Establishing experimental systems for studying the replication biology of *Providence virus*

Thesis

Submitted in the fulfilment of the requirements for the degree of

Doctor of Philosophy
(Microbiology)

In the

FACULTY OF SCIENCE
RHODES UNIVERSITY

By

CHERYL TRACY WALTER

December 2008

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

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

When I first applied to complete a PhD at Rhodes University, the footnote at the bottom of the application went something like this: “a PhD is a serious undertaking and you should carefully consider this application before submitting it”. Well, they weren’t being melodramatic! One thing that anyone who has completed a PhD will tell you is that you cannot do this on your own. Besides completing the research and writing up process, I am here today because of the fantastic support from my friends and colleagues. When the chips were down and I didn’t know what to do, my friends were always there to support and listen to me and when I couldn’t do it all at work, had bad days or got stuck on something, my trusty lab mates were always there. Four years later and so much has happened since then. The “*Providence virus* project” started off with the cDNA sequence of the PrV genome, a little anti-capsid antiserum and a very nervous PhD student but has since provided a fascinating account on this particular tetravirus, and in general, virus phylogenetics and viral genome processing. Below, I hope to mention everyone who made this a reality and who have helped me to grow throughout this process.

My supervisor (since my BSc (Hons) degree, 2002), Prof. Rosie Dorrington for all of your support towards me over the past 7 years and the amazing possibilities and opportunities you made available to me – thank you. Dr. Pablo de Felipe and Prof. Martin Ryan, our collaborators from the University of St. Andrews: two amazing scientists and gentlemen. I hope we get to work together again. Fiona Pringle: you continue to inspire me and I know you are sorely missed in the research arena. Thank you for all your help and cheerful emails. Susan van Dyk and Morgan White: two of the most special people anyone could ask for as friends. I’m going to miss you so much. Prof. Greg Blatch and Dr. Caroline Knox: thank you for your calming words of advice and encouragement and as well serving as referees for me in countless applications. Jo-Anne de le Mare, for being such a wise and understanding friend through thick and thin. Mrs Val Hodgson and Jo-Anne for proof reading my thesis – your help was invaluable especially considering the time constraints involved, thank you. Mr James Short for optimising the immunofluorescence experiments involving the anti-p40 antiserum and the MG8 cells – you’re a star! To all past and present lab 417/Dorrington Research Group members, in particular; Dr. Arno Venter, Dr. Mez Jiwaji and Mrs. Sally Gardener - you will always be an intrinsic part of my academic self. I wish you al the happiness and success you will undoubtedly create for yourselves in the future. Finally, my sincere gratitude to the National Research Foundation for funding me so generously for 5 years – I would not have been able to study without your financial support.

ABSTRACT

Providence virus (PrV) is a member of the *Tetraviridae*, a family of small, positive sense, single-stranded RNA viruses, which characteristically infect the midgut tissue of heliothine larvae. PrV is the only known tetravirus that replicates in cultured insect cells. The virus comprises a monopartite genome resembling members of the genus *Betatetravirus* with the capsid precursor protein undergoing autoproteolytic cleavage at its C-terminus consistent with other tetravirus capsid precursor proteins.

Analysis of viral cDNA predicted the presence of three potential overlapping gene products (from 5` to 3`): (1) p130, a protein of unrecognized nucleotide or amino acid homology with a 2A-like processing site at its N-terminus; (2) p104, the replicase ORF, which was found to be phylogenetically related to tombus-and umbraviruses replicases. The presence of a read-through stop signal in the p104 ORF was proposed to produce and amino terminal product with a predicted MW of 40 kDa (p40) and (3) the capsid protein precursor (81 kDa) which has two 2A-like processing sites at its N-terminus. Metabolic radiolabelling of viral translation products in persistently infected MG8 cells and *in vitro* translation of the individual ORFs were performed in order to analyse the expression of PrV gene products. p130 was translated with no evidence of 2A-like processing. Two products of 40 kDa and 104 kDa were translated from the p104 ORF, indicating that the read-through stop signal was likely to be functional. Finally, the capsid protein precursor ORF produced a major translation product of 68 kDa corresponding to the capsid protein precursor as well a peptide of 15 kDa that was attributed to the activity of the second 2A-like site at the N-terminus of the p81 ORF.

The subcellular distribution of viral RNA (vRNA) and p40 in MG8 cells was investigated using immunofluorescence and biochemical fractionation. The results showed that p40/p104 and vRNA accumulated in polarized, punctate structures in some but not all MG8 cells and in some cases, co-localization was observed. This thesis concludes that PrV is a novel tetravirus with significant similarities plant carmo-like viruses that should be re-classified at the family level.

LIST OF ABBREVIATIONS

GENERAL ABBREVIATIONS

(-)	negative sense viral genome
(+)	positive sense viral genome
$3D^{POL}$	RNA-dependent RNA polymerase of picornaviruses
BHK	Baby Hamster Kidney cell line 21
BLAST	Basic Local Alignment Search Tool
bp	base pair
BrUTP	5` bromouridine 5` triphosphate sodium salt
BSA	bovine serum albumin
CPE	cytopathic effect
cDNA	complementary DNA
DAPI	4', 6-diamidino-2-phenylindole
DdDp	DNA-dependent DNA polymerase
DNA	deoxyribonucleic acid
ds	double-stranded
DTT	dithiothreiotol
EDTA	ethylenediaminetetraacetic acid
eGFP	enhanced green fluorescence protein
eIF	eukaryotic initiation factor
ER	endoplasmic reticulum
FB33	*Helicoverpa zea* fat body cell line 33
FCS	foetal calf serum
g	gravitational force
FITC	fluorescein isothiocyanate
GST	glutathione S transferase
HeLa	HeLa S3 immortalised cervical cancer cell line
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
hr	homologous region
IgG	immunoglobulin G
IRES	internal ribosome entry site
kb	kilobase
kcal/mol	kilocalorie per mole
kDa	kiloDalton
LTR	long terminal repeat
MG8	*Helicoverpa zea* midgut cell line 8
mg/ml	milligram per millilitre
ml	millilitre
MOTC	microtubule-organizing centre
n	number of specimens in sample
NCBI	National Centre for Biotechnology Information
NLS	nuclear localization signal
nm	nanometre
nt	nucleotide
$OD_{600\ nm}$	optical density at 600 nm
ORF	open reading frame

PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PMSF	phenylmethylsulphonyl fluoride
RBS	ribosome binding site
RdRp	RNA-dependent RNA polymerase
RNA	ribonucleic acid
RNAi	RNA interference
rRNA	ribosomal RNA
RT	reverse transcriptase
S	Svedberg unit
Sarkosyl	N-sodium lauryl sarcosinate
SARS	severe acute respiratory syndrome
SD	Shine-Delgarno sequence
SDS	sodium dodecyl sulphide
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
*Se*1	*Spodoptera exigua* embryonic cell line 1
*Sf*9	*Spodoptera frugiperda* ovarian cell line 9
si	small interfering
ss	single-stranded
TBS	TBS/Tween- 20
TBS	tris-buffered saline
TDPPs	template-dependent polynucleotide polymerases
Tris	tris-2-amino-2-(hydroxymethyl)-1,3-propandiol
tRNA	transfer RNA
URI	uniform resource identifier
UTR	untranslated region
v/v	volume per volume
VDAC	voltage-dependent anion channel
VLPs	virus-like particles
VPg	genome-linked viral protein
vRNA	viral RNA
w/v	weight per volume
µg	microgram
µg/ml	microgram per millilitre
µl	microlitre

VIRUS NAME ABBREVIATIONS

AcMNPV	*Autographa californica multiple nuclear polyhedrosis virus*
AeV	*Antheraea eucalypti virus*
BMV	*Brome mosaic virus*
CIRV	*Carnation Italian ringspot virus*
CMV	*Cytomegalovirus*
CNV	*Cucumber necrosis virus*
CrPV	*Cricket Paralysis virus*
CSVF	*Cholera swine fever virus*
CymRSV	*Cymbidium ringspot*
DpTV	*Dendrolimus punctatus tetravirus*

DpV	*Dasyichira pudibunda virus*
DtV	*Darma trima virus*
ECMV	*Encephalomyocarditis virus*
EeV	*Euprosterna elaeasa virus*
FHV	*Flock House virus*
FMDV	*Foot-and-mouth disease virus*
HaSV	*Helicoverpa armigera stunt virus*
HCV	*Hepatitis C virus*
HSV	*Herpes Simplex virus*
IBDV	*Infectious bursal disease virus*
IFV	*Infectious flacherie virus*
IPNV	*Infectious pancreatic necrosis virus*
Nζ V	*Nudaurelia* ζ virus
NψV	*Nudaurelia* ψ virus
NβV	*Nudaurelia capensis β virus*
NωV	*Nudaurelia capensis* ω *virus*
PCRPV	*Pelargonium chlorotic ring pattern virus*
PenuNPV	*P. nuda* nucleopolyhedrovirus
PiV	*Pseudoplusia includens virus*
PMV	*Panicum mosaic virus*
PnPV	*P. nuda* picorna-like virus
PrV	*Providence virus*
PSIV	*Plautia stali intestine virus*
PV	*Poliovirus*
PxV	*Philosamia cynthia x ricini virus*
SBV	*Sacbrood virus*
SV40	*Simian vacuolating virus* 40
TaV	*Thosea asigna virus*
TBSV	*Tomato bushy stunt virus*
TnV	*Trichoplusia ni virus*
TYMV	*Turnip yellow mosaic virus*

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
Literature Survey

1.1. INTRODUCTION

It is surprising how little is known about most of the classified viruses that infect insects, when one considers the fact that the total number of insect species is likely greater than the rest of the classes of eukaryotes combined (Rossmann and Tao, 1999). Insects comprise over 80 % of the existing species on the planet and viruses infect most orders of insects including both beneficial and pest species. The first noted case of insect viral disease can be traced back as far as the 16th and 17th centuries where scientific records exist describing the “wilting” disease of silkworms, indicating that insect viruses have been of interest for the last four centuries (Fields *et al*., 1996).

A number of virus families are known to infect insects. These include both DNA and RNA viruses as well as retroviruses (Fauquet *et al*., 2005). The latest phylogenetic classification of these virus families and examples of their hosts are summarized in Table 1.1. Interestingly, only a small number of the families listed have a host range limited only to insects. These include the *Ascoviridae*, *Baculoviridae*, *Tetraviridae* and the *Polydnaviridae* (Fauquet *et al*., 2005). Furthermore, only the *Tetraviridae* have a limited host range consisting of a single insect order, whereas the *Polydnaviridae* and *Baculoviridae* infect a broad range of insect host families.

Table 1.1. List of virus families known to infect insects

Genome	Family	Examples of hosts
dsDNA	*Poxviridae, Iridoviridae, Baculoviridae, Polydnaviridae* (*Ichnoviridae* and *Bracoviridae*) and the *Ascoviridae*	Red Hairy caterpillar, *Spodoptera* spp. and the Codling moth.
dsRNA	*Reoviridae* and *Birnaviridae*	Silk worms, *Spodoptera* spp. and *Drosophila* spp.
ssDNA	*Circoviridae* and *Parvoviridae*	Silk worms and yellow fever mosquito
(-)ssRNA	*Rhabdoviridae* and *Bunyaviridae*	*Drosophila* spp. and mosquitoes
(+)ssRNA	*Togoviridae, Flaviviridae, Picornaviridae, Nodaviridae, Tetraviridae* and *Dicistroviridae*	Pine Emperor moth, Cotton bollworm, various cricket species, Gypsy moth and *Drosophila* spp.
RT-RNA	*Metaviridae*	*Drosophila* spp.

Key: ss = single stranded; ds = double stranded; (-) = negative sense; (+) = positive sense; RT = reverse transcription. (Fauquet *et al*., 2005)

Relatively little research has focused on the fundamental analysis of small, unenveloped insect RNA viruses; although much work has been focused on viruses with potential medical or agricultural applications. A classic example of an agriculturally important family of viruses is the *Baculoviridae*, a family of large dsDNA viruses. The type member, *Autographa californica multiple nucleopolyhedrosis virus* (AcMNPV) and its natural host, the alfalfa looper, *Autographa californica* (Fauquet *et al.,* 2005) have been extensively studied with regards to their agricultural impact, and in the last three decades, significant advances in eukaryotic vector research (O' Reilly *et al.* 1992; Possee, 1997 and Kost *et al.,* 2005).

The *Tetraviridae* are a family of icosahedral, non-occluded viruses that have a host range limited to a single order of insects, the Lepidoptera. They are structurally closely related to the *Nodaviridae*, a family of small RNA viruses known to infect both animals and insects (Johnson *et al*., 1994). In terms of genome processing strategies, tetraviruses also display similarities to the recently taxonomically partitioned picorna-like family, *Dicistroviridae* and more distantly to the *Picornaviridae*, a diverse family that infects animals and perhaps also insects (Johnson *et al*., 1994; Mayo, 2002; Fauquet *et al*., 2005). The survey serves to describe the *Tetraviridae* and provides information on these viruses which served as motivation for this body of work. Thereafter, this survey discusses two

other (+)ssRNA virus families and the important similarities that exist between them and the *Tetraviridae*.

1.2. THE *TETRAVIRIDAE*

The first members of the *Tetraviridae* to be discovered were found some forty years ago in South Africa and infected the larvae of the Pine Emperor moth, *Nudaurelia cytherea capensis* (Hendry *et al*., 1968). In 1974, five small viruses were purified from moribund *N. capensis* larvae (Juckes, 1974). The viruses were named *Nudaurelia* α, β, γ, δ, and ε *viruses* in order of increasing prevalence. Structural analysis on *Nudaurelia capensis β virus* showed that this virus exhibited a unique *T*=4 icosahedral symmetry, making it the first official member of a new family of viruses, the *Tetraviridae* (Olson *et al*., 1990). More than 10 years after the initial characterization of NβV, another dominant virus was isolated from *N. capensis* larvae, the *Nudaurelia* ω *virus* (NωV), which had biophysical properties similar to those of NβV but differed by the presence of a second genomic RNA strand (Hendry *et al*., 1985).

NωV virus became the type virus for a second genus within the *Tetraviridae*, the ω-like genus, and was joined by a second similar virus from Australia, *Helicoverpa armigera stunt virus* (HaSV) by Hanzlik *et al*., in 1993. HaSV was isolated from a laboratory-bred colony of the cotton bollworm, *Helicoverpa armigera* (Hanzlik *et al*., 1993). In 2004, a third omegatetravirus was discovered infecting larvae of *Dendrolimus punctatus*, a member of the *Lasiocampidae* family (Yi *et al*., 2005) and was thus named *Dendrolimus punctatus tetravirus* (DpTV).

These and other potential members of the *Tetraviridae*, with the distinctive *T*=4 symmetry, containing either a monopartite or bipartite genome, were separated into two genera: the beta-like now known as the genus *Betatetravirus* and omega-like now known as the genus *Omegatetravirus* (Hanzlik and Gordon, 1997). Viruses in the family *Tetraviridae* are (+)ssRNA, unenveloped particles with an average diameter of 40 nm and

display an icosahedral geometric symmetry (Hendry and Agrawal, 1994; Fauquet *et al*., 2005).

