2012).

1.7.1.3 Ribosome shunting

Unlike IRESs, ribosome shunting is 5'-end-dependent and allows for scanning-independent translation of downstream ORFs. This is achieved by the ribosome physically shunting or translocating over the 5' UTR, due to a stem-loop structure near the start codon (Fig. 1.6). RNA viruses that utilize shunting include the (-)-ssRNA Sendai respirovirus and the dsRNA Nelson bay- and Avian orthoreovirus (de Breyne *et al*., 2004; Racine and Duncan, 2010). This mechanism of initiation was, however, first described in the pararetrovirus (dsDNA-RT) Cauliflower mosaic virus (CaMV) where a 600 nt leader sequence forms a complex stem-loop structure upstream of the

ORF to be translated (Schmidt-Puchta et al., 1997). In this example, the 80S ribosome dissociates from the mRNA after translation of a short upstream ORF, leaving the 40S subunit transiently associated therewith (by retaining certain eIFs). This subunit then translocates to the proximal mRNA at the 3' end of the stem-loop structure to initiate translation of the downstream ORF (Ryabova *et al.,* 2006).

1.7.1.4 Reinitiation

As aforementioned, upon termination of translation of larger ORFs, the ribosomal subunits and eIFs dissociate from the mRNA. However, after translation of small ORFs (that are smaller than 30 codons in length), the 40S ribosome subunit does not immediately dissociate from the RNA as it would for a large ORF. Instead, the 40S subunit continues to scan until another start codon is reached (Ryabova *et al.,* 2006). In the case of reinitiation after translation of larger ORFs, special signals or specific trans-acting protein factors are required (Firth and Brierley, 2012). Members of the (+)-ssRNA *Caliciviridae,* including noroviruses, are known to employ reinitiation strategies during translation. These viruses produce a minor capsid protein VP2 with a specific site, a termination upstream ribosomal binding site (TURBS). TURBS are often 40-90 nts long, and contains a short sequence motif (motif 1; UGGGA) that is complementary to the RNA component of the 40S subunit. The ribosome therefore remains associated with the RNA after termination. Recruitment of eIFs to re-form the pre-initiation complex is, however, required before translation of another ORF can commence. The downstream ORF is therefore not located immediately next to the previously translated ORF (Firth and Brierley, 2012). Reinitiation differs from leaky scanning in that it does not overlook start codons and it follows after the termination of translation.

1.7.2 Elongation and termination of translation

The elongation step of canonical translation entails a peptidyl tRNA occupying the P-site of the 80S ribosome and an empty A-site, that awaits an aminoacyl tRNA. Once the aminoacyl tRNA binds to the codon exposed in the A-site, a peptide bond is formed between both tRNA molecules, catalysed by the ribosomal peptidyl transferase enzyme (reviewed in Jackson *et al.,* 2010). During this process, the ribosome continues moving in a 5' to 3' direction with respect to the mRNA. As a result, the peptidyl tRNA now moves into the E-site and the aminoacyl tRNA to the P-site, leaving the A-site empty again, ready to accept new aminoacyl tRNAs. This process is

repeated until a stop codon (UGA/UAG/UAA) is encountered (Jackson *et al.*, 2010). Viruses may employ mechanisms that alter canonical elongation and termination to enhance their protein translation. The mechanisms that have been described to date include frameshifting, stop codon read through and the processing of 2A sites (Schneider and Mohr, 2003).

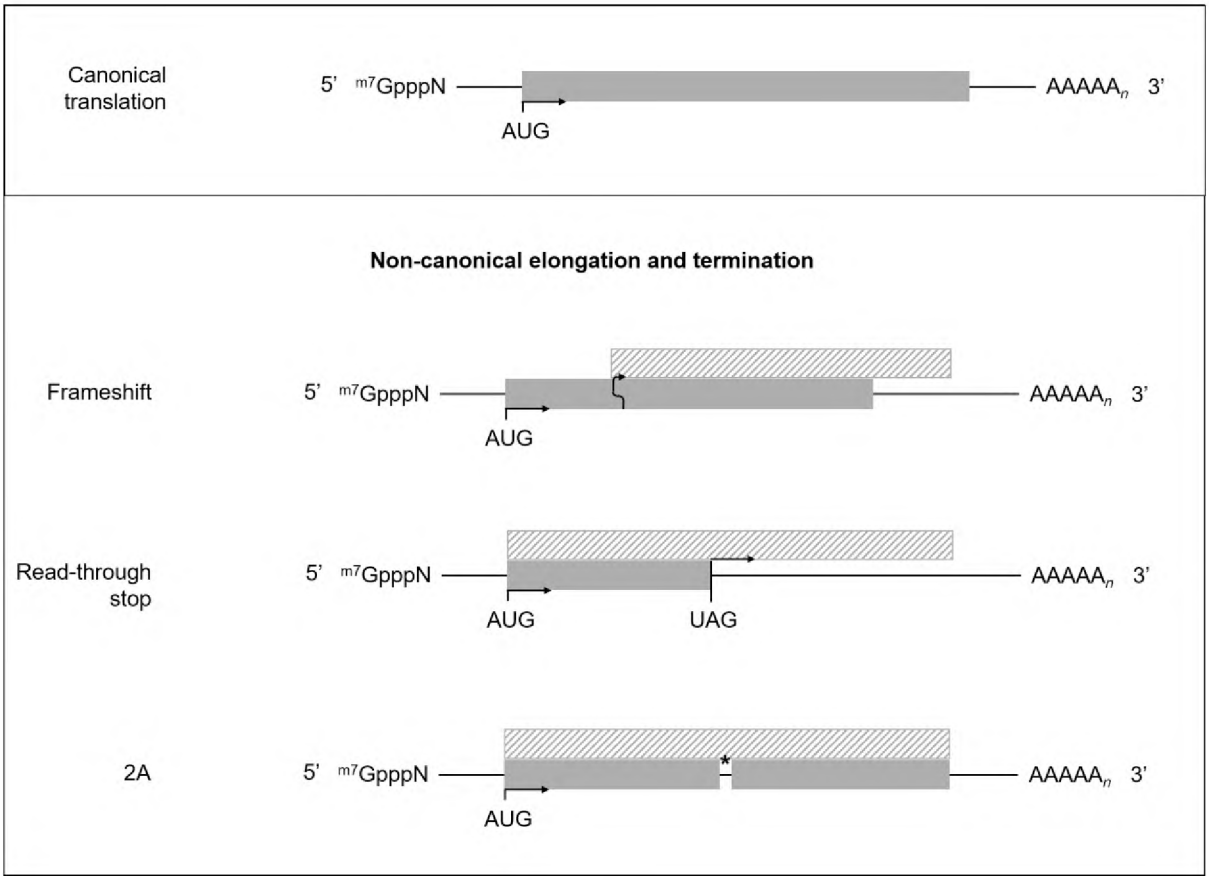

Figure 1.7: Mechanisms of non-canonical translation elongation and termination employed by RNA viruses. Arrows depict the initiation of protein synthesis and/or continuation of translation. * - 2A processing site. Solid blocks depict the predominantly translated ORF, whereas patterned blocks depict either additional ORFs translated or carboxyl-terminal extensions of the predominant ORF. (Adapted from Firth and Brierley, 2012).

1.7.2.1 Frameshifting

Frameshifting occurs at both transcriptional and translational levels. The RdRps of non-segmented (-)-RNA viruses, for example measles (Subfamily *Paramyxovirinae*) and Ebola (Family *Filoviridae*), are capable of "slipping" during transcription of vRNA resulting in the addition or removal of bases (Atkins *et al.,* 2016). Translational (or ribosomal) frameshifting involves the shifting of the ribosome during translation to

change the reading frame of the protein; and then continue translation to produce a different protein in the new frame (Fig. 1.7). This usually occurs when the ribosome encounters a secondary structure and/or a slippery sequence as stimulatory elements. When the ribosome encounters a hairpin or pseudoknot structure, it pauses briefly and is physically shifted backwards (Fig. 1.8A) (Plant, 2012). Slippery sequences, on the other hand, are heptanucleotide sequences with the sequence X_XXY_YYZ where XXX represents any three identical nucleotides, Y an A/U, and Z an A/C/U (Giedroc and Cornish, 2009; Firth and Brierley, 2012).

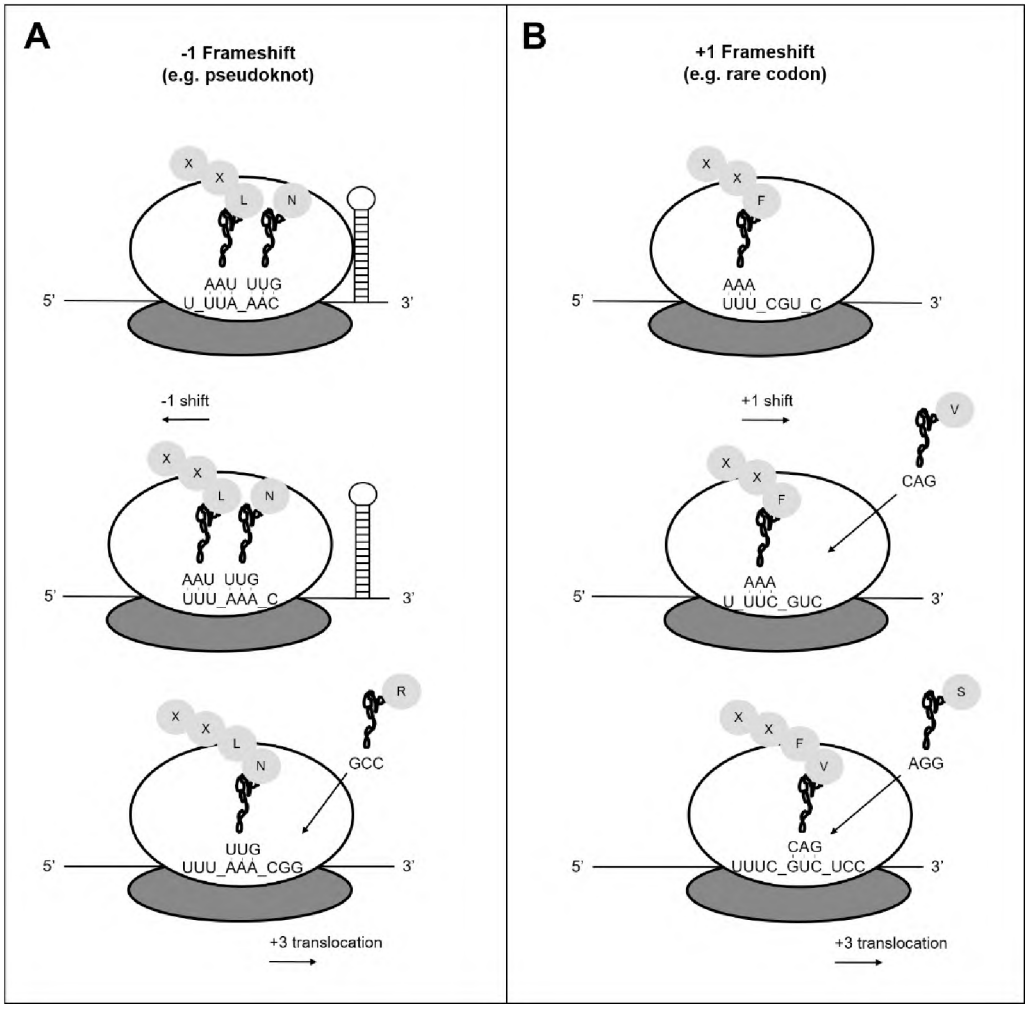

Figure 1.8: Schematic representation of ribosomal frameshifting. The 80S ribosomal subunit is shown, translating mRNA using amino acid-linked tRNAs. **Panel A**: Ribosome shifts back a frame (-1) as it encounters a pseudoknot. **Panel B**: Ribosome shifts forward a frame (+1) as it encounters a slippery sequence. After shifting, the ribosome then translocates (+3) to introduce a new mRNA codon into the A site, and further continue with elongation. X – any amino acid; L – leucine; N – asparagine; R – arginine; F – phenylalanine; V – valine; S –serine. (Adapted from http://viralzone.expasy.org/all_by_protein/860.html).

When the slippery sequence codes for rare tRNA codons (for example CGU in Influenza A virus), the ribosome is thought to pause for some time. During this time, the ribosome occasionally advances forward resulting in a +1 frame shift (Fig. 1.8B) (Atkins *et al.,* 2016). It has also been shown that the incoming rare aminoacyl tRNA is capable of binding out of frame when the slippery sequence is encountered affecting the frame of the protein product (Plant, 2012). Viruses utilizing translational frameshifting often use both stimulatory elements for optimal expression. Translational frameshifting is used by many (+)-ssRNA viruses to express their RdRps (for example the umbraviruses) as well as other proteins, such as NS1 of Japanese encephalitis virus, TF of alphaviruses or 2B of picornaviruses. Porcine reproductive and respiratory syndrome virus (Family *Arteriviridae)* is a (+)-ssRNA that exhibits a -2 frame shift in the production of its accessory proteins (from ORF1TF) in a frame different to the RdRp (Fang *et al*., 2012).

1.7.2.2 Read-through stop codons

This mechanism entails the non-canonical translation of one of the three stop codons, during which a non-cognate tRNA base pairs with the stop codon at the acceptor-site of the ribosome (Murgola *et al.,* 1985). By doing so, the RNA out-competes binding of the ribosomal release factor. Consequently, an extended polypeptide is produced (Fig. 1.7). As discussed in Section 1.4.2, the PrV REP ORF is expressed as two proteins due to read-through of a UAG stop codon (Walter *et al.,* 2010).

Base pairing of the mRNA and tRNA can occur in a conventional or unconventional manner, depending on the tRNA composition. Conventionally, mutations in the anticodon sequence result in a Watson-Crick pair (A-T/U, G-C). Unconventional base pairing on the other hand, occurs by natural non-sense suppression during which the stop codon is read as a sense codon by normal cellular tRNAs (Murgola *et al.,* 1985; Eggertsson and Söll*,* 1988). The exact mechanism by which this natural suppression occurs is not known. Standard amino acids may be inserted for UAA (glutamine) (Kuchino *et al*., 1987), UAG (tyrosine, glutamine, leucine) (Beinz and Kubli*,* 1981; Kuchino *et al*., 1987; Valle *et al*., 1987) or UGA (tryptophan, cysteine, arginine) (Hirsch, 1971; Feng *et al*., 1990; Baum and Beier*,* 1998). In addition, non-standard amino acids can be inserted for UAG (pyrrolysine) (Srinivasan *et al.,* 2002) or UGA (selenocysteine) (Berry *et al.,* 1991).

The efficiency of the read-through is affected by the downstream nucleotide sequence (especially the 3' adjacent nucleotide), as well as secondary RNA structures such as pseudoknots. According to the surrounding sequence homology, these codons are classified as type I, II, or III (Firth and Brierley, 2012). Type I (UAA, UAG) and II (UAA, UGA), read-through stop codons are followed by a cytidine (C) nucleotide and are known as weak stop codons. It should be noted that the UAA stop codon read-through is known to occur less than UAG or UGA. The 3' adjacent codon distinguishes type I (glutamine, CAA/CAG) from type II (CGG/CUA) (reviewed in Beier *et al.,* 2001). Furthermore, type I read-through stop codon sequences contain the consensus CARYYA where Y represents a pyrimidine and R representing a purine (Skuzeski *et al.*, 1991). Type III read-through stop codons (UAG, UGA) display a guanine after the stop codon, with the 3' adjacent codon most often encoding either for glycine (GGG) or valine (GUA) (reviewed in Beier *et al.,* 2001).

1.7.2.3 2A processing sites

This mechanism is mediated by an amino acid motif, D(V/I)EXNPGP, where X represents any amino acid that is positioned downstream of other amino acids making the 2A processing site, on average, 30 aa in length (Brown and Ryan, 2010). When the ribosome translates the vRNA, the formation of a peptide bond between the last two residues of this motif (GP) is inhibited with 90 to 99% efficiency (Luke *et al,* 2008). As result, two discrete proteins are produced from a single ORF. One of the proteins has G at its C-terminus and the second protein a P at its N-terminus. An example is seen in the translation of proteins 2A-2B of cardioviruses, and p17-VCAP of tetraviruses (Luke *et al.*, 2008; Walter *et al.,* 2010). Also, as discussed in Section 1.4.2, PrV encodes three 2A-like processing sites, two of which known to be active. The peptide bond can still form between these residues, however this occurs with between 1 to 10% efficiency. An example is seen in the production of the p130 protein of PrV, whereas "cleavage" products p17 and p113 are predominantly produced (Walter *et al.,* 2010).

1.8 Translation of subgenomic RNA

All of the mechanisms discussed thus far can occur on vRNA as well as sgRNA. The latter is produced from the vRNA during replication by one of three possible

mechanisms: internal initiation, premature termination or discontinuous transcription (Miller and Koev, 2000).

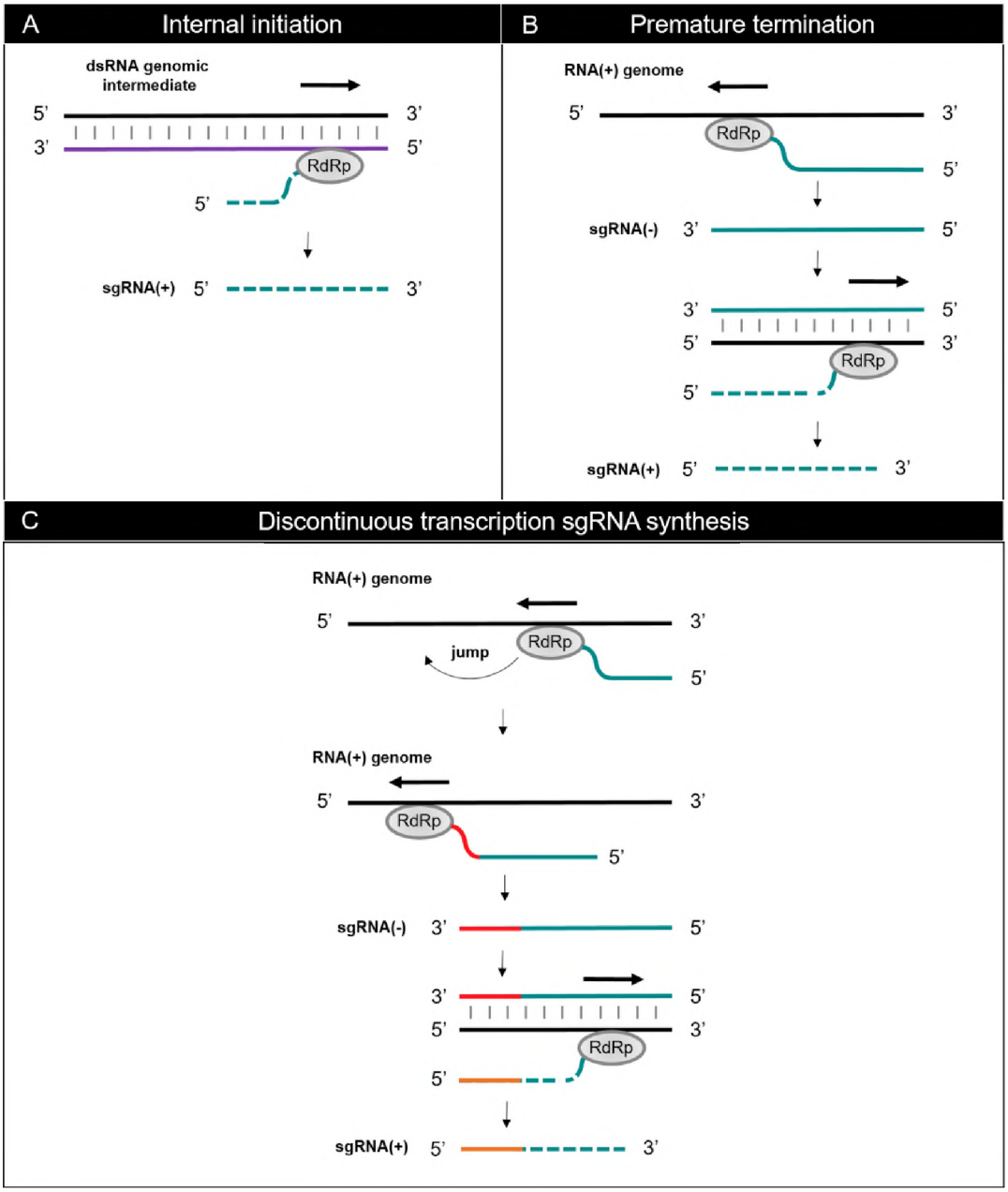

Figure 1.9: Schematic representation of mechanisms for subgenomic RNA synthesis. (A) Internal initiation, during which the viral RdRp initiates synthesis of sgRNA during replication of genomic RNA. **(B)** Premature termination, as per the name, involves early termination of (-)-RNA synthesis. This (-)-RNA then serves as template for (+)-sgRNA synthesis. **(C)** Discontinuous template synthesis occurs by an altered mechanism premature termination, involving shunting of the RdRp. (Adapted from http://viralzone.expasy.org/all_by_species/1876.html)

The sgRNAs of (+)-ssRNA viruses generally express structural or movement proteins required in intermediate and later stages of infection. As mentioned earlier, monopartite tetraviruses are known to produce sgRNA for the translation of capsid proteins (Gordon *et al.,* 1999; Pringle *et al.*, 2003). During internal initiation, the REP initiates transcription of (+)-sgRNA from (-)-vRNA. This is initiated when the RdRp recognises and then binds the promoter sequence to initiate transcription (Fig. 1.9) (Miller and Koev, 2000). An example of this is seen in the expression of proteins B1 and B2 of the *Nodaviridae.* These ORFs overlap with each other and the RdRp (protein A) at the 3' terminus of vRNA1 (Chen *et al.,* 2009). Premature termination, however, occurs during (-)-vRNA synthesis to form a subgenomic (-)-vRNA strand which later acts as template for end-to-end (+) strand synthesis (Fig. 1.9) (White, 2002). An example of this is observed in the expression of the sgRNAs which respectively encode the movement protein and the viral suppressor of vRNA silencing of the Tomato bushy stunt virus (*Tombusviridae*) (Choi *et al.*, 2001).

The third mechanism uses discontinuous RNA synthesis while generating the (−)-strand RNA templates (Fig. 1.9), and occurs by the premature termination model where the RdRp is shunted (Miller and Koev, 2000). This mechanism has been described in viruses of the order *Nidovirales* (Gorbalenya *et al.,* 2006).

1.9 Problem statement

The three small putative ORFs of HaSV, p11, p15 and p8, are hypothesized to function in the infectious virus life cycle, since the size and position of the p11 and p15 ORFs is conserved in both HaSV and DpTV. There is, however, no evidence to show that these three proteins are expressed in virus-infected insect cells. This study aimed to investigate the expression of these three ORFs in HaSV-infected *H. armigera* larvae.

1.10 Objectives

1. Investigate the sequence conservation of the p11, p15 and p8 ORFs in HaSV
2. Determine whether p11, p15 and p8 are expressed in HaSV-infected *H. armigera* larvae
3. Investigate the mechanisms by which p11, p15 and p8 may be expressed and their potential function in the infectious viral lifecycle

Chapter 2:
Materials and methods

2.1 Bacterial culture conditions

All *Escherichia coli* cells were grown overnight at 37°C on Luria Bertani (LB) agar plates (1.0% peptone, 0.5% yeast extract, 0.5% NaCl and 1.5% bacterialogical agar) or in LB broth supplemented with 100 µg/µL ampicillin (Amp) or 50 µg/mL kanamycin (Kan), 7 µg/mL gentamicin (Gent) and 10 µg/mL tetracycline (Tet) as required. The *E. coli* DH5α strain (Hanahan, 1983) was used to propagate recombinant plasmids. Competent *E. coli* DH5α cells were produced by the RF1/RF2 method as described by Hanahan *et al.* (1991) and transformed accordingly. For heterologous overexpression of proteins, the *E. coli* BL21 (DE3) strain (Studier and Moffatt, 1986), which contains the T7-RNA polymerase under the control of the *lac*UV5 promoter was used. Competent *E. coli* BL21 (DE3) cells were produced using the $CaCl_2$ method of Swords (2003), and transformed as described by Sambrook *et al.* (1989).

2.2 Insect cell culture conditions

Spodoptera frugiperda (Sf9) cells (Vaughn *et al.*, 1977), further referred to as *S. frugiperda* cells, were maintained in complete TC100 insect medium (Lonza) (containing 10% [v/v] Foetal Bovine Serum (FBS) and 1% [v/v] Penicillin/Streptomycin) at 28°C without CO_2. Unless otherwise stated, all tissue culture experiments were conducted using the TC100 medium described above. Cells were maintained by diluting 1:3 [v/v] with fresh medium, upon reaching 80% confluence after 4 days.

2.3 Purification of HaSV from infected *H. armigera* larvae

Larvae infected with HaSV were obtained from David Penkler (Rhodes University) and virus was purified according to Penkler *et al.* (2016). Larvae were homogenized in capsid buffer (50mM sodium acetate, 250mM sodium chloride and 5mM ethylenediamine tetra-acetic acid, pH 5.0). For every gram of larvae, 2 mL capsid buffer was used for homogenization. The cell debris was pelleted at 10000 x g for 20 min at 4°C, after which the virus was pelleted from the supernatant through a 30% sucrose cushion at 100000 x g for 4 h at 4°C. The pellet was re-suspended in 500 µL capsid buffer overnight at 4°C. The virus was then further purified using a 10-40%

sucrose gradient, centrifuged at 140000x g for 90 min at 4°C. After extracting the virus band, the sample was concentrated by centrifugation at 140000 x g for 4 h at 4°C and re-suspended in 150 µL capsid buffer (supplemented with 10% glycerol). Western blot analysis was performed on the sample (as described in Appendix A) to verify the presence of HaSV.

2.4 Detection of small proteins in HaSV-infected *H. armigera* larvae

H. armigera eggs produced by a field-caught colony of *H. armigera,* were generously provided by Kudzai Mtambanengwe and Caroline Knox (Rhodes University). Once hatched, the larvae were reared at room temperature on cabbage until they reached the second instar growth stage. At this point the larvae were transferred onto an artificial diet (1.5% agar, 12.4% soy flour, 8.6% wheat germ, 7.2% yeast, 0.5% nipagen, 0.5% ascorbic acid, 0.2% sorbic acid, 0.2% propionic acid and 0.02% phosphoric acid). Three control larvae were reared, as described, in a different laboratory by Mr. Mtambanengwe. Purified virus (Section 2.3) was diluted to 10^6 virus particles per millilitre (vp/mL), and 2 µL was inoculated per 2 cm² of diet surface area (i.e. 2 x 10^3 vp per well). Three days post infection, the larvae were collected and frozen at -80°C.

Control (uninfected) and infected frozen larvae were thawed and then their heads and tails removed. The remaining biomass was weighed and washed with 100 µL phosphate buffered saline buffer (PBS; 140 mM NaCl, 2.7 mM KCl, 10 mM Na_2PO_4, 1.8 mM KH_2PO_4, pH7.4) prior to homogenisation. A total of 10 µL of a 1:1 (v/v) 0.1 M sodium phosphate buffer (0.1 M NaH_2PO_4, 0.1 M Na_2HPO_4, pH 7.4) and cOmplete™ Mini ethylenediaminetetraacetic acid (EDTA) free protease inhibitor cocktail (Roche, Cat. No. 11836170001) mix was used per mg of larval biomass. The animals were homogenized using sterile mortar and pestles. The samples were mixed with 4x sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) sample buffer (40% glycerol, 240 mM Tris-HCl, 8% SDS, 0.04% bromophenol blue, 5% beta-mercaptoethanol, pH 6.8), resolved by SDS-PAGE and subjected to western blot analysis as described in Appendix A. Detection was performed using a mouse anti-p11 (diluted 1:1000) and rabbit anti-p23 (diluted 1:50000) antibody (Table 2.1). The membranes were then stripped for 30 seconds (0.1% SDS, 0.2 M glycine, pH 2.5),

blocked overnight and probed with a rabbit anti-VCAP antiserum (diluted 1:10000; Mendes *et al.*, 2015) as described in Appendix A.

Table 2.1: Antibodies used throughout this study.

Primary				
Antibody name	**Source**	**Comment**	**Application**	**Reference**
Anti-VCAP	Rabbit	Raised against purified HaSV	Western – 1:10000	Mendes *et al.*, 2015
Anti-His6	Mouse	Peptide antibody (HHHHHH)	Western – 1:10000	Roche Applied Biosciences, Cat. No. 1922416001
Anti-FLAG®	Mouse	Peptide antibody (DYKDDDDK)	IF – 1:500	Milipore, Cat. No. MAB3118
Anti-p11	Mouse	Peptide antibody (HRKDHRRLEQQKAA)	Western – 1:1000 IF – 1:1000	This study
Anti-p23	Rabbit	Raised against p15-p8 fusion protein	Western – 1:50000 IF – 1:5000	This study
Secondary				
Antibody name	**Source**	**Comment**	**Application**	**Reference**
Anti-rabbit HRP conjugate	Goat	HRP-linked IgG (H+L)	Western – 1:20000	Advansta, Cat. No. R-05072-500
Anti-mouse HRP conjugate	Goat	HRP-linked IgG (H+L)	Western – 1:10000	Advansta, Cat. No. R-05071-500
Anti-mouse Alexa fluor® 546 conjugate	Goat	546 fluorophore-linked IgG (H+L)	IF – 1:500	Invitrogen™, Cat. No. A-11030
Anti-rabbit Alexa fluor® 633 conjugate	Goat	633 fluorophore-linked IgG (H+L)	IF – 1:500	Invitrogen™, Cat. No. A-21070

IF - immunofluorescence; HRP - horseradish peroxidase; H+L - heavy and light chain.