1.2.1. Tetravirus capsid structure

The first structural information on tetravirus particles was obtained by cryo-electron microscopic analysis of NβV and NωV particles (Olson *et al*., 1990). At 3.2 nm resolution, each face of the virus particle has distinct structures at points of symmetry protruding approximately 4 nm above the contiguous surface (Olson *et al*., 1990). However, the NβV face is distinguishable from NωV as it is more planar with deeper grooves between the faces and has three prominent pits on each triangular face (Johnson *et al*., 1994) (Figure 1.1).

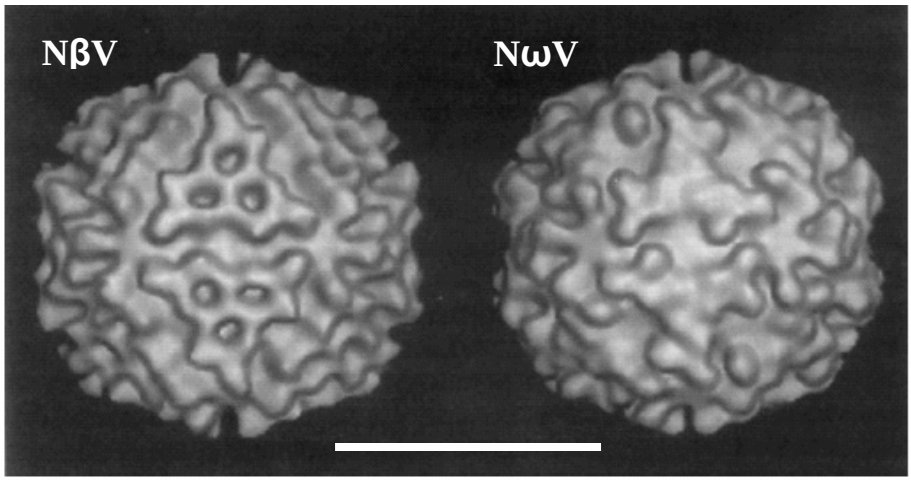

Figure 1.1. Image reconstruction of frozen-hydrated *Nudaurelia* viruses by cryo-electron microscopy. The bar represents 25 nm. Taken from Johnson *et al*., 1994.

Tetravirus capsids consist of multiple domains created by 240 protein subunits each adapting one of four slightly different conformations to form an icosahedral particle (Munshi *et al*., 1996). The classical, highly conserved eight-stranded antiparallel β-barrel motif that forms the contiguous shell of all icosahedral virus particles is also found in the tetraviruses (Olson *et al*., 1990; Johnson *et al*., 1994; Munshi *et al*., 1996; Helgstrand *et al*., 2004). Interestingly, an immunoglobulin C-like fold is

positioned on the outer capsid surface, inserted between two strands of this shell domain (Munshi *et al*., 1996). A third structural domain, the internal helical domain, is situated below the surface of the capsid shell and is composed of a pentameric bundle of coiled helices predicted to play a role in the membrane translocation of viral RNA (vRNA) (Munshi *et al*., 1996; Helgstrand *et al*., 2004).

During the assembly process, the provirion structure is characterized as being larger, rounder and more porous than the mature virion although relatively stable at a higher pH, and each capsid precursor is cleaved to form an infectious particle (Canady *et al*., 2000). The cleavage mechanism and resultant mature particle containing both β and γ proteins will be described later, with reference to particular examples of tetraviruses.

1.2.2. Phylogenetic organization of the *Tetraviridae*

Currently, the complete genome sequences of three betatetraviruses, namely NβV, *Thosea asigna virus* (TaV), *Euprosterna elaeasa virus* (EeV); and two omegatetraviruses, HaSV and DpTV, are known (Pringle *et al*., 1999; Gorbalenya *et al*., 2002; Gordon *et al*., 1995; Hanzlik *et al*., 1995; Yi *et al*., 2005). In addition, the capsid coding sequences of one other omegatetravirus, NωV, (du Plessis *et al*., 2005) and a further betatetravirus, *Providence virus* (PrV), have also been reported (Pringle *et al*., 2003).

The *Tetraviridae* are grouped into two genera according to their genome organization (Fauquet *et al*., 2005). The betatetraviruses are monopartite, with the replicase and capsid coding sequences present on a single genomic RNA and the capsid precursor is translated from a subgenomic RNA (Figure 1.2) (Gordon *et al*., 1999; Pringle *et al*., 2003). NβV is considered the type member for this particular genus (Fauquet *et al*., 2005). The omegatetraviruses are characterized by the encapsidation of

two genomic RNAs: RNA1 encodes the replicase and three small overlapping ORFs (Gordon *et al*., 1995; Yi *et al*., 2005), while RNA2 encodes the capsid precursor protein as well as a smaller overlapping ORF designated p17 (Figure 1.2) (Hanzlik *et al*., 1995; du Plessis *et al*., 2005; Yi *et al*., 2005). NωV is considered the type virus for this particular genus (Fauquet *et al*., 2005).

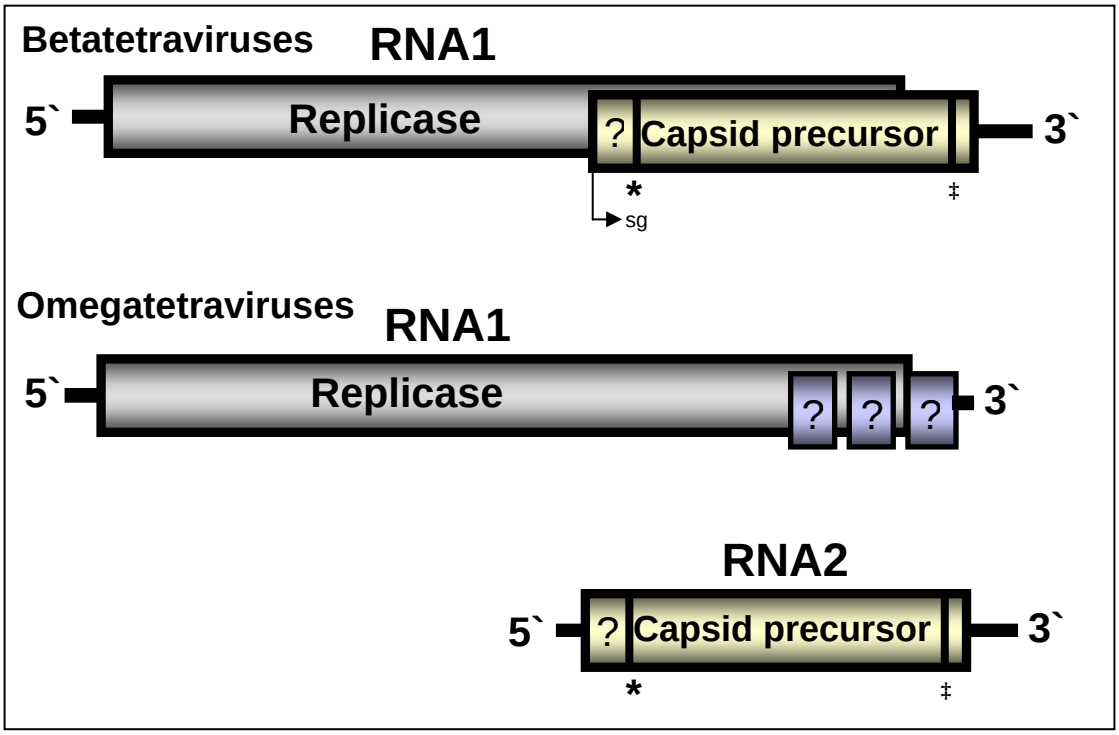

Figure 1.2. Genome organization and translation products of the viruses belonging to the genus *Betatetravirus* and genus *Omegatetravirus*. The site of subgenomic RNA initiation of the capsid precursor is indicated by the arrow. UTRs are represented by horizontal lines. Questions marks represent theoretical ORFs. The * represents the end of a PEST-like sequence in a theoretical ORF and the site of a potential protein cleavage event. The ‡ represents the site of an autoproteolytic cleavage event occurring after particle formation resulting in large and small capsid proteins (Adapted from Fauquet *et al.,* 2005).

The current classification of the betatetraviruses, as well as the *Tetraviridae* in general, is currently under review (A.E. Gorbalenya pers. comm., 2008). Analyses of the replicase sequences, and in particular the core motif arrangement of the RNA-dependent RNA polymerase (RdRp), has raised questions on the validity of a phylogenetic arrangement which was based on genome structure and particle appearance (A.E. Gorbalenya pers. comm., 2008).

1.2.3. The omegatetraviruses

Three species have been classified within the genus *Omegatetravirus*, including HaSV, NωV and DpTV (Fauquet *et al*., 2005). To date, no omegatetraviruses have been shown to replicate in insect cell lines and only the cDNA of HaSV has been shown to be infectious (Bawden *et al*., 1999; Gordon *et al*., 2001).

RNA1 encodes the viral replicase (187 kDa) consisting of the methyltransferase, helicase and RdRp domains as well as three small potential ORFs which, if translated, would result in proteins of 11, 15 and 8 kDa proteins, respectively (Gordon *et al*., 1995; Yi *et al*., 2005). Whether or not these small predicted peptides function in the virus lifecycle is not known. Phylogenetic analysis of the RdRp domain of RNA1 and the presence of methyltransferase and helicase domains, groups HaSV and DpTV with members of the alpha-like superfamily (Gordon *et al*., 1995; Gorbalenya *et al*., 2002; Yi *et al*., 2005).

RNA2 is translated into the capsid precursor protein often described as 'VCap' or protein α (Hanzlik *et al*., 1995; Agrawal and Johnson, 1992; Yi *et al*., 2005). During particle maturation, VCap is autoproteolytically cleaved at a scissile bond to produce the large capsid protein of 62 – 64 kDa (protein β) and a small capsid protein of 6 – 8 kDa (protein γ) (Agrawal and Johnson, 1995; Taylor *et al*., 2002; Yi *et al*., 2005). The ORF at the N-terminal of RNA2 of HaSV has been shown to contain a large percentage of the amino acids, P, E, S and T (Hanzlik *et al*., 1995) and is thus considered as a PEST-like protein sequence. Such sequences have been found in proteins which are often found to be rapidly degraded in a cellular environment (Rogers *et al*., 1986). Expression of p17 within bacterial heterologous systems lead to the production of hexagonal tube-like structures similar to those described in the expression of plant movement proteins (Hanzlik *et al*., 1995). The heterologous expression of the homologous p17 proteins of both TaV and DpTV did not result in the

accumulation of tube-like multimers, although dimers and trimers were observed upon translation of the DpTV p17 coding sequence in a bacterial expression system (Zhou *et al*., 2008).

1.2.4. The betatetraviruses

At present this genus comprises ten official members (Fauquet *et al*., 2005), with a further two potential members (Walter *et al*., 2008). Table 1.2 lists these twelve betatetraviruses, the host lepidopteran family and accession numbers of their genome sequences.

Table 1.2. The ten confirmed and two tentative members of the genus Betatetravirus

Species name	Abbreviation	Host family	Reference	Genome sequence accession #
Antheraea eucalypti virus	AeV	*Saturniidae*	Grace and Mercer, 1965	n.d
Darma trima virus	DtV	*Limacodidae*	Reinganum *et al*., 1978	n.d
Dasyichira pudibunda virus	DpV	*Lymantriidae*	Greenwood and Moore, 1984.	n.d
Euprosterna elaesa virus	EeV	*Limacodidae*	Zeddam *et al*. u.d.	AF461742
Nudaurelia capensis β virus	NβV	*Saturniidae*	Gordon *et al*., 1999	AF102884
Philosamia cynthia x ricini virus	PxV	*Saturniidae*	Reinganum *et al*., 1978	n.d
Providence virus	PrV	*Noctuidae*	Pringle *et al*., 2003	AF548354
Pseudoplusia includens virus	PiV	*Noctuidae*	Chao *et al*., 1983	n.d
Thosea asigna virus	TaV	*Limacodidae*	Pringle *et al*., 1999	AF062037
Trichoplusia ni virus	TnV	*Noctuidae*	Morris *et al*., 1979	n.d
Nudaurelia ψ virus*	NψV	*Saturniidae*	Walter *et al*., 2008	n.d
Nudaurelia ζ virus*	NζV	*Saturniidae*	Walter *et al*., 2008	n.d

Key: n.d = not determined, u.d = unpublished data, * = potential members of the genus *Betatetravirus*

The size of the betatetravirus genome ranges between 5.7 and 6.6 kb and typically contains two open reading frames (Figure 1.2). The first open reading frame encodes the replicase protein, while the second encodes the capsid precursor protein (Fauquet *et al*., 2005). The capsid precursor protein is translated from a subgenomic copy of the gene and is autoproteolytically cleaved to form large and small capsid proteins of about 55 - 60 kDa and 6 - 8 kDa, respectively (Gordon *et al*., 1999; Pringle *et al*., 1999; 2001; 2003).

1.2.4.a. *Nudaurelia β virus*

The NβV, the type member, encapsidates a monocistronic genome of 6625 nucleotides with the 5' proximal ORF of 5778 nucleotides encoding a protein of 215 kDa containing the three functional domains characteristic of a viral replicase: the methyltransferase, helicase and RdRp domains (Gordon *et al*., 1999). The replicase gene covers 87 % of the NβV genomic RNA, terminating at a stop codon over nucleotides 5868 - 5870. The capsid precursor ORF is located from nucleotides 4039 - 5874 with no intervening sequence between the replicase and VCap ORFs. A surprising feature of the genome of NβV is the degree of overlap - 1827 nucleotides, which represents virtually the entire capsid gene (Gordon *et al*., 1999).

The size of the coding sequence of the subgenomic RNA is 1.9 kb and there are no other significant ORFs on the NβV sequence, including the 636 nucleotide stretch between the end of the capsid gene and the beginning of the tRNA-like structure at nucleotide 6514 (Gordon *et al.*, 1999). The molecular weight of the major coat protein has been reported as 58.4 kDa with a minor protein of 8 kDa and the N-terminal of this precursor protein appears to be blocked – a common feature amongst the tetraviruses (Gordon *et al*., 1999).

1.2.4.b. *Thosea asigna virus*

Thosea asigna virus (TaV), another well-studied member of the betatetraviruses, has a substantially different gene presentation at the 3` proximal region of its genome with a PEST-like protein, p17, being translated from the subgenomic copy of the capsid precursor ORF (Pringle *et al*., 1999). This protein, although not detected in purified TaV particles, is cleaved first from the capsid precursor protein of virus-like particles (VLPs). Next, autoproteolytic cleavage of the large and small capsid proteins occurs, a processing event demonstrated in VLPs using in a baculovirus-mediated heterologous expression system (Pringle *et al*., 2001). This initial cleavage event, leading to the release of p17 from the large and small capsid proteins, is mediated by a 2A-like cleavage site - the first (highly efficient) report of a 2A-like mediated 'cleavage event' in a tetravirus (Donnelly *et al*., 2001b).