2.5 Viral RNA extraction and reverse transcription (RT)-PCR

Viral RNA (vRNA) was extracted from HaSV using the RNA Shield™ Purification Kit (Zymo Research, Cat. No. R1100) according to the manufacturer's instructions. Of the purified RNA, 360 ng was used as template for cDNA synthesis using the QuantiTect® Reverse Transcription Kit (Qiagen, Cat. No. 205311) according to the manufacturer's instructions. The primer MDB1 (5' GGA TCC GTT GCG ACA TGT GCG AAG 3') was annealed at 59°C for 30 seconds. The reaction was then transferred to 42°C for 30 min, after which the reverse transcriptase (RT) was inactivated by incubation at 95°C for 3 min. A PCR was performed to amplify a 996 bp region corresponding to the p11-p15-p8 region using primers MDB1 and MDB2 (5' ACT AGT GGG ACC ACC TTG TTT ACG 3'). Hifi Taq (Kapa Biosystems, Cat. No. KK2101) was used for the PCR as per the manufacturer's instructions. The cycling parameters are detailed in Table B2 of Appendix B6. Samples were analysed by agarose gel electrophoresis (AGE) using a 1% agarose gel in Tris-acetate-EDTA buffer (40 mM Tris, 20 mM glacial acetic acid, 1 mM ethylenediaminetetraacetic acid, pH 7.6). Electrophoresis was performed at 70 V

for one hour, and the DNA in the gel stained with Sybr® Safe DNA Gel Stain (Invitrogen™, Cat. No. S33102).

2.6 Cloning and sequence analysis of cDNA

The PCR product obtained from the RT-PCR reaction (Section 2.5) was gel-purified using the Isolate II PCR and Gel kit (Bioline, Cat. No. BIO-52059). After purification, the DNA was "A-tailed" with KAPA Taq ReadyMix (KAPA Biosystems). The DNA and the KAPA Taq ReadyMix was mixed in a 1:1 (v/v) ratio, and incubated at 72°C for 15 min. Samples were then incubated on ice for 5 min, after which nucleotides were removed using the Zymo Clean & Concentrator kit (Cat. No. D4004). The PCR product was then ligated into the pDrive cloning vector (Qiagen, Cat. No. 231122). The cloning process is described in Appendix B. Nine plasmids containing a DNA insert were analysed by Sanger sequencing (South African Institute for Aquatic Biodiversity) using the Genetic Analyzer 3500 (Applied Biosystems™). For each of the 9 clones, a forward and reverse sequencing reaction was performed using pUC/M13F (5' CCC AGT CAC GAC GTT GTA AAA CG 3') and pUC/M13R (5' AGC GGA TAA CAA TTT CAC ACA GG 3') primers. The sequences obtained were "cleaned up" and trimmed using CLC Main Workbench version 6.7.1 software, and aligned with HaSV_RNA1 (GenBank accession number KX423453) using ClustalW (Online-http://www.genome.jp/tools/clustalw). The HaSV_RNA1 sequence was further analysed using the NCBI ORF Finder tool (Online-https://www.ncbi.nlm.nih.gov/orffinder). The additional ORF identified was then subjected to BLAST analysis (Online-https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins) and PROSITE (Online-http://prosite.expasy.org) to identify potential motifs. The secondary RNA structure was analysed using the CLC Main Workbench software.

2.7 Heterologous expression of proteins in *E. coli* BL21 (DE3) cells

Proteins were expressed in *E. coli* BL21 (DE3) cells (Studier and Moffatt, 1986) using expression vectors based on pT7-7, which encodes a T7 RNA polymerase promoter, a ribosome binding site (RBS) and a hexa-histidine (6xHis) tag at the 5' terminal of the multiple cloning cassette. The plasmid pMDB4 represents a construct expressing 6XHis-p11-FLAG (Appendix B1). The second construct used in this study, pGH17 (Hughes, 2013) expresses a 6XHis tagged p15-p8 fusion protein (Table 2.2).

Table 2.2: Plasmids used throughout this study.

Plasmid name	Gene	Backbone	Antibiotic resistance	Organism used for expression	Figure	Reference
pAV8F	HaSV RNA1	pKG4	Amp	-	Appendix B2, B3, B5	Venter, 2001
pCW56	EGFP	pCW31	Kan/Amp	Sf9	-	Walter, 2008
pT7-7-6xHis-XbaI	-	pT7-7	Amp	*E. coli* BL21 (DE3)	Appendix B1	Short, 2010
pGH17	p15-TCS-p8	pT7-7-6xHis-XbaI	Amp	*E. coli* BL21 (DE3)	Figure 3.8A	Hughes, 2013
pGH2B	p11-p15-EGFP	pCW56	Kan/Amp	Sf9	Appendix B3, B4	Hughes, 2013
pGH6	p11-EGFP	pCW56	Kan/Amp	Sf9	-	Hughes, 2013
pGH43	p11-ECFP	pGH33	Amp	Sf9	-	Hughes, 2013
pJA1	2A-EGFP	pCW56	Kan/Amp	Sf9	Appendix B2	Awando, unpublished data
pFastBac™ Dual	Bacmid donor plasmid	N/A	Gent/Amp	*E. coli* DH10Bac™, Sf9	Appendix B5	Invitrogen™
pMDB2	p11-p15-p8	pFastBac™ Dual	Gent/Amp	*E. coli* DH10Bac™, Sf9	Appendix B5	This study
pMDB3	p11-FLAG	pUC57	Amp	-	Appendix B1	This study
pMDB4	6xHis-p11-FLAG	pT7-7-6xHis-XbaI	Amp	*E. coli* BL21 (DE3)	Appendix B1	This study
pMDB9	FLAG-p11-EGFP	pJA1	Kan/Amp	Sf9	Appendix B2	This study
pMDB10	FLAG-p11-p15-EGFP	pGH2B	Kan/Amp	Sf9	Appendix B3	This study
pMDB20	ΔFLAG-p11-p15-EGFP	pGH2B	Kan/Amp	Sf9	Appendix B4	This study

EGFP - enhanced green fluorescent protein; ECFP - enhanced cyan fluorescent protein; 2A - 2A-like processing site derived from Equine rhinitis A virus (ERAV); TCS - thrombin cleavage site; 6xHis – hexahistidine peptide tag; FLAG – FLAG peptide tag; Amp - ampicillin; Kan - kanamycin; Gent – gentamycin; Sf9 – *Spodoptera frugiperda* (Sf9) cells.

The following protocol was used for the expression of proteins in *E. coli* BL21 (DE3) cells transformed with pMDB4 or pGH17, respectively. Expression vectors were transformed into *E. coli* BL21 (DE3) cells and grown overnight at 37˚C on LB Agar plates supplemented with Amp. Single colonies were inoculated into 5 mL LB broth supplemented with Amp, and cultured overnight at 37˚C with agitation. Of these cultures, 0.5 mL was inoculated into 50 mL LB broth containing Amp and grown at 37˚C with agitation until an OD_{600} of 0.4-0.6 was reached. Expression of proteins was induced by adding isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.1 mM. Cells were grown for a further 2 (for pMDB4) or 4 h (for pGH17). One millilitre samples were removed prior to induction (T_0), and at hourly intervals post induction (T_1-T_4), measuring the OD_{600} at each time interval. These samples were pelleted by centrifugation at 10000 x g for 10 min after which the pellets were re-suspended in equal volumes of PBS buffer and 4x SDS-PAGE sample buffer. The re-suspension volume was determined by the OD_{600} (100 µL per OD_{600} of 1.0) in order to ensure equal cell number in all the samples. The samples were subjected to SDS-PAGE (Appendix A) with a 15% resolving gel, and visualized by staining with Coomassie Brilliant Blue.

2.8 Production and testing of anti-p11 peptide antiserum

The region spanning amino acid (aa) residues 77-90 (HRKDHRRLEQQKAA) was identified as a putative antigenic determinant of p11 using the Welling antigenicity (Welling *et al.,* 1985) and Hopp-Woods hydrophilicity plots (Hopp and Woods, 1981). An additional cysteine (C) residue was included at the N-terminus of the synthetic peptide to allow conjugation to keyhole limpet hemocyanin (KLH) and the resulting peptide was used for immunization to generate mouse polyclonal antiserum (GenScript).

To validate affinity of the antiserum for p11, the pMDB4 expression vector was used to heterologously express the 6xHis-p11-FLAG protein (Section 2.7). These attempts were all unsuccessful as no protein expression was detected by SDS-PAGE analysis (data not shown). As an alternative, overnight cultures of *E. coli* BL21 (DE3) cells transformed with pMDB4 were grown in 50 mL LB broth containing Amp as well as 1% glucose. These cultures were grown as before, until the desired OD_{600} of 0.4 was reached. The cells were then pelleted by centrifugation at 2000 x g for 10 min and

each pellet was re-suspended in 50 mL LB broth containing Amp but lacking glucose. IPTG was added to final concentration of 0.1 mM to induce the expression of p11. The OD_{600} was recorded at 0 (T_0), 1 (T_1) and 2 h (T_2) after induction (in triplicate). At each time point, cells were collected for analysis as described in Section 2.7. Samples were analysed by SDS-PAGE and western blot analysis (Appendix A). The primary antibodies used to detect the 6xHis-p11-FLAG protein included mouse anti-His6 monoclonal antibodies (1:500 dilution, Roche Applied Biosciences, Cat. No. *1922416001*) and mouse anti-p11 antiserum (1:1000 dilution). The secondary antibody was a HRP-linked goat anti-mouse IgG (H+L) used at a dilution of 1:10000 (Advansta, Cat. No. R-05071-500).

2.9 Production and testing of anti-p23 antiserum

Anti-p23 antiserum was generated using a 6XHis-p15-p8 fusion protein expressed in *E. coli* BL21 (DE3) cells transformed with pGH17 (Section 2.7). The protocol used to express and purify p15-p8 was described by Hughes (2013). After induction of protein expression and cell collection by centrifugation, the cell pellets were re-suspended in native binding buffer (0.1 M NaH_2PO_4, 0.01 M Tris-HCl, pH 8) at a 10:1 [v/v] ratio of cells to buffer. The cOmplete™ Mini EDTA-free protease inhibitor cocktail (Roche, Cat. No. 11836170001) was added according to manufacturer's instructions, with lysozyme (20 µg/mL), and the cells incubated on ice for 30 min. Treated cell pellets were sonicated using a hand-held 3 mm probe (Vibra Cell™) at 40% amplitude for twenty 10 second bursts. Urea was added to a final concentration of 8 M, and the sonicated samples incubated for 2 h with agitation in order to solubilize the p15-p8 protein. The samples were then centrifuged at 15000 x g for 30 min and 5 mL aliquots of the supernatant were stored at -80°C for subsequent purification. The 23 kDa 6xHis-tagged polypeptide was purified using the denaturing nickel-affinity protocol for HisTrap™ HP kit (Amersham Biosciences). The eluted protein fractions were concentrated and exchanged in to a denaturing buffer containing 2 M urea, using an Amicon Ultra-15 centrifugal filter unit (10 kDa cut-off, Millipore). A sample of each fraction was collected throughout the purification and concentration procedure and analysed by SDS-PAGE (Appendix A). A Thermo Scientific™ NanoDrop Lite Spectrophotometer was used to measure the final concentration of protein at 280 nm (0.461 mg/mL). The purified protein was sent to Professor Dirk Bellstedt (Department of Biochemistry, University of Stellenbosch) for raising the polyclonal rabbit antiserum.

The urea was removed from the protein by dialysis before the immunisation of rabbits. Antiserum was collected at 0 and 41 days post immunization with the purified 6xHis-p15-p8 protein.

A dot blot was used to optimise detection over a range of protein concentrations (0.01 µg, 0.1 µg and 1 µg) and dilutions of the 0 day (1:1000) and 41 day antiserum (1:20000, 1:50000 and 1:100000) (Appendix B). Native BL21 (DE3) cell-free lysate was included as a negative control. After dotting, the binding of the primary antibody was detected using HRP-linked goat anti-mouse IgG (H+L) used at a dilution of 1:10000 (Advansta, Cat. No. R-05072-500). A working concentration of 0.1 µg 6XHis-p15-p8 at an antiserum dilution of 1:50000 was subsequently used for western blot analysis. The same antiserum dilution was employed for the detection of p15-p8 in infected larvae (Section 2.4).

2.10 Expression of proteins in *S. frugiperda* cells

S. frugiperda cells were transfected with pCW56, the parent plasmid, and subsequent daughter plasmids encoding regions of interest (Table 2.2). These plasmids encode an enhanced hybrid Autographa californica multiple nuclear polyhedrovirus (hr/AcMNPV) promoter that control the expression of the regions of interest (Walter, 2010). Optimal expression conditions were determined using the pGH2B construct (Table 2.2; Fig. 2.1), encoding the native p11-p15 fused to EGFP at the C-terminus. Cells were transfected at a density of $1x10^5$ cells/mL on 13 mm glass cover slips using the non-liposomal transfection reagent FuGENE® HD (Promega, Cat. No. E2311), with small modifications to the manufacturer's instructions. For each transfection, 500 ng plasmid DNA, 3 µL of FuGENE® and 20 µL serum-free Grace's insect medium (Gibco™, Cat. No. 11300-027) was used per well. This mix was incubated for one hour at room temperature to permit the formation of lipid-DNA complexes. The medium was then aspirated from each well, and the transfection mixture added in a drop-wise manner. After a 10 minute incubation at 28°C with gentle agitation, 250 µL of fresh TC100 medium was added. Following a 2 hour incubation at 28°C, 1 mL of TC100 medium was added per well. Protein expression was tested at 24, 36 and 48 h. It was decided to express proteins for 36 hour in subsequent experiments. After protein expression, cells were washed with PBS buffer, fixed with 4% paraformaldehyde for

15 min, and treated with permeabilization buffer (5% goat serum, 1% Triton x-100, 10% sucrose in PBS buffer, pH7.4) for 20 min.

Immunofluorescence of expressed protein in cells transfected with pGH2B, pGH6 (p11-EGFP), pMDB9 (FLAG-p11-2A-EGFP), pMDB10 (FLAG-p11-p15-EGFP) and pMDB20 (ΔFLAG-p11-p15-EGFP) (Table 2.2; Fig. 2.1) was detected with primary antibodies diluted in permeabilization buffer: mouse anti-p11 (1:500, 1:1000, 1:2500, and 1:5000), rabbit anti-p23 (1:500, 1:1000, 1:5000 and 1:10000) and mouse anti-FLAG® monoclonal antibodies (Milipore, Cat. No. MAB3118) at 1:500. Incubation with the primary antibody was 90 min, with gentle rocking. Buffer containing the primary antibodies was aspirated and replaced with buffer containing the secondary antibody for 30 min with gentle rocking. The secondary antibodies include goat anti-mouse AlexaFluor® 546 (Invitrogen™, Cat. No. A-21070) and/or goat anti-rabbit AlexaFluor® 633 conjugate secondary antibodies (Invitrogen™, Cat. No. A-21070), each diluted 1:500 in permeabilization buffer.

The cells were then washed thrice with permeabilization buffer for 10 min each, with the second wash containing 1 µg/mL DAPI (Sigma, Cat. No. D8417). Coverslips were mounted using Dako fluorescent mounting medium (Cat. No. S3023) and analysed using a Zeiss LSM 780 laser scanning confocal microscope using the x63 objective.

2.11 Construction of a recombinant baculovirus for expression of p11, p15 and p8 in *S. frugiperda* cells

The Bac-to-Bac® Baculovirus Expression System (Invitrogen™) was used as an alternative to plasmid-based expression in *S. frugiperda* cells due to the high levels of expression associated with recombinant baculoviruses. A recombinant bacmid (pMDB2) was created (as described in Appendix B5) and transformed into *E. coli* DH10Bac™ cells (Bac-to-Bac® manual, Invitrogen™). Transformants were screened by blue-white screening (Appendix B). After 36 h incubation at 37˚C, white colonies for both pFastBac™ Dual and pMDB2 were re-streaked onto MacConkey agar (Biolab). Thereafter, the white colonies were inoculated into 5 mL of LB broth supplemented with Kanamycin, Gentamycin and Tetracycline, and cultured for 36 h at 37°C with agitation. DNA was extracted from these cell cultures using the ZR Bac DNA Miniprep Kit (Zymo research) and recombinant bacmids were analysed for positive transposition events by PCR using two different primer sets; pUC/M13F-GH10R (5'

AAG CTT TCA GAC GAG CGT GAT TGC CT 3') and MDB1-pUC/M13R. Reactions were carried out using the KAPA Taq ReadyMix as per the manufacturer's instructions (KAPA Biosystems, Cat. No. KK1006), and cycling parameters as detailed in Tables B6 and B7 in Appendix B.

Recombinant bacmids were transfected into *S. frugiperda* cells using FuGENE® as described by Short *et al.* (2012). The pFastBac Dual^{TM} bacmid was used to produce control baculovirus (pFastBV), and the pMDB2 bacmid to produce a p11-p15-p8 encoding baculovirus (pMDB2 BV). Cells were seeded at a density of 8 x 10^5 cells per well in a 6-well plate and allowed to settle for 60 min. During this time the transfection mixtures were prepared; 800 ng bacmid DNA, 12 µL FuGENE® was added to sufficient serum-free Grace's insect medium (to make a total volume of 200 µL). The mixture was incubated at room temperature for 30 min. The medium was aspirated from each well and the transfection mixture added in a drop-wise manner to the cells. An additional 200 µL of serum-free Grace's insect medium was added to the cells before a 2 hour incubation at 28˚C, after which 1 mL of TC100 medium was added. Following incubation at 28˚C for four days, the budded virus was harvested as the passage 1 (P1) virus stock in the supernatant and used for subsequent plaque-purification and propagation.

For plaque purification, cells were seeded at a density of 8.5 x 10^5 cells per well in a 6-well plate and allowed to settle for 60 min. The medium was aspirated from the cells and the respective wells infected with 500 µL of a ten-fold serial dilution (10^{-1}-10^{-6}) of the P1 viral stock (from the previous paragraph) prepared in TC100 medium. The cells were incubated for 60 min at 28˚C with gentle agitation. During this time, a 1:1 mixture of 2% agarose and TC100 medium was prepared. Tissue culture grade agarose (Sigma) was dissolved in sterile triple distilled water by heating to just below boiling point and then incubated at 42˚C until use. TC100 medium was also warmed up to 42˚C. After the infection, the medium was aspirated and 2 mL of a 1:1 mixture of 2% agarose and TC100 medium was added to each well and the agarose permitted to set for 10 min. Plates were incubated for 5 days, after which the overlays in each well were stained with 500 µL thiazolyl blue tetrazolium bromide (0.1% in PBS, pH 7.4). Individual plaques were picked using sterile P1000 pipette tips and re-suspended in 500 µL TC100 medium.

The virus titre was estimated by calculating the number of plaque forming units per mL (pfu/mL) as described below:

$$\text{pfu/mL} = \frac{\text{number of plaques}}{\text{dilution factor x mL of inoculum}}$$

To amplify the virus titre (10^6-10^9 pfu/mL), the re-suspended plaques were used to re-infect *S. frugiperda* cells. Cells were seeded at a density of 2 x 10^6 cells per well and allowed to settle for 60 min. Each well was infected with 500 µL of the re-suspended virus and incubated at 28°C with gentle agitation (55 rpm) for 1 hour. A further 3 ml TC100 medium was added to each well and cells were incubated at 28°C for 4 days before harvesting. The cells were re-suspended in the spent medium, pelleted by centrifugation at 2000 x g for 5 min and the supernatant retained as the passage 2 (P2) virus stocks. This process was repeated once more, ultimately yielding passage 3 (P3) virus stocks. The resulting pellets were re-suspended in 150 µL PBS and 50 µL 4 x sample buffer for subsequent SDS-PAGE and western blot analysis using rabbit anti-p23 antibodies at a 1:50000 dilution (Appendix A).

Infection of *S. frugiperda* cells with recombinant baculovirus was confirmed by monitoring the cell morphology using a Nikon TMS inverted light microscope (x10 objective). Uninfected cells were compared to pFast BV and pMDB2 BV infected cells. The identity of the infecting baculovirus in P2 and P3 supernatants was confirmed using PCR. Supernatants from uninfected *S. frugiperda* cells, and *S. frugiperda* cells infected with pFast BV (P3) were included as negative controls. Baculoviral DNA was extracted with a proteinase K treatment where 100 µl of virus was mixed with proteinase K (1.64 µg/µl final concentration) and 10x proteinase K buffer (10mM EDTA, 5% SDS and 100mM Tris-HCl, pH8.0). The DNA was isolated using the Zymo Clean & Concentrator kit (Cat. No. D4004) as per manufacturer's instructions. PCR was executed using primers that flank p11, GH10F (5' CCA TGG GCG AAG CCG GAA ACC 3') and GH3R (5' CCG GTA CGA GCG TGA TTG CCT GTC TTG C 3'), using cycling parameters detailed in Table B3 of Appendix B6, and products analysed by AGE (Section 2.5).

Chapter 3:
Results

3.1 Introduction

It is not known whether the p11, p15 and p8 proteins found at the 3' terminal of HaSV RNA1 are expressed during HaSV infection. Protein expression is traditionally investigated using immunological approaches with antibodies directed against the proteins of interest. Baillie (Rhodes University, unpublished data) attempted to generate antibodies against recombinant p11. First, she attempted to overexpress the native p11 coding sequence (CDS) in *E. coli* BL21 (DE3) cells and attempted to purify it. These attempts were however, unsuccessful. The inability to detect p11 expression was attributed the viral codon bias. Baillie also used RT-PCR to amplify the p11 and p15 coding region from HaSV vRNA, and analysis showed 100% sequence conservation amongst four cDNA clones (Baillie, unpublished data). Considering the lack of proofreading mechanisms in RNA virus replication (Lauring and Adino, 2010), the observed sequence conservation suggests that the proteins encoded by this sequence likely have a potential function in the viral life cycle.

In the absence of suitable antibodies, Hughes (2013) investigated the expression of the small ORFs using transient expression of EGFP fusion proteins in *S. frugiperda* cells. Plasmids encoding EGFP-tagged p11 and p11-p15 were successfully expressed, and localized to punctuate structures in the cytoplasm. Co-expression of these proteins with organelle markers showed that these proteins localised to the Golgi apparatus and secretory vesicles (Hughes, 2013). Hughes also showed that p11-EGFP and p11-p15-EGFP co-localise with the HaSV REP, and proposed that if these small proteins were expressed in infected cells, they were likely to be involved in viral replication. Attempts were made to express p11-p15-p8-EGFP, however no fluorescence was detected (Hughes, 2013). It could not be determined whether p8 was expressed at levels that were too low for detection or not at all.

In this study, the first objective was to confirm the conservation of the small ORFs, p11, p15 and p8 in infectious HaSV. Next, the expression of the small ORFs was investigated in HaSV infected *H. armigera* larvae. To achieve this, antiserum capable

of detecting these small proteins were generated and their affinity for their target proteins validated. These antibodies were then used to detect the small proteins in HaSV infected larvae. This study further aimed to elucidate the mechanism by which these proteins are expressed and their potential function in the infectious viral lifecycle.

3.2 Investigating the conservation of small ORFs in HaSV

Purification of virus particles from infected larvae was confirmed by SDS-PAGE and visualization of the proteins with Coomassie staining. Protein bands migrating at 71 kDa and 63 kDa, corresponding in predicted molecular weight of the capsid precursor and major capsid protein, respectively, were observed (Fig.3.1A, lane 1). Western blot analysis using antibodies directed against the HaSV VCAP (anti-VCAP) confirmed that these bands were the viral capsid protein (Fig.3.1A, lane 2). Viral RNA was purified from these virus particles and subjected to RT-PCR using primers flanking the 5' terminus of p11 (MDB1) and 3' terminus of p8 (MDB2) on vRNA1 (Fig. 3.1C). A band of approximately 1000 bp, corresponding in size to one expected if all three small ORFs encoding p11, p15 and p8 was amplified, was observed in the reaction charged with vRNA + RT (Fig. 3.1B, lane 4). A corresponding band was seen in the pAV8F control sample, where the plasmid encodes vRNA1 cDNA (Fig. 3.1B, lane 3). This band was not present in the no template DNA and vRNA – RT samples (Fig. 3.1B, lane 1-2).

The PCR product was gel-purified and cloned into the pDrive vector. A total of 9 recombinant clones were sequenced, and the resulting cDNA sequences were aligned with HaSV_RNA1 (GenBank accession number KX423453). This sequence, obtained from the pAV8F plasmid (Venter, 2001), was used as the reference sequence for the alignment as it served as the PCR template for the expression vectors generated by Hughes (2013), as well as this study (Table 2.1). Two cDNA clones (pMDB12 and pMDB18) exhibited point mutations in their p11 and p8 coding sequences. For pMDB18, a silent c.105A>G mutation occurred in the p11 CDS (Fig. 3.1C). Also in this clone, was a c.77C>T mutation in the p8 CDS, resulting in a p.L26P change. This introduces a 3 aa long helix adjacent to the strand (TVRV) (Fig. C1, Appendix C). For pMDB12, a c.167T>C point mutation also occurred in the p8 CDS, resulting in a p.P56L change (Fig. 3.2C), disrupting a predicted helix (Fig. C1, Appendix C).

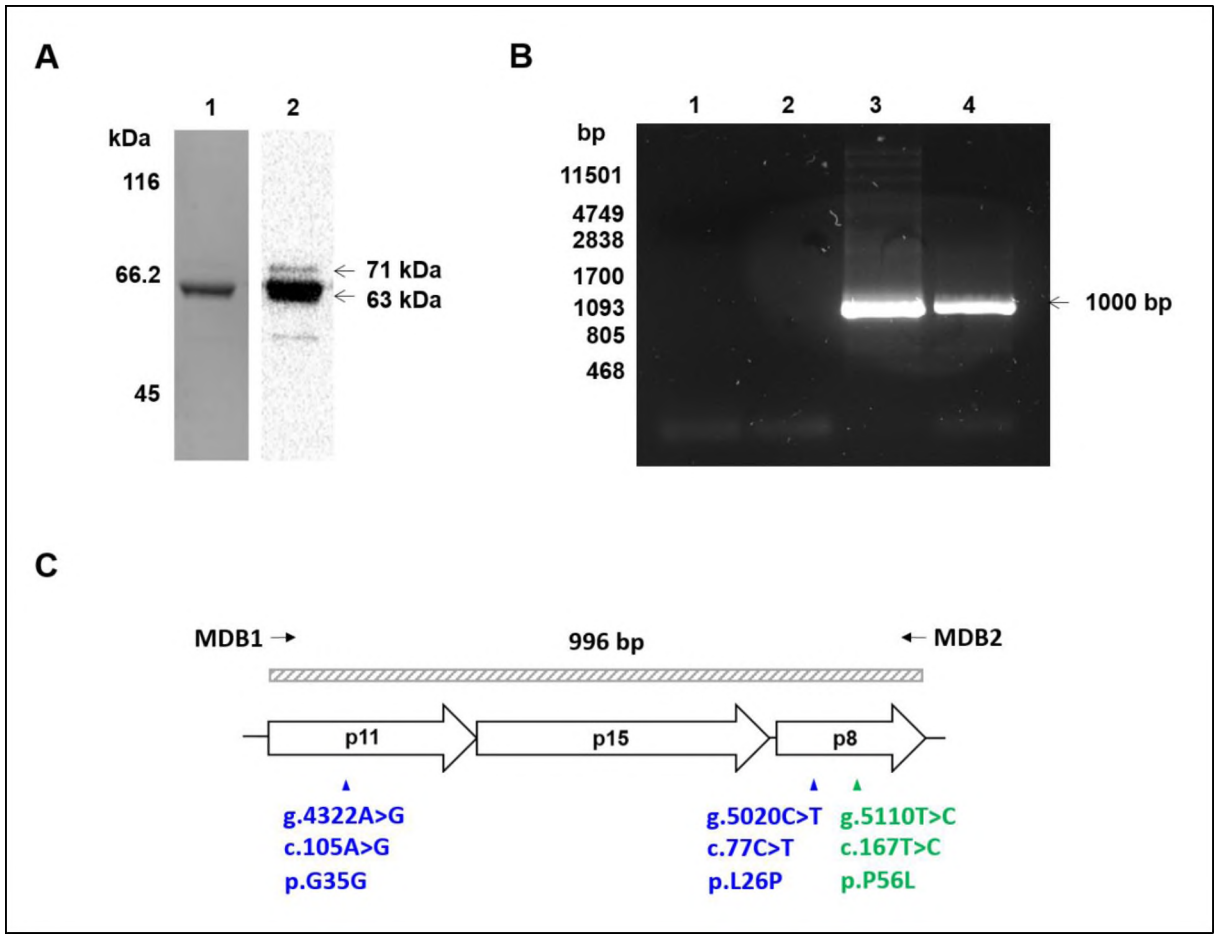

Figure 3.1: Sequence conservation of the small ORFs in HaSV vRNA1. A: Virus particles purified from HaSV infected larvae, resolved by SDS-PAGE and visualised by Coomassie staining (Lane 1) followed by western blot analysis using anti-VCAP antiserum at a dilution of 1:10000 (Lane 2). Arrows indicate the capsid precursor and capsid sizes detected. **B:** RT-PCR amplification of the p11-p15-p8 coding sequence from HaSV vRNA1, PCR products were resolved by AGE. Lane 1: No DNA template; Lane 2: vRNA - RT; Lane 3: pAV8F; Lane 4: vRNA + RT. **C:** Schematic representation of mutations observed in small ORFs of HaSV. The sequences obtained from 9 clones were aligned to the current sequence of RNA1 (GenBank accession number KX423453). The patterned grey bar represents the size of the PCR product (amplified using primers MDB1 and MDB2). Two clones, pMDB12 (green) and pMDB18 (blue) exhibited point mutations, with positions of the mutated nucleotides indicated with respect to the genomic (g.) vRNA and coding (c.) sequence. Positions of amino acids affected by the point mutations are indicated (p.) with respect to individual ORFs.

All seven remaining cDNA clones were identical to the HaSV_RNA1 sequence, leading to the conclusion that these clones were most likely representative of the majority of vRNAs in the virus population.

3.3 Production of polyclonal peptide antiserum for detecting p11

Hughes (2013) attempter to raise polyclonal antiserum directed against aa residues 3-16 of p11 (Fig. 3.2). This mouse antiserum was unable to detect p11 generated in insect and mammalian *in vitro* transcription-translation (TnT®) coupled systems or in HaSV infected larval gut tissue. When the [^{35}S]-methionine labelled translation products from the *in vitro* transcription-translation (TnT®) coupled systems were analysed by autoradiography, a band of the size corresponding to p11 was observed in the insect system. The inability of the anti-p11 antiserum to detect this band suggests that the antiserum was ineffective. Hughes (2013) attributed this to the non-antigenic nature of the p11 region selected for generating the antibodies (Fig. 3.2).

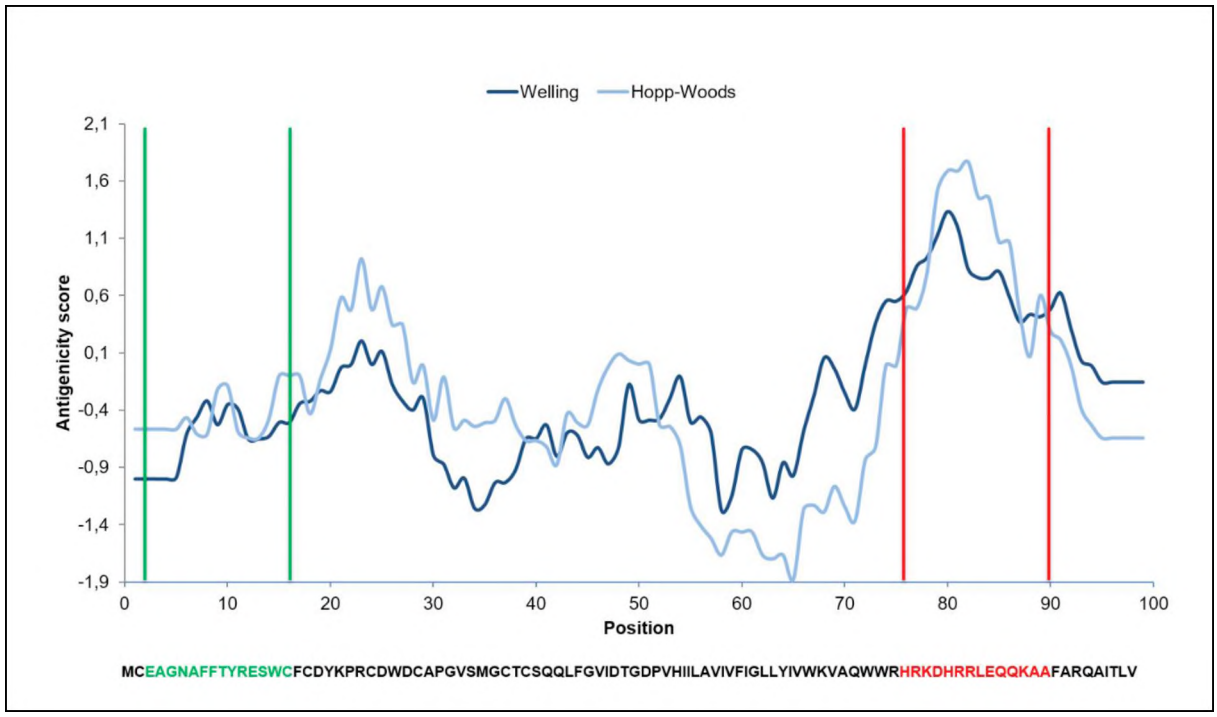

Figure 3.2: Antigenicity plots of p11 showing the region selected for peptide antibody production. The Welling (dark blue) and Hopp-Woods (light blue) plots show hydrophilic antigenicity scores as positive values, and hydrophobic scores as negative. The amino acid sequence is positioned on the x-axis, with the region selected for peptide antibody production depicted in red. The region previously identified by Gareth Hughes (2013) is indicated in green.

In order to assess the antigenicity of p11, the Welling and Hopp-Woods methods were used. The Welling scale ranks the amino acid residues of the protein according to their occurrence in known antigenic regions of proteins, while the Hopp-Woods scale ranks them according to their water solubility (Welling *et al.*, 1985; Hopp and Woods, 1981).

According to both algorithms, residues 77-90 of p11 were predicted to have the highest antigenicity with scores of 1.336 and 1.767 for the Welling and Hopp-Woods plots, respectively (Fig. 3.2).

Residues 19-32 are also predicted to exhibit some antigenicity, although less than residues 77-90 (Scores of 0.206 and 0.922 for the Welling and Hopp-Woods plots, respectively). Any score above zero is considered hydrophilic, and therefore antigenic, with an increase in score representing an increase in likelihood of surface exposure. The region spanning residues 77-90 was therefore chosen for the target peptide and for subsequent antiserum production.

3.3.1 Heterologous expression of p11 in *E. coli*

Since it was unknown whether p11 would be expressed during the infectious virus lifecycle, it was decided to test the affinity of the mouse polyclonal peptide antiserum against recombinant p11 expressed in *E. coli* BL21 (DE3) cells (further referred to as *E. coli*). The expression construct (pMDB4) carried a synthetic sequence encoding the p11 coding sequence (CDS), codon optimized for expression in *E. coli* in a T7-7 expression vector. The p11 sequence was flanked by an N-terminal hexa-histidine (6xHis) tag and a C-terminal FLAG-tag (Fig. 3.3A).

In the Coomassie stained gel (Fig. 3.3B), a band migrating at 6.7 kDa was observed in the crude extract collected after 2 h (T2) from *E. coli* cells transformed with pMDB4 (6xHis-p11-FLAG). Since the band was absent in cells transformed with pT7-7 (control vector), it could have represented the 6xHis-p11-FLAG protein, migrating at a smaller size than expected. The same extracts were subjected to western blot analysis with antibodies directed against p11, as well as anti-His6 antibodies to detect expression of the 6xHis-p11-FLAG protein.

A purified 6xHis-tagged p17 protein (Mendes *et al.*, 2015) was included as a positive control for detection with the anti-His6 antibodies. A band migrating at approximately 24 kDa, corresponding to the p17 protein, was detected (Fig.3.3C), however the anti-His6 antibodies did not specifically detect the 6xHis-p11-FLAG protein (pMDB4; Fig. 3.3C). Instead, bands migrating at 35 kDa and 11 kDa were detected in all lanes, including *E. coli* cells that were expressing p17 or those that were transformed with

the control vector (Fig. 3.3). Thus the anti-6xHis antibodies failed to detect the 6xHis-p11-FLAG fusion protein. A similar result was obtained with the anti-p11 antiserum, where non-specific bands migrating at 50 kDa, 35 kDa and 11 kDa were detected in all the lanes (Fig. 3.3). Since the anti-His6 antibodies detected the positive control (p17), but nothing in the 6xHis-p11-FLAG protein extracts, it was concluded that the band observed at 6.7 kDa did not represent the 6xHis-p11-FLAG recombinant protein, and that this protein was not produced at sufficient levels for detection.

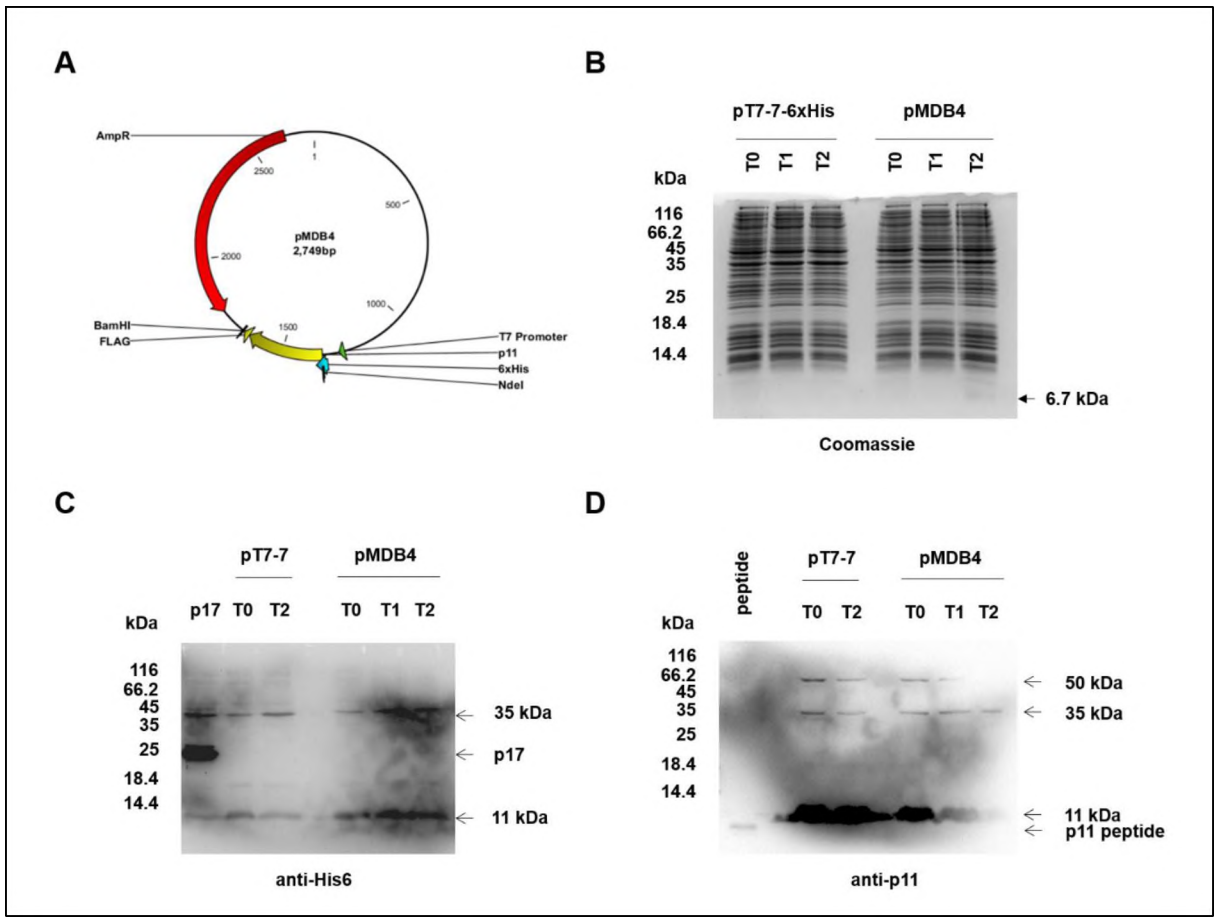

Figure 3.3: Expression of 6xHis-p11-FLAG in *E. coli* BL21 (DE3). (**A**) Plasmid map of pMDB4, the recombinant bacterial expression vector used for expression of the 6xHis-p11-FLAG protein. Amp R: *β*-lactamase**;** 6xHis-hexa-histidine tag. **(B)** SDS-PAGE analysis of crude protein extracts of *E. coli* BL21 (DE3) cells transformed with pT7-7 (vector) and pMDB4 (6xHis-p11-FLAG) prior to-induction (T0), one hour or (T1) or two h (T2) post induction with 0.1 mM IPTG. **(C)**: Western blot analysis of crude protein extracts derived from *E. coli* BL21 (DE3) cells transformed with pT7-7 (vector) and pMDB4 (6xHis-p11-FLAG), using mouse anti-His6 antibodies (1:10000), and 6xHis-p17 (a viral protein encoded on HaSV vRNA2) included as a positive control for detection. **(D)**: Western blot analysis using anti-p11 peptide antiserum (1:1000), with 10 ng of purified p11 peptide included as a positive control for detection. The molecular weight standards are indicated on the left of the gels and arrows on the right of the image indicate the sizes of the fragments detected.

The integrity of the plasmid construct used for the expression of 6xHis-p11-FLAG had previously been validated by Sanger sequencing, thus the lack of expressed protein was not due to an error in the plasmid construct. It was, however, possible that the expression of the p11 protein may be toxic to the *E. coli* cells. To test whether this was the case, the optical density (OD_{600}) of an *E. coli* cell culture transformed with the 6xHis-p11-FLAG expressing construct (pMDB4) was compared to cells transformed with the control vector (pT7-7) over a period of two h.

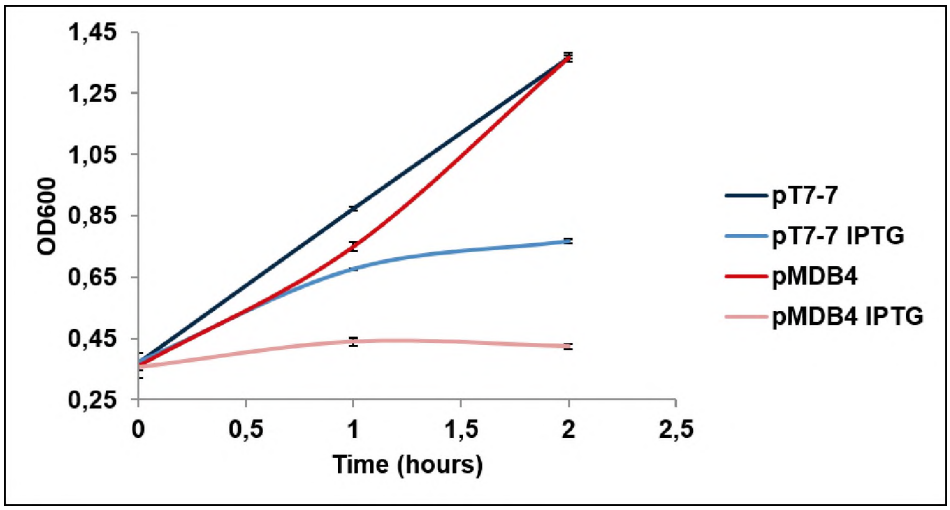

Figure 3.4: Optical density of *E. coli* BL21 (DE3) cultures expressing 6xHis-p11-FLAG. Cultures were grown until an OD_{600} of 0.3 was reached, at which point protein expression was induced using 0.1 mM IPTG. The OD_{600} was further recorded at 1 and 2 h after induction. Cells transformed with the control vector (pT7-7) and expression vector (pMDB4), both uninduced and induced (IPTG) were included. Error bars at every time point represents the standard deviation over triplicate readings.

The *E. coli* cells transformed with both pT7-7 and pMDB4, but not induced, exhibited a linear increase in OD_{600} over time (Fig. 3.4). The net effect of IPTG on the OD_{600} is illustrated by cells transformed with pT7-7, where the addition of IPTG results in a lower OD_{600} 1 and 2 h post induction than that in the uninduced samples (Fig. 3.4). When cells transformed with pMDB4, the OD_{600} of the induced pMDB4 sample was lower than that cells transformed with pT7-7. In fact, the OD_{600} had not increased significantly 2 h post induction (Fig. 3.4). This indicated that the cells had not grown after the induction of 6xHis-p11-FLAG expression, suggesting that the expression of 6xHis-p11-FLAG inhibited bacterial cell proliferation and that this protein may be toxic to *E. coli* cells. For this reason, the p11 protein was expressed in insect tissue culture cells instead.

3.3.2 Immunofluorescent detection of p11 in *S. frugiperda* cells

In light of the problems encountered with heterologous expression of p11 in *E. coli,* the p11 protein was expressed in *S. frugiperda* cells to test the anti-p11 antiserum affinity. *S. frugiperda* cells were transfected with pGH6 (Fig. 3.5), an expression vector encoding p11 fused to EGFP at its C-terminus (Hughes, 2013). The p11-EGFP localized to cytoplasmic punctate structures (Fig. 3.6A), whereas the backbone of pGH6, pCW56 encoding EGFP only (Fig. 3.5), displayed both cytoplasmic and nuclear EGFP fluorescence (Fig. 3.6A).

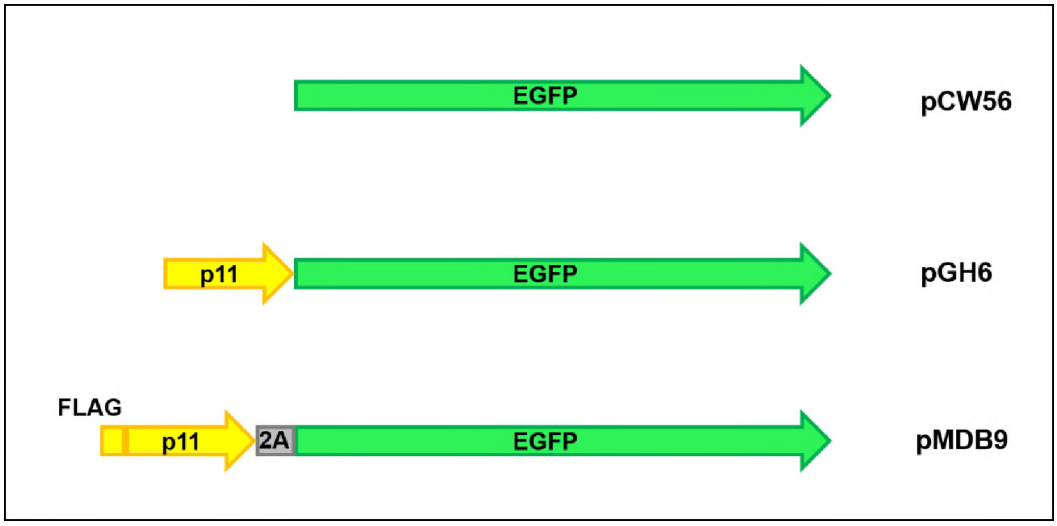

Figure 3.5: Schematic representation of p11 insect expression vectors. All vectors share the pCW56 plasmid as backbone, with inserts illustrated occurring downstream of an Autographa californica multiple nuclear polyhedrovirus promoter. EGFP - enhanced green fluorescent protein; FLAG – FLAG peptide tag; 2A – 2A-like processing site derived from Equine rhinitis A virus (ERAV). The cloning strategies used to create these plasmids can be found in Appendix B. Further details can be found in Table 2.2.

Although a background level of fluorescence was observed in all cells, potentially due to non-specific binding, the anti-p11 antiserum did not detect p11-EGFP, as no increased fluorescence was observed in cells expressing p11-EGFP (Fig. 3.6B). A possible explanation for this result was that the EGFP, fused to the C-terminus of p11, was masking the antigenic region used to raise the antibodies. To address this problem, the pMDB9 expression vector was therefore created.

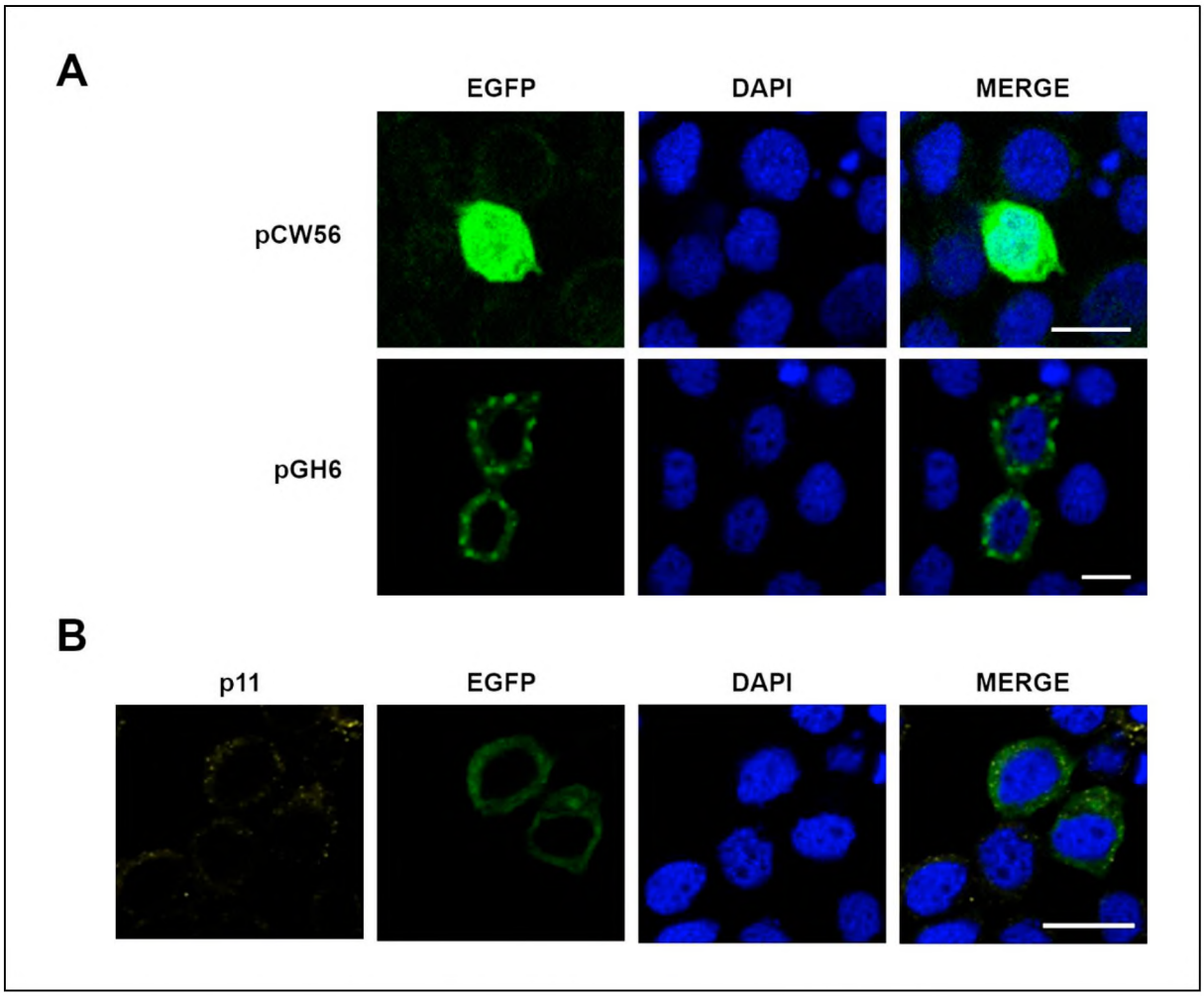

Figure 3.6: Expression of p11-EGFP in *S. frugiperda* cells. A: Cells were transfected with pCW56 (EGFP) or pGH6 (p11-EGFP). **B:** Cells transfected with pGH6 was probed with anti-p11 antiserum (1:1000). Scale bars represent 10 µm.

pMDB9 encodes p11 fused to a FLAG tag at its N-terminus, with an in-frame Equine rhinitis A virus (ERAV) 2A-like processing site, followed by EGFP at its C-terminus (Fig. 3.5). The activity of the ERAV 2A-like site would result in differential levels of expression of FLAG-p11-2A-EGFP and FLAG-p11 / EGFP, at an expected ratio of 1:10 (Donnely *et al.,* 2001). *S. frugiperda* cells transfected with pJA1, the vector plasmid used to create pMDB9, encoding 2A-EGFP displayed both cytoplasmic and nuclear EGFP fluorescence (Fig. 3.7A). This was consistent with the fluoresence observed for EGFP (pCW56; Fig. 3.6A).

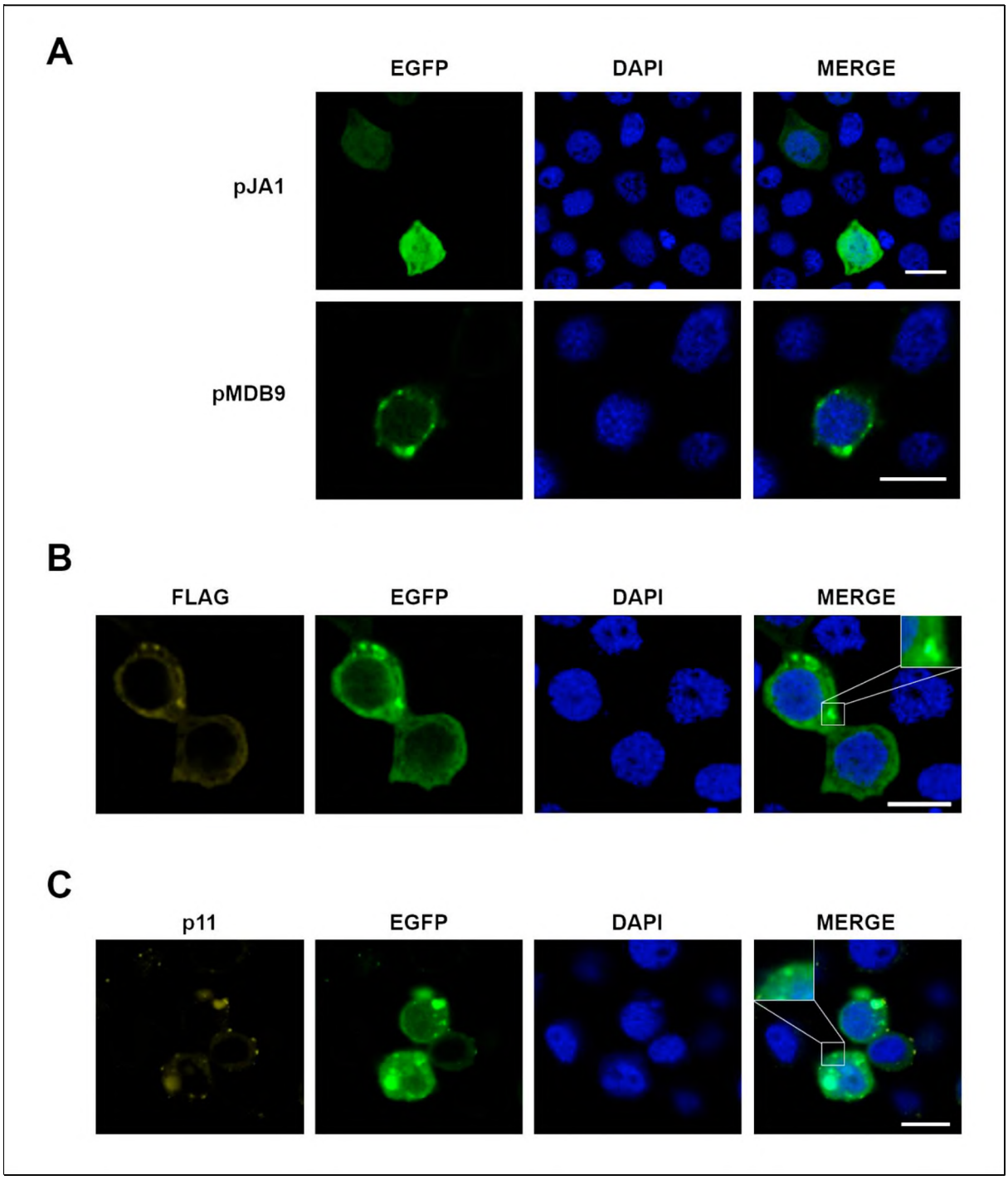

Figure 3.7: Expression of FLAG-p11-2A-EGFP in *S. frugiperda* cells. **A:** Cells were transfected with pJA1 (2A-EGFP) or pMDB9 (FLAG-p11-2A-EGFP), where 2A represents an Equine rhinitis A virus (ERAV) 2A-like processing site. **B and C:** Cells transfected with pMDB9 was probed probed with **(B)** anti-FLAG antibodies (1:500) and **(C)** anti-p11 antiserum (1:1000). Magnified boxes indicate punctate structures showing perfect overlap and scale bars represent 10 µm.

The distribution of the EGFP changed when FLAG-p11 was introduced into pJA1 (pMDB9; Fig. 3.5). Here, EGFP fluorescence was observed predominantly as punctate cytoplasmic structures but some fluorescence, attributed EGFP alone, was also

observed throughout the cytoplasm and nucleus (Fig. 3.7A). The punctate structures observed in cells transfected with pMDB9 were similar to those that were observed in cells transfected with pGH6 (Fig. 3.6A), indicating sub-optimal activity of the 2A-like site. Theoretically, only 10 % of the proteins expressed should be tagged with EGFP, and the remaining 90% should be untagged (Donnely *et al.,* 2001; Short, 2010). It was therefore likely that the presence of p11 upstream of the 2A-like site may have interfered with processing activity.

S. frugiperda cells transfected with pMDB9 were then probed with anti-FLAG and anti-p11 antiserum, respectively. As observed before, FLAG-p11-2A-EGFP localised to cytoplasmic punctate structures, with some diffuse fluorescence observed in the cytoplasm and nucleus (Fig. 3.7B and C). The signal obtained with anti-FLAG antibodies and anti-p11 antiserum overlapped with the fluorescent punctate structures distributed throughout the cytoplasm (Fig. 3.7B and C). Thus it was concluded that the anti-p11 antiserum was capable of recognizing p11.

3.4 Production of polyclonal antiserum for the detection of p15 and p8

The expression vector pGH17, encodes a synthetic sequence for the expression of a p15-p8 fusion peptide that was codon-optimised for expression in *E. coli.* The p15 and p8 coding sequences were fused in-frame and separated by a thrombin cleavage site (6 aa in length) (Hughes, 2013). This sequence is tagged at the N-terminal with 6xHis-tag and occurred downstream of a T7 RNA polymerase promoter and a ribosome binding site (Fig. 3.8A). The pGH17 expression vector was transformed into *E. coli* cells, and expression of the p15-p8 protein was induced by the addition of IPTG. A 23 kDa protein was detected in cell-free extracts 2 and 4 h after induction (Fig. 3.8B, lane 2-3). This band was absent in uninduced cells (Fig. 3.8A, lane 1), leading to the conclusion that this 23 kDa represented the p15-p8 peptide.