The p17 protein is proposed to play a non-structural role, since TaV VLPs produced by recombinant baculoviruses still assembled in the absence of the p17 ORF (Pringle *et al*., 2001). This sequence of protein processing and the presence of a processed and unblocked N-terminus of the capsid precursor protein (Pringle *et al*., 2001) are characteristics not shared by other tetraviruses. Another tetravirus, EeV, shares a high amino acid sequence homology of both the replicase and capsid protein precursor proteins with those of TaV (Gorbalenya *et al*., 2002). The positioning of replicase and capsid precursor protein ORFs in these two viruses show far less overlap (Figure 1.3) than that observed in NβV (Figure 1.2) (Gordon *et al*., 1999). Furthermore, both TaV and EeV lack any identifiable helicase and methyltransferase domains within their respective replicase ORFs. The RdRp domain within the replicase ORF of both of these viruses is unusual in that the motif arrangement is permuted, unlike the overwhelming majority of viruses where the canonical motif presentation characterizes the RdRp arrangement (Gorbalenya *et al*., 2002). The significance of this motif permutation in the RdRp domains of these two viruses is discussed in more detail later in this work.

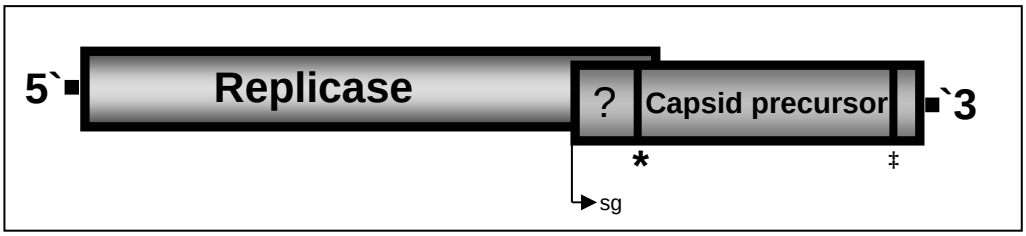

Figure 1.3. Genome organization of the betatetraviruses, TaV and EeV. The schematic diagram represents (to scale) the positioning and relative size of the replicase and capsid precursor ORFs. The size and positioning of these two ORFs within the genomes of TaV and EeV are virtually identical. The site of subgenomic RNA initiation of capsid precursor is indicated by the arrow and is in a +1 position relative to the RdRp reading frame. UTRs are represented by horizontal lines. The question mark refers to the PEST-like sequence within a theoretical ORF. The * represents the site of a 2A-like co-translational processing signal and the ‡ represents the site of the autocatalytic cleavage event.

The unusual processing signals, ORF positioning, permuted RdRp motif arrangement and PEST-like sequences make TaV and EeV divergent members of the genus *Betatetravirus* which may warrant a potential reclassification of the *Tetraviridae* based on factors other than the capsid homology approach alone.

1.2.4.c. *Providence virus*

In 2002, Pringle and co-workers in Providence, USA, discovered a unique tetravirus persistently infecting a midgut cell line (MG8) derived from the corn earworm (*Helicoverpa zea*) (Pringle *et al*., 2003). The virus, named *Providence virus* (PrV), is the first tetravirus able to replicate in tissue culture and purified virions are also able to infect two other insect cell lines, one from *H. zea* fat body (FB33) and one from the larval tissue of beet armyworm larvae, *Spodoptera exigua* (*Se*1) (Pringle *et al*., 2003). PrV virions are typical icosahedral particles that encapsidate both the genomic (6.4 kb) and subgenomic (2.5 kb) RNAs. Negative sense replication intermediates of the PrV genome were also uncovered (Pringle *et al.,* 2003), a feature so far unproven in other tetraviruses. Virions have an average diameter of 40 nm and an external capsid appearance resembling NβV and TaV (Pringle *et al*., 2003).

PrV capsid particles were shown to be composed of both the major (60 kDa) and minor capsid (8 kDa) proteins. As observed in TaV, the N-terminus of the PrV major capsid protein is also unblocked (Pringle *et al*., 2003) unlike most of the other tetraviruses (Fauquet *et al*., 2005). PrV encodes a capsid protein precursor of 81 kDa with a 2A-like 'cleavage' sequence ('PrV-2A_3') at its N-terminus, in a similar position to TaV (Pringle *et al*., 2003; Luke *et al*., 2008). Co-translational processing of the capsid precursor results in the production of a 13 kDa protein, which has not been detected in virus particles (Pringle *et al*., 2003). In addition, a second 2A-like 'cleavage' sequence (PrV-2A_2) is present in the 13 kDa protein coding sequence (Luke *et al*., 2008). This coding sequence is positioned at the N-terminal of the capsid protein precursor and has a high P, E, S and T amino acid content (Pringle *et al*., 2003) similar to the suspected PEST-like sequences of HaSV (Hanzlik *et al*., 1995), DpTV (Yi *et al.*, 2005), and TaV (Pringle *et al.*, 2001).

Attempts by Taylor *et al*. (2005) to produce VLPs using a baculovirus-mediated heterologous expression system were unsuccessful when the capsid protein precursor and capsid protein alpha (virus capsid before autoproteolytic cleavage) were individually expressed in this system. Heterologously expressed capsid protein precursor was marginally more stable than capsid protein alpha alone, even though both proteins proved highly insoluble and were expressed at very low levels (Taylor *et al*., 2005). The roles of the 2A-like processing in the regulation and expression of viral proteins will be discussed in detail later in this survey.

Further unpublished studies on PrV revealed an interesting genome organisation and a highly divergent replicase sequence (F.M. Pringle and L.A. Ball, pers. comm., 2005). At the 5` end of the genome, a large ORF spanning from nucleotides 45-3708 which, if translated, would produce a protein with a predicted molecular weight of 130 kDa, exists. This protein contains a 2A-like 'cleavage' sequence at its N-terminus (PrV-2A_1) and the processing ability of this sequence has been shown to be active *in vitro*, using a reporter-based system (Luke *et al.,* 2008). Furthermore, neither

the nucleotide nor amino acid sequences show any homology with other deposited sequences in the National Centre for Biotechnology Information (NCBI) database.

The replicase ORF is positioned between nucleotides 1027 and 3775 and, if translated, should produce a protein of 104 kDa in a +1 frame relative to p130 (F.M. Pringle and L.A. Ball, pers. comm., 2005). Analysis of the amino acid sequence of this ORF revealed no identifiable methyltransferase or helicase domains. Comparison of this amino acid sequence with non-redundant sequences in the NBCI databases aligned p104 with replicase and RdRp sequences associated with (+)ssRNA plant viruses belonging to the *Tombusviridae* and *Umbraviridae* families (F.M. Pringle pers. comm., 2005). Repeated sequence analysis of the p104 ORF revealed the presence of an in-frame amber stop codon (UAG), a feature not previously observed in other tetraviruses.

The discovery of *Providence virus* has not only added more uncertainty to the phylogenetic classification of the *Tetraviridae,* but also provides an opportunity to study a renewable source of tetravirus in the context of capsid formation, assembly and disassembly, replication and coupled mechanisms associated with genome transcription and translation, the real-time genome processing of viral translation products and finally the site of replication within the infected cell line(s). Virtually nothing is known about the replication of tetraviruses nor the host and accessory viral factors involved in the life cycle. The sections below describe viral RNA-dependent RNA polymerases, their interactions with host cell membranes and proteins as well as a survey of two model viral replication systems: the *Picornaviridae* and *Tombusviridae*. Finally, the remainder of this survey serves to create a foundation to understanding the gene expression and protein processing strategies employed by the tetravirus, *Providence virus*.

1.3. RNA-DEPENDENT RNA POLYMERASES

The template-dependent polynucleotide polymerases (TDPPs) that replicate cellular and viral genomes can be described as one of the few classes of enzymes which bridge the gap between eukaryotic, prokaryotic and viral replication (Gorbalenya *et al*., 2002). These enzymes are usually present as one of three to four enzymatic domains in an ORF broadly referred to as a replicase protein. The DNA genomes of cellular organisms and the majority of DNA viruses are replicated by a DNA-dependent DNA polymerase (DdDp). Positive strand RNA viruses encompass over one-third of all virus genera and include numerous important pathogens such as severe acute respiratory syndrome (SARS) coronavirus, *Hepatitis C virus* (HCV) and *Ebola virus* (Ahlquist *et al*., 2003). The RNA genomes of these viruses are classified as (+)ssRNA and (-)ssRNA viruses, or dsRNA viruses and retroviruses that reverse transcribe their genomes using a reverse transcriptase (Fauquet *et al*., 2005).

The RdRp is the only enzyme universally conserved in all of the thousands of known non-satellite RNA viruses (Ahlquist, 2002). The replication complexes of positive sense viruses usually contain one or more viral proteins. Additional domains corresponding to RNA helicase, methyltransferase and proteinase activities may be present (van der Heijden and Bol, 2002). RdRps are used to replicate the genomes of viruses with no DNA stage and RNA-dependent DNA polymerase (RdDp and reverse transcriptase) is used only by viruses with a DNA stage in the life cycle (Baltimore, 1971). RdRps share multiple sequence motifs that are conserved across all three major virus superfamilies and crystal structures of RdRps of positive strand RNA and dsRNA viruses show structural similarity not only to each other, but also to DNA-dependent RNA and DNA polymerases and reverse transcriptases (Ahlquist, 2002). All of the above-mentioned polymerases share a tertiary structure similar to that of a cupped right hand, with three subdomains termed palm, thumb, and fingers (Ollis *et al*., 1985). This structural and functional conservation

implies that palm subdomains of all TDPPs may have evolved from a common ancestor (Gorbalenya *et al*., 2002).

1.3.1. Structure of RNA-dependent RNA polymerases

Poch and co-workers (1989) originally proposed a core polymerase module whose four well-ordered motifs (motifs A to D) in a region of 120 - 210 amino acids were essential for RNA polymerization (Poch *et al*., 1989). This group also went on to compile a loosely based phylogenetic map of viruses based on their polymerase sequences. This lead to the proposal of the essential nature of these four motifs as well as the relatedness of the RNA-dependent RNA polymerases to DNA polymerases which probably originate from a common, putative RNA polymerase ancestor, encompassing all current polymerases (Poch *et al*., 1989). This work was explored further by Koonin (1991) and later Koonin and Dolja (1993) who identified a total of eight motifs (I - VIII) in the RdRps of positive-stranded RNA viruses, although only three (IV, V and VI) showed complete conservation among all RdRps, which overall can share very little sequence identity amongst closely related viruses (Johnson *et al*., 2001). These three highly conserved motifs; IV, V and VI (Koonin and Dolja, 1993), correspond to motifs A, B and C, originally discovered by Poch *et al*., (1989) thus making them ideal for basic analysis between RdRp sequences.

The palm domain structure of the RdRp is especially conserved and contains the four sequence motifs (A - D) preserved in all RNA and DNA polymerases (Ahlquist, 2002). The palm subdomain consists of a four-stranded antiparallel β-sheets with two α-helices packed beneath the sheet, and comprises the four ordered motifs A – D (reviewed in Gorbalenya *et al*., 2002). Motif A contains two aspartate residues separated by four or five residues and is responsible for structural features in the polymerase resulting in catalysis, while motif C (GDD), the polymerase active site, contains an Asp-Asp dipeptide, which is often

preceded by a glycine residue. Motif B, the ribose-binding pocket, contains an asparagine that contributes to the discrimination between dNTPS and NTPs and thus determines whether DNA or RNA is synthesized (reviewed in Gorbalenya *et al*., 2002). This asparagine residue is present at the same structurally homologous position in all plant and animal RNA virus polymerases and only differs in RNA baceteriophages resulting in an elevated mutation frequency in this group (Korneeva and Cameron, 2007). Figure 1.4 depicts a typical RdRp compared to other polymerase enzymes all encoding the four highly conserved motifs, A - D and if present, motif E, annotated on each structure.

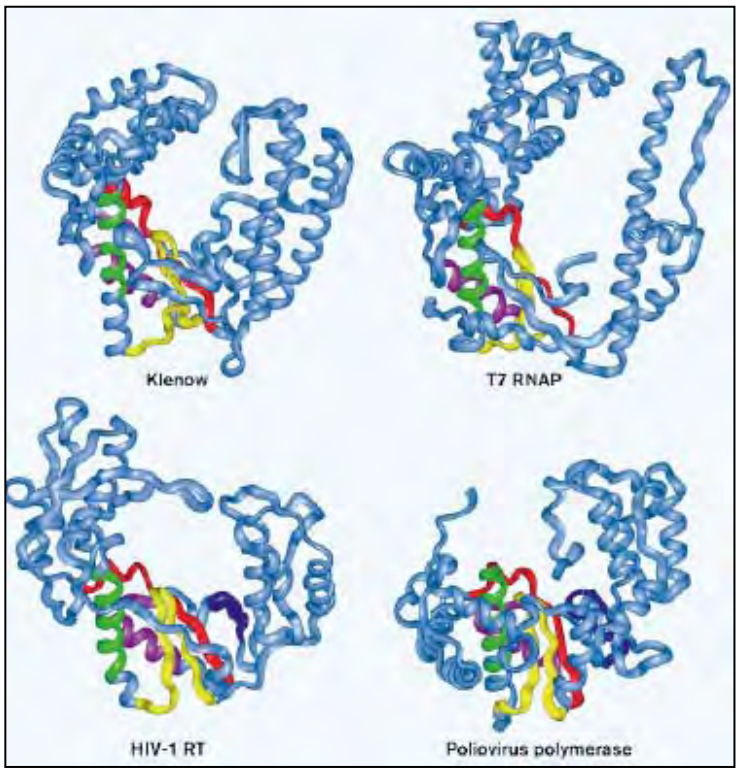

Figure 1.4. Polymerase modules in representative structures from each of the four categories of polymerases. Structures of a Klenow polymerase (a DNA polymerase), T7 RNA polymerase (a DNA-dependent RNA polymerase), *Human immunodeficiency virus* 1 reverse transcriptase (an RNA-dependent DNA polymerase) and *Poliovirus* polymerase (a RNA-dependent RNA polymerase). The thumb is positioned on the right of each structure. The structural/sequence motifs are colour coded: A = red; B = green; C = yellow; D = purple; E = navy blue. Taken from Hansen *et al*. (1997).

1.3.2. Subgroup classification of viruses based on conserved domains within the replicase protein

Sequence analysis of viral RdRps has been used to phylogenetically cluster RNA viruses to the supergroup level (Koonin, 1991; Koonin and Dolja, 1993). The existence and diversity of other enzymatic modules within the replicase protein, such as helicase, guanylyltransferase or methyltransferase and proteinase activities are also used to distinguish members of the three supergroups (reviewed in van der Heijden and Bol, 2002).

Typically, RNA viruses that replicate in the cytoplasm of the host cell contain RdRps that have RNA guanylyl- and methyltransferase domains. RNA helicase domains are found in the sequence of most RdRps (Koonin and Dolja, 1993) although larger quantities of these translated domains are usually present relative to the RdRp (van der Heijden and Bol, 2002). All RNA viruses are classified into either supergroup I, II or III. Supergroup I, the picorna-like viruses contain a replicase expressing a type I polymerase, type 1, 2 or 3 helicase, no guanylyltransferase /methyltransferase and a cysteine or serine-like proteinase. Supergroup II is split into two major lineages. The first, the carmo-like viruses, contain a replicase expressing only a type II polymerase. The second clade, the flavi-like viruses also contain a type II polymerase in addition to a type 2 helicase, type 2 guanylyltransferase /methyltransferase and a serine-like proteinase. Finally supergroup III, the alpha-like viruses, contain a replicase expressing a type III polymerase, a type 1 helicase and type 1 guanylyltransferase /methyltransferase (reviewed in van der Heijden and Bol, 2002). The core RdRp component of any virus requires host proteins in order to be active: often post-translational modification of the RdRp within the host cellular environment as well as its particular location or micro-environment within the cell are all essential requirements for polymerase activity. The sections below describe some of these requirements and modifications, using well-studied (+)ssRNA virus RdRps and their replication complexes as examples.