Hughes (2013) showed that p23 was insoluble in *E. coli* cells so the 6xHis-p15-p8 protein was purified by denaturing affinity chromatography. The crude protein extract was loaded onto charged nickel affinity columns and a prominent band corresponding to the p23 band was detected in the first two elution fractions (Fig. 3.8C, lanes 8 and 9). The 23 kDa protein co-purified with low levels of endogenous *E. coli* proteins,

indicated by bands of varying sizes observed in the eluted fractions (Fig. 3.8C, lanes 8 and 9).

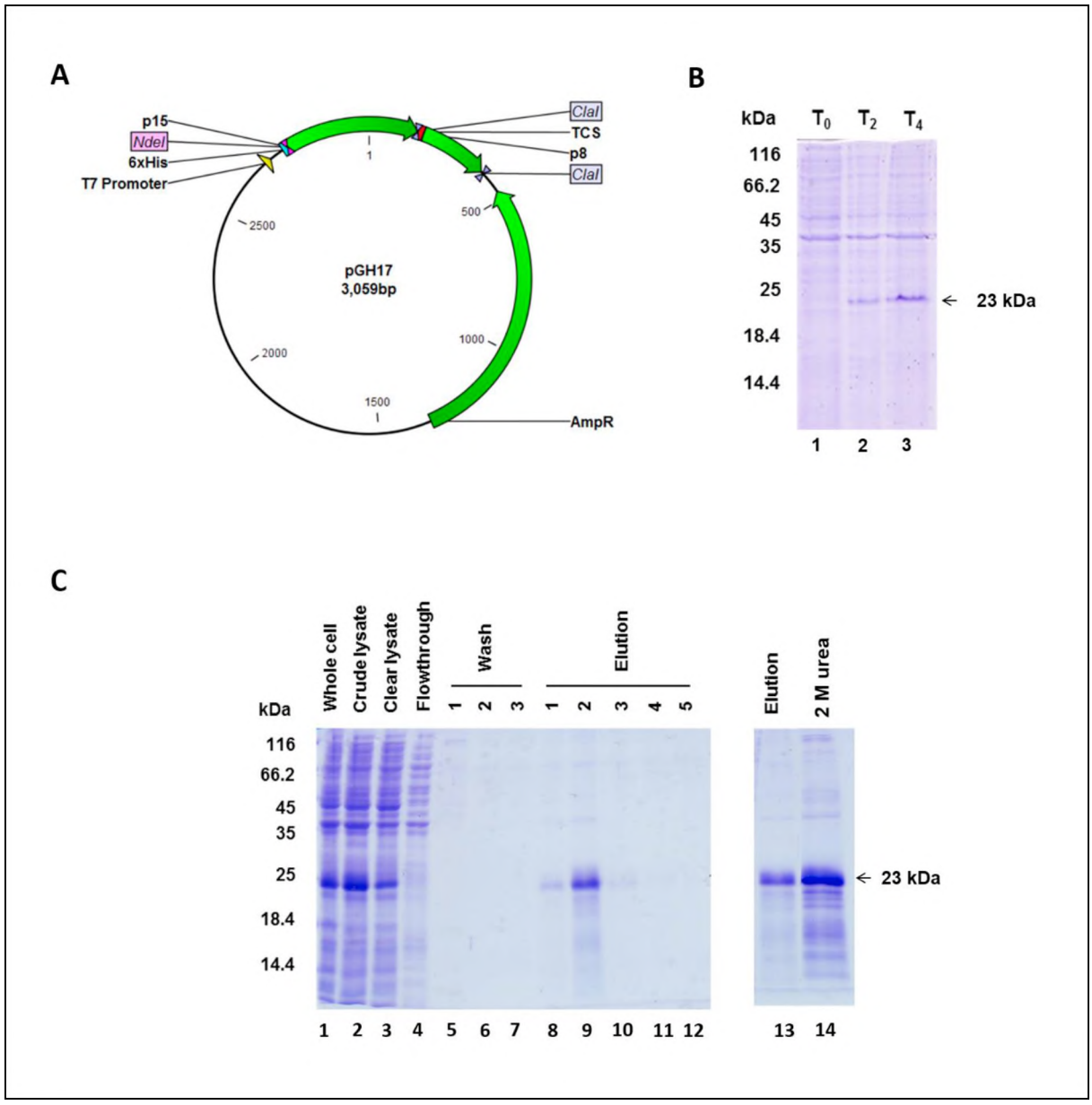

Figure 3.8: Heterologous overexpression of the p15-p8 fusion protein in *E. coli* BL21 (DE3). A: Plasmid map of pGH17, the recombinant bacterial expression vector used for overexpression of the 23 kDa 6xHis-p15-p8 peptide. TCS - thrombin cleavage site; AmpR - ampicillin resistance. **B**: SDS-PAGE analysis showing heterologous overexpression of the 6xHis-p15-p8 peptide in *E. coli* BL21 (DE3) cells over 4 h, with samples removed pre-induction (T0), and every 2 h post induction (T2 and T4). **C**: SDS-PAGE analysis showing the denaturing Ni-affinity purification (lane 1-12), and buffer exchange and concentration (lane 13-14) of the 6xHis-p15-p8 peptide. The concentrated protein in lane 14 was exchanged from a denaturing buffer containing 8 M urea to one with 2 M urea. The molecular weight standards are indicated on the left of the images and arrows on the right indicate the purified protein.

The purified p23 protein was concentrated and the buffer exchanged from 8 M to 2 M urea to prevent toxicity to the rabbit during the immunisation process. The purified p23 remained soluble in buffer containing 2 M urea (Fig. 3.8C, lane 14), and so this sample was used to immunize a rabbit to raise polyclonal antiserum.

A dot blot analysis was used to test the affinity of the p23 antiserum to the 6xHis-tagged p15-p8 peptide as well as to optimise the antiserum dilution (Appendix D). Thereafter, the affinity of the antiserum for p23 was confirmed by western blot analysis. A band corresponding in size to the purifed p23 was detected by the 41 day antiserum (Fig. 3.9). This band was also observed in crude protein extracts derived from *E. coli* cells transormed with pGH17, but absent in cell-free extracts of cells transformed with pMV2, a plasmid that results in the expression of HaSV 6xHis-p17 (Fig 3.9). It was therefore concluded that the anti-p23 antiserum successfully detected p23.

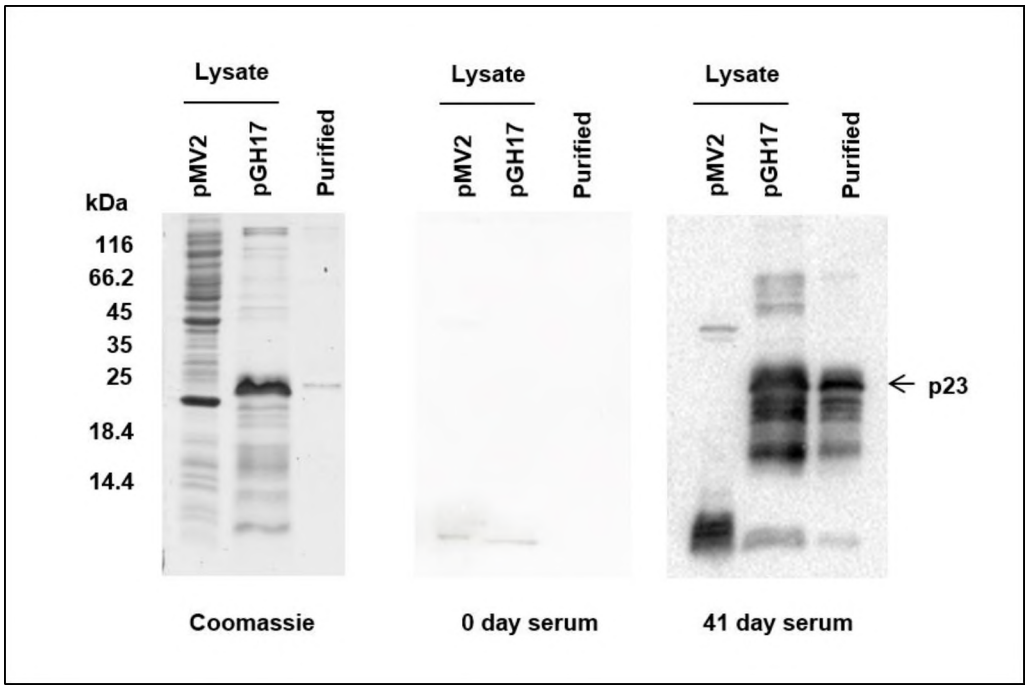

Figure 3.9: Detection of p15-p8 fusion protein with anti-p23 antiserum. SDS-PAGE (Coomassie) and western blot analysis showing the affinity of the antiserum for the 23 kDa fusion peptide. Affinity was validated by testing the 0 day antiserum and 41 day antiserum against *E. coli* BL21 (DE3) cell-free lysates expressing 6xHis-p17 (pMV2) and 6xHis-p15-p8 (pGH17), as well as 0.1 µg purified peptide (purified). Antiserum was diluted 1:50000. The molecular weight standards are indicated on the left of the gels and arrow on the right of the image indicate the p23 band.

From this body of work, it was concluded that the anti-p11 and anti-p23 antiserum could successfully detect p11 and p15, respectively. Both antisera were therefore used to determine whether the p11 and p15 proteins are expressed in HaSV infected *H. armigera* larvae.

3.5 Detection of small proteins in HaSV infected *H. armigera* larvae

A field caught colony of *H. armigera* was used to produce eggs, which were hatched and second instar larvae were infected with HaSV. Infected larvae were stunted when compared to uninfected larvae, a characteristic of a productive HaSV infection. As HaSV is known to infect the larval midgut (Brooks *et al.,* 2002), the heads and tails of control and infected larvae were removed before weighing and subsequent homogenization for western blot analysis.

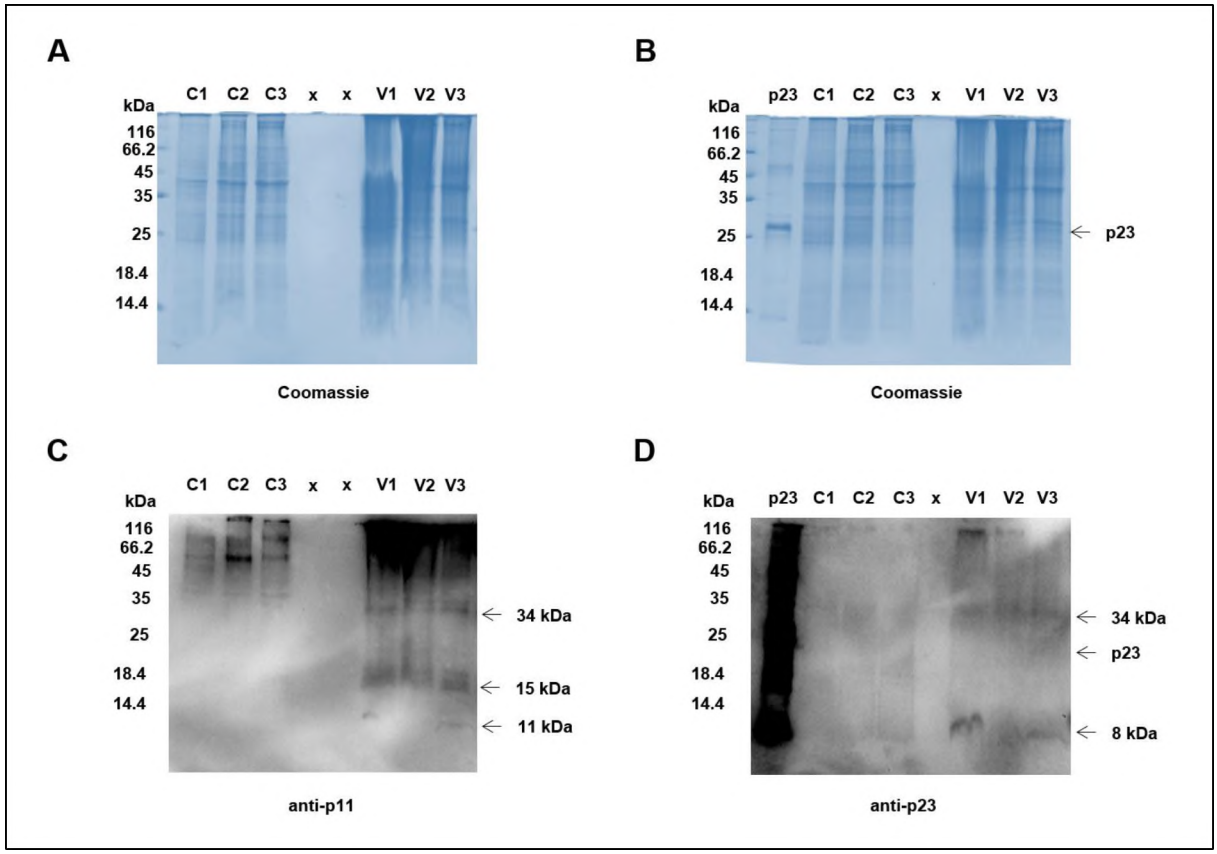

Figure 3.10: Detection of small proteins in HaSV infected *H. armigera* larvae. SDS-PAGE **(A and B)** and western blot analysis **(C and D)** was performed on three uninfected larvae (C1-3) and three HaSV infected larvae (V1-3). Anti-p11 antiserum was used at 1:1000 dilution **(C)**, and anti-p23 antiserum used at 1:50000 **(D)**. The migration of molecular mass markers is indicated on the left of the gels, and arrows on the right indicate the sizes of detected proteins.

Protein extracts derived from these larvae were resolved by SDS-PAGE and stained with Coomassie to estimate protein concentration and ensure equal protein loading (Fig. 3.10A and B). Western blot analysis was conducted with anti-VCAP antiserum. The mature HaSV capsid protein (p64) was detected in two (V2 and V3) of the three infected larvae (Appendix E), confirming that at least these two larvae were productively infected with the virus. Next, western blot analysis was performed with anti-p11 and anti-p23 antisera. Protein bands migrating at 34 kDa were observed in infected larvae with both antisera (Fig. 3.10C, V1-V3 and Fig. 3.10D, V1-V3). This band was not detected in the uninfected larvae (Fig. 3.10C, C1-C3 and Fig. 3.10D, C1-C3). In addition, protein bands migrating at 11 kDa and 17 kDa were detected by the anti-p11 antiserum (Fig. 3.10C, V1-V3), while a band migrating at 8 kDa was observed in the infected larvae probed with anti-p23 antiserum (Fig. 3.10D, V1-V3).

In an attempt to investigate the origin of the 17 kDa band, the complete HaSV vRNA1 sequence was reanalyzed. The NCBI ORF Finder tool was used to screen the vRNA1 sequence for additional ORFs, and one putative ORF (named ORF2) was predicted upstream of, and overlapping with, p11 in a -1 frame (Fig. F1A; Appendix F). ORF2 was predicted to be 84 aa long, which converts to 9.7 kDa protein if expressed. In addition, this protein was not found in DpTV (data not shown). BLAST analysis of ORF2 did not yield any significant results; the best match being acyl-CoA dehydrogenase of *Burkholderia cenocepacia* (GenBank accession number WP_006495643.1) at 34% identity with an E-value of 5.5. No motifs were predicted on this ORF2 using the PROSITE tool (data not shown). Furthermore, analysis of the ORF2 amino acid sequence revealed 8 P, 7 S and 7 T residues, residues often affecting the migration of proteins during SDS-PAGE. The possibility of cross reactivity of the anti-p11 antiserum with this potential protein was investigated by aligning the peptide region of p11, used for anti-p11 production, with the translated ORF2. Of the 14 residues, 3 were identical, 4 showed strong similarity and 1 was weakly similar (Fig. F1B; Appendix F). It was therefore possible that the band detected at 17 kDa with the anti-p11 antibody may have represented ORF2.

The 8 kDa protein was not detected by the anti-p11 antiserum, and the 11 kDa protein was not detected by the anti-p23 antiserum. It was therefore concluded that both anti-p11 and anti-p23 antisera detected a 34 kDa band in infected larvae; the size of this

band corresponding to a potential p11-p15-p8 polyprotein. This indicated the expression of these small proteins during infection of *H. armigera* larvae with HaSV. Following the detection of the small proteins expressed from HaSV vRNA1, the study aimed to investigate the mechanisms by which these proteins are being expressed. Together, these data led to the conclusion that p11, p15 and p8 were indeed expressed in HaSV-infected larvae, and therefore play a role in the infectious virus lifecycle. The detection of bands corresponding to p11 and p8 together with the 34 kDa polyprotein, but no p15 raised interesting questions about mechanism by which these ORFs are translated from RNA1.

3.6 Elucidating the mechanism by which p11, p15 and p8 are expressed

3.6.1 Construction of a recombinant baculovirus expressing p11-p15-p8

Hughes (2013) reported that plasmid-based expression of the small ORFs did not yield the levels of protein in *S. frugiperda* cells that were required for the detection of these proteins by western blot analysis. For this reason, a recombinant baculovirus (BV) system was developed in an attempt to generate the high levels of protein expression that are traditionally associated with baculoviral systems (Invitrogen™; Van Oers, 2011). The system consists of two major components, the pFastBac™ Dual donor plasmid and the baculovirus shuttle vector or bacmid (bMON14272). The expression cassette encoding the proteins of interest are cloned into pFastBac™, which utilises an AcMNPV polyhedrin (polH) or p10 promoter for expression. The donor plasmid also encodes the Herpes Simplex Virus thymidine kinase (HSVtk) and Simian virus 40 (SV40) polyadenylation signals that allow efficient transcription termination and polyadenylation of mRNA (Cole and Stacy, 1985; Westwood *et al.,* 1993). Furthermore, the vector has Tn7R and Tn7L transposon elements flanking the insert that allow for site specific transposition into the attachment Tn7 (attTn7) site on the bacmid of DH10Bac™ cells (Luckow *et al.*, 1993). These cells also contain a helper plasmid (pMON7124) encoding the transposase that is required for transposition.

The p11-p15-p8 expression cassette, carrying the native Kozak sequence upstream of the p11 AUG, was inserted into the pFastBac™ Dual donor plasmid creating pMDB2. The transposition of the p11-p15-p8 cassette from pMDB2 was identified by blue-white screening. Recombinant bacmid DNA was extracted from selected colonies and recombination was confirmed by PCR analysis using two sets of primers, each

containing a p11-specific primer (Fig.3.11A). The primer set pUC/M13F-GH10R would yield a predicted 2461 bp product if the p11-p15-p8 cassette had successfully transposed onto the bacmid and no product if the native bacmid or pFastBac^{TM} Dual had transposed (Fig.3.11A). Similarly, the primer set MDB1-pUC/M13R would yield a 1630 bp product for pMDB2 and no product for the control bacmids (Fig.3.11A).

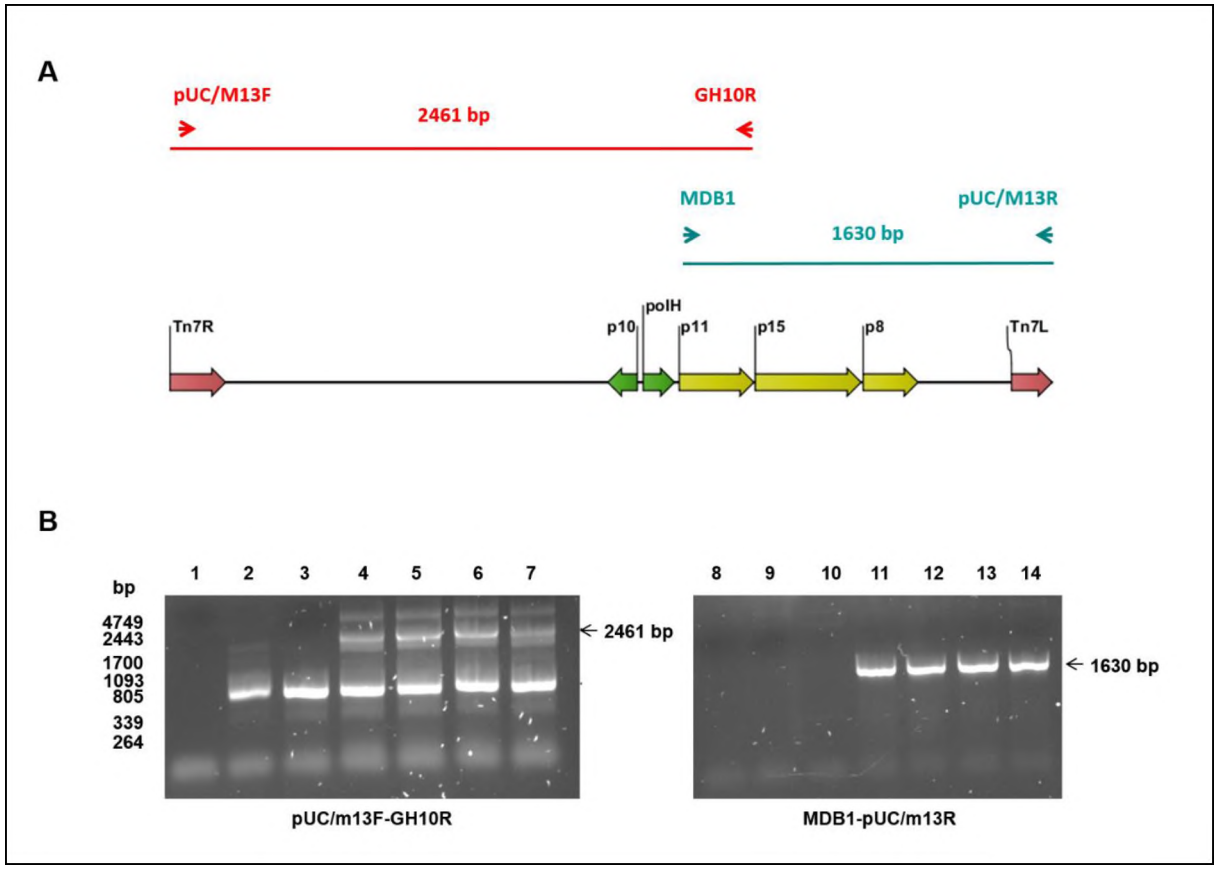

Figure 3.11: PCR verification of transposition into the bacmid DNA. A: Schematic representation of the transposition event resulting in the recombinant bacmid DNA, with primer binding sites and predicted product sizes indicated above. **B:** AGE analysis of the PCR products obtained for the native bacmid (lanes 2 and 9), bacmid with pFastBac Dual^{TM} (lanes 3 and 10) and bacmid with pMDB2 insert (lanes 4-7 and 11-14). A no DNA template control was included for each primer set (lanes 1 and 8). The molecular weight standards are indicated on the left of the agarose gel and arrows on the right of the image indicate the resulting fragment sizes obtained.

A band corresponding to the expected size (2461 bp) was observed after PCR amplification with primer pair pUC/M13F-GH10R in DNA from colonies originating from cells transfected with pMDB2 (Fig. 3.11B, lanes 4-7). This band was absent in the native bacmid and pFastBac^{TM} Dual transformed controls (Fig. 3.11B, lanes 1 and 2-3, respectively). The PCR reactions for the MDB1-pUC/M13R primer set confirmed

this, with a PCR product at 1630 bp for the pMDB2 bacmids that are not seen in any of the control bacmids (Fig. 3.11B). Some non-specific bands were detected in the PCR reactions and these were attributed to non-specific binding of the primers. In the PCR reactions with the pUC/M13F-GH10R primer set, an undesired prominent band was observed at 800 bp for all the bacmid samples (Fig. 3.11B). As GH10R is specific for p11 sequence, it was likely that this non-specific band resulted from binding of the forward primer elsewhere in the bacmid sequence.

The pFastBac^{TM} Dual and pMDB2 bacmids were then transfected into *S. frugiperda* cells and the budding viruses harvested (pFast BV and pMDB2 BV). These were plaque-purified and propagated to a higher titre of 10^9 pfu/mL. Crude extracts were prepared from cells that had been infected with passage 2 (P2; budding virus yielding P3), and analysed by SDS-PAGE and western blot analysis using anti-p23.

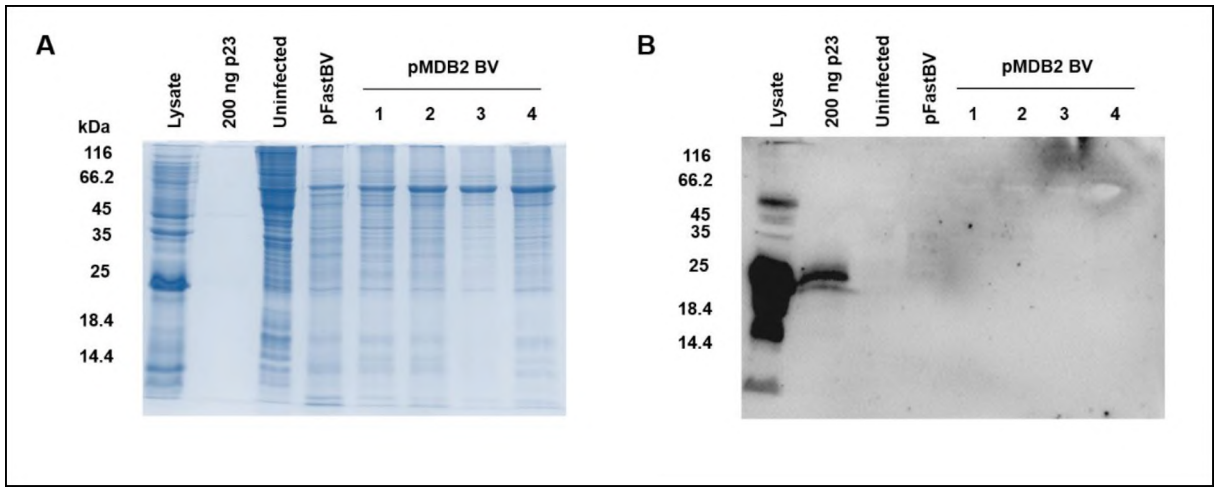

Figure 3.12: Evaluation of protein expression in recombinant baculovirus infected *S. frugiperda* cells. SDS-PAGE **(A)** and western blot analysis **(B)** showing the protein expression following a 4 day infection with passage 2 of pFast BV and pMDB2 BV. The 41 day bleed was used at a 1:50000 dilution. As a control for detection, *E. coli* BL21 (DE3) cells expressing p23, as well as the purified p23 protein, were included. The molecular weight standards are indicated on the left of the gels.

There was no visible difference between the pFast BV and pMDB2 BV infected cells (Fig. 3.12A). The anti-p23 antiserum detected p23 in an *E. coli* cells free lysate (Lysate), and a purified sample (p23) (Fig. 3.12B). These results suggest that p11-p15-p8 was not expressed in baculovirus-infected *S. frugiperda* cells.

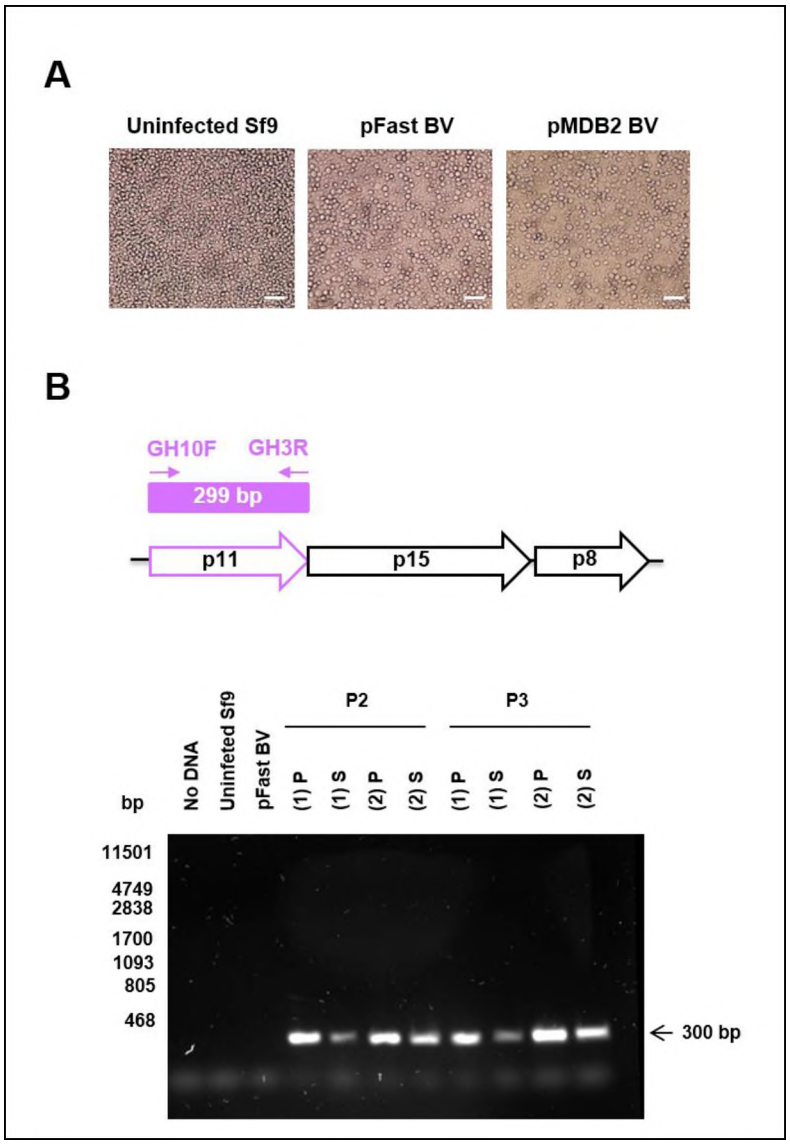

Figure 3.13: Validating recombinant baculovirus infectivity and integrity. A: Uninfected, pFast BV, and pMDB2 BV infected *S. frugiperda* cells under 10x magnification. Scale bars represent 50 µm. B: PCR performed on DNA extracted from uninfected, pFast BV-, and pMDB2 BV infected *S. frugiperda* cells. DNA was extracted from both supernatants (S) and pellets (P) collected for two plaques (1 and 2) at passage 2 and 3 (P2 and P3), and subjected to PCR using p11-specific primers (GH10F and GH10R). The molecular weight standards are indicated on the left of the gel and arrow on the right of the image indicate the resulting fragment sizes obtained.

Further experiments focussed on validating the infectivity of the recombinant baculoviruses. This was done by monitoring *S. frugiperda* cells infected with P2 recombinant baculoviruses for lysis by light microscopy. The uninfected *S. frugiperda* cells remained intact and became visibly overgrown (Fig. 3.13A). In comparison, both the pFast BV and pMDB2 BV infected cells exhibited a mixture of lysed and un-lysed cells (Fig. 3.13A). Furthermore, the cells that were not lysed had a larger diameter

than the uninfected cells (Fig. 3.13A). These results suggested that the BVs were present and were also infective.

The next step involved confirming presence of the p11-p15-p8 expression cassette in the infected *S. frugiperda* cells. Here, PCR was performed using p11-specific primers (Fig. 3.13B). PCR amplification was performed on total DNA extracted from budded P2 and P3 BV (supernatant), and BV infected *S. frugiperda* cells. DNA extracted from uninfected *S. frugiperda* cells and budded pFast BV were included as controls. A band corresponding to the size of p11 (300 bp) was observed for the DNA extracted from the cells and supernatants for both passages of pMDB2, that was absent in the control reactions (Fig. 3.13B). Therefore, the pMDB2 BV in addition to being infective, still encoded the p11-p15-p8 expression cassette. However, it did not express the p11, p15 or p8 proteins.

3.6.2 Detection of p11 and p15 in *S. frugiperda* cells

Since baculovirus expression of p11-p15-p8 cassette was unsuccessful, the mechanism of expression of these small proteins was investigated using plasmid-based expression systems in *S. frugiperda* cells.

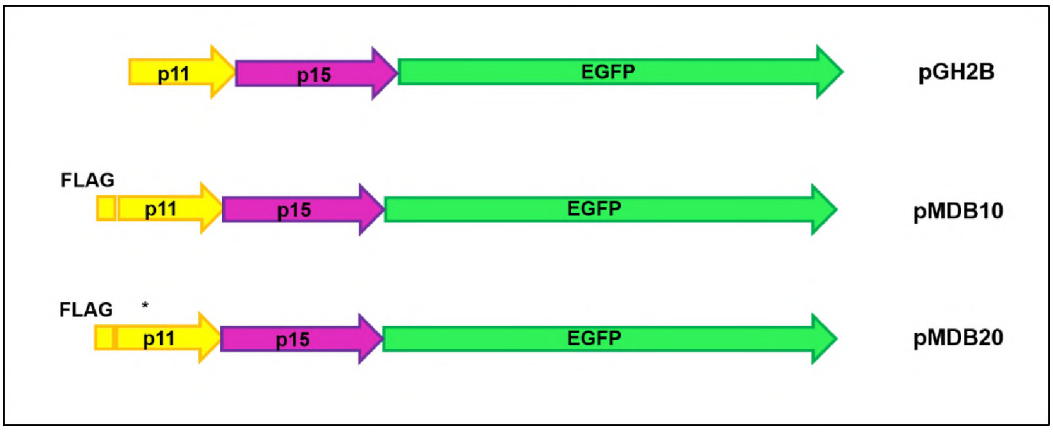

Figure 3.14: Schematic representation of p11 and p15 insect expression vectors. All vectors share the pCW56 plasmid as backbone, with inserts illustrated occurring downstream of an Autographa californica multiple nuclear polyhedrovirus promoter. * - mutated site; EGFP - enhanced green fluorescent protein; FLAG – FLAG peptide tag. The cloning strategies used to create these plasmids can be found in Appendix B. Further details can be found in Table 2.2.

The affinity for p15 expressed in *S. frugiperda* cells was tested, and detection optimized over a range of antiserum dilutions. Cells were transfected with pGH2B, a

plasmid encoding the native p11-p15 CDS with EGFP fused to the C-terminal of p15 (Fig. 3.14). Punctate EGFP (representing p11-p15-EGFP) structures were observed throughout the cytoplasm (Fig. 3.15). When stained with the anti-p23 antiserum, fluorescence was observed in cells expressing p11-p15-EGFP, but not in cells lacking EGFP fluorescence (Fig. 3.15). Since immunofluorescence was observed only in cells expressing p11-p15-EGFP, it was concluded that the anti-p23 antiserum successfully detected p15.

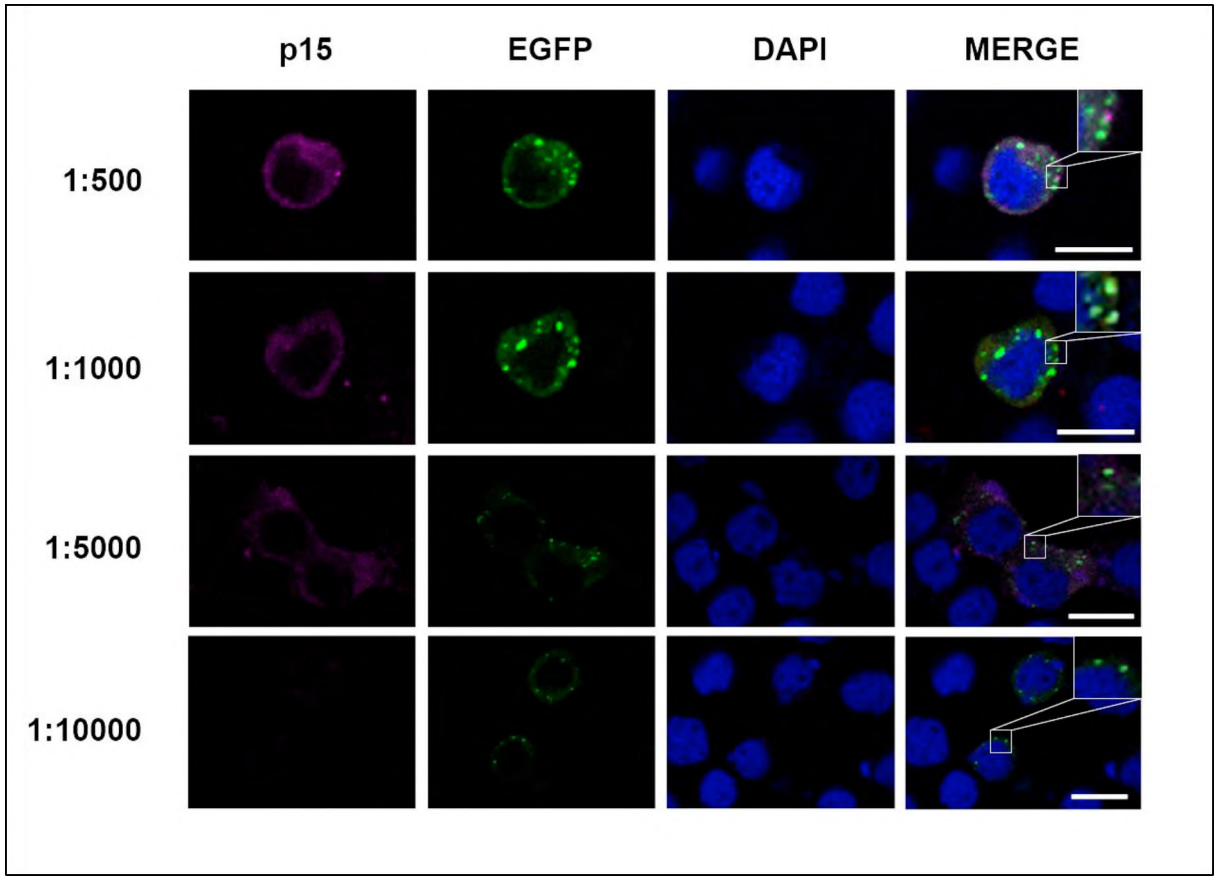

Figure 3.15: Immunofluorescent detection of p15 in *S. frugiperda* cells using the anti-p23 antiserum. Cells were transfected with pGH2B and probed with anti-p23 antiserum at a dilution of 1:500, and a dilution series (1:1000, 1:5000 and 1:10000) was then used to optimize detection. Magnified boxes indicate punctate structures showing no overlap, and scale bars represent 10 µm.

Unexpectedly, punctate structures were not detected with the anti-p23 antiserum, with uniform distrubution of immunofluorescence throughout the cytoplasm observed (Fig. 3.15). The same cytoplasmic distribution was observed when the antiserum was diluted (Fig. 3.15, 1:1000-1:10000). No overlap was observed between the punctate structures observed for the EGFP fluorescence and those observed for p23.

A construct encoding p11-p15-p8-EGFP was also transfected and probed using both anti-p11 and anti-p23 antibodies, however no EGFP or immunofluorescence was detected in any cells (data not shown).

To investigate the expression of p11 and p15, a construct encoding FLAG-p11-p15-EGFP (pMDB10) was created (Fig. 3.14) with the aim of performing immunoprecipitation experiments using anti-GFP antibodies, and subsequent western blot analysis using anti-p11 and anti-p23 antiserum. These attempts were, however, unsuccessful. Though the anti-GFP antibodies successfully coupled to the beads, no p11 or p15 proteins were detected, possibly due to levels of protein expression being too low for binding and subsequent detection (data not shown). This lead to immunofluorescence analysis, and subsequent mutation, of the pMDB10 construct.

Cells were transfected with the pMDB10 expression vector, and probed with mouse anti-p11 and rabbit anti-p23 antisera. The anti-p23 antiserum detected proteins uniformly throughout the cytoplasm of cells exhibiting EGFP fluorescence, as seen previously in Fig. 3.15, with some punctate EGFP structures observed. These structures overlapped with punctate structures detected with the anti-p11 antiserum (Fig. 3.16A). The perfect overlap of p11 with EGFP (Fig. 3.16A) indicated potential expression as a p11-p15 polyprotein, and/or as discrete p11 and p15 since both are targeted to the secretory vesicles. The results obtained from the western blot analysis of infected larvae indicated that all three of the small proteins are likely to be expressed as a single polyprotein, or with respect to p11 and p8, as individual proteins.

To determine whether translation of p15 is dependent upon p11 translation, a frame shift mutation was introduced into p11 CDS of pMDB10, shifting the p15 CDS to the +1 frame relative to p11 (Fig. 3.14). This mutated construct would express a 55 aa FLAG-tagged p11 peptide, where the first 42 amino acids were that of p11 (Appendix G). Since anti-p11 antibodies were raised against a peptide corresponding to aa residues 77-90 of p11, the anti-p11 antiserum would not detect any protein whereas the anti-FLAG antibody would detect the new 55 aa protein. The detection of EGFP fluorescence would indicate p15 expression independently of p11 translation.

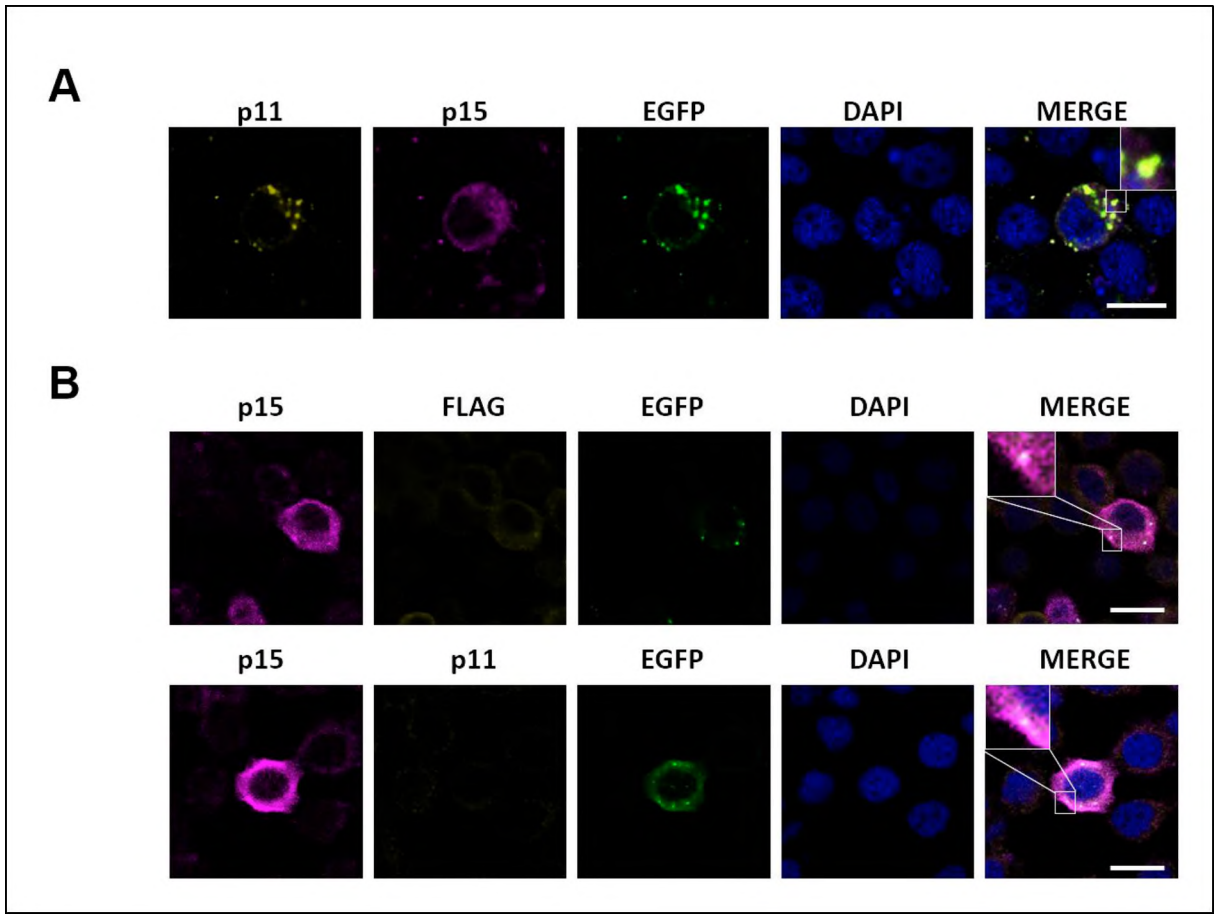

Figure 3.16: Evaluating the dependency of p15 expression on p11 in *S. frugiperda* cells. Cells transfected with **(A)** pMDB10 encoding FLAG-p11-p15-EGFP, and **(B)** pMDB20 encoding a mutated FLAG-p11-p15-EGFP, with p15 occurring in a +1 frame to p11, probed with anti-p11 and anti-p23 antisera at dilutions of 1:1000 and 1:5000, respectively. Anti-FLAG antibodies used at a dilution of 1:500. Magnified boxes indicate punctate structures showing perfect overlap and scale bars represent 10 µm.

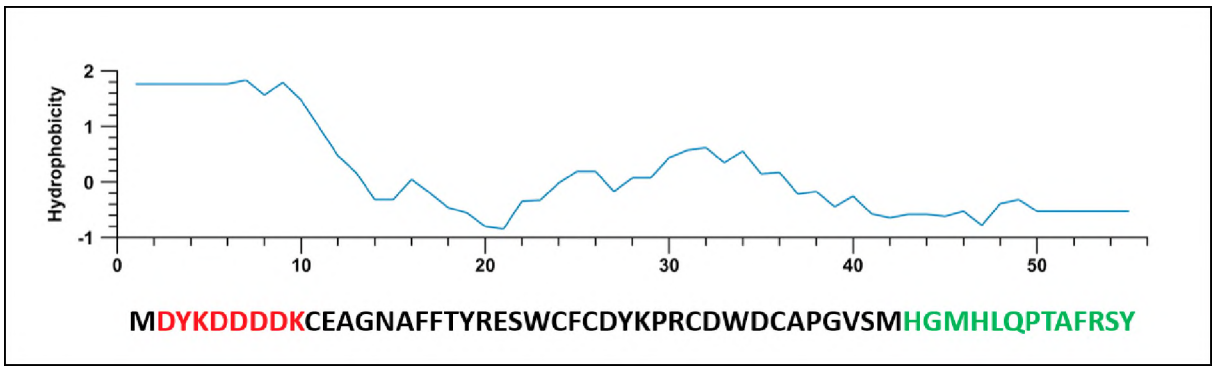

Figure 3.17: Hydrophobicity plot of p11 produced by pMDB20. Hopp-Woods plot show hydrophilic antigenicity scores as positive values, and hydrophobic scores as negative. The amino acid sequence is positioned on the x-axis, with the FLAG-tag depicted in red and region affected by the mutation in green.

When this construct was transfected into *S. frugiperda* cells, the anti-p23 antiserum detected p15 in transfected cells that were also expressing EGFP (Figure 3.16B). The distribution of p15 was consistent with that observed previously with pGH2B (Fig. 3.15) and pMDB10 (Fig. 3.16A). A small number of punctate structures were visible through a uniform distribution of fluorescence in the cytoplasm (Fig. 3.16B).

No fluorescence was detected with cells stained with either the anti-p11 antiserum or anti-FLAG antibodies. The latter observation was unexpected. The hydrophobicity of the mutant protein that would be generated by the frame shift was therefore investigated using a Hopp-Woods hydrophobicity plot. The FLAG-tag exhibited scores ranging from 1-2, and residues 30-35 scores between 0-0.5. The remainder of the protein was hydrophobic with scores below zero (Fig. 1.17). This change in the protein could have resulted in this aberrant observation. Alternatively, it is possible that the truncated FLAG-p11 protein degraded due to a change in stability. Preliminary results suggested that p15 could be expressed independently of p11.

Chapter 4:
Discussion

4.1 Sequence conservation of small ORFs

The HaSV_RNA1 sequence is currently one of two known sequences for RNA1. The second sequence was reported by Gordon *et al.* (GenBank accession number NC_001981.1; 1995). The latter sequence was not used as reference for alignment, due to the availability of the plasmid (pAV8F) from which the HaSV_RNA1 sequence was obtained. This plasmid pAV8F was the template from which all the expression vectors were generated and to which all the antisera were generated. In addition, sequencing results confirm that this sequence was from an infectious virus. It should therefore be possible to detect the small proteins, if expressed during infection, using antisera generated against proteins generated using the pAV8F sequence.

The data in this report support previous findings that show absolute amino acid conservation of p11 and p15 between four cDNA clones and the HaSV_RNA1 sequence (Baillie, unpublished data). The larvae that were used for virus purification were infected with HaSV that has undergone multiple passages through larvae since the Baillie study (2008). Considering the error prone nature of RNA virus replication (Lauring and Adino, 2010), as well as the conservation of the size and the position of both p11 and p15 in DpTV, this implies a conserved function for the p11 and p15 proteins in omegatetraviral infection. Sequencing analysis further revealed single point mutations in the p8 protein in two of the nine cDNA clones, which either introduced or disrupted a predicted alpha helix. This would alter the tertiary structure of the protein. The potential effect on the function of the protein is discussed in Section 4.4 of the discussion.

4.2 Expression of small proteins during infection

In order to investigate whether the small proteins were expressed during infection, suitable antibodies were required for detection. It was therefore decided to generate polyclonal antisera capable of recognizing the small proteins by using a p11 peptide, and a p15-p8 fusion protein expressed in *E. coli.* The affinity of these antisera for their

target proteins was tested before attempting detection of the small proteins in HaSV-infected *H. armigera* larvae.

4.2.1 Production and testing polyclonal antiserum for detection of p11

When selecting the antigenic region of p11, two regions were identified. The region spanning residues 19-32 of p11 was identified as an antigenic determinant of this protein, however this region was less antigenic than residues 77-90. Apart from its lower antigenicity, this region was not chosen for subsequent antiserum production as it encodes two cysteine residues that would interfere with the conjugation of the antigenic peptide to KLH that is used to enhance antiserum production (Lane *et al.,* 1982).

To test the affinity of anti-p11 for its target protein, attempts were made to overexpress p11 in *E. coli*. Previous attempts at this were unsuccessful, likely due to codon bias in the *E. coli* cells. The tRNA abundance in an organism, *E. coli* in this case, modulates gene expression allowing selective expression of endogenous genes over those that are introduced (Holm, 1986; Kurland, 1991). The failure of Baillie's (2008) approach highlighted the importance of codon optimization prior to gene expression studies. Attempts at overexpressing codon optimized p11 in order to validate anti-p11 affinity, were however, unsuccessful. Investigation into the problem included monitoring the optical density, representing bacterial growth, following induction of p11 protein expression with IPTG. The effect of IPTG on bacterial growth was also investigated, so that the net effect of p11 could be determined. As expected, the IPTG lowered the growth of both the control and p11-expressing cells (Dvorak *et al.,* 2015), however the reduction in the growth rate of p11-expressing cells was marked. There was no increase in the growth of cells transformed with the p11-encoding plasmid, suggesting that the 6xHis-p11-FLAG protein was toxic in bacterial cells.

Certain steps were taken to further express p11 in *E. coli*, including the initial addition of small amounts of glucose to the growth medium of transformed cells, in order to prevent leaky expression of p11. This would prevent protein production until the cells were in log phase, at which stage they were transferred to glucose free expression medium, and expression induced with IPTG. Further approaches to improve expression could have involved reduced concentrations of IPTG, and temperatures

used for induction (Saïda *et al.,* 2006). Alternatively, this challenge could be overcome by using *E. coli* strains containing the pLysS or pLysE plasmids that encode and constitutively express the T7 lysozyme. This reduces the level of basal expression of target genes by controlling T7-RNA polymerase activity in the cells (Saïda *et al.,* 2006).

As mentioned in Chapter 1, p11 has a predicted transmembrane domain spanning residues 52-69 (Hughes, 2013). Transmembrane proteins often induce toxic effects when expressed in bacteria, attributed to their destabilization of membranes as transmembrane proteins often form channels (Montigny *et al.,* 2004). Channels and porins, common transmembrane proteins, comprise of multiple alpha-helices or beta-barrels, respectively (Rost, 2001). The p11 protein, despite encoding a predicted transmembrane domain, forms a single beta-strand (Hughes, 2013). The exact mechanism by which p11 inhibits bacterial cell growth is therefore not known.

In view of the problems encountered with heterologous expression of p11 in *E. coli* BL21 (DE3) cells, it was decided to turn to the *S. frugiperda* expression system to test the affinity of the anti-p11 antiserum. The anti-p11 antiserum was unable to detect p11 of p11-EGFP, but could detect p11 in a construct expressing FLAG-p11-2A-EGFP. It is likely that the EGFP masked the region against which the antiserum was generated, since it was located only 9 residues from the C-terminus of p11. Furthermore, immunofluorescence data showed perfect overlap between anti-p11 and EGFP fluorescence, indicating that the full read-though product FLAG-p11-EGFP was detected. The data in this report suggest that the translation product of the 2A-like site physically negated the effect of EGFP on detection of p11. This 20 aa 2A-like sequence is predominantly hydrophilic and located between the C-terminus of p11 and N-terminus of EGFP, effectively permitting detection of p11.

Though the 2A-like site allowed detection of p11, the activity of the 2A site was not optimal. Cells transfected with 2A-EGFP displayed both cytoplasmic and nuclear EGFP fluorescence, consistent with that of EGFP only. This distribution changed when the FLAG-p11 expression casette was introduced, with EGFP observed predominantly as punctate cytoplasmic structures. Some diffuse fluorescence was still observed throughout the cytoplasm and nucleus. The punctate structures observed with pMDB9 were similar to those observed in cells that had been transfected with p11-EGFP,

indicating suboptimal activity of the 2A-like site. The ERAV 2A-like site is an oligopeptide sequence that mediates ribosome stop and carry on during translation effectively producing individual proteins from a polyprotein sequence as described in Chapter 1 (Donnely *et al.*, 2001; Luke *et al.*, 2008). Theoretically, only 10% of the proteins expressed should have been tagged with EGFP if the 2A site was functional (Donnely *et al.*, 2001; Short, 2010). This means that the EGFP distribution of pMDB9 should have resembled that of pJA1 (2A-EGFP) more closely than pGH6 (p11-EGFP). It was therefore likely that the presence of p11 upstream of the 2A-like site may have interfered with processing activity. A study by Minskaia and Ryan (2013) revealed that sequences up- and downstream of the Foot-and-mouth disease virus 2A and "2A-like" site of Thosea asigna virus, affected the efficiency thereof. Despite the suboptimal activity of the 2A site, this data presented the first evidence for the successful detection of p11.

4.2.2 Production and testing polyclonal antiserum for detection of p15 and p8

Continuing on from a previous study by Hughes (2013), a p15-p8 fusion protein was expressed and purified from *E. coli* BL21 (DE3) cells. Some endogenous *E. coli* proteins co-purified with the 23 kDa protein, a problem that is common when using Ni-affinity chromatography of proteins expressed in *E. coli.* These contaminating proteins are likely to have clustered histidine residues or have biologically relevant metal binding sites within the native proteins (Robichon *et al.*, 2011). The anti-p23 antiserum was therefore generated against p23 as well as these contaminating native *E. coli* proteins, however was not considered a major issue as the antisera was to be used in experiments with insect and not bacterial cell lysates. The affinity of the antiserum for p23 was then investigated by western blot analysis. A band at 23 kDa was detected in the purified protein sample, that was absent in a cell free extract of cells expressing HaSV p17. It was therefore concluded that the anti-p23 antiserum was successful in detection of p23. The antiserum was also shown to detect p15 in immunofluorescence analysis, discussed in Section 4.4.

4.2.3 Detecting small proteins in HaSV-infected *H. armigera* larvae

After validating the affinity of antisera for the small proteins, western blot analysis was performed using both anti-p11 and anti-p23, respectively, on crude extracts of HaSV-infected *H. armigera* larvae. Bands migrating at approximately 34 kDa, the sum of p11,

p15 and p8, were observed in infected larvae with both antisera. This indicated that the three small proteins were in fact expressed, and that they were expressed as a polyprotein. In addition, bands migrating at approximately 11 kDa and 8 kDa were detected by the anti-p11 and anti-p23 antisera, respectively. This raised the possibility of post translational cleavage or discrete expression of these proteins in the infected larvae.

A band migrating at approximately 17 kDa was detected by the anti-p11 antiserum. This band could not represent p17 as it was not raised against p17 and is known to migrate at an approximate size of 24 kDa. This is due to the high number of P, E, S and T residues in p17. These amino acids are often post translationally modified (e.g. phosphorylated) and such modifications physically alter the rate of migration during SDS-PAGE (Rogers, 1986). An additional ORF was identified (ORF2) that occurs upstream of and overlaps with p11 in a -1 frame. The size of this ORF is 9.7 kDa, however due to a high P, S and T content, it may appear 5-7 kDa larger during SDS-PAGE analysis. This is similar to what is observed for the p17 protein. The possibility of anti-p11 antiserum cross reacting with this protein was investigated by aligning the p11 peptide against ORF2. Some similarities were observed, making it difficult to rule out the possibility of the band detected at 17 kDa with anti-p11 may have represented this ORF2 protein. The function of this protein, if it is produced, could not be determined by preliminary bioinformatic analysis. Further analysis is therefore required to assign a function to this protein.

HaSV infection of the larvae was confirmed by western blot analysis using antiserum targeting the capsid protein of the virus. The capsid protein was detected in two of the three larvae. This may indicate different stages of infection, with VCAP expression representing the later stages of infection. There was no correlation between the levels of small protein expression to that of VCAP, as the intensity of the small protein bands were comparable in all the infected animals. This suggests that the small proteins are produced and active throughout the viral life cycle.

4.3 Mechanism by which small proteins are expressed

Western blot analysis of crude extracts of infected larvae indicated that all three of the small proteins p11, p15 and p8 could be expressed as a single polyprotein or as

individual proteins. To produce the polyprotein, a read-through stop mechanism of translation is required. The type I consensus sequence does not correlate with the sequences surrounding the p11 or p15 stop codons so this type of read-through mechanism is not expected to occur during HaSV infection (Hughes, 2013).

As previously mentioned, the replicase of HaSV is alphavirus-like (Dorrington *et al.*, 2011). The UGA stop codon in type II read-through sequences is followed by a 3' adjacent arginine (CGG) codon in several plant and animal alphaviruses (Li and Rice, 1989). Although p11 possesses a UGA stop codon, it is unlikely to be the mechanism of translation employed by HaSV as the p15 start codon occurs immediately after the stop codon. Type III read-through stop codons (UAG, UGA) display a guanine after the stop codon (reviewed in Beier *et al.,* 2001), however this is not the case for p11 or p15, nor is it the case for p8, though the stop codon of p15 (UAA) and the start codon of p8 are separated by GGG. Therefore, the type III mechanism of read-trough translation is not thought to occur in the translation of these ORFs. This may suggest that p11-p15-p8 are produced by a novel, as yet undescribed, type of stop codon read-through.