1.3.3. Associated proteins and replication complexes of RNA-dependent RNA polymerases

Although certain RdRps have polymerase activity as individual proteins, they may still require additional viral polypetides for complete functionality and have often been shown to oligomerize the catalytic components of the replication complex (Dye *et al*., 2005). The polymerase itself may require a protein primer in order to be active (termed primer-dependent initiation) or, in certain virus families, can be primer-independent (Buck, 1996). Differentiation between the two types of polymerases and their modes of initiation can be determined by structural analysis of the template binding site within the polymerase structure (Ferrer-Orta *et al*., 2006).

An example of RdRp self-interaction has been shown for the nodavirus, *Flock House Virus* (FHV), which functions as a multimer rather than a monomer, in one or more steps during genome replication (Dye *et al*., 2005). The model picornavirus, *Poliovirus* (PV), is another example of polymerase self-interaction based on the number of regions involving polymerase-polymerase interactions observed in the crystal structure of $3D^{POL}$, the polymerase component of PV (Hansen *et al*., 1997). Oligomerization of the RdRp is important for its functioning since RNA binding and polymerization activities are highly cooperative with respect to the concentration of polymerase (Pata *et al*., 1995).

The single catalytic centres of a replication complex are also often biochemically modified in order to be active (Ahlquist *et al*., 2003). Such is the case with FHV whose RdRp requires glycerophospholipids for activity *in vitro* (Wu *et al*., 1992). This discovery has since been supported by findings that phosphorylation of RdRps is often required to regulate some aspects of their function (Jakubiec and Jupin, 2007). Phosphorylation is a reversible post-translational modification seen in an increasing number of viral non-structural proteins and is a means of modulating protein properties such as stability, enzymatic activities, subcellular localization etc. (Jakubiec and Jupin, 2007). Phosphorylation of the replicase

component of the Tymovirus, *Turnip yellow mosaic virus* (TYMV), was documented when the protein was expressed both individually and in the context of an infection (Jakubiec *et al*., 2006). Phosphorylation of the TYMV RdRp showed a role in the infectivity of the virus but interestingly, not in the interaction between viral replication proteins (Jakubiec *et al*., 2006).

Direct interactions between host proteins and viral replication components are essential for the relatively gene-poor viruses to continue propagation (Kushner *et al., 2003)*. Many regulatory and other host cell proteins have been identified as playing a key role in virus replication and one such study was performed by Kushner and co-workers in 2003, whereby the entire genome of the *Brome Mosaic virus*-permissive *Saccharomyces cerevisiae* was screened for the up- or down-regulation of genes during infection with BMV (Kushner *et al*., 2003). The results both confirmed previous studies of this nature and introduced new theories and cellular pathways involved during viral infection of a (simple) eukaryote.

Replication was found to be down-regulated by host genes such as histone modification enzymes, glutathione synthesis, RNA polymerase regulation genes, RNA cap-binding protein and ubiquitin modification. In contrast, genes involved in organelle morphology, phosphorylation, transcription, elongation and phospholipid biosynthesis were found to be up-regulated during BMV viral replication (Kushner *et al*., 2003).

The role of some of the host cell proteins involved in the replication of the FHV in *Drosophila* cells has been studied (Kampmueller and Miller, 2005; Weeks and Miller, 2008). Analyses showed the involvement of the chaperone, heat shock protein 90, as a key host protein that facilitates viral RNA replication by FHV sequestering the chaperone pathway to assemble its RNA replication complexes on mitochondrial membranes (Kampmueller and Miller, 2005). Interestingly, parallel infection of FHV in *S. cerevisiae* requires the heat shock protein 70 co-chaperone, YDJ1, (Weeks and Miller, 2008) highlighting how these viruses adapt to different hosts. This was achieved in part by utilizing different, perhaps more

metabolically available, host factors in their replication life cycle (Weeks and Miller, 2008).

1.3.4. Anchoring of the RdRp to accessible intracellular membranes

The self-interaction, protein modification and host protein sequestering by a viral replication complex often involves the host cell undergoing morphological and biochemical changes (Marcotte *et al*., 2007). Importantly, RdRps and their associated replication complexes must interact with a host intracellular membrane for the following reasons: (i.) physical support to the RNA/protein complexes, (ii.) concentrating and compartmentalizing of the components, (iii.) supplying essential lipids that are required for RNA synthesis and (iv.) providing attachment of viral RNA during unwinding (Lee *et al*., 2006). It is also speculated that the membranous structures formed during virus replication could shield this process, as well as the RdRp and potential dsRNA replication intermediates, from host antiviral mechanisms (Jonczyk *et al*., 2007). It is because of this association with intracellular membranes that all RNA viruses produce replication complexes made up of both host and cellular factors and are located at a host intracellular membrane normally provided by a cellular organelle (Ahlquist *et al*., 2003).

Recent evidence points to the use of aggresomes and auotophagy within the host cell in order for the infecting virus to generate sites for viral replication (Wileman, 2006). The use of these two physical characteristics of a eukaryotic cell as a means of concentrating and localizing viral factors has been described in dsDNA viruses (reviewed in Wileman, 2006). Only a few ssRNA viruses, including PV, have been shown to be dependent (e.g. virus yield, pH sensitivity) on the presence of cellular proteins directly involved in the formation of either of these two subcellular formations (Wileman, 2006; Netherton *et al*., 2008). Due to limited information on the role of aggresomes and autophagy in the lifecycles of other ssRNA viruses available at present, the impending discussion of replicase complex formation and localization will centre mainly on interaction and anchoring

of replication complexes with an intracellular membrane using well studies (+)ssRNA plant viruses as examples.

1.4. TRANSCRIPTIONAL AND TRANSLATIONAL REGULATORY MECHANISMS IN RNA VIRUSES

1.4.1 Translational control in plant viruses and associated host factors

Approximately 20 % of plant (+)RNA viruses have genomic (and subgenomic) RNAs with a 5`-cap and poly-(A) tail, thus requiring a closed loop conformation of these viral transcripts in order for efficient translation to occur (Dreher and Miller, 2006). The closed loop conformation is achieved by sequestering various cellular translation factor proteins such as elongation initiation factor 4E (eIF4E), eIF1A and eIF4G although the arrangement and presence of these translation factors has been found to be very different in infected plant cells compared to that seen in the standard cellular mRNA pool (Dreher and Miller, 2006). The binding of 40S ribosomal subunit of the host's translation machinery to the 5`-end of the vRNA is mediated by the eIF3 and eIF4 groups (Dreher and Miller, 2006). The eIF1 group factors are required to promote the 'scanning step' necessary to find the start codon (reviewed in McCarthy, 2000), as well as potentially, to act as a molecular bridge between the 5` and 3` RNA ends (Dreher and Miller, 2006).

A number of 5` to 3` bridging mechanisms exist in positive strand viral RNAs, but these are variable in nature due to varying 3` structures and sequences adopted by the virus genome (Dreher and Miller, 2006). Virus genome-linked proteins or 'protein primers' such as those employed by the *Picornaviridae* are often used as replacements in cap-independent translation (Buck, 1996). Regulation of cap-independent translation by members of the *Tombusviridae* is often controlled by sequences in the 3`-UTR. These sequences form stem-loop structures and rRNA-binding

tracts as a means of recruiting ribosomes to these enhancer signals (Dreher and Miller, 2006). Interestingly, pseudoknot structures in this region play a vital role in re-initiation of translation resulting from leaky stop codons (Feng *et al*., 1992), which themselves play an essential regulatory role in the translation of replication proteins of the *Tombusviridae* (Skuzeski *et al*., 1991).

If present, a poly-A tail acts as a potent translational enhancer and also stabilizes the RNA (Dreher and Miller, 2006). Translation is also enhanced by the aminoacylatable tRNA-like structures found at the 3`-end of many plant virus RNAs. These structures enhance the production of protein because they bind tightly to eIF1A. Once translation has been optimized by one of these mechanisms, translational recoding is often utilized as a means of expressing more than one gene, and in asymmetric quantities (Dreher and Miller, 2006; Ahlquist, 2006).

A common method of translation recoding besides leaky scanning, missense codon reading and frameshifting is the presence of a read-through stop codon, usually a UAG or UGA, found in frame with a coding sequence and used to express asymmetric levels of the 5` versus the 3` coding sequence (Skuzeski *et al*., 1991; Beier and Grimm, 2001). Alternatively, the stop codon is position in the capsid protein ORF, where extensions of structural proteins are essential for infectivity (Schmitt *et al*., 1992). When found in a particular sequence context, the stop codon does not result in termination of translation but rather the addition of a tyrosine, glutamine, tryptophan or leucine at this codon, thus continuing with the peptide elongation and not “hopping” over the stop codon to re-initiate at a similar codon as was originally proposed (Beier and Grimm, 2001). These misread stop codons are found in certain plant, animal and (rarely) insect virus families (Skuzeski *et al*., 1991; Harrell *et al*., 2002) as well as tissue-specific full-length protein in eukaryotic translation (Dalphine *et al*., 1997; Beier and Grimm, 2001). A scan of the NCBI nucleotide database by Harrell *et al*. (2002), revealed that at least 23 different virus genera had confirmed read-through stop signals in their genomes, yet only one insect

virus species, *Plautia stali intestine virus* (PSIV) of the Cripavirus genus (*Dicistroviridae*), was predicted to contain one of these highly conserved recoding signals (Harrell *et al*., 2002).

Virus read-through stop signals have been categorized into three groups based on: the 3` sequence immediately downstream of the stop, the type of stop codon and the frequency of occurrence within certain virus groups (Beier and Grimm, 2001). An alternative categorization of these signals has also been compiled based on the 3` sequence downstream of the leaky stop codon (Harrell *et al*., 2002). More data is now available on the *cis*- and *trans*-acting elements responsible for the frequency of readthrough in viral, prokaryotic and eukaryotic sequences (Craigen *et al*., 1985; Beier and Grimm, 2001; Harrell *et al*., 2002). A basic stop signal component has been described for many in-frame stop codons characterized mainly in plant viruses but also including examples found in animal viruses such as the *Reoviridae,* a number of Alphaviruses, gammaretroviruses and epsilonretroviruses, and many encephalitis-causing virus species (Beier and Grimm, 2001; Harrell *et al*., 2002). The most common functional and also effective read-through stop signal was characterized by the nucleotide sequence UAG CAR YYA where Y represents a pyrimidine base and R, a purine base (Skuzeski *et al*., 1991; Harrell *et al*., 2002)

When present in the retroviruses, these signals are thought to function better in the presence of short RNA hairpin-loop/pseudoknot structures in the sequence downstream of the stop codon (Feng *et al*., 1992). In prokaryotes, the read-through stop signal has been shown to operate in conjunction with a frameshift after the stop signal creating alternative proteins beyond this point, which are dependant on the nature of the structural motif encoded by the re-initiated 3`-UTR (Craigen *et al*., 1985). Not only do read-through stop signals moderate translation levels of the C-terminally extended protein, but have also been found to be a pre-requisite in virion assembly in the case of luteoviruses, where the longer capsid protein is essential for particle formation (Schmitt *et al*., 1992). This in

turn, facilitates the transmission and infectivity of these viruses during the course of infection (Schmitt *et al*., 1992).

1.4.2. Translation initiation, control and modification in insect RNA viruses

In general, animal viruses exhibit less tightly controlled translational control than plant and insect viruses (Dreher and Miller, 2006). However they have to adopt a more 'aggressive' means of evading and sequestering immunomodulatory factors and transcription and/or translation factors, respectively. Alternatively, some insect viruses such as the *Dicistroviridae* employ an internal ribosome entry site (IRES) at the 5`-UTR of ORF2 and contained within the intergenic region of the two cistrons (Masoumi *et al*., 2003; reviewed in Martínez-Salez *et al*., 2008). Translation resulting from ribosome entry from this structure does not require the activities of eIF1, eIF1A or the cap-binding protein eIF4 (McCarthy, 2000). An additional advantage of such a secondary RNA structure guiding the binding of the 40S ribosomal subunit and a pseudoknot structure occupying the P-site, is that these viruses can initiate translation using alternative start codons such as CUG, CUU or ACG, without competing for host initiation factors or initiator tRNA (RNA_i^{Met}) (McCarthy, 2000; Sasaki and Nakashima, 1999). These initiation codons do not function in a scanning-dependent manner during translation and are found with IRES structures, which direct the host ribosome to the start of translation (Sasaki and Nakashima, 1999).

Interestingly, IRES elements have been implicated in the direct transfer of ribosomes from the initial binding site on the 5`-UTR to the second or third AUG. The reasons for the "dropping" of an accessible in-frame AUG have yet to be uncovered, but is thought that cell-specific factors could facilitate an IRES conformation allowing for translation of products from an alternative upstream AUG (Sarnow, 2003). The polypetides translated from the ORFs of the *Dicistroviridae* are post-translationally processed via

2A-like ‘cleavage’, a mechanism similar to that observed in the picornaviruses and certain members of the genus *Betatetravirus* (Rossmann and Tao, 1999; McCarthy, 2000; Pringle *et al*., 2001; Luke *et al*., 2008). In contrast, the tetraviruses make use of overlapping reading frames, subgenomic RNAs or bicistronic genomes in order to translate the various ORFs of their genomes (Hendry and Agawal, 1994; Gordon *et al*., 1999).

An interesting co-translational modification that is well documented in the picorna- and tetraviruses is the presence of a functional 2A-like polypeptide ‘cleavage site’ usually positioned at the N-terminal of a protein coding sequence (Donnelly *et al*., 2001b; Luke *et al*., 2008). 2A-like processing is characterized by a co-translational ribosome skipping event which can be 90 - 95 % efficient *in vitro* (Donnelly *et al*., 2001a; Luke *et al*., 2008). 2A-like ‘cleavage’ occurs by ribosomonal stuttering over a prolyl-tRNA (within a consensus sequence), which is stearically hindered, resulting in the hydrolysis of the ester bond between the nascent peptide and $tRNA^{Gly}$ (Ryan *et al*., 1999; Donnelly *et al*., 2001). The end products of this processing are small quantities of a full length protein resulting in immediate re-initiation of translation at the downstream glycine residue of the consensus sequence, and large quantities of a small peptide and large protein (Ryan and Drew, 1994). The core 2A-like sequence (DxExNPG↓P), where x represents any amino acid, is highly conserved within the Cardiovirus and Apthovirus genera of the *Picornaviridae*, some members of the *Dicistroviridae* and the Tetravirus genus Betatetravirus, type C rotaviruses (Donnelly *et al*., 2001b), and more recently discovered in eukaryotic sequences usually embedded in non-LTR retrotransposon regions within the genome (Donnelly *et al*., 2001b; Heras *et al*., 2006). Most of these sequences have been found to be active *in vitro* (Luke *et al*., 2008) and function in the picornaviruses by mediating the 2A and 2B protein partitioning during translation, while the 2AB precursor is also functional during the virus lifecycle (Ryan and Drew, 1994).