Since baculovirus expression of the p11-p15-p8 cassette was unsuccessful, the mechanism of expression of these small proteins was further investigated using plasmid-based expression systems. Attempts at detecting the small proteins expressed in *S. frugiperda* cells by western blot analysis was unsuccessful, leading to immunofluorescence analysis on transfected cells in an alternative approach. Since it was not known whether the anti-p23 antiserum could detect p15 only, the affinity for p15 expressed in *S. frugiperda* cells was first tested and detection optimized over a range of antiserum dilutions. The antiserum successfully detected p15 in p11-p15-EGFP expressing cells. While punctate structures were observed with anti-p23 fluorescence, a diffuse distribution of fluorescence was predominantly observed throughout the cytoplasm. This contrasts with the data collected with anti-p11, antiserum where predominantly punctate structures were observed in the cytoplasm. The distribution of fluorescence did not change when more dilute anti-p23 antiserum was used, therefore the signal was not considered to have resulted from non-specific binding.

This diffuse distribution was also observed in cells expressing FLAG-p11-p15-EGFP, with some punctate EGFP structures overlapping with fluorescence of both antisera. Considering that p11-p15 is produced as a polyprotein, it is possible that p11 serves to anchor p15 to the membranes of secretory vesicles (Hughes, 2013). When p15 is then individually produced, either by leaky scanning or more likely post-translational cleavage (requiring the action of a host protease), the cytoplasmic distribution of p15 may change. This possibility is however, contradicted by the lack of a 15 kDa band in the analysis of the proteins from infected larvae.

The dependency of p15 expression on that of p11 was therefore investigated, done by shifting the frame of p15 with respect to that of p11 in FLAG-p11-p15-EGFP. Preliminary results showed that p15 was still produced, with a similar distribution as before, suggesting that p15 could be produced independently of p11. While a truncated p11 was produced, the start codon of p11 (upstream of the FLAG tag) and an internal methionine in p11 could still initiate translation. If the initiation of p15 translation in this experiment was dependent on the start of p11, a mechanism of reinitiation would have been required to produce p15.

Table 4.1: Comparing the Kozak sequences of the three small ORFs in HaSV to the insect consensus sequence.

ORF	Kozak sequence						
	C	*A/G	*A/C	C/G	**AUG**	A/G	*C
p11	C	G	A	C	**AUG**	U	G
p15	C	U	G	A	**AUG**	U	C
p8	A	G	C	G	**AUG**	C	U

Sequences in **bold** - start codons for initiation of translation. * - More important nucleotides in the Kozak sequence (Nakagawa *et al.*, 2008). Adapted from: Cavender, 1987; Cavender *et al.,* 1991.

The efficiency of translation initiation is determined by the Kozak sequence (Kozak, 1986; Nakagawa *et al.*, 2008). This sequence is recognized by the ribosome as the translational start site. A comparison of the Kozak sequence of p11 to the insect consensus sequence (Cavender, 1987; Cavender *et al.,* 1991) suggests that the p11 Kozak will favour initiation of translation compared to that of p15 and p8 (Table 4.1).

The Kozak sequence for p8 is similar to that of the consensus (Table 4.1), however the last nucleotide differs (U not C) and does not favour initiation of translation. If p15 expression is dependent on p11, the Kozak sequences could explain why only p11 and p8 were detected as individual proteins in infected larvae. This hypothesis would require scanning after p11 termination and reinitiation (at p8) mechanisms to occur, despite the fact that no TURBS were detected, where p11 would be produced discretely (without stop codon read through to p15).

In a study investigating the host specificity of HaSV infection, Bawden *et al.* (1999) performed a number of northern blots using RNA-probes created from the 3' termini of both vRNA1 and vRNA2. RNA smaller than RNA2 can be seen in the infected larvae but not in mock infected larvae (Fig.1 in Bawden *et al.*, 1999). Similar results were observed in northern blots on total RNA extracted from *S. frugiperda* cells, transfected with vRNA1 and vRNA2, where a definitive band can be seen below vRNA2 (Fig. 3 in Bawden *et al.*, 1999). The authors did not, however, comment on these bands. Further, they concluded that no sgRNA was detected without elaborating on this statement. Upon inspection, these bands could potentially represent sgRNA that was approximately 1000 nts in length.

Similar to HaSV, alphanodaviruses (family *Nodaviridae*) have a bipartite genome, with a larger vRNA1 and smaller vRNA2. The vRNA1 encodes protein A that houses the RdRp domain and vRNA2 encodes the viral capsid protein. Interestingly, these viruses are known to produce a sgRNA molecule (RNA3) from the 3' end of vRNA1, encoding two ORFs for proteins B1 and B2 (Ball and Johnson,1999; Schneemann *et al.*,1998). If HaSV does produce sgRNA encoding the small proteins, it is likely to occur via an internal initiation mechanism, similar to that of alphanodaviruses. Alternatively, it could also occur via a premature termination mechanism. However, this would have to be determined experimentally, by northern analysis using probes generated against negative sense of RNA1.

It is likely that HaSV exploits cap-dependent initiation of translation for the REP, as it exhibits a 5' cap and a 3' tRNA-like structure that is known to promote circularization of vRNA (Mans *et al.*, 1991). If HaSV produces sgRNA to express the small proteins, and ORF2 is not included therein, translation of the small proteins could occur without

any frameshift mechanism required. However, if ORF2 is included in the sgRNA, or if no sgRNA is produced, a frameshift would be required for initiation of translation of the small proteins.

Though mechanisms of reinitiation may be involved in the discrete expression of these proteins, this mechanism requires translation to stop and restart. Considering the p11, p15 and p8 ORFs overlap with the REP ORF, this could not be the case for initiation of expression, in the event that the small proteins are not expressed via sgRNA. No IRESs or pseudoknots have been predicted to form in this region (Hughes, 2013), indicating the need for either a hairpin loop structure or a slippery sequence for frameshifting. However, no slippery sequences were detected in residues surrounding the start codons of the small proteins, indicating that the small (<30 nt) predicted hairpin loop structure occurring 3' of the start codon in p11 would have to be sufficient to cause the +2 frameshift (Hughes, 2013). This scenario is, however, highly unlikely. If ORF2 is expressed, this would require a +1 frameshift with respect to the REP, and a further +1 shift to express p11-p15-p8. However, analysis on the secondary RNA structure showed that the start codon of ORF2 occurs within a hairpin loop structure, suggesting that this is not the case.

4.4 Potential function in the infectious viral lifecycle

Initial attempts at elucidating the mechanism of expression included using recombinant baculovirus-based approach. This system is renowned for its high levels of protein expression when compared to plasmid-based systems, however, no proteins were detected in recombinant baculovirus-infected *S. frugiperda* cells. Though a very rare occurrence, this is often associated with protein toxicity in cells (Van Oers *et al.,* 2015). These results support those that were obtained for p11 expression in *E. coli* BL21 (DE3), suggesting that p11 is toxic to cells. Similar results were observed in attempts to express and detect p11-p15-p8-EGFP in *S. frugiperda* cells. No EGFP was observed and in addition, no signal was detected in the cells when they were probed with anti-p11 and anti-p23 antisera. Hughes (2013) attributed the lack of EGFP detected to low or no levels of p8 being expressed. The detection of an 8 kDa protein in the western blot analysis of HaSV-infected larvae, contradicts this hypothesis. Failure in detecting p11 and p15 suggested that EGFP, when fused to the C-terminus of both p11 and p15, affected the potential function of the proteins in *S.*

frugiperda cells. Based on failed attempts thus far to express p11 in *E. coli*, and p11-p15-p8 by recombinant baculovirus in *S. frugiperda* cells, both lacking C-terminal tags, the inability to express the small proteins supported the toxicity of p11.

There are seven cysteine (C) residues in p11, grouped as either CxC or CxxxC (where x is any amino acid). Some of these residues are conserved in both HaSV and DpTV (Hughes, 2013). This may be important at a functional level. Unlike the residues in p8, these do not occur in CxxC, so they are not zinc finger domains that bind DNA. While these residues do not match known motifs, it is possible that these residues may be involved in protein binding. Hughes also showed that p11-EGFP and p11-p15-EGFP co-localised with the HaSV REP in *S. frugiperda* cells, and proposed that these small proteins likely formed part of the replication complex. It is therefore interesting to speculate that p11 interacts with host proteins in order to protect the replication complex.

There are three CxxC motifs in p15, often occurring in zinc-fingers, however none of which have been found to match known zinc-finger motifs (James Short, unpublished data). One of these motifs (CRRC) is known to be associated with inhibitors of Toll-like receptor (TLR) 9-like proteins, an innate pattern-recognition receptor in mammalian cells, recognizing uncapped DNA (Takuechi and Akira, 2007). The TLR3 and TLR7 recognize dsRNA and ssRNA, respectively, however none of these TLRs are known to occur in insect cells. Toll-9 of *Drosophila*, however, shares structural and functional similarities with mammalian TLRs, exhibiting leucine-rich repeats flanked by cysteine-rich motifs (Bilak *et al.,* 2003). It is also interesting to note that all three of these TLRs localize to cytoplasmic vesicles including that of the secretory pathway (Takuechi and Akira, 2007). It is therefore possible for p15 to interact with the host immune response, however this would have to be experimentally determined.

A possible zinc-finger domain between Cys^{44} to Cys^{62} (C/X2/C/X11/C/X2/C) was found in p8 (James Short, unpublished data). It is therefore likely to act as nucleic acid binding protein, where it can either bind vRNA and interestingly, act as an RNA silencing suppressor (RSS) protein of RNA interference (RNAi)-mediated antiviral defence employed by the host immune system. This is the primary response against viral infection in insect cells, in which Dicer-2 recognizes and cleaves viral dsRNA into

small interfering (si)RNAs (Gammon and Mello, 2013). Known viral RSSs include the B2 protein of nodaviruses (e.g. Wuhan Nodavirus), which B2 protein silences the RNAi-mediated response by sequestering the dsRNA (Qi *et al.,* 2011).

One of the mutations detected in the p8 CDS of a cDNA clone was located within this region, with residue 56 mutated from L to P. While both are hydrophobic amino acids, proline is more hydrophilic than leucine. If this region in p8 has a role in binding nucleic acids, then this mutation may affect the hydrophobicity and consequently the functionality, of the hydrophobic pockets that are a characteristic of zinc finger domains (Brown, 2005). Though p8 likely functions to bind nucleic acids, the mutations detected in two of the cDNA clones, and further lack of conservation among omegatetraviruses questions the absolute necessity of this protein for infection.

Chapter 5:
Conclusions and future work

This study aimed to investigate the expression of the small proteins during HaSV infection. To date, no studies of this nature have been performed on the omegatetraviruses. To achieve this, antisera capable of detecting p11, p15 and p8 were generated, their affinity for respective target proteins validated, and used to detect the small proteins in infected larvae. It was evident that the proteins were produced as a polyprotein, and that the p11 and p8 proteins were detected as discrete proteins as well. Production of the polyprotein likely involve a novel mechanism of stop codon read through, however, the exact mechanism by which p11 and p8 can be produced as discrete proteins has yet to be elucidated.

Preliminary immunofluorescence studies on p15 expression suggest that p15 may also be expressed as a discrete protein, independent of p11. This is however contradicted by results from the analysis of the proteins isolated from infected larvae, as no band corresponding to the size of p15 was detected with anti-p23 antiserum. The Kozak sequence for p11 favours initiation of translation over that of p15, and also favours the expression of p8 over that of p15. It is therefore possible that p15 occurs only as a polyprotein, with p11 anchoring the polyprotein to the membranes of secretory vesicles.

The functions of these proteins have not yet been determined, however results suggest that p11 exhibits a detrimental effect on both bacterial and insect host cells. Since these proteins are known to co-localize with the REP in *S. frugiperda* cells, they are likely to form part of the replication complex. It is likely that p11, p15 and p8 interact with the host immune system, protecting the vRNA during replication. Before this report, HaSV was known to express only three proteins; REP, p17 and VCAP. The results obtained in this study showed that HaSV expressed three additional proteins during infection. In addition, this work has identified a new putative protein, ORF2. Considering its relatively small genome size, less than 8 kb in total, the complexity of HaSV is remarkable.

Though this study has shown the expression of the small proteins in HaSV-infected larvae, many questions remain unanswered. Future work will therefore aim to further explore the mechanisms by which these proteins are expressed. This will be done by firstly determining whether HaSV produces sgRNA by northern blotting experiments. Furthermore, the start codon, and the methionine residue within p11, will be deleted to determine if p15 can truly be expressed independently of p11. Immunoprecipitation of the small proteins could also be conducted, followed by mass spectrometry to definitively determine the size of the proteins. Both EGFP fluorescence, and the antisera generated in this study will be used to detect p11 and p15 in such an experiment. These antisera can further be used to detect the small proteins in HaSV-infected *S. frugiperda* cells to verify the subcellular localization thereof, and will also allow the monitoring of these proteins throughout the infectious cycle. Knockout studies could also be performed to further determine the functions, and necessity of these proteins in the infectious viral life cycle.

Further elucidation of the mechanisms employed by HaSV to produce the small proteins, and their function in the infectious life cycle, may provide novel insights into the replication of other tetraviruses and possibly other insect RNA viruses. This is of particular importance, as most epidemic outbreaks are cause by insect borne RNA viruses. Well known examples are classified under the family *Flaviviridae*, including Zika, Dengue, West Nile and tick-borne encephalitis and yellow fever viruses (Barba-Spaeth *et al.,* 2016). In order to limit the damage caused by these outbreaks, vaccines are primarily developed and employed as a preventative measure. This involves injection of a viral antigen or attenuated virus at low concentrations, allowing the body to naturally generate antibodies capable of neutralizing infecting virus particles (Fenner *et al.*, 2013).

There are, however, many challenges associated with vaccines. One of the biggest concerns arise from virus mutations, especially seen in RNA viruses due to their error prone replicases. For this reason, vaccines must be adapted frequently as seen in annual flu-vaccines (Carrat *et al.*, 2007). That being said, even if we are able to keep up with the viral mutation trends, there are some cases in which vaccines can exacerbate infection. This phenomenon is referred to as antibody-dependent enhancement of infection, and is seen in viruses targeting macrophages and

granulated cells of the immune system (Tirado and Yoon, 2003). Viruses that do so include the previously mentioned flaviviruses, and the Sindbis virus (Subfamily *Alphavirus*). Alternative treatments for infection of these viruses as prevention is not guaranteed. An understanding of the virus-host interactions is therefore imperative, as this will enable us to identify new potential drug targets.

Appendix A
Western blot analyses

Samples were mixed with 4x sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) sample buffer (40% Glycerol, 240 mM Tris-HCl pH 6.8, 8% SDS, 0.04% bromophenol blue, 5% beta-mercaptoethanol). After boiling for 10 minutes, the sample was subjected to SDS-PAGE using a 15% resolving, and 4% stacking gel, electrophoresed in the Mini-PROTEAN 3 electrophoresis system (Biorad®) for 90 minutes at 150V. One gel was visualised by Coomassie staining (Laemmli, 1970), while another was transferred to an Immobilon®-P polyvinylidene difluoride membrane (Millipore) using the Mini Trans-blot electrophoretic transfer cell system (Biorad®) as per the manufacturer's instructions at 350 mA for 90 minutes. To confirm successful transfer of the proteins onto the membrane and to mark the positions of the molecular weight markers, Ponceau-S (Salinovich and Montelaro, 1986) was used to stain the membrane for 5 minutes. The water soluble stain was then washed off before continuing with the blocking step.

The membrane was incubated overnight in 5% non-fat milk powder in Tris-buffered saline (TBS, 50 mM Tris-HCl, 150 mM NaCl, pH 7.5) containing 0.1% Tween-20. Primary antibodies, diluted in TBS-Tween containing 1% bovine serum albumin (BSA) lyophilised pH7 (Biowest), were applied to the membrane and incubated for 90 minutes at room temperature. Following a set of three 15 minute washes with TBS-Tween, the membrane was incubated in secondary antibodies conjugated to horseradish peroxidase (HRP) for 30 minutes. The membrane was then washed thrice for 15 minutes with TBS-Tween, and Western Bright ECL HRP substrate (Advansta, Cat. No. K-12045-D20) was applied for subsequent chemiluminescence detection using the ChemiDoc™ XRS+ (Bio-rad).

Appendix B
Construction of plasmids

Unless otherwise stated, cloning was performed in the following manner. PCR was conducted using Hifi Taq (Kapa Biosystems, Cat. No. KK2101) as per the manufacturer's instructions, with the cycling parameters detailed below for respective primer sets. The product obtained was then gel purified using the Isolate II PCR and Gel kit (Bioline, Cat. No. BIO-52059). After purification, the sample was mixed with KAPA Taq ReadyMix (KAPA Biosystems), in a 1:1 (v/v) ratio, and incubated at 72°C for 15 minutes followed by 5 minutes on ice. An ultimate purification step was performed using the Zymo Clean & Concentrator kit (Cat. No. D4004), followed by subsequent ligation of the product into pDrive (Qiagen, Cat. No. 231122) using T4 DNA ligase (Promega, Cat. No. M1801) following manufacturer's instructions. Plasmids were transformed and blue white screening conducted, using 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) and isopropyl β-D-1-thiogalactopyranoside (IPTG) according to the manufacturer's instructions (VWR Peqlab, Cat. No. PEQL-37-2610 and PEQL-37-2010). DNA was extracted from white colonies according to a method by Berghammer and Auer (1993), and screened for correct inserts using *Eco* RI, a restriction site flanking the multiple cloning site (MCS) of pDrive. Plasmid DNA containing the correct inserts were purified using the QIAprep Spin Miniprep Kit (Qiagen, Cat. No. 27106), as per the manufacturer's instructions. Finally, plasmids were subjected to Sanger sequencing by the South African Institute for Aquatic Biodiversity using the Genetic Analyzer 3500 (Applied Biosystems™). Thereafter, the inserts were sub-cloned into the respective expression vectors.

Appendix B1:

A synthetic p11 encoding sequence was codon optimized for expression in *E. coli* (IDT codon optimization tool). The p11 sequence was further modified to express a C-terminal FLAG tag (DYKDDDDK). An N-terminal *Nde* I and C-terminal *Bam* HI site were included for downstream sub-cloning, and this was then generated through GenScript (pMDB3). This insert was then sub-cloned from the pUC57 vector, using *Nde* I and *Bam* HI, into the bacterial T7-7-6xHis-XbaI expression vector (pMDB4).

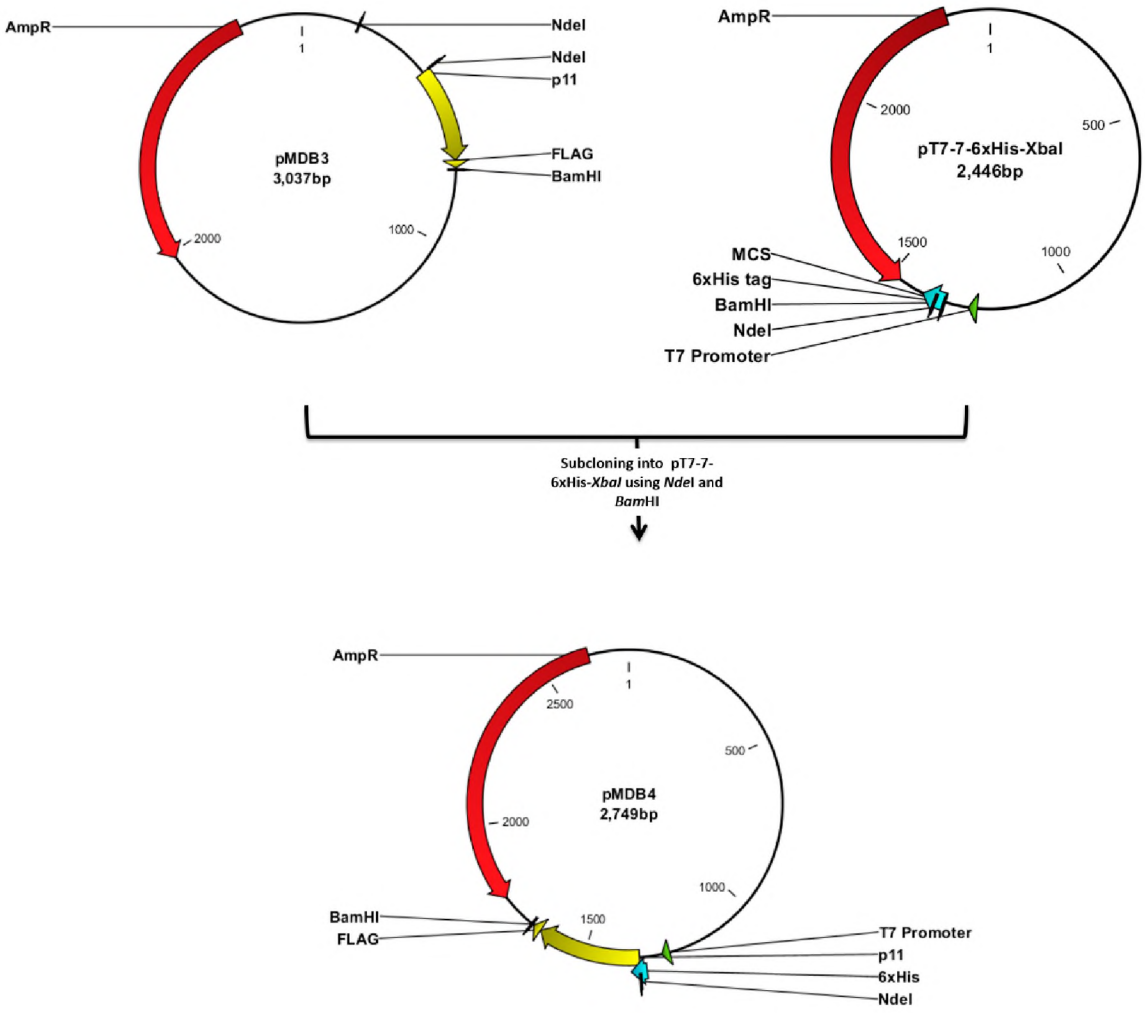

Figure B1: Schematic representation of the construction of pMDB4. Amp R: *β*-lactamase**;** 6xHis-hexa-histidine tag.

Appendix B2:

Using the primer set MDB7 and MDB8, p11 was amplified from pAV8F (encoding the full length RNA1 of HaSV). The MDB7 primer added an N-terminal FLAG tag to p11, downstream of the native p11 Kozak sequence. The FLAG-p11 expression cassette was then sub-cloned from the pDrive clone (pMDB7) into pJA1 by exploiting the *Bam* HI and *Hind* III sites included in the primers.

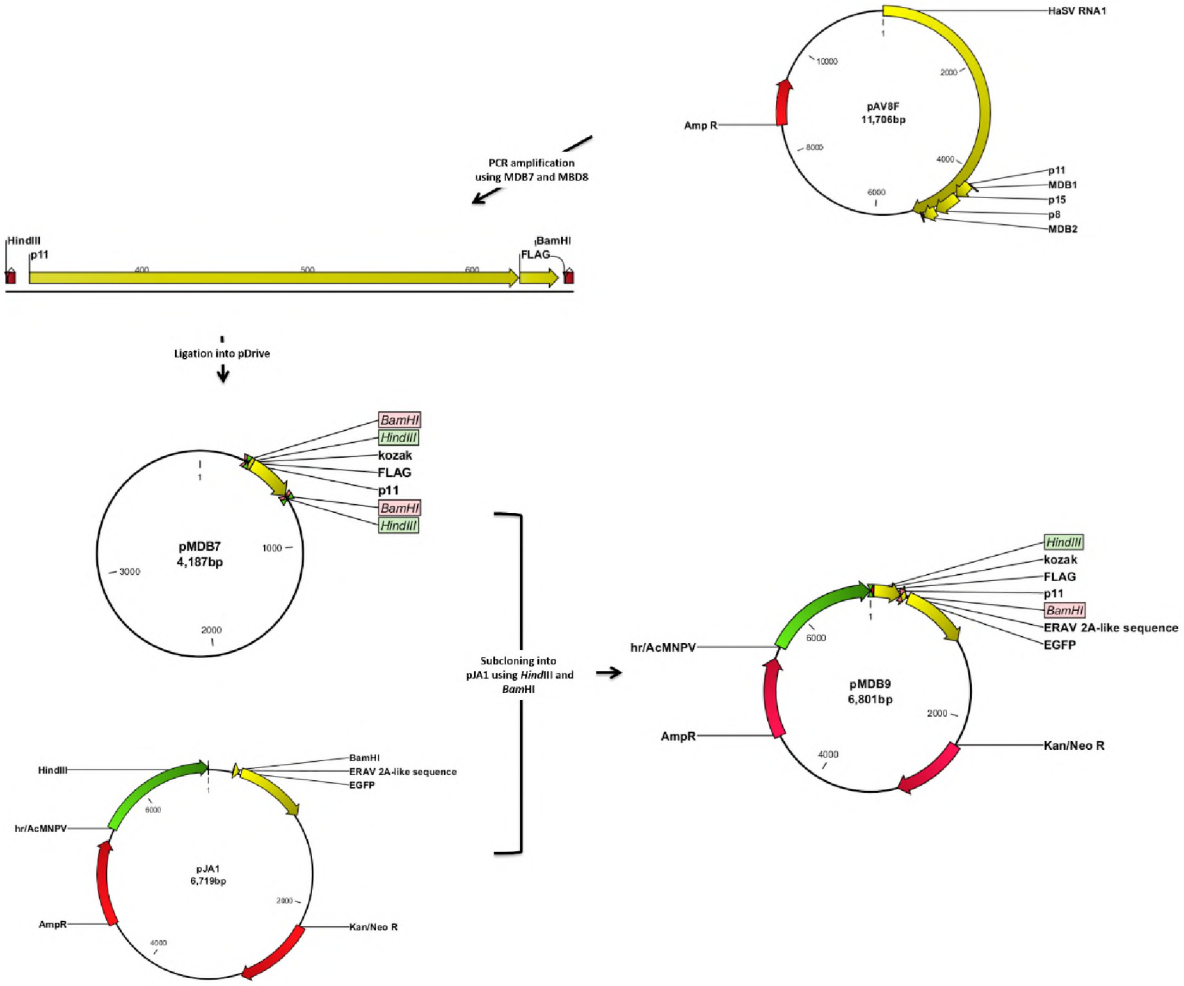

Figure B2: Schematic representation of the construction of pMDB9. Amp R: *β*-lactamase, EGFP: enhanced green fluorescent protein, hr/AcMNPV Promoter: hybrid Autographa californica multiple nuclear polyhedrovirus promoter, Kan/Neo R: neomycin phosphotransferase, ERAV 2A-like sequence: 2A-like processing site derived from Equine rhinitis A virus (ERAV).

Appendix B3:

Using the primer set MDB7 and GH2R, p11-p15 was amplified from pAV8F (encoding the full length RNA1 of HaSV). Prior to sub-cloning, the pGH2B vector was digested using *Hind* III, and treated with Shrimp Alkaline Phosphatase (SAP) as per the manufacturer's instructions (New England Biolabs, Cat. No. M0371S). This effectively eliminated the risk of vector self-ligation. The p11-p15 expression cassette was then sub-cloned from the pDrive clone (pMDB8) into the SAP-treated pGH2B using *Hind* III. To confirm ligation of the insert, PCR was performed using the MDB7-GH2R primer set.

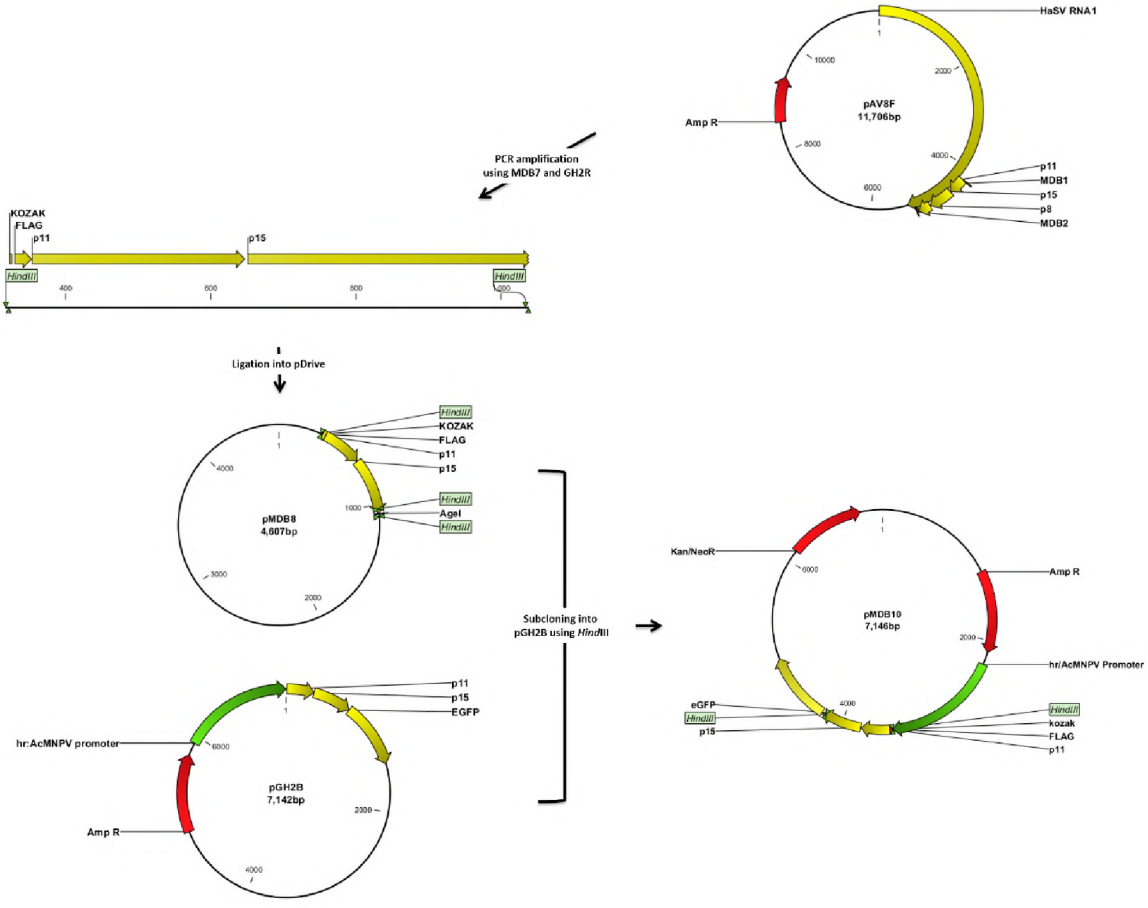

Figure B3: Schematic representation of the construction of pMDB10. Amp R: *β*-lactamase, EGFP: enhanced green fluorescent protein, hr/AcMNPV Promoter: hybrid Autographa californica multiple nuclear polyhedrovirus promoter, Kan/Neo R: neomycin phosphotransferase.