As described for the tetraviruses earlier (Section 2.4.3), these sequences are situated at the N-terminal ends of the capsid precursor ORF and are functional *in vitro* (Pringle *et al*., 2001, Luke *et al*., 2008). These 2A-like sequences could represent an alternative processing mechanism of the capsid precursor polyprotein. Apart from the obvious involvement of host translation factors in insect virus genome replication and the few regulatory mechanisms discussed above, what other *cis*-acting viral factors control insect virus gene expression? Transcriptional and translational control mechanisms have been studied in-depth using the nodavirus, FHV, as a model insect virus and PV as a model animal virus, yet the most progressive research to date on virus replication, replication complexes and genome processing has been conducted on (+)ssRNA plant viruses, in particular the *Tombusviridae* (Rajendran and Nagy, 2003; Panaviene *et al.,* 2003; Panavas *et al*., 2005a; Nagy and Pogany, 2006). During the course of this study, it was shown that PrV uses at least one expression strategy utilised by this family of viruses during genome replication, thus qualifying the inclusion of the well-characterized replication strategies employed by the *Tombusviridae* in the discussion below.

1.5. THE *TOMBUSVIRIDAE*

The *Tombusviridae* family of viruses are a group of small, (+)ss monopartite RNA viruses, which infect a number of agriculturally important plant species (Rubino and Russo, 1998). Much research has been dedicated to studying the viral proteins involved in genome replication of these viruses, in particular the interaction of the RdRp and its associated viral accessory protein at the site of replication (Scholthof *et al*., 1995; Panavas *et al*., 2005; Batten *et al*., 2006). The burgeoning field of plant virology partially owes its success to the ease of manipulation of these viruses in the laboratory, the robustness of infection and the availability of the model eukaryotes *S. cerevisiae* and *Arabidopsis thaliana* as permissive hosts for many of the species in this family (Weber-Lotfi *et al*., 2002; Panaviene *et al*., 2004 and Jonzcyk *et al*., 2007). In addition, the

availability of RNAi technology for fully sequenced host genomes allows for a rapid, systematic identification of host factors involved in the replication of these viruses. These tombusviruses employ a complex multicomponent approach of gene duplication (Lin *et al*., 2007) and a selection of these mechanisms are summarized below.

1.5.1. Genome organization, translation products and processing

Tombusvirus genomes are characterized by five ORFs encoding proteins for the capsid (~41 kDa), cell-to-cell movement (~22 kDa), suppression of gene silencing (~19 kDa), and two proteins required for genome replication (~33 kDa and ~92 kDa) (Figure 1.5) (Russo *et al*., 1994). The two replicase proteins (both located at the very 5` end of the genome), often referred to as the accessory/auxiliary protein and replicase protein, are necessary for genome replication within the host plant cell (Scholthof *et al*., 1995). The auxiliary protein, 25 - 40 kDa in molecular weight, is translated first from the multicistronic genome, followed by leaky read-through translation of the replicase component (85 - 100 kDa) in a much lower (approximately 20 fold) concentration (Scholthof *et al*., 1995).

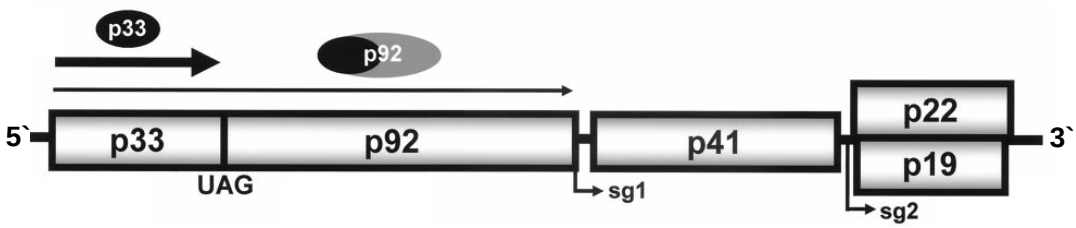

Figure 1.5. **Genome organization of a model tombusvirus, *Tomato Bushy Stunt virus* (TBSV).** The ORFs are represented by the open rectangles with the name (and molecular weight in kDa) of the product in each box. The characteristic read-through stop codon is labelled UAG and the sites of subgenomic RNA initiation are indicated by the bent arrows. The replicase proteins, p33 and p92 are indicated by black and grey ovals, respectively, while arrows below the ovals represent the relative proportions of each product. UTRs are represented by horizontal lines (Taken from Monkewich *et al*., 2005).

Read-through translation in the replicase ORFs of the most well-studied tombusviruses, *Tomato Bushy Stunt virus* (TBSV) and *Cucumber necrosis virus* (CNV), occurs at a frequency of about 5 -10 % in plant cells and 10 % in yeast (Scholthof *et al*., 1995; Panaviene *et al*., 2004). Both N-terminal ('accessory') and full length (RNA polymerase) read-through products are essential for virus replication (Scholthof *et al*., 1995; Weber-Lotfi *et al*., 2002) with species-specific interaction between the two proteins in TBSV shown to be essential for replicase assembly and thus infectivity (Rajendran and Nagy, 2006). The auxiliary protein of CNV, which has no recognizable enzymatic activity or recognizable catalytic motifs (Panaviene *et al*., 2003) is phosphorylated by a membrane-bound plant kinase (Shapka *et al*., 2005). The modified version of this accessory protein is one of the *cis*-acting components responsible for replicase and RNA template interactions during infection (Shapka *et al*., 2005). The regulation of full-length replicase translation by various recoding mechanisms is in fact a highly conserved characteristic in both (+) RNA viruses and retroviruses. Although different mechanisms are employed in the control of expression of these proteins, the ratios of replication-related proteins leading to what is essentially polymerase down-regulation implies a functional requirement in not only plant (and animal) alphaviruses but also retroviruses and arteriviruses (Ahlquist, 2006).

The control of tombusvirus genome transcription and translation is regulated by at least two highly unusual *cis*-acting RNA elements, one of which is unique in that it regulates RNA replication (Sit *et al*., 1998; Pantaleo *et al*., 2004; Monkewich *et al*., 2005; Lin *et al*., 2007). This particular element is an internal RNA element (RII) comprising a helical stem-loop structure found within the read-through section of p92 as well as a functionally essential C-C mismatch in the loop region (Figure 1.6). This component is dispensable late in the RNA replication cycle and is thus thought to play a crucial role in template recruitment and/or replicase complex assembly (Monkewich *et al*., 2005). RII, which is only active on the positive strand, is essential for virus infectivity and as has been found with other secondary RNA elements that come into contact with

ribosomes, is functionally inhibited by the translation of the viral RNA *in vivo* (Figure 1.6) *(*Monkewich *et al*., 2005).

Recently, a multicomponent model comprising RNA-RNA interactions has been proposed for the premature termination of tombusvirus transcription (Lin *et al*., 2007). Premature termination of transcription of genomic vRNA is proposed to occur in a number of ssRNA virus families and the result of this RdRp halt in activity is that the 3` truncated minus strands are used for templates in subgenomic mRNA transcription (Sit *et al*., 1998). A complicated system of RNA-RNA interactions and long-distance base pairing results in a genomic RNA structure primed for premature termination of transcription before the subgenomic RNA sequences are reached (Lin *et al*., 2007)

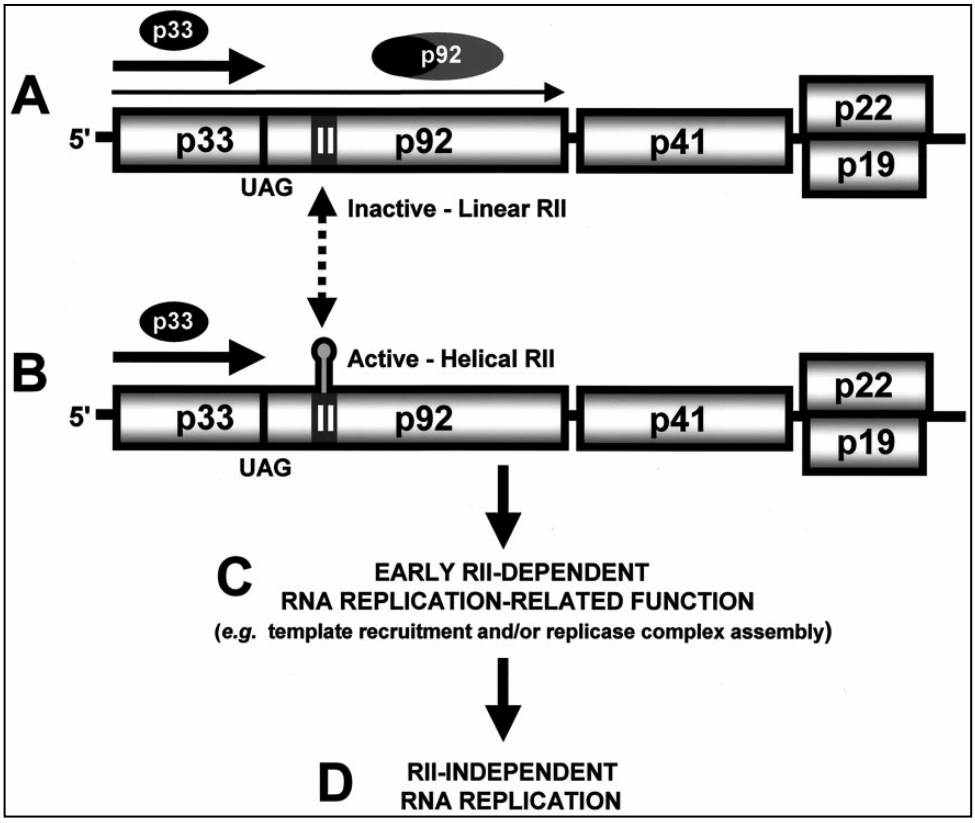

Figure 1.6. A model of the role of the RII replication element during tombusvirus replication and the effect of translation on this internal RNA element. A: The (infrequent) translation of p92 results in the inactivation of the RII replication element. B: When not undergoing translation the RII replication element takes on its secondary structure and is thus active. The dotted, double-headed arrow reinforces the dynamic state of an active/inactive RII replication element C: Activities and roles of active RII

replication element. D: later in the lifecycle of the virus, the RII replication element is no longer required for successful genome replication (Taken from Monkewich *et al*., 2005).

In summary, the two models discussed above, represent a small portion of data available on the transcriptional and translational control mechanism employed by these viruses. It is important to compare these mechanisms with those utilised by other RNA families and integrate this knowledge within other spheres of the tombusvirus lifecycle, in particular, the subcellular localization and components of tombusvirus replication complexes.

1.5.2. Subcellular localization and functional roles of replication proteins

Scholthof and co-workers were the first group in the field of tombusvirus replication to show that both the auxiliary and replicase proteins of the type member, TBSV, were not only asymmetrically expressed but also membrane-associated (Scholthof *et al*., 1995). Subsequently the sites of replication of many tombusvirus species, as well as the roles of their replication proteins, have been characterized (Jonczyk *et al*., 2007). These subcellular sites often vary between virus species, but the localization signals can be altered to create an alternative yet authentic site of replication (Jonczyk *et al*., 2007). What advantages might be conferred at each variable subcellular site of replication for viral replication is currently unknown since the exact role of a membrane surface in RNA replication complexes has yet to be elucidated (Jonczyk *et al*., 2007).

The two replication proteins of *Carnation Italian ringspot virus* (CIRV) are targeted to the mitochondria by an import-independent signal-anchor mechanism relying on two transmembrane segments and multiple recognition signals present in the N-terminal of the accessory protein, p36 (Weber-Lotfi *et al*., 2002). Both of these proteins localize to the mitochondria independently of one other (Pantaleo *et al*., 2004). Within

the overlapping sequence of the auxiliary and replicase proteins there exists an essential RNA-binding domain that has been shown to play different roles, depending on which protein it originates from, during virus replication (Panaviene *et al*., 2003). In CIRV, p36 has also been shown to play a role in stabilizing and targeting template RNA (by the presence of an RNA-binding motif) to the replication site in conjunction with the replicase, p92 (Pantaleo *et al*., 2004).

The auxiliary (p48) and replicase (p112) proteins of another tombusvirus, *Panicum mosaic virus* (PMV), are not only unusual in their substantially larger molecular weights compared to the analogous proteins of other tombusviruses, but have also been shown to play a crucial role in accumulation of newly synthesized virus (Batten *et al*., 2006). The potential targeting of these replication proteins to the peroxisome within infected host cells has also been suggested in PMV (Batten *et al*., 2006). In CNV, p33 interacts with other p33 molecules in yeast cells and facilitates the formation of multimolecular complexes. These complexes include p33, p92, viral RNA and host factors which are then targeted to the site of replication – namely the peroxisomal membrane (Panavas *et al*., 2005a; reviewed by Nagy and Pogany, 2006). This emphasises the essential role of these nonstructural proteins in the replication lifecycle of these viruses as well as in other virus families (Chen and Ahlquist, 2000).

The discoveries on the use of two replication proteins and their roles and interactions during the vRNA replication process cumulatively represent one of the few models published to date that describes, not only the viral factors involved in a ‘pre-targeted’ replication complex, but also the ratios of accessory to replicase protein involved (Panavas *et al*., 2005). One could assume, based on the literature, that the major driving force shuttling replication proteins and/or pre-formed complexes to a subcellular membrane is the (often ambiguous) targeting signal present on one or more of the replication proteins.

This conclusion is supported by the results obtained by Jonczyk *et al*. (2007), who replicated TBSV in a yeast strain with the *PEX3* and *PEX19*

genes deleted from its genome. The products of these two genes are essential for peroxisome biogenesis in yeast (Jonczyk *et al*., 2007). Replication of TBSV in the peroxisome-minus yeast strain suggested that TBSV replicons were targeted to and were functional at the endoplasmic reticulum (E.R.) at a fidelity and activity level comparable to that in wild type yeast (Jonczyk *et al*., 2007).

These replication proteins serve as attractive niches of interest in the investigation of virus replication, host cell-virus interactions and possibly targets in the popular field of silent interference (si) RNA studies. In order to understand more about virus replication and the evolution of this process between related virus families, one should understand the roles that replication proteins and the replication event itself plays on the lifecycle of viruses. These processes are often coupled to one another and are tightly regulated by a variety of factors during the lifecycle of the virus. A family which is well studied in almost all aspects of its life cycle is the *Picornaviridae*. Insight into the processes used by members of this family in the replication, host cell interactions and genome processing ability are summarized as a means of supplementing and diversifying this review on (+)ssRNA virus replication and genome processing.

1.6. THE *PICORNAVIRIDAE*

The *Picornaviridae*, which are among the smallest (+)ssRNA viruses known, comprise one of the largest and most important families of medically and agriculturally significant viruses (Fauquet *et al*., 2005). Many picornaviruses cause disease in humans and animals. PV, the prototypic picornavirus, has caused severe poliomyelitis in the past and brought financial burden to many developing countries attempting to eradicate the disease (Bedard and Semler, 2004). Human rhinovirus causes one of the most widespread viral diseases, the common cold, for which there is still no cure or cost-effective prevention (Bedard and Semler, 2004). *Foot-and Mouth Disease virus* (FMDV), a highly

contagious apthovirus, is the most economically important disease of domesticated cloven hoof animals causing regular outbreaks and loss of substantial income (Ferrer-Orta *et al*., 2004). The *Picornaviridae* are divided into seven genera based on similarities in virion physical properties, RNA sequence and genome organization (Bedard and Semler, 2004). The seven genera are enteroviruses, rhinoviruses, cardioviruses, apthoviruses, hepatoviruses, parechoviruses and teschoviruses (Fauquet *et al*., 2005).

Surprisingly for such a large family of animal viruses, this family has been shown to infect relatively few species of insects and lately, and most of the few documented members that infect insects are now classified in distinct taxonomic clusters outside the family *Picornaviridae* (Christian *et al*., 1999). In 2000, *Cricket Paralysis virus* (CrPV), *Infectious Flacherie virus* of silkworms and other invertebrates were classified as unassigned members of the *Picornaviridae*, but have since been grouped into the new *Dicistroviridae* family made up of the one genus, *Cripavirus* (Masoumi *et al*., 2003;) and the floating genus, *Inflavirus*, which has yet to be grouped in a particular family (Mayo, 2002; Fauquet *et al*., 2005). To add to the uncertainty of the phylogenetics of these picorna-like viruses, it has been discovered that CrPV, has a wide range of insect hosts in the wild, and is most likely responsible for the etiology of many viral diseases associated with a variety of insect species exhibiting similar symptoms (Fauquet *et al*., 2005).

More and more picorna-like viruses have been discovered in recent times and appear to make up a (structurally, genomically and pathogenically) diverse group of viruses (Wu *et al*., 2002; Fujiyuki *et* al., 2004; Fauquet *et al*., 2005). IFV lethally infects the ficus transparent wing moth, *Perina nuda*, and has been the subject of many classification studies based on its unusual genome arrangement, since the monocistronic genome is considered more similar to mammalian picornaviruses (Isawa *et al.,* 1998). The lethal disease, termed flacherie, has since been demonstrated to be the result of a mixed infection consisting of the dicistrovirus, *P. nuda*

picorna-like virus (PnPV) and the baculovirus, *P. nuda* nucleopolyhedrovirus (PenuNPV) (Wang *et al*., 1998; 1999).

Phylogenetic analysis of the insect virus *Ectropis obliqua* picorna-like virus has suggested that it is most closely related to PnPV and bioinformatic analysis of the RdRp alone showed that these two viruses along with Sacbrood virus (SBV) and IFV, should actually be classified in a group distinct from the dicistroviruses (Wang *et al*., 2004). Many such examples of phylogenetic ambiguity have arisen from analyses of these picorna-like viruses grouped in the *Dicistroviridae* and Inflaviruses and consequently classification of the *Dicistroviridae* is still ongoing (Fauquet *et al*., 2005).

1.6.1. Genome organization

Picornaviruses contain a 7.0 - 8.5 kb (+)ssRNA genome (Bedard and Semler, 2004; Fauquet *et al*., 2005). The genomic RNA is polyadenylated at the 3`-terminus and has a genome-linked protein (VPg) situated at the 5`-terminus, as opposed to the 7-methyl guanosine cap structure often found on the genome of RNA viruses (Bedard and Semler, 2004). The PV genome (7433 nt) is translated into a polypeptide of 2207 amino acids. This polyprotein is characteristically cleaved at three key positions followed by further cleavage to form 11 stable proteins (Figures 1.7 and 1.8) (Moore *et al*., 1985) by 2 types of viral proteinases, the 3C and 2A proteinases. The number of proteinases involved in cleavage of the polyprotein precursor is dependent on the genus of the virus (Bedard and Semler, 2004).

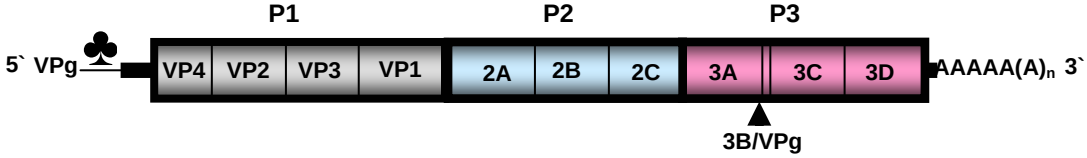

Figure 1.7. Genome organization of a model picornavirus, PV. The ultimate protein products of the proteolytic processing of P1 - P3 are colour coded according to their precursor origin. VPg= genome-linked protein, ♣ = clover leaf RNA structure. UTRs are represented by thick black horizontal lines (Adapted from Bedard and Semler, 2004).

The polyprotein is initially cleaved into three precursor proteins simply labelled P1, P2 and P3 (Figure 1.7). These three precursor proteins are cleaved into four, three and four end products respectively (Fauquet *et al*., 2005) with the intermediates playing vital roles in the virus lifecycles before they are processed further (Aminev *et al*., 2003; Marcotte *et al*., 2007). In members of the Apthovirus and Cardiovirus genera, a leader protein (L) exhibiting proteinase activity is translated from the very 5`-end of the polyprotein and is cleaved off the neighbouring P1 early in the polyprotein processing cascade (Strebel and Beck, 1986). This protein serves to shut off host cell translation in infected cells (Devaney *et al*., 1988). Precursor P1 is the capsid intermediate and further cleavage results in the production of the three or four capsid proteins. Cleavage products of P2 and P3 function in the shutting off of host cells translation, RNA replication (RNA-dependent RNA polymerase), protein processing, nuclear targeting of the viral genome, host cell membrane rearrangement (to increase permeability), the formation of virus-induced cytoplasmic vesicles and immune response interference (Bedard and Semler, 2004, Lee *et al*., 2006). These proteins and cleavage intermediates will be discussed further, with particular reference to the role they play in genome replication and translation of the viral RNA.

1.6.2. Polypeptide processing

The post-translational processing of the picornavirus polyprotein is performed by one or more of the viral-encoded proteinases (reviewed in Bedard and Semler, 2004). The first cleavage event of several picornaviruses is the cleavage at the junction of P1 and P2 precursors proteins, mediated by proteinase 2A, resulting in the production of the P2-P3 precursor and the P1 region of the polyprotein. The autocatalytic cleavage of the L protein from the N-terminus of the polyprotein varies in the order of processing from different protein intermediates between the genera of the *Picornaviridae*, yet its cleavage is universally mediated by the 3C proteinase (Bedard and Semler, 2004). The remaining secondary

cleavage events within the viral protein precursors are carried out by the 3C viral proteinases (and its precursor, 3CD) (Bedard and Semler, 2004). Once 3C has self-cleaved it is responsible for the cleavage of the P2 and P3 proteins, as well as the cleavage of several host cell proteins involved in particularly mammalian cell transcription (Bedard and Semler, 2004). A summary of this process and the function of the resultant cleavage products are represented in Figure 1.8.

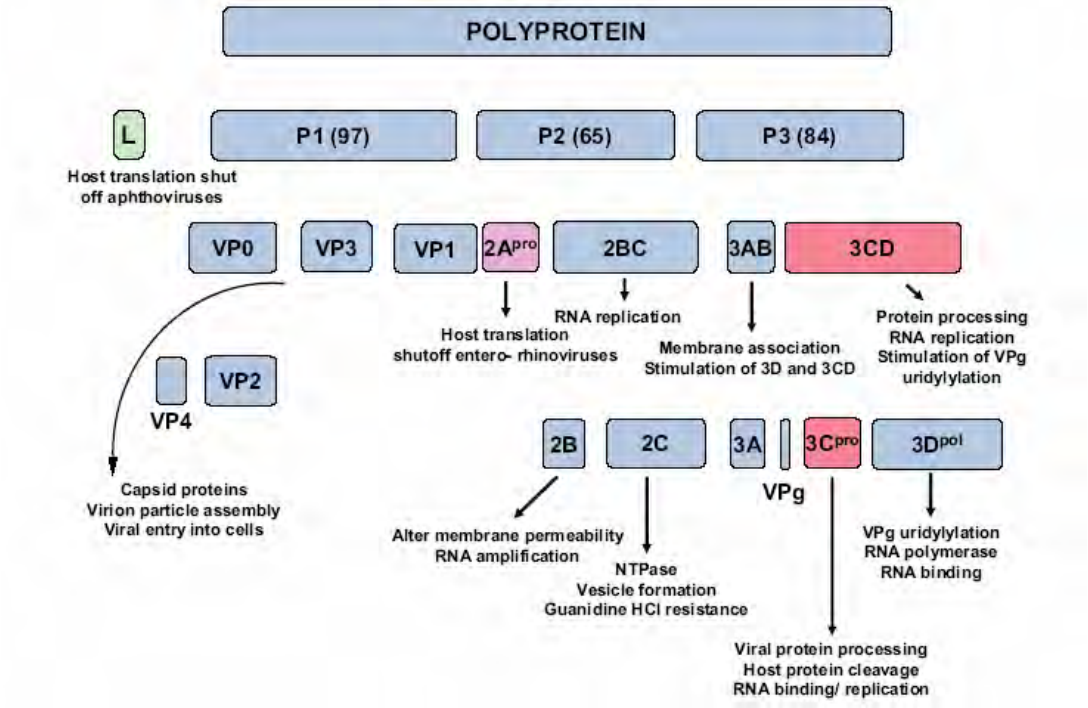

Figure 1.8. Picornavirus polyprotein processing cascade. The functions of mature and precursor proteins are highlighted and the proteinase responsible for each step shown (Taken from Bedard and Semler, 2004).

During these cleavage events, a mature virion is formed from pentameric intermediates of capsid protein and finally matures into an infectious icosahedral particle (Rossmann and Tao, 1999). Interestingly, picorna-like insect viruses have a completely different gene arrangement compared to mammalian picornaviruses (Rossmann and Tao, 1999). The four structural proteins are typically positioned at the 3` terminus of the genome and form dimer intermediates during maturation of the virus particles (Rossmann and Tao, 1999).

1.6.3. Replication and translation of vRNA

In picornaviruses the RdRp, named $3D^{pol}$, is required for positive and negative strand RNA synthesis. During viral replication, $3D^{pol}$ catalyzes chain elongation during viral synthesis and production of the protein primer, VPg-pU-pU, which is necessary for the initiation of viral RNA replication in a process described as VPg uridylylation (Lee *et al*., 1977). This is followed by the production of intermediate negative sense viral RNA, whereby the 3`-poly A tract acts as an initiation site. The ratio of positive to negative sense RNA is almost always asymmetrical and tightly controlled during the replication process (Agol *et al*., 1999). As is the case with other positive sense RNA genomes, the negative sense RNA in turn acts a template for the production of positive sense RNA, which is later packaged into virions (Bedard and Semler, 2004).

Picornavirus genome replication is characterized as being primer-dependent; in other words dependant on the uridylylated protein (VPg) linked to the 5`-end that initiates translation (Lee *et al*., 1977). The VPg is cleaved off during the early stages of infection by an unknown host cellular protein. This in turn uncovers the highly structured 5`-UTR that contains the IRES necessary for translation initiation (Bedard and Semler, 2004). This element also allows CAP- and eIF-4G-dependent modes of viral RNA replication while shutting off the host transcription processes (Aminev *et al*., 2003). The IRES element of a picornavirus is a relatively long (300-400 nt) and highly structured region of the RNA genome exhibiting a complex secondary and tertiary structure (Masoumi *et al*., 2003).

The IRES elements of picornaviruses have been proposed to contain multiple protein binding domains, which act as RNA chaperones and regulate IRES activity via the modification of cellular proteins by virus-encoded proteins (Belsham and Sonenberg, 2000). It is unlikely that translation initiation in picornaviruses occurs by other means, such as the scanning mechanism or read-through stop signals (none of which have been discovered in this family), due to the presence of this cap-

independent internal initiation of protein synthesis (Belsham and Sonenberg, 2000).

Early in infection, the viral mRNAs have to compete with the host, not so much for ribosomes, but for the limited pool of eukaryotic initiation factors (eIFs) that mediate the recruitment of ribosomes to viral mRNAs. The host protein, eIF4G, a component of eIF4F, is cleaved by the viral leader protein or 2A-like protease (depending on the virus), and used to form a pre-initiation complex together with $tRNA^{Met}$, 40 S ribosome subunits and the translation initiation factors eIF2 and eIF3 (Belsham and Sonenberg, 2000). Picornavirus IRES elements bind intact or truncated forms of eIF-4F which aid in the recruitment of 40 S ribosomal subunits allowing initiation and elongation of translation of the viral genome to occur (Bedard and Semler, 2004; Sarnow, 2003). At least two types of IRES structures have been characterized in the picornaviruses, and the type of IRES will determine the host cell factors involved in the translation of the viral genome (Bedard and Semler, 2004). Subsets of these factors have been shown to be involved in translation initiation, translation fidelity and quantity of the resultant viral translation product (Bedard and Semler, 2004).

1.6.4. Location and mechanisms of replication complexes

Much of the replicative processing of PV has been unravelled because of the once medical significance of this particular picornavirus and, more recently, other members of the family such as FMDV (Ferrer-Orta *et al*., 2004; 2006). In general, picornaviruses have been shown to localize their replication complexes on the membranes of the Golgi body and E.R. i.e. derived from components of the cellular anterograde membrane traffic system (Rust *et al*., 2001).

Once 3ABCD has been processed into 3AB by the 3C protease, it can functionally direct itself and the associated 3D to the cytoplasmic face of the ER, the location of the replication complex (Netherton *et al*., 2008).

2BC and 2C are also components of this complex and are targeted to membranes by individual hydrophobic domains (Netherton *et al*., 2008). With the number of viral proteins, precursor molecules and other RNA and/or protein interactions that occur during the lifecycle of a picornavirus, it is not surprising that the replication complex has yet to be fully dissected and described.

The active genome replication of PV induces the rearrangement of intracellular membranes into characteristic vesicles which then assemble into an RNA replication complex (Egger *et al*., 2000). Characterization of these vesicles revealed that ER membranes are initially used, followed by the incorporation of Golgi bodies into the complexes later in the infection process (Egger *et al*., 2000). In addition, replicating RNA is required for the formation of a functional PV replication complex (Egger *et al*., 2000), which supports previous findings that the picornavirus genome must first be translated before it can be replicated (Novak *et al*., 1994). In PV-infected HeLa cells, de-capsidation is rapid and it is suspected that the viral genome, and not the entire capsid, is transported to the perinuclear regions for translation (Egger and Bienz, 2005). Studies on this phenomenon in HeLa cells have revealed a series of molecular shuttling within PV-infected cells. The viral genome undergoes extensive viral translation at the E.R within viral-induced vesicles (Egger and Bienz, 2005). All newly synthesized proteins and RNA not involved in the replication complexes are then rapidly shuttled towards the nucleus, following the anterograde membrane pathway to their final location:- the Golgi complex within the microtubule-organizing centre (MOTC), where particle assembly presumably occurs (Egger and Bienz, 2005).

FMDV replication occurs next to the nucleus and close to the Golgi complex, but the membranes from this organelle are not part of the replication complex (Knox *et al*., 2005), suggesting that FMDV uses another source of intracellular membranes in order to anchor its replication complex. Alternatively, certain organelle-specific markers are extruded from these adapted membranes during the membrane rearrangement by

FMDV, explaining the lack of cross-reactivity with Golgi-specific markers and antibodies (Knox *et al*., 2005). Research conducted by Moffat *et al*. (2005) reiterated the invasiveness of subcellular organization by the picornaviruses, a factor which is often the leading cause of cell death and cytopathology during infection (Egger *et al*., 2000). FMDV precursor protein 2BC was shown to block the expression of cellular proteins on the cell surface, effectively disrupting the secretory pathway as a possible immune evasion strategy (Moffat *et al*., 2005), furthermore highlighting the multifunctional nature and importance of precursor proteins in the picornavirus lifecycle.