Appendix B4:

The pMDB10 expression vector was mutated so as to introduce a +1 frame shift of p15 with respect to p11. This was done by using the DNA Pol I, Large Fragment (Klenow). It exhibits a 3'-5' exonuclease activity (the 5'-3' exonuclease activity has been removed), and still maintains 5'-3' polymerase activity. The pMDB8 plasmid contains a single *Nco* I site within p11 (position 121-127 from the start of FLAG). This plasmid was then digested accordingly, producing a 5' overhang, and gel purified. Blunting of the overhangs was achieved by using the 5'-3' polymerase activity of Klenow (Promega, Cat. No. M220A), and the resulting linearized plasmid was then ligated and transformed as described previously (pMDB19). Sub-cloning was then conducted as per the creation of pMDB10, ultimately creating pMDB20.

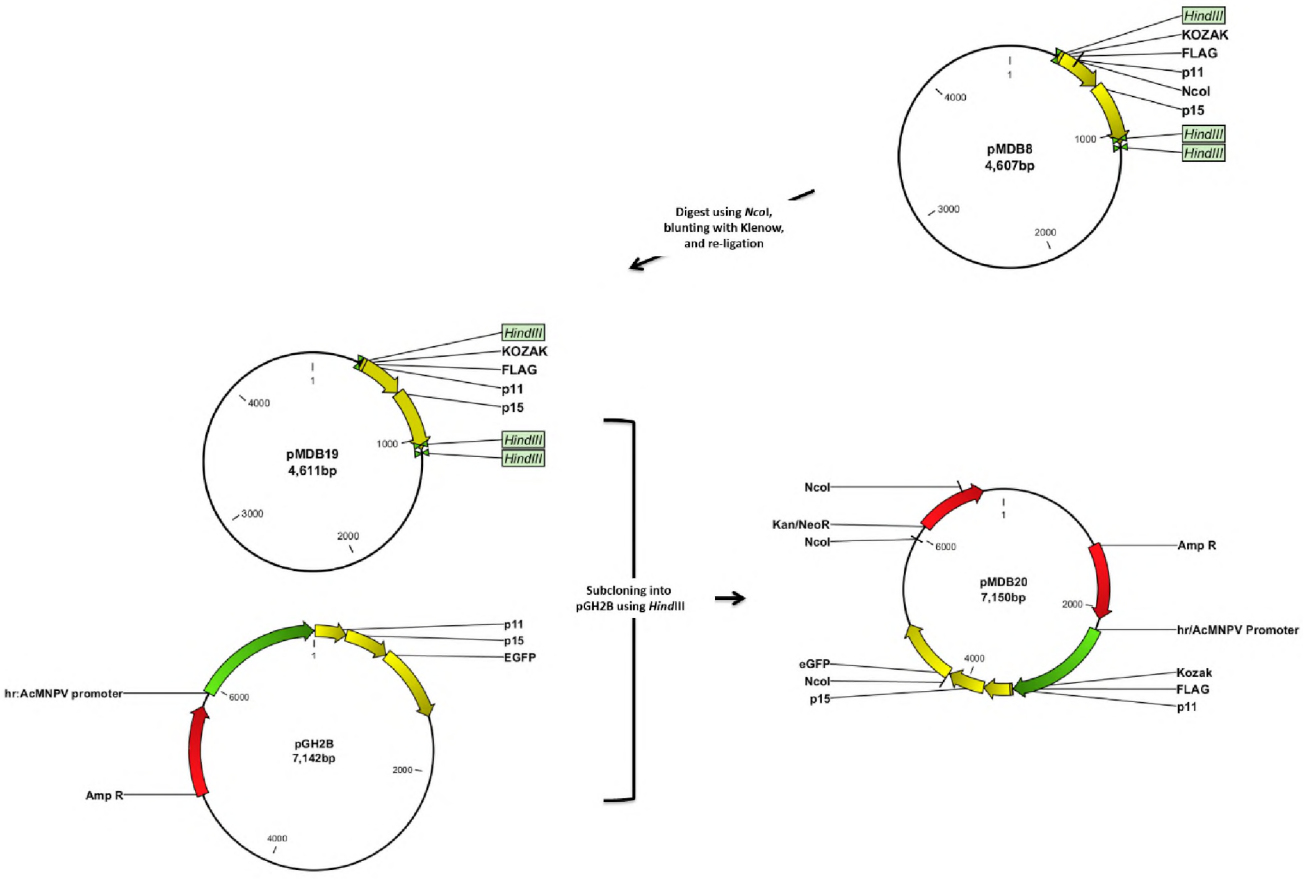

Figure B4: Schematic representation of the construction of pMDB20. Amp R: *β*-lactamase, EGFP: enhanced green fluorescent protein, hr/AcMNPV Promoter: hybrid Autographa californica multiple nuclear polyhedrovirus promoter, Kan/Neo R: neomycin phosphotransferase.

Appendix B5:

The primer set MDB1 and MDB2 was used to amplify p11-p15-p8 from pAV8F (encoding the full length RNA1 of HaSV). This insert was ultimately sub-cloned into a pFASTBac™ Dual plasmid by using the *Bam H*I and *Spe* I sites included in the respective primers.

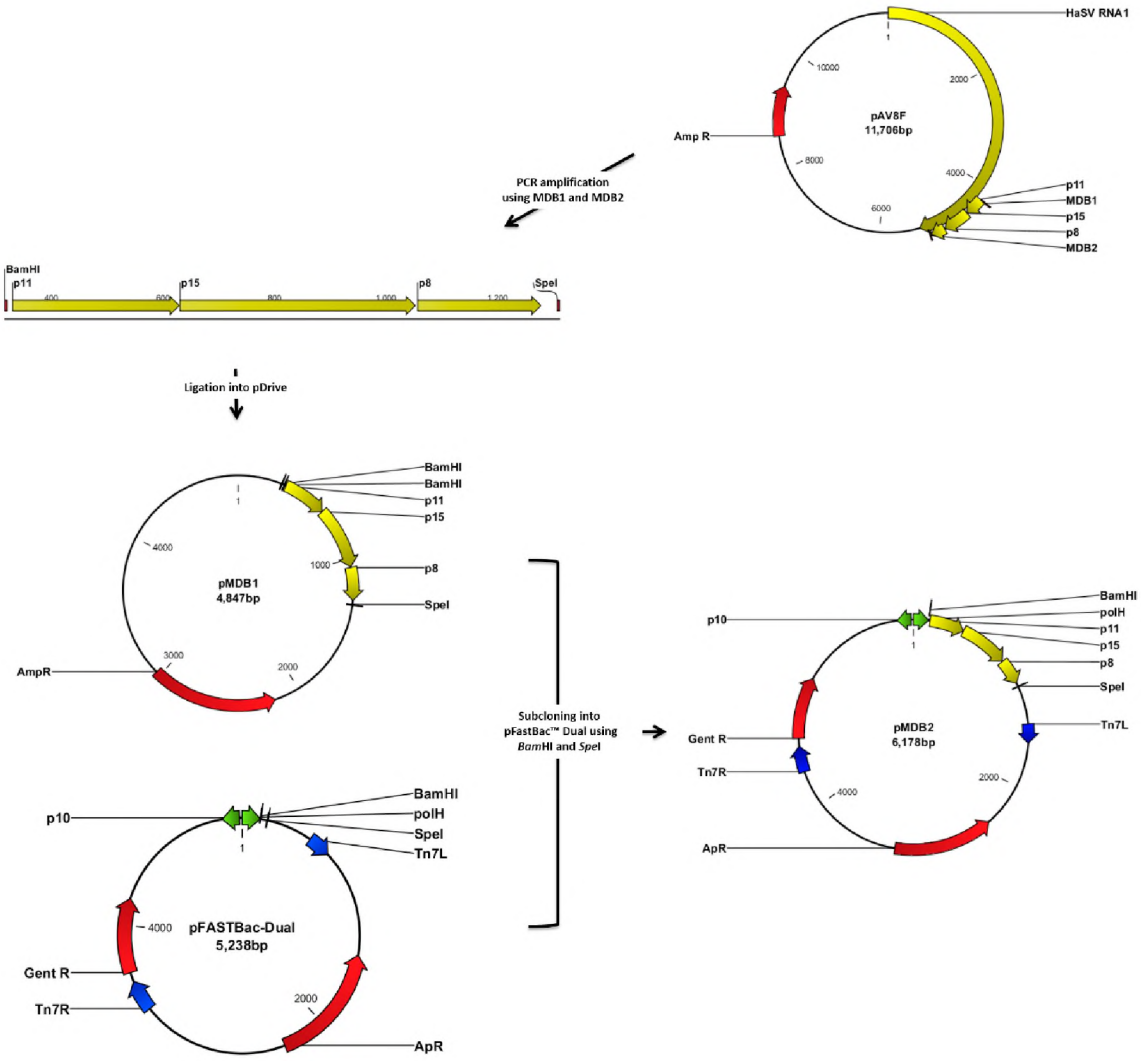

Figure B5: Schematic representation of the construction of pMDB2. Gent R- gentamycin resistance; AmpR- ampicillin resistance; Tn7R and Tn7L- ends of the transposon; p10 and polH- promoters; EGFP- enhanced green fluorescent protein.

Appendix B6:

Primers used for PCR

Table B1: List of primers used throughout this study.

Primer	Sequence (5'-3')
MDB1	GGATCCGTTGCGACATGTGCGAAG
MDB2	ACTAGTGGGACCACCTTGTTTACG
MDB7	AAGCTTCGACATGGACTACAAAGACGATGACGACAAGTGCGAAGCCGGAAAC
MDB8	GGATCCACGAGCGTGATTGCCTGTCTTGC
GH2R	ACCGGTTTGAGTAGCCTTTGGATGCGTCG
GH3R	ACCGGTACGAGCGTGATTGCCTGTCTTGC
GH10F	CCATGGGCGAAGCCGGAAACC
GH10R	AAGCTTTCAGACGAGCGTGATTGCCT
pUC/M13F	CCCAGTCACGACGTTGTAAAACG
pUC/M13R	AGCGGATAACAATTTCACACAGG

PCR cycling parameters

Table B2: Amplification of p11-p15-p8 from pAV8F using the MDB1-MDB2 primer set.

Stage	Temperature (˚C)	Time	Number of cycles
Initial denaturation	94	3 min	1
Denaturation	94	15 sec	20
Annealing	58	30 sec	
Extension	72	45 sec	
Final extension	72	5 min	
	4	∞	

Table B3: Amplification of p11 from pMDB2 BV DNA using the GH10F-GH3R primer set.

Stage	Temperature (˚C)	Time	Number of cycles
Initial denaturation	94	3 min	1
Denaturation	94	15 sec	30
Annealing	63.2	30 sec	
Extension	72	45 sec	
Final extension	72	10 min	
	4	∞	

Table B4: Amplification of p11 from pAV8F using the MDB7-MDB8 primer set.

Stage	Temperature (˚C)	Time	Number of cycles
Initial denaturation	94	3 min	1
Denaturation	94	15 sec	20
Annealing	64.4	30 sec	
Extension	72	45 sec	
Final extension	72	5 min	
	4	∞	

Table B5: Amplification of p11-p15 from pAV8F using the MDB7-GH2R primer set.

Stage	Temperature (˚C)	Time	Number of cycles
Initial denaturation	94	3 min	1
Denaturation	94	15 sec	20
Annealing	60.4	30 sec	
Extension	72	45 sec	
Final extension	72	5 min	
	4	∞	

Table B6: Amplification of insert from pMDB2 bacmid using the pUC/M13F-GH10R primer set.

Stage	Temperature (°C)	Time	Number of cycles
Initial denaturation	94	3 min	1
Denaturation	94	15 sec	25
Annealing	55	30 sec	
Extension	72	45 sec	
Final extension	72	10 min	
	4	∞	

Table B7: Amplification of insert from pMDB2 bacmid using the MDB1-pUC/M13R primer set.

Stage	Temperature (°C)	Time	Number of cycles
Initial denaturation	94	3 min	1
Denaturation	94	15 sec	25
Annealing	56	30 sec	
Extension	72	45 sec	
Final extension	72	5 min	
	4	∞	

Appendix C

The figure below shows the Hopp-Woods hydrophobicity plots of p8 of HaSV_RNA1, pMDB12 and pMDB18, as well as the predicted secondary structures therein.

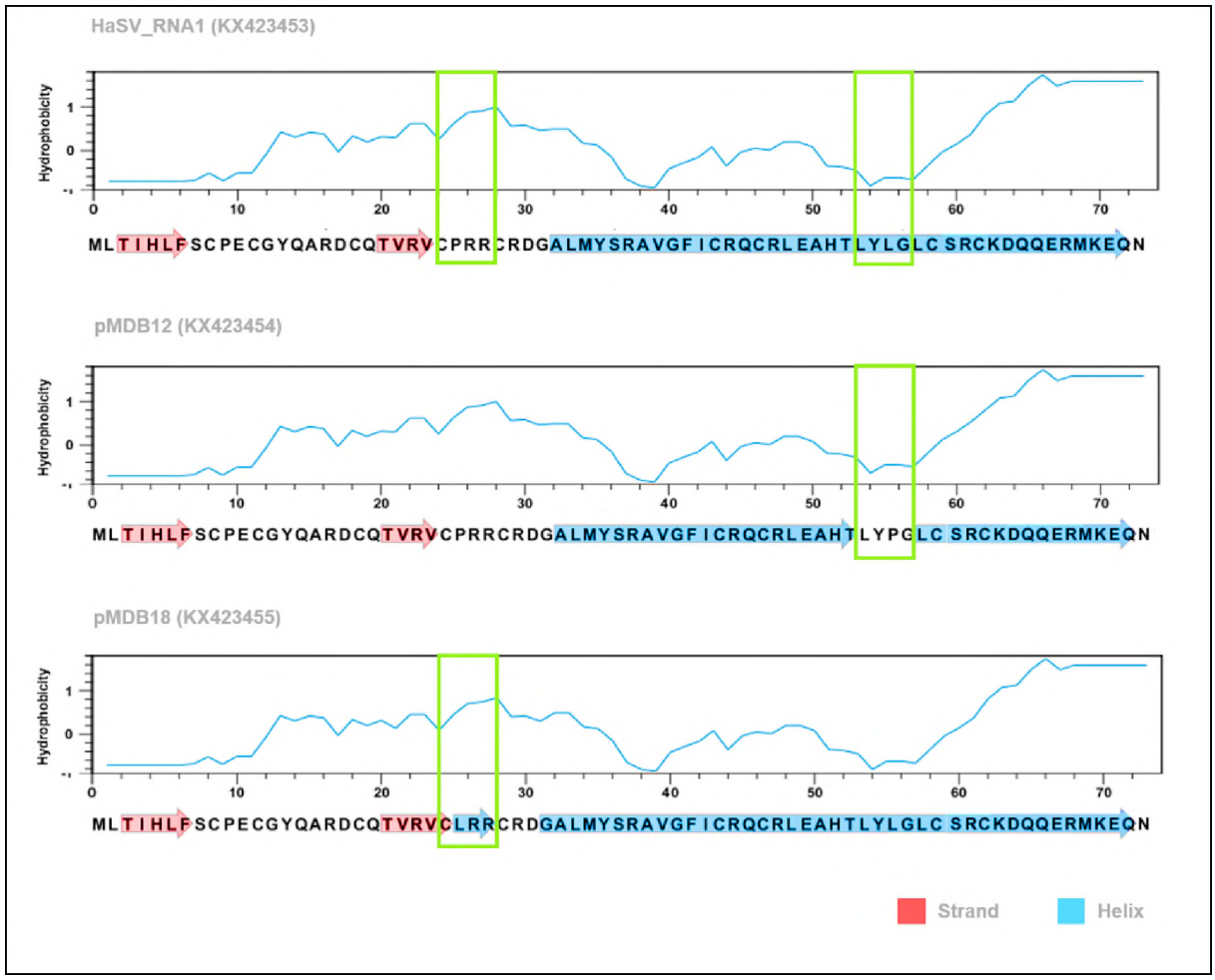

Figure C1: Predicted protein secondary structures for p8 of HaSV_RNA1, pMDB12 and pMDB18. Hopp-Woods plots show hydrophilic antigenicity scores as positive values, and hydrophobic scores as negative. The amino acid sequence is positioned on the x-axis, with predicted secondary structures depicted in red (strands) and blue (helices). Secondary protein structures were determined by CLC Main Workbench software. The region demarcated in green indicates the residues affected by the respective point mutations.

Appendix D

A dot blot was used to optimise detection of anti-p23 over a range of purified protein concentrations (0.01 µg, 0.1 µg and 1 µg) and dilutions of the 0 day (1:1000) and 27 day antiserum (1:50000, 1:10000 and 1:1000). After dotting, the membrane was treated as previously described in Chapter 3. A working concentration of 0.1 µg at an antibody dilution of 1:1000 was deemed optimal, and was subjected to SDS-PAGE and subsequent western blot analysis.

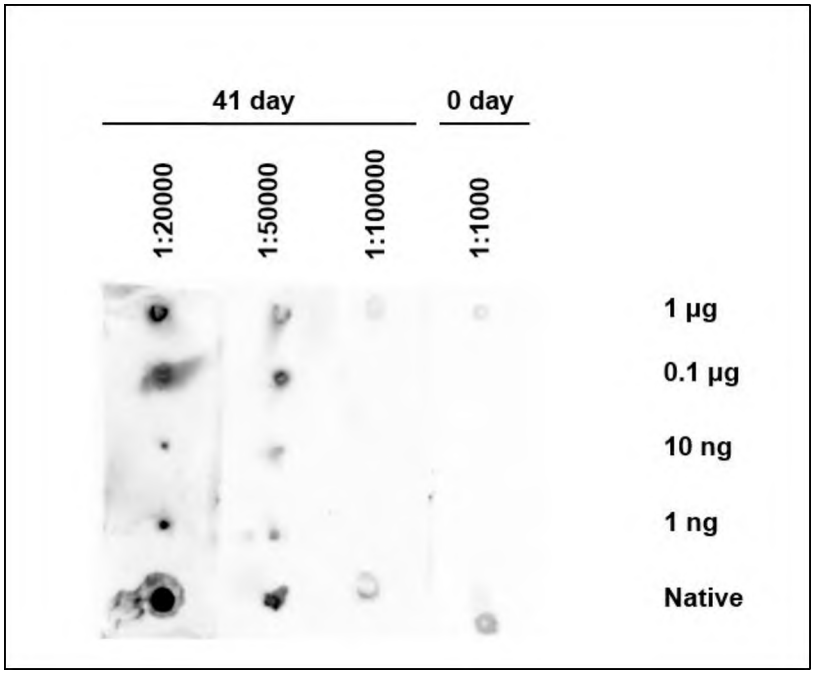

Figure D1: Preliminary optimization of anti-p23 against the p15-p8 fusion protein. Dot blot optimization testing different antibody dilutions (1:1000, 1:10000 and 1:50000) against varying protein concentrations (1 µg, 0.1 µg and 0.01 µg), and a native *E. coli* BL21 (DE3) cell-free lysate.

After probing with anti-p11, the membrane was stripped (0.1% SDS, 0.2 M glycine, pH 2.5) and re-probed using and anti-VCAP (1:10000) antibodies to verify infection.

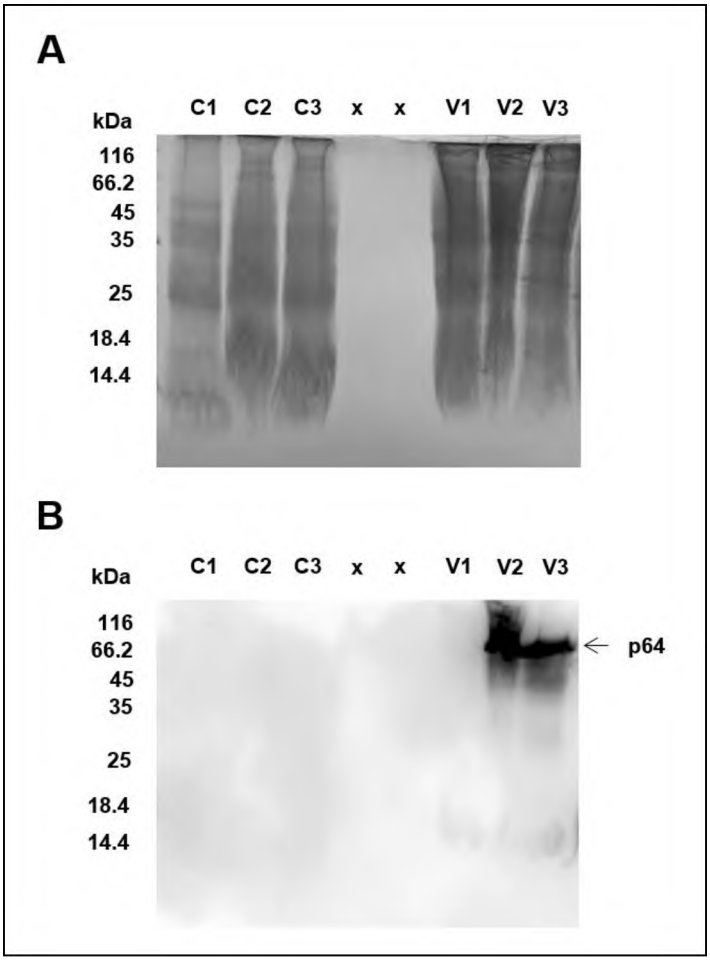

Figure E1: Detection of capsid in HaSV infected *H. armigera* larvae. SDS-PAGE **(A)** and western blot analysis **(B)** was performed on three uninfected larvae (C1-3) and three HaSV infected larvae (V1-3). Anti-VCAP antiserum was used at a 1:10000 dulution. The molecular weight standards are indicated on the left of the gels and arrow on the right of the image indicate the size of the band detected.

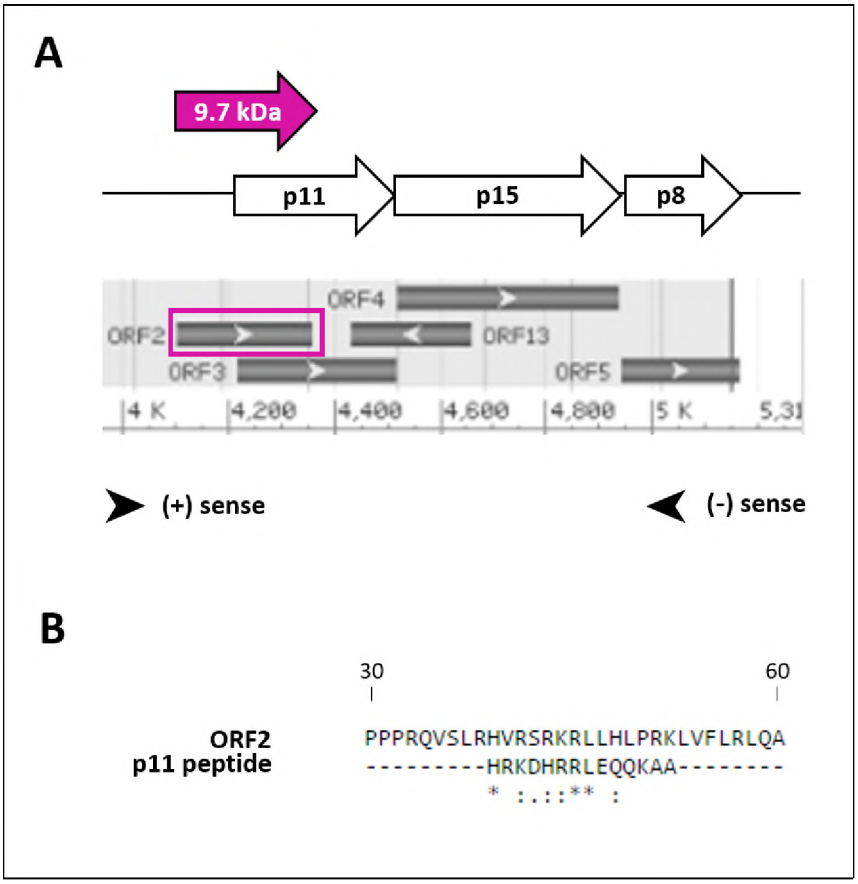

Figure F1: Prediction of an additional ORF on RNA1. (A) ORFs overlapping with the 3' end of HaSV RNA1, as predicted using the NCBI ORF Finder tool (Online- https://www.ncbi.nlm.nih.gov/orffinder). The additional ORF is indicated in purple (ORF2), with a predicted size of 9.7 kDa. The arrows indicate the sense of the ORFs; (+) – positive; (-) – negative, and nucleotide positions on RNA1 are indicated. **(B)** ClustalW alignment of the p11 peptide chosen for antiserum production, with ORF2. ***** - Conserved residue, **:** - residues exhibiting strong similarity, **.** – residues exhibiting weak similarity. Amino acid positions are indicated ontop of the figure.

The frame of p15 was shifted +1 with respect to p11 to make pMDB20. This mutated construct will express a 55 amino acid (FLAG-tagged) protein, with the first 42 amino acids of the native p11.

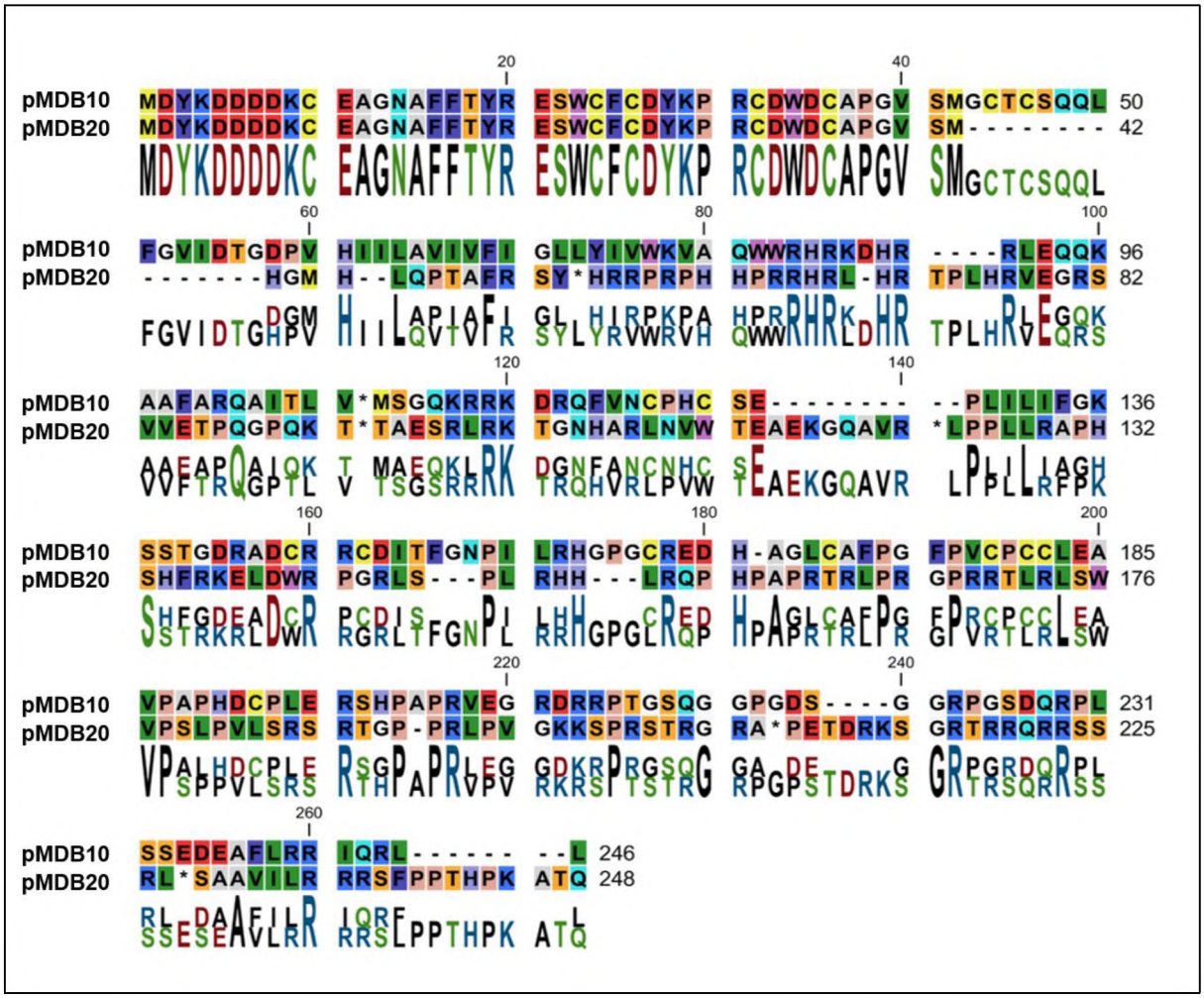

Figure G1: Alignment of the translation products of pMDB10 and pMDB20. The sequences of pMDB10 and pMDB20 were translated and aligned using CLC Main Workbench. * - stop codon.