Another picornavirus, *Encephalomyocarditis virus* (EMCV), a member of the genus Cardiovirus, employs a very different localization (Aminev *et al*., 2003). A distinct nuclear localization signal (NLS) present on the N-terminal end of the $3D^{POL}$ is active, directing the RNA replication of the genome to the host nucleus (Aminev *et al*., 2003). Additionally, it was found that the proteins, 2A and precursor, 3BCD, inhibited cellular mRNA transcription but not rRNA transcription (which is required for functional ribosome formation) and that the precursor, 3BCD, was localized to the nucleus before undergoing further proteolytic processing (Aminev *et al*., 2003).

Many small insect virus species have picornavirus-like biophysical properties (Moore *et al*., 1985), but their relationships with established members of the picorna-like virus lineage have only recently begun to be established (Wu *et al*., 2002; Wang *et al*., 2004). The comparative biophysical and replication studies of viruses from different families and lineages is leading to a better understanding of the (co- and divergent) evolution of insect viruses and their respective modes of replication within the host cell, which are often highly adaptable to various changes within the host cell environment.

1.7. PROBLEM STATEMENT

The discovery of *Providence virus* provided an opportunity to study tetravirus replication *in vivo.* Virtually nothing is known about the replication of tetraviruses nor the host and auxiliary viral factors involved in their life cycle. The results achieved described in the accompanying chapters serve to shed light on some of these processes, further bridging the ever-narrowing knowledge gap between plant, animal and insect ssRNA viruses as well as the general viral replicative mechanisms within the context of a (+)ssRNA virus.

1.8. AIM AND BROAD OBJECTIVES

The overall aim of this work was to establish experimental systems for studying the replication biology of PrV. The broad objectives were as follows:

1. Bioinformatic analysis of the PrV genome and its translation products
2. Creation of antiserum towards PrV translation products
3. *In vivo* and *in vitro* analysis of PrV translation products
4. Analysis of the distribution of PrV replication proteins and viral RNA in PrV-infected MG8 cells

CHAPTER 2

Bioinformatic analysis of the PrV genome and its translation products

2.1. INTRODUCTION

In 2003, Pringle *et al*. reported the discovery of a new betatetravirus, PrV in a *Helicoverpa zea* midgut cell line. A partial cDNA of 704 nt was later annotated by Pringle to be part of a non-structural ORF of 2748 nt, the replicase gene following submission of the entire capsid protein precursor ORF sequence. N-terminal sequencing of the major and minor capsid proteins of PrV revealed that, like TaV, the capsid protein precursor (p81) is cleaved twice: it is processed at its N-terminus by a 2A-like processing signal, and at the C-terminus by the conserved autoproteolytic event that occurs during particle maturation (Pringle *et al*., 2003).

A year later, the cDNA sequence of what is believed to be the entire PrV genome (6155 nt) had been determined. Three 2A-like processing signals were discovered in the genome: one in the N-terminal region of a large, unidentified ORF named p130 and the remaining two at the N-terminus of p81, the capsid protein precursor. The discovery of an in-frame stop codon, proposed to be a read-through stop signal in the replicase coding sequence in all the cDNA clones, and an RdRp domain showing conservation with (+)ssRNA plant virus RdRps were some of the unpublished preliminary results obtained by Pringle during this period (F.M Pringle and L.A. Ball pers. comm., 2005). The aim of the research described in this chapter was to confirm Pringle's initial analysis of the PrV cDNA sequence and to use various bioinformatic tools to further annotate the genome, secondary RNA and predicted protein structures and translation modification signals associated with the individual ORFs.

2.2. METHODS

The nucleotide sequence of the p104 ORF of PrV was confirmed using the dideoxy sequencing method. The procedure used and substitutions in the p104 sequence are described in Appendix 3. The uniform resource identifiers (URIs) utilised in this Chapter and dates of when the webpages and links hosted at these locations were accessed, are listed in Appendix 4. The accession numbers of the nucleotide and amino acid sequences are referred to in text and listed in Appendix 5.

Using the online National Centre for Biotechnology Information (NCBI) database, the three main ORFs of the PrV genome, namely the capsid precursor ORF (p81), the putative replicase ORF (p104) and p130 were submitted as both nucleotide and amino acid sequences and compared to other non-redundant sequences deposited in the database. This was performed using the Basic Local Alignment Search Tool (BLAST) program (Altschul *et al*., 1997). In the alignment of nucleotide and protein sequences, BLASTn and BLASTp programs, which align the query with 'somewhat similar sequences' in the NCBI database, were each used respectively in the analysis of individual ORFs.

Identification of amino acid residues that are potentially phosphorylated *in vivo* was conducted using the NetPhos V2 program, and by limiting the search to identify only candidate tyrosine, phenlyalanine and tryptophan residues in the p40/p104 coding sequence, with probability scores of over 0.95 considered statistically relevant. The online tool, TargetP was used to predict the subcellular localization of PrV peptides by identifying and assigning the probability of the following functional sequences: chloroplast transit peptides, mitochondrial targeting peptides and secretory pathway signal peptides.

The online programs, ClustalW v. 1.82 (Thompson *et al*., 1994) and phylogenetic tree viewer TreeView v. 1.6.6 9Page, 1996), were used to align amino acid sequences, measure phylogenetic distances and produce

phylograms based on this input sequence. ClustalW is a multiple sequence alignment algorithm which generates enough information in the identities, similarities and divergence existing between sequences such that a phylogenetic relationship can be calculated (Pearson and Lipman, 1988; Mackey *et al*., 2002). The graphic interface version of ClustalW, ClustalX (Larkin *et al.,* 1997), was used to predict bootstrap confidence values of the alignment files generated by ClustalW. A total of 1000 bootstrap trials were conducted on each submitted alignment file and 111 used as a random number generator seed.

In the phylogenetic analyses of capsid precursor proteins, the gap extension penalty was set at 2 and gap creation penalty set at 8, whereas in the analyses of RdRp sequences the gap penalty was increased to 2.5 and the gap opening penalty decreased to 5. The phylogram was generated from the aligned sequences using the neighbour-joining method (Saitou and Nei, 1987). The Java-based program TreeView was used to produce an output phylogram based on sequence and parameters set in ClustalW. The algorhythm, MFOLD and its server-supported graphic interface, were used to predict secondary structures of both the 5` and 3` UTRs of PrV. The MFold program designed by Michael Zuker (Matthews *et al*., 1999; Zuker, 2003) is a widely used bioinformatic tool for predicting the most likely, energy-efficient folding of both DNA and RNA. MFOLD has successfully been used in the accurate prediction of both viral (see Collier *et al*., 2002 and Baird *et al*., 2006 for examples) and cellular IRES structures because of the algorithms' high rate of accuracy in determining the minimum free energy constraints and thus secondary structures (Matthews *et al*., 1999). This is achieved by the incorporation of, on average, 73% of all known base pair combinations in each of the most likely theoretical structures that are computed and not by an averaging of general base pairing combinations (Matthews *et al*., 1999). Nucleic acid sequences longer than 750 nt contain up to 86 % of known base pair combinations since the longer the sequence, the higher the level of confidence in structure prediction (Matthews *et al*., 1999). The single strandedness of a molecule is also predicted on a statistical level in all

significant predicted structures, thus confirming the overall probability of the existence of such structures existing in nucleic acids and also the lack of any structured region under varying free energy values (Zuker, 2003).

Furthermore, the minimum free energy difference (MFED) of each structure can easily be determined by subtracting the free energy calculated on randomly shuffled input sequence from the free energy predicted for the structure of the original sequence. Values (expressed as a percentile of the original sequence) in the range of 15-20% are considered to represent accurate, sequence-specific secondary structures and not a randomised prediction on the natural folding properties of the RNA (Simmonds *et al*., 2008). In this study, the most likely secondary structures were predicted using the recommended default parameter settings and used in further interpretations.

2.3. RESULTS AND DISCUSSION

2.3.1. Genome organization and theoretical protein products

The PrV cDNA sequence was obtained from Dr. Fiona M. Pringle and Prof. L. Andrew Ball (University of Alabama, Department of Microbiology, Birmingham, Alabama, USA). Analysis of potential ORFs encoded by the PrV genome suggested the presence of 3 major ORFs, the first two overlapping extensively (Figure 2.1). Nearest to the 5`-end of the genome, extending from nts 45 – 3708, is a large theoretical ORF of 3663 nt in length, the translation product of which has a theoretical molecular weight of 130 kDa, hereafter referred to as p130 in text. The translational start of p130 could be at either of two downstream methionine (AUG) codons at nts 139 - 141 or nts 354 - 356, which like the first AUG, are both in good context for translation. This was based on a comparison of its Kozak sequence with the invertebrate consensus sequence (Caverner, 1987; Kozak, 1987). The first AUG positioned at nts 45 - 47 has a Kozak sequence of acaaAUGg, whereas the second AUG positioned at nts 354 –

356 has a Kozak sequence of ggagAUGa. The invertebrate consensus sequence from nts -4 to +4 is reported as: ca(a/g)(c/g)AUG(g/a) (Caverner, 1987).

Figure 2.1. Genome organization of PrV. ORF and additional processing features of the PrV genome as originally annotated by Pringle *et al*. (F.M. Pringle, pers. comm.) included the p130, the replicase and capsid precursor ORF. The three 2A-like processing sites are indicated by stars. The in-frame stop codon within the replicase is indicated by the circle containing the strike-through. The vertical dashed line in the capsid precursor ORF indicates the site of autoproteolytic cleavage of the capsid precursor protein. The bent arrow indicates the start of subgenomic RNA transcription.

Positioned at nts 1027 to 3775 is a second ORF (2748 nt) encoding a protein with a predicted molecular weight of 104 kDa in a +1 frame relative to p130 (Figure 2.1). The p104 ORF (hereafter be referred to as p104 or the replicase coding sequence in text) has a Kozak sequence of uuguAUGu, which is not in a particularly good context for initiation of translation, compared to the optimal invertebrate Kozak consensus sequence stated above. An in-frame UAG stop codon is present in the p104 ORF at position 2104 - 2107 nt, repeatedly detected during sequence analysis of different cDNA clones, confirming the presence of the UAG sequence (F.M. Pringle, pers. comm., 2005). Termination of translation at this stop codon would produce a protein with a predicted molecular weight of 40 kDa (p40). Lastly, the capsid precursor ORF begins just 4 nt after the end of the p104 ORF and is 2265 in length (3779 - 6044 nt) with a Kozak sequence of gaucAUGc. This ORF is translated from a subgenomic copy of RNA into an 81 kDa capsid protein precursor protein (Pringle *et al*., 2003).

2.3.2. The read-through stop signal in p104

The analysis of the reported in-frame stop codon, UAG, and the 6 nucleotides downstream, revealed that PrV contains a read-through stop signal with the highly prevalent, conserved sequence of: UAG CAA CUA typical of a group 1 set of sequences that constitute 80 % of the known stop signals (Harrel *et al*., 2002). Group 1 is composed entirely of (+)ssRNA plant viruses belonging to the *Tobamovirus*, *Benyivirus* and *Pomovirus* genera and one unassigned potex-like virus, Botrytis virus F (Harrel *et al*., 2002). Although the PrV stop signal is best categorized within group 1, it has a divergent C nucleotide at the start of the second codon (nt +4), making it the only member in group I with this sequence despite its otherwise perfect homology to the consensus sequence (UAG CAA UU/CA) (Harrel *et al*., 2002). The effects of substituting this nucleotide with a C tested both *in vivo* and *vitro* did not affect the activity levels of the read-through stop signals in this particular group (Skuzeski *et al*., 1991; Harrel *et al*., 2002). If this read-through stop signal in p104 was indeed functional, large amounts of a 40 kDa (p40) protein would be translated as well as small quantities (approximately one tenth to one twentieth relative to p40) of a read-through product with a predicted molecular weight of 104 kDa (p104).

2.3.3.a. Bioinformatic analysis of the PrV replicase ORF

Based on the prediction that PrV replicates in the cytoplasm, a search was conducted for the consensus sequence of the enzymatic domains of a RNA helicase and guanylyl- and/or methyltransferase and proteases activities in the p104 amino acid sequence. Only an RdRp catalytic domain in the C-terminus of this coding sequence was positively identified. Consensus sequences of the additional enzymatic domains normally associated with (+)ssRNA viruses that replicate in the host cytoplasm are apparently absent. The carmo-like viruses of supergroup II are the only group of (+)ssRNA viruses that do not contain other additional motifs besides the essential RdRp (reviewed in van der Heijden and Bol, 2002). By inference

PrV could be loosely classified within the carmo-like supergroup based on the lack of additional enzymatic domains within the replicase coding sequence. These motifs are also absent in the replicases of two other betatetraviruses, namely TaV and EeV (Gorbalenya *et al*., 2002*)*.

2.3.3.b. Identification of the RdRp domain and its conserved motifs in p104

Initial analysis of the PrV genome involved confirming the sequence of the replicase ORF and presence of motifs A - C encompassing the core RdRp domain as originally stipulated by Poch *et al*. (1989), as well as the semi-conserved motif G (Koonin, 1991). The core motifs were found to have the same permuted arrangement as discovered in two other members of the genus *Betatetavirus*, as well as in certain dsRNA viruses (Gorbalenya *et al*., 2002). The core motif arrangement in the amino acid sequence of p104, the replicase, is as follows: C (721 - 723) - A (755 - 761) - B (774 - 783), with the numbers in brackets corresponding to the amino acid position within the p104 coding sequence (Figure 2.2). The D motif, if present in the RdRp domain of PrV, lacks a conserved sequence, and it has been reported that this motif is not always readily recognizable in RNA viruses (Gorbalenya *et al*., 2002).

The putative G motif [$_{547}$SDPETNLPIVA<u>K</u>G$_{559}$] is not highly conserved and overlaps with the second of the two K residues (underlined in the sequence above) that usually flank this particular motif (Figure 2.2) (Gorbalenya *et al*., 2002). The putative motif found in the PrV RdRp domain, although divergent, is likely a G motif, since conservation of amino acids downstream of this stretch of amino acids is very similar to the two type members of the birnaviruses as annotated by Gorbalenya *et al*. (2002). Motif G in PrV is also situated an additional 35 amino acids upstream of the C motif, relative to the same motif positioning in the Birnaviruses, TaV and EeV (Figure 2.2) (Gorbalenaya *et al*., 2002), possibly leading to a new structural variant of the palm, fingers and thumb model of the RdRp protein

in its tertiary structure state or an evolutionary footprint, where the motif is not functional within the RdRp domain.