References

Agrawal, D.K. and Johnson, J.E. (1995). Assembly of the T54 Nudaurelia capensis ω virus capsid protein, post-translational cleavage, and specific encapsidation of its mRNA in a baculovirus expression system, *Virology*, **207**:89–97.

Allende, R., Lewis, T.L., Lu, Z., Rock, D.L., Kutish, G.F., Ali, A., Doster, A.R. and Osorio, F.A. (1999). North American and European porcine reproductive and respiratory syndrome viruses differ in non-structural protein coding regions, *Journal of General Virology*, **80**(2):307-315.

Atkins, J.F., Loughran, G., Bhatt, P.R., Firth, A.E. and Baranov, P.V. (2016). Ribosomal frameshifting and transcriptional slippage: From genetic steganography and cryptography to adventitious use, *Nucleic Acids Research*, p.gkw530.

Ball, L. A., and Johnson, K. L. (1999). Reverse genetics of Nodaviruses, *Advances in Virus Research*, **53**:229–244.

Barba-Spaeth, G., Dejnirattisai, W., Rouvinski, A., Vaney, M.C., Medits, I., Sharma, A., Simon-Lorière, E., Sakuntabhai, A., Cao-Lormeau, V.M., Haouz, A. and England, P. (2016). Structural basis of potent Zika–dengue virus antibody cross-neutralization, *Nature*, **536(7614)**:48-53.

Baum, M., and Beier, H. (1998). Wheat cytoplasmic arginine tRNA isoacceptor with a U*CG anticodon is an efficient UGA suppressor *in vitro, Nucleic Acids Research*, **26**:1390-1395.

Bawden, A.L., Gordon, K.H.J., and Hanzlik, T.N. (1999). The specificity of Helicoverpa armigera stunt virus infectivity, *Journal of Invertebrate Pathology*, **74**:156–163.

Beier, H., and Grimm, M. (2001). Misreading of termination codons in eukaryotes by natural nonsense suppressor tRNAs, *Nucleic Acids Research,* **29**:4767-4782.

Beinz, M., and Kubli, E. (1981). Wild-type $tRNA^{Tyr}_{G}$ reads the TMV RNA stop codon, but Q base-modified $tRNA^{Tyr}_{Q}$ does not, *Nature*, **294**:188-190.

Berry, M.J., Banu, L., Chen, Y., Mandel, S.J., Kieffer, J.D., Harney, J.W., and Larsen, P.R. (1991). Recognition of UGA as a selenocysteine codon in type I deiodinase requires sequences in the 3′ untranslated region, *Nature,* **353**:273-276.

Bilak, H., Tauszig-Delamasure, S., and Imler, J.L. (2003). Toll and Toll-like receptors in *Drosophila, Biochemical Society Transactions,* **31(3):**648-651.

Blakqori, G., van Knippenberg, I., and Elliot, R.M. (2009). Bunyamwera orthobunyavirus S-segment untranslated regions mediate poly(A) tail independent translation, *Journal of Virology*, **83**:3637-3646.

Brooks, E.M., Gordon, K.H.J., Dorrian, S.J., Hines, E.R., and Hanzlik, T.N. (2002). Infection of its Lepidopteran host by the Helicoverpa armigera stunt virus (*Tetraviridae*), *Journal of Invertebrate Pathology,* **80**:97–111.

Brown, R.S. (2005). Zinc finger proteins: getting a grip on RNA. *Current Opinion in Structural Biology*, **15**(1):94-98.

Brown, J.D., and Ryan, M.D. (2010). Ribosome ''skipping'': ''stop-carry on'' or ''StopGo'' translation. In: *Recoding: Expansion of decoding rules enriches gene expression*, pp. 101–121. Edited by Atkins, J.F., and Gesteland, R.F. Springer, Heidelberg.

Butcher, S. J., Grimes, J. M., Makeyev, E. V., Bamford, D. H. and Stuart, D. I. (2001). A mechanism for initiating RNA-dependent RNA polymerization, *Nature,* **410:**235–240.

Canady, M. A., Tsuruta, H. and Johnson, J. E. (2001). Analysis of rapid, large-scale protein quaternary structural changes: Time-resolved x-ray solution scattering of Nudaurelia capensis ω Virus (NωV) Maturation, *Journal of Molecular Biology,* **311**:803-814.

Canady, M.A., Tihova, M., Hanzlik, T.N., Johnson, J.E., and Yeager, M. (2000). Large conformational changes in the maturation of a simple RNA virus, Nudaurelia capensis ω virus (NωV), *Journal of Molecular Biology,* **299**:573–584.

Carrat, F. and Flahault, A. (2007). Influenza vaccine: the challenge of antigenic drift, *Vaccine*, **25(39):**6852-6862.

Cavender, D.R. (1987). Comparison of the consensus sequence flanking translational start sites in *Drosophila* and vertebrates, *Nucleic Acids Research,* **15**:1353–1361.

Cavender, D.R., and Ray, S.C. (1991). Eukaryotic start and stop translation sites, *Nucleic Acids Research,* **19**:3185-3192.

Chao, Y.C., Scott, H.A. and Young, S.Y. (1983). An icosahedral RNA virus of the soybean looper (*Pseudoplusia includens*), *Journal of General Virology,* **64**:1835-1838.

Chen, L.J., Su, Y.C. and Hong, J.R. (2009). Betanodavirus non-structural protein B1: A novel anti-necrotic death factor that modulates cell death in early replication cycle in fish cells. *Virology*, **385**(2), pp.444-454.

Choi, I.R., Ostrovsky, M., Zhang, G. and White, K.A. (2001). Regulatory activity of distal and core RNA elements in tombusvirus subgenomic mRNA2 transcription, *Journal of Biological Chemistry*, **276**(45):41761-41768.

Christian, P.D., Dorian, S.J., Gordon, K.H.J., and Hanzlik, T.N. (2001). Pathology and properties of the tetravirus Helicoverpa armigera stunt virus, *Biol. Control*, **20:**65–75.

Christian, P.D., Murray, D., Powell, R., Hopkinson, J., and Hanzlik T.N. (2005). Effective control of a field population of *Helicoverpa armigera* by using the small RNA virus Helicoverpa armigera stunt virus (*Tetraviridae*: *Omegatetravirus*), *Journal of Economic Entomology*, **98(6):**1839-1847.

Cole, C.N., and Stacy, T.P. (1985). Identification of sequences in the Herpes Simplex virus thymidine kinase gene required for efficient processing and polyadenylation, *Molecular and Cellular Biology*, **5:**2104-2113.

De Breyne, S., Monney, R.S. and Curran, J. (2004). Proteolytic Processing and Translation Initiation two independent mechanisms for the expression of the Sendai virus Y proteins, *Journal of Biological Chemistry*, **279(16)**:16571-16580.

Decroly, E., Ferron, F., Lescar, J., and Canard, B. (2012). Conventional and unconventional mechanisms for capping viral mRNA. *Nature Reviews Microbiology*, **10(1):**51-65.

Dinman, J.D. (2012). Mechanisms and implications of programmed translational Frameshifting, WIREs RNA. doi: 10.1002/wrna.1126.

Donnelly, M. L. L., Hughes, L. E., Luke, G., Mendoza, H., Ten Dam, E., Gani, D. and Ryan, M. D. (2001). The "cleavage" activities of Foot-and-mouth disease virus 2A site-directed mutants and naturally occurring "2A-like" sequences, *Journal of General Virology,* **82:**1027-1041.

Dorrington, R.A., and Short, J.R. (2010). Chapter 12: Tetraviruses, In: *Insect Virology*, pp. 283-305. Edited by Asgari, S., and Johnson, K.N., Caister Academic Press, Norfolk, UK.

Dorrington, R.A., Gorbalenya, A.E., Lauber, C., Gordon, K.H.J., and Ward, V.K. (2011). Tetraviridae, In: *Virus Taxonomy: Classification and Nomenclature of Viruses: Ninth Report of the International Committee on Taxonomy of Viruses*, pp. 1091-1102. Edited by King, A.M.Q., Adams, M.J., Carstens, E.B, and Lefkowitz, E.J., Elsevier, San Diego.

Du Plessis, L., Hendry, D.A., Dorrington, R.A., Hanzlik, T.N., Johnson, J.E. and Appel, M. (2005). Revised RNA2 sequence of the tetravirus, Nudaurelia capensis ω virus (NωV), *Archives of Virology.* **150:**2397–2402.

Dvorak, P., Chrast, L., Nikel, P.I., Fedr, R., Soucek, K., Sedlackova, M., Chaloupkova, R., Lorenzo, V., Prokop, Z. and Damborsky, J. (2015). Exacerbation of substrate toxicity by IPTG in Escherichia coli BL21 (DE3) carrying a synthetic metabolic pathway, *Microbial Cell Factories*, **14(1):**1.

Eggertsson, G., and Söll, D. (1988). Transfer ribonucleic acid-mediated suppression of termination codons in *Escherichia coli*, *Microbiological Reviews,* **52(3)**:354–374.

Fang, Y., Treffers, E.E., Li, Y., Tas, A., Sun, Z., Van Der Meer, Y., De Ru, A.H., Van Veelen, P.A., Atkins, J.F., Snijder, E.J. and Firth, A.E. (2012). Efficient− 2 frameshifting by mammalian ribosomes to synthesize an additional arterivirus protein, *Proceedings of the National Academy of Sciences*, **109(43)**:E2920-E2928.

Feng, Y.X., Copeland, T.D., Oroszlan, S., Rein, A., and Levin, J.G. (1990). Identification of amino acids inserted during suppression of UAA and UGA termination codons at the gag-pol junction of Moloney Murine Leukemia Virus, *Proceedings of the National Academy of Sciences of the United States of America*, **87(22)**:8860-8863.

Fenner, F.J., McAuslan, B.R. and Mims, C.A. (2013). *The Biology of Animal Viruses*, pp. 543-586, Elsevier.

Firth, A.E., and Brierley, I. (2012). Non-canonical translation in RNA viruses, *Journal of General Virology*, **93:**1385–1409.

Fütterer, J., Rothnie, H. M., Hohn, T. and Potrykus, I. (1997). Ricetungro bacilliform virus open reading frames II and III are translated from polycistronic pregenomic RNA by leaky scanning, *Journal of Virology*, **71**:7984–7989.

Gammon, D.B., and Mello, C.C. (2013). RNA interference-mediated antiviral defence in insects, *Current Opinion in Insect Science*, **8:**111-120.

Genty, P. (1976). Etude morphologique et biologique d'un lépidoptère défoliateur du palmier à huile en Amérique latine Darna metaleuca, Walker, *Oléagineux,* **31**:99–107.

Giedroc, D. P., and Cornish, P. V. (2009). Frameshifting RNA pseudoknots: structure and mechanism, *Virus Research,* **139**:193-208.

Gohara, D. W., Crotty, S., Arnold, J. J., Yoder, J. D., Andino, R. and Cameron, C. E. (2000). Poliovirus RNA-dependent RNA polymerase (3Dpol): structural, biochemical, and biological analysis of conserved structural motifs A and B, *Journal of Biological Chemistry*, **275:**25523–25532.

Gorbalenya, A.E., Pringle, F.M., Zeddam, J.-L, Luke, B.T., Cameron, C.T., Kalmakoff, J., Hanzlik, T.N., Gordon, K.H.J., and Ward, V.K. (2002). The palm subdomain-based active site is internally permuted in viral RNA-dependent RNA polymerases of an ancient lineage, *Journal of Molecular Biology*, **324**:47–62.

Gorbalenya, A.E., Enjuanes, L., Ziebuhr, J. and Snijder, E.J. (2006). Nidovirales: evolving the largest RNA virus genome. *Virus research*, **117**(1):17-37.

Gordon, K.H.J., Johnson, K.N., and Hanzlik, T.N. (1995). The larger genomic RNA of Helicoverpa armigera stunt tetravirus encodes the viral RNA polymerase and has a novel 3'-terminal tRNA-like structure, *Virology,* **208**:84-98.

Gordon, K. H. J., Williams, M. R., Hendry, D. A., and Hanzlik, T. N. (1999). Sequence of the genomic RNA of *nuduarelia* β Virus (*Tetraviridae*) defines a novel virus genome organisation, *Virology,* **258**:42–53.

Grace, T.D.C. and Mercer, E.H. (1965). A new virus of the *Saturniid Antheraea eucalypti* Scott, *Journal of Invertebrate Pathology*, **7**:41-244.

Greenwood, L.K. and Moore, N.F. (1984). Pathogenicity of a small RNAvirus of insects determined by a virulent virus rather than a susceptible host, *Microbiologica*, **7**:97-102.

Hanahan, D., Jessee, J., and Bloom, F.R. (1991). Plasmid transformation of *Escherichia coli* and other bacteria, *Methods in Enzymology,* **204:**63-113.

Hanzlik, T.N., Dorrian, S.J., Gordon, K.H.J. and Christian, P.J. (1993). A novel small RNA virus isolated form the cotton bollworm, *Helicoverpa armigera*, *Journal of General Virology,* **74**:1805-1810.

Hanzlik, T.N., Dorrian, S.J., Gordon, K.H.J., and Christian, P.D. (1993). A novel small RNA virus isolated from the Cotton Bollworm, *Helicoverpa armigera, Journal of General Virology,* **74**:1805-1810.

Hanzlik, T.N., Dorrian, S.J., Johnson, K.J., Brooks, E.M., and Gordon, K.H.J. (1995). Sequence of RNA2 of the Helicoverpa armigera Stunt Virus (*Tetraviridae)* and bacterial expression of its genes, *Journal of General Virology,* **76**:799–811.

Harrell, L., Melcher, U., and Atkins, F.J. (2002). Predominance of six different hexanucleotide recoding signals 39 of read-through stop codons, *Nucleic Acids Research,* **30**:2011–2017.

Hendry, D.A., Hodgson, V, Clark, R., and Newman, J. (1985). Small RNA viruses coinfecting the Pine Emperor Moth (*Nudaurelia cytherea capensis*), *Journal of General Virology,* **66**: 627-632.

Hirsch, D. (1971). Tryptophan transfer tRNA as the UGA suppressor, *Journal of Molecular Biology*, **58**:439-458.

Holm, L. (1986). Codon usage and gene expression, *Nucleic Acids Research*, **14(7)**:3075-9087.

Hughes, G.E. (2013). An investigation into the expression and subcellular localisation of three small open reading frames on Helicoverpa armigera stunt virus RNA1, M.Sc thesis, Rhodes University, South Africa.

Imataka, H., Gradi, A., and Sonenberg, N. (1998) A newly identified N-terminal amino acid sequence of human eIF4G binds poly(A)-binding protein and functions in poly(A)-dependent translation, *EMBO J.* **17:**7480–7489.

ICTV Taxonomy Release. (2011). Available online: http://www.ictvonline.org/virusTaxonomy.asp?msl_id=26.

Invitrogen. (2013). *Bac-to-Bac Baculovirus Expression System 10359 Revision A.0*, pp. 1-80, Life Technologies Corporation, California, USA.

Jacks,T. and Varmus,H.E. (1985) Expression of the Rous sarcoma virus *pol* gene by ribosomal frameshifting, *Science*, **230**, 1237–1242.

Jackson, R. J. (2005). Alternative mechanisms of initiating translation of mammalian mRNAs, *Biochemical Society Transactions*, **6**:1231-41.

Jackson, R.J., Hellen, C.U. and Pestova, T.V. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation, *Nature Reviews Molecular Cell Biology*, **11**(2):113-127.

Jiwaji, M., Short, J.R. and Dorrington, R.A. (2016). Expanding the host range of small insect RNA viruses: Providence virus (*Carmotetraviridae*) infects and replicates in a human tissue culture cell line, *Journal of General Virology*, **97(10)**:2763-2768.

Kieft, J. S. (2008). Viral IRES RNA structures and ribosome interactions, *Trends in Biochemical Sciences*, **33(6):**274-83.

Komar, A. A., and Hatzoglou, M. (2010). Cellular IRES-mediated translation: the war of ITAFs in pathophysiological states. *Cell Cycle* **10**:229-240.

Koonin, E. V. (1991). The phylogeny of RNA-dependent RNA polymerases of positive strand RNA viruses, *Journal of General Virology*, **72:**2197– 2206.

Kozak, M. (1986). Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes, *Cell,* **44**:283-292.

Kuchino, Y., Beier, H., Akita, N., and Nishimura, N. (1987). Natural UAG suppressor glutamine tRNA is elevated in mouse cells infected with Moloney Murine Leukemia Virus, *Proceedings of the National Academy of Sciences of the United States of America*, **84(9):**2668-2672.

Kurland, C.G. (1991). Codon bias and gene expression, *FEBS 09993*, **285(2)**:165-169.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4*, Nature*, **227***:680-685.*

Lauring, A. S. and Andino, R. (2010). Quasispecies theory and the behaviour of RNA viruses, *PLoS Pathogens*, **6(7)**:e1001005.

Lane, H.C., Shelhamer, J.H., Mostowski, H.S. and Fauci, A.S. (1982). Human monoclonal anti-keyhole limpet hemocyanin antibody-secreting hybridoma produced from peripheral blood B lymphocytes of a keyhole limpet hemocyanin-immune individual, *The Journal of Experimental Medicine*, **155(1):**333-338.

Li, G.P., and Rice, C.M. (1989). Mutagenesis of the in-frame opal termination codon preceding nsP4 of Sindbus Virus: Studies of translational read-through and its effect on virus replication, *Journal of Virology,* **63**:1326-1337.

Luckow, V.A., Lee, C.S., Barry, G.F., and Olins, P.O. (1993). Efficient generation of infectious recombinant baculoviruses by site-specific transposon-mediated insertion of foreign genes into a baculovirus genome propagated in *Escherichia coli*, *Journal of Virology*, **67**:4566-4579

Luke, G.A. (2012). *Translating 2A research into practice*. INTECH Open Access Publisher.

Luke, G.A., de Felipe, P., Lukashev, A., Kallioinen, S.E., Bruno, E.A., and Ryan, M.D. (2008). Occurrence, function and evolutionary origins of '2A-like' sequences in virus genomes, *Journal of General Virology*, **89:**1036–1042.

Mans, R.M., Pleij, C.W., and Bosch, L. (1991). tRNA-like structures, *European Journal of Biochemistry,* **201(2)**:303-324.

Mendes, A., Vlok, M., Short, J.R., Motsui, T., and Dorrington, R.A. (2015). An encapsidated viral protein and its role in RNA packaging by a non-enveloped animal RNA virus, *Virology,* **476**:323–333.

Minskaia, E. and Ryan, M.D. (2013). Protein coexpression using FMDV 2A: effect of "linker" residues, *BioMed research international*, *2013*.

Montigny, C., Penin, F., Lethias, C., Falson, P. (2004). Overcoming the toxicity of membrane peptide expression in bacteria by upstream insertion of Asp-Pro sequence, *Biochimica et Biophysica Acta,* **1660**:53–65.

Moore, N.F., Reavy, B., and King, L.A. (1985). General characteristics, gene organization and expression of small RNA viruses of insects, *Journal of General Virology,* **66**:647–559.

Moore, D. (2001). Chapter 6: Population regulation of palm pests, In: *Insects on palms*, pp. 305-307, CABI Publishing, London, UK.

Morris, T.J., Hess, R.T and Pinnock, D.E. (1979). Physiochemical characterization of a small RNA virus associated with baculovirus infection in *Trichoplusia ni*, *Intervirology,* **11**: 238-247.

Munshi, S., Liljas, L., Cavarelli, J., Bomu, W., McKinney, B., Reddy, V. and Johnson, J. E. (1996). The 2.8 Å structure of a T = 4 animal virus and its implications for membrane translocation of RNA. *Journal of Molecular Biology,* **261:**1-10.

Murgola, E.J. (1985). tRNA, Suppression, and the code, *Annual Review of Genetics*, **19(1)**:57-80.

Nakagawa, S., Niimura, Y., Gojobori, T., Tanaka, H., and Miura, K. (2008). Diversity of preferred nucleotide sequences around the translation initiation codon in eukaryote genomes, *Nucleic Acids Research,* **36**:861-871.

Olson, N. H., Baker, T. S., Johnson, J. E. and Hendry, D. A. (1990). The three-dimensional structure of frozen-hydrated *Nudaurelia capensis* beta virus, a *T* = 4 insect virus. *Journal of Structural Biology,* **105:**111-122.

Penkler, D.L., Jiwaji, M., Domitrovic, T., Short, J.R., Johnson, J.E. and Dorrington, R.A. (2016). Binding and entry of a non-enveloped *T* = 4 insect RNA virus is triggered by alkaline pH, *Virology*, **498**:277-287.

Picard-Jean, F., Tremblay-Létourneau, M., Serra, E., Dimech, C., Schulz, H., Anselin, M., Dutilly, V., and Bisaillon, M. (2013). RNA 5′-end Maturation: A Crucial Step in the Replication of Viral Genomes, *Current Issues in Molecular Virology - Viral Genetics and Biotechnological Applications*, Prof. Victor Romanowski (Ed.), InTech, DOI: 10.5772/56166.

Plant, E.P. (2012). *Ribosomal frameshift signals in viral genomes*. INTECH Open Access Publisher.

Poch, O., Sauvaget, I., Delarue, M. and Tordo, N. (1989). Identification of four conserved motifs among the RNA-dependent polymerase encoding elements, *EMBO Journal,* **8:**3867–3874.

Pringle, F.M., Johnson, K.N., Goodman, C.L., McIntosh, A.H. and Ball, L.A. (2003). Providence virus: a new member of the *Tetraviridae* that infects cultured insect cells, *Virology,* **306**:359-370.

Qi, N., Cai, D., Qiu, Y., Xie, J., Wang, Z., Si, J., Zhang, J., Zhou, X., and Hu, Y. (2011). RNA Binding by a novel helical fold of B2 protein from Wuhan Nodavirus mediates the suppression of RNA interference and promotes B2 dimerization, *Journal of Virology*, **85(18)**:9543–9554.

Racine, T., and Duncan, R. (2010). Facilitated leaky scanning and atypical ribosome shunting direct downstream translation initiation on the tricistronic S1mRNA of avian reovirus, *Nucleic Acids Research*, **38**:7260–7272.

Reinganum, C., Robertson, J.S. and Tinsley, T.W. (1978). A new group of RNA viruses from insects, *Journal of General Virology,* **40**:195-202.

Robichon, C., Luo, J., Causey, T.B., Benner, J.S., and Samuelson, J.C. (2011). Engineering *Escherichia coli* BL21 (DE3) derivative strains to minimize *E. coli* protein contamination after purification by immobilized metal affinity chromatography, *Applied and Environmental Microbiology*, **77**:4634–4646.

Rogers S, Wells R, Rechsteiner M (1986). Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis, *Science*, **234**(4774): 364–8.

Rost, B. (2001). Review: protein secondary structure prediction continues to rise, *Journal of structural biology*, **134(2):**204-218.

Ryabova, L.A., Pooggin, M.M., and Hohn, T. (2006). Translation reinitiation and leaky scanning in plant viruses*, Virus Research*, **119**:52-62.

Salinovich, O., and Montelaro. R.C. (1986). Reversible staining and peptide mapping of proteins transferred to nitrocellulose after separation by sodium dodecylsulfate-polyacrylamide gel electrophoresis, *Analytical biochemistry,* **156(2):**341-347.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). *Molecular cloning: A laboratory manual, 2nd edition,* Cold Spring Harbor Laboratory, New York.

Saïda, F., Uzan, M., Odaert, B., and Bontems, F. (2006). Expression of highly toxic genes in *E. coli*: Special strategies and genetic tools, *Current Protein and Peptide Science,* **7(1):**47-56.

Schneemann, A., Reddy, V., and Johson, J.E. (1998). The structural and function of Nodavirus particles: A paradigm for understanding chemical biology, *Advances in Virus Research,* **50**:381–446.

Schneider, R.J., and Mohr, I. (2003). Translation initiation and viral tricks, *TRENDS in Biochemical Sciences*, **28(3)**:130-136.

Short, J.R. (2010). An investigation into the replication biology of Helicoverpa armigera stunt virus, PhD thesis, Rhodes University, South Africa.

Short, J.R., Knox, C., and Dorrington, R.A. (2010). Subcellular localization and live-cell imaging of the Helicoverpa armigera stunt virus replicase in mammalian and *Spodoptera frugiperda* Cells*, Journal of General Virology*, **91**:1514–1523.

Short, J.R., and Dorrington, R.A. (2012). Membrane targeting of an alpha-like tetravirus replicase is directed by a region within the RNA-dependent RNA polymerase domain, *Journal of General Virology*, **93:**1706–1716.

Short, J.R., Nakayinga, R., Hughes, G.E., Walter, C.T., and Dorrington, R.A. (2013). Providence virus (family: *Carmotetraviridae*) replicates vRNA in association with the golgi apparatus and secretory vesicles, *Journal of General Virology*, **94**:1073–1078.

Skuzeski, J.M., Nichols, L.M., Gesteland, R.F., and Atkins, J.F. (1991). The signal for a leaky UAG stop codon in several plant viruses includes the two downstream codons, *Journal of Molecular Biology,* **218**:365-373.

Slack, J., and Arif, B.M. (2007). The baculoviruses occlusion-derived virus: Virion structure and function, *Advances in Virus Research*, **69**:99-165.

Speir, J. A. and Johnson, J. E. (2008). Tetraviruses. In: Encyclopaedia of Virology, pp 27-37, Edited by Mahy, B. W. J. and Van Regenmortel, M. H. V., Elsevier, USA.

Srinivasan, G., James, C.M., and Krzycki, J.A. (2002). Pyrrolysine encoded by UAG in Archaea: Charging of a UAG-decoding specialized tRNA, *Science*, **296(5572)**:1459-1462.

Studier, F.W. and Moffatt, B.A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes, *Journal of Molecular Biology,* **189:**113-130.

Studier, F.W., Rosenberg, A.H., Dunn, J.J. and Dubendorf, J.W. (1990). Use of T7 RNA polymerase to direct expression of cloned genes, *Methods in Enzymology,* **185:**60-85.

Swords, W.E. (2003). Chemical transformation of *E. coli*, *Methods in Molecular Biology*, **235**:49-53.

Takeuchi, O., and Akira, S. (2007). Recognition of viruses by innate immunity, *Immunological Reviews*, **220**:214-224.

Tirado, S.M.C. and Yoon, K.J. (2003). Antibody-dependent enhancement of virus infection and disease. *Viral immunology*, **16(1):**69-86.

Tomasicchio, M., Venter, P.A., Gordon, K.H.J., Hanzlik, T.N., and Dorrington, R.A. (2007). Induction of apoptosis in *Saccharomyces cerevisiae* results in the spontaneous maturation of tetravirus procapsids *in vivo*, *Journal of General Virology,* **88**:1576–1582.

Touriol, C., Bornes, S., Bonnal, S., Audigier, S., Prats, H., Prats, A. C. and Vagner, S. (2003). Generation of protein isoform diversity by alternative initiation of translation at non-AUG codons, *Biology of the Cell,* **95**:169–178.

Van Oers, M.M. Pijlman, G.P., Vlak, J.M. (2015). Thirty years of baculovirus-insect cell protein expression: from dark horse to mainstream technology, *Journal of General Virology*, **96**:6-23.

Valle, R.P., Morch, M.D., and Haenni, A.L. (1987). Novel amber suppressor tRNAs of mammalian origin, *The EMBO Journal*, **6(10)**:3049-3055.

Vaughn, J.L., Goodwin, R.H., Tompkins, G.J., and McCawly, P. (1977). The establishment of two cell lines from the insect *Spodoptera frugiperda* (Lepidoptera: Noctuidae), *In Vitro,* **13**:213-217.

Vaughn, J.L., Goodwin, R.H., Tompkins, G.J., and McCawly, P. (1977). The establishment of two cell lines from the insect *Spodoptera frugiperda* (Lepidoptera: Noctuidae), *In Vitro* **13**:213-217.

Walsh, D., Mathews, M.B., and Mohr, I. (2013). Tinkering with Translation: Protein Synthesis in Virus-Infected Cells, *Cold Spring Harbor Perspectives in Biology,* **5**:a012351.

Walter, C.W., Pringle, F.M., Nakayinga, R., de Felipe, P., Ryan, M.D., Ball, A.L., and Dorrington, R.A. (2010). Genome organization and translation products of Providence virus: Insight into a unique tetravirus, *Journal of General Virology,* **91**:2826–2835.

Westwood, J.A., Jones, I.M., and Bishop, D.H.L. (1993). Analyses of alternative poly(A) signals for use in baculovirus expression vectors, *Virology*, **195**:90-93

White, K.A., (2002). The premature termination model: a possible third mechanism for subgenomic mRNA transcription in (+)-strand RNA viruses, *Virology*, **304**(2):147-154.

Yi, F., Zhang, J., Yu, H., Liu, C., Wang, J. and Hu, Y. (2005). Isolation and identification of a new tetravirus from *Dendrolimus punctatus* larvae collected from Yunnan Province, China, *Journal of General Virology,* **86**:789–796.

Zeddam, J.-L., Gordon, K. H. J., Lauber, C., Felipe Alves, C. A., Luke, B. T., Hanzlik, T. N., Ward, V. K. and Gorbalenya, A. E. (2010). Euprosterna elaeasa virus genome sequence and evolution of the Tetraviridae family: emergence of bipartite genomes and conservation of the VPg signal with the dsRNA Birnaviridae family, *Virology*, **397:**145–154.