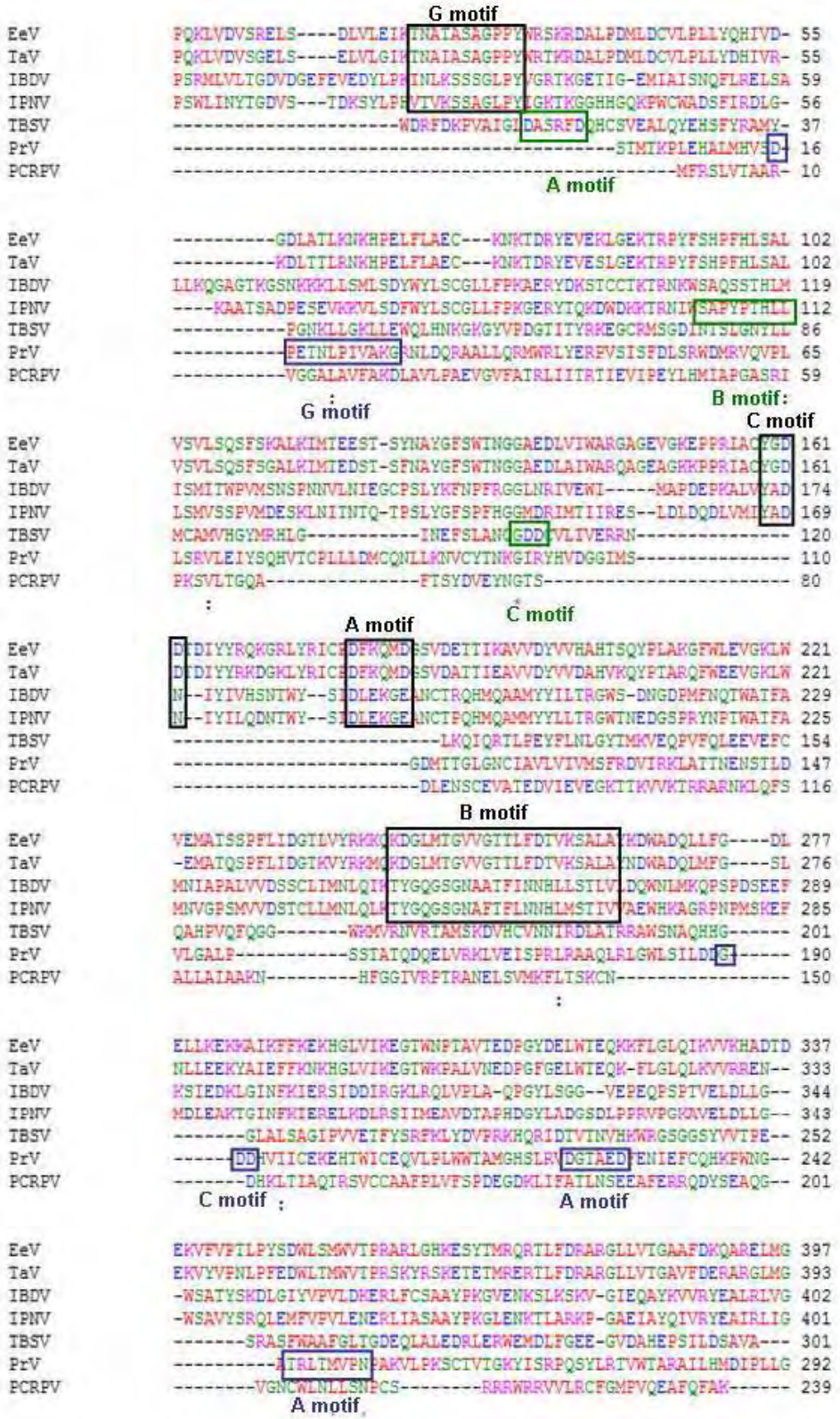

Figure 2.2. The permuted motif arrangement in the PrV RdRp aligned to that of the tetraviruses EeV and TaV, the birnaviruses, IBDV and IPNV and the tombusviruses, TBSV and PCRPV. The A – C and G motifs of EeV, TaV, IBDV and IPNV are indicated by the black boxes and the same motifs in PrV, using blue boxes. Motifs A-C of TBSV have been boxed in green. The same motifs in PCRPV have not been labelled due to high sequence divergence in this virus and little literature to support the identification of the RdRp domain in this putative replicase ORF. "*" refers to an exact match of amino acids in the alignment column; ":" refers to conserved substitutions in the column and "." refers to semi-conserved substitutions in the column. IPNV: *Infectious Pancreatic Necrosis virus*. IBDV: *Infectious Bursal Disease virus*. The alignment was generated, under default parameters, using the ClustalW program.

2.3.4. Evolutionary relationships between PrV and other tetraviruses

BLASTn analysis of the nucleotide sequence of p130, p40 and full length p104 revealed no significant matches in the database. The capsid precursor protein was analysed by submitting amino acid sequence starting at Pro123 (the proline residue after the second 2A-like processing site) of the p81 ORF and extending to the final amino acid of the coding sequence. BLASTp analysis of the capsid precursor sequence showed significant sequence homology with the other tetravirus capsid precursor ORFs, as was expected based on the results achieved by Pringle *et al*. (2003). Pringle *et al*. (2003) analysed the same stretch of protein using the GAP program within the GCG software package and a comparison of results in this study using BLASTp vs. this earlier publication remain virtually unchanged. The E-values and percentages calculated in this study are updated by the inclusion of the EeV and DpTV capsid precursor sequences. Percentage identities recorded in descending order of significant homologies were: DpTV (38 %); HaSV (36 %); NωV (35 %); TaV (26 %), NβV and EeV (25 %). The accession numbers of the proteins used in this analysis are listed in Appendix 5.

Although PrV is considered a betatetravirus based on its moncistronic genome, subgenomic copy of the capsid precursor and outer capsid particle appearance (Fauquet *et al*., 2005), its capsid precursor sequence has a higher percentage identity with the omegatetraviruses versus the betatetravirsuses (36.3 % vs. 25.3 %; n=3). The implications of this ambiguous capsid precursor sequence conservation between species of

two genera are discussed later in light of phylogenetic analysis of these particular sequences.

Analysis of the amino acid sequences of p130 and p40 using BLASTp revealed no significant similarity to any non-redundant amino acid sequence in the database. Similar analysis of p104 revealed significant homologies with the replicase proteins of members of the (+)ssRNA virus families, *Umbraviridae* and *Tombusviridae* (Table 2.1).

Table 2.1. Comparison of the PrV p104 replicase amino acid sequence with the ten most homologous individual sequences (either replicases or putative replicases, and regions thereof) in the database

Protein (and accession number)	Virus species (and family)	Percentage identity	E-value	Score (Bits)
Replicase (YP05295.1)	*Pelargonium chlorotic ring pattern virus* (tombusvirus)	26	3×10^{-30}	138
RdRp (NP619714.1)	*Groundnut rosette virus* (umbravirus)	24	3×10^{-30}	137
RdRp (AAB81639.1)	*Carrot mottle mimic virus* (umbravirus)	26	4×10^{-30}	137
89K protein BAG13035.1	*Melon necrotic spot virus* (tombusvirus)	26	4×10^{-28}	130
P111 (ACA57840.1)	*Maize chlorotic mottle virus* (tombusvirus)	26	1×10^{-27}	129
p84 -putative RdRp (NP619751.1	*Oat chlorotic stunt virus* (tombusvirus)	26	2×10^{-27}	129
RdRp (ABV66257.1)	*Opium poppy mosaic virus* (umbravirus)	25	2×10^{-27}	128
RdRp (NP733848.1)	*Tobacco bushy top virus* (umbravirus)	24	5×10^{-27}	127
Putative replicase (NP619718.1)	*Maize chlorotic mottle vein virus* (tombusvirus)	26	5×10^{-27}	127
RdRp (AAU20330.1)	*Pea enation mosaic virus-2* (tombusvirus)	26	1×10^{-26}	126

Interestingly, the first 100 significant identities originate from the sequence of replicase proteins belonging to (+)ssRNA plant viruses. In particular, all of these plant viruses are grouped into the families *Umbraviridae* and *Tombusviridae*. Regions of high conservation exist between the PrV replicase and these plant virus replicases, and the significance of the conservation is substantial with up to 500 amino acids of sequences

aligned to a significant percentage of similarity. The RdRp region of p104 showed significant conservation with the RdRp domains of all of the ten most similar sequences. To measure the similarities of the amino acid sequence of only the RdRp domain of p104 PrV replicase with protein sequences in the database, the 292 amino acids of the RdRp sequence, located from amino acids 533 - 827 of the replicase protein was submitted to BLASTp for analysis (Table 2.2).

Table 2.2. Comparison of the PrV RdRp domain amino acid sequence with the five most homologous individual sequences in the database

Protein (and accession number)	Virus species (and family)	Percentage identity	E-value	Score (Bits)
p89 (BAB82513.1)	*Melon necrotic spot virus* (tombusvirus)	28	2×10^{-19}	100
RdRp (AAB81639.1)	*Carrot mottle mimic virus* (umbravirus)	25	3×10^{-19}	99.8
87 kDa replicase protein (YP052925.1)	*Pelargonium chlorotic ring pattern virus* (tombusvirus)	27	1×10^{-18}	97.6
p84 (NP862835.1)	*Pea stem necrosis virus* (tombusvirus)	27	3×10^{-18}	96.3
p87 (ACF35783.1)	*Pelargonium line pattern virus* (tombusvirus)	26	5×10^{-18}	95.5

As observed in the BLASTp analysis of the whole protein, significant conservation between the PrV RdRp domain and sequences in the database was only observed with the RdRp domains and replicases (or predicted replicases) of other (+)ssRNA plant viruses belonging to the *Umbraviridae* and *Tombusviridae*. Reshuffling of the PrV RdRp permuted sequence into a classical linear A-B-C motif arrangement, achieved by relocating amino acids 755 - 802 (motifs A and B) between amino acid 705 and 706 (before the C motif), did not increase the bias of the amino acid homology between this RdRp domain and other insect or animal virus RdRp/replicase sequences in the database. Moreover, this reshuffling only decreased the score and E-value of the alignments with other plant virus RdRp domains, which is significant since the plant viruses showing greatest homology with the PrV RdRp domain do not exhibit motif

permutation as observed here in PrV and a select number of animal dsRNA viruses (Gorbalenya *et al*., 2002).

2.3.5. Phylogenetic analysis of the RdRp domain and capsid precursor protein versus other related (+)ssRNA viruses

The genetic diversity of PrV is so far very evident, based on factors such as a significant replicase sequence identity with both the full length replicases and RdRp domains of (+)ssRNA plant viruses, in particular the *Tombus*- and *Umbraviridae* (Tables 2.1 and 2.2) versus the minimal conservation observed with the homologous gene sequences of other beta- or omegatetraviruses. To further characterise this apparent divergence, phylogenetic analyses of both the capsid precursor and replicase gene of PrV versus the analogous sequence in other tetraviruses and a proposed phylogenetic outlier, *Cricket paralysis virus* (CrPV), were conducted using ClustalW, and phylograms generated from these aligned sequences using the neighbour-joining tree builder, TreeView (Figure 2.3).

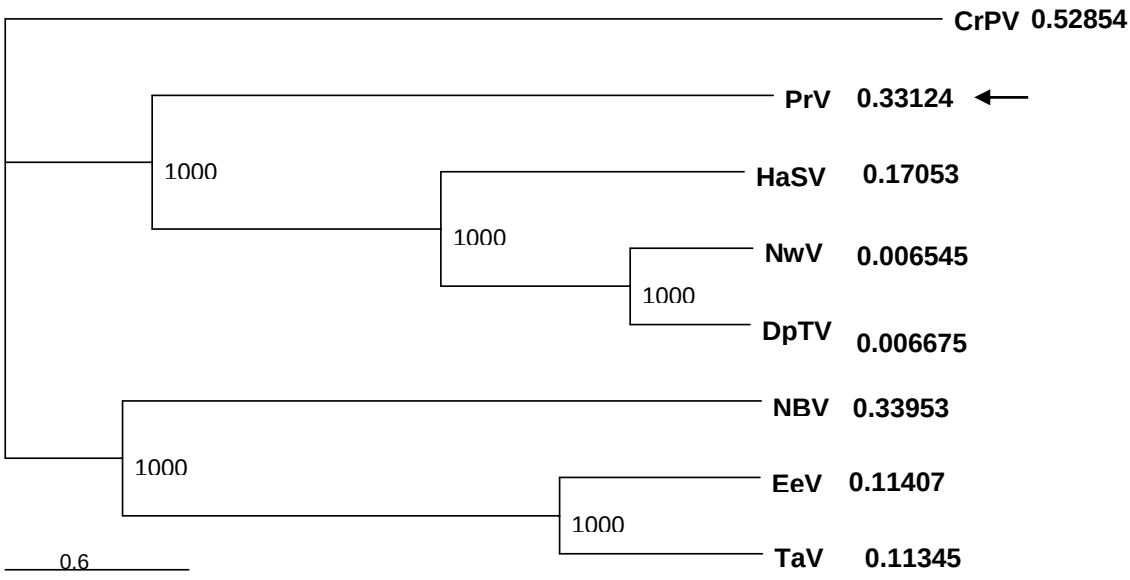

Figure 2.3. Phylogram of aligned tetravirus capsid precursor sequences. The capsid precursor amino acid sequences of NωV, NβV, DpTV, PrV, TaV, EeV and structural protein of the genetic outlier, CrPV, were aligned in ClustalW and a phylogram generated by the progam, TreeView using a gap extension penalty of 2 and gap creation penalty of 8. The phylogenetic distance of each sequence is represented next to the sequence name and the distance scale in the bottom left corner of the phylogram. Bootstrap values are indicated at the nodes of each branch The positon of PrV is indicated by the arrow.

The position of these viruses based on their phylogentic relationships suggested that PrV probably evolved as an ancestral offshoot from the omegatetraviruses. It can be argued that the two betatetraviruses, EeV and TaV, co-evolved from a common unidentified ancestor, and that the two omegatetraviruses, NωV and DpTV, have only recently diverged from a comon ancestor 'younger' in phylogenetic terms than both HaSV and PrV. This data suggests that capsid precursor gene of PrV has not evolved as recently from an ancestor or intermediate relative to NωV and DpTV. The bootstrap values generated from this particular phylogram do support this hypothesis, with the branch arrangement in Figure 2.3, the most probable genetic relationship between the capsid sequences used in this study under the stringencies stated. This is reflected by the 100 % (1000 out of 1000) confidence levels in the clustering predicted in this phylogram. This data only represents a phylogenetic model of the evolutionary relationships of these six tetraviruses and does not take into account the divergence of other sequence information in these candidates or the possible relationships with other unsequenced tetraviruses.

The RdRp domain, rather than the entire diverse replicase sequence of PrV, was analysed in a comparable manner. Using the minimal RdRp sequence defined by Gorbalenya *et al*. (2002) to align and analyse TaV, EeV, IBDV and IPNV RdRp motifs, the 292 amino acids of RdRp domain located from amino acids 533 - 825 of PrV p104 were used to analyse the phylogenetic relationship of PrV relative to these other RdRp sequences. The RdRp sequences of HaSV and NβV were added to the alignment as well as an evolutionary outlier, CrPV. The equivalent region of the replicase gene of *Pelargonium chlorotic ring pattern virus* (PCRPV) was also used in this phylogenetic analysis because of its known conservation with the RdRp sequence of PrV.

Since the other RdRp sequences used in the analysis are not permuted, the A, B and C motifs were identified and an equal number of amino acids included upstream and downstream of the core domain resulting in a ~400 amino acid RdRp-containing sequence. The RdRp motif sequences of

IBDV, IPNV, TaV, EeV, PrV, PCRPV, HaSV, NβV and CrPV were aligned using ClustalW and the phylogram generated using TreeView and the neighbour-joining (N-J) method (Figure 2.4). The phylogenetic relationship between the nine RdRp amino acid sequences revealed a complex relationship between the sequences, as well as some surprising implications on the evolutionary branches and distances between these viruses (Figure 2.4).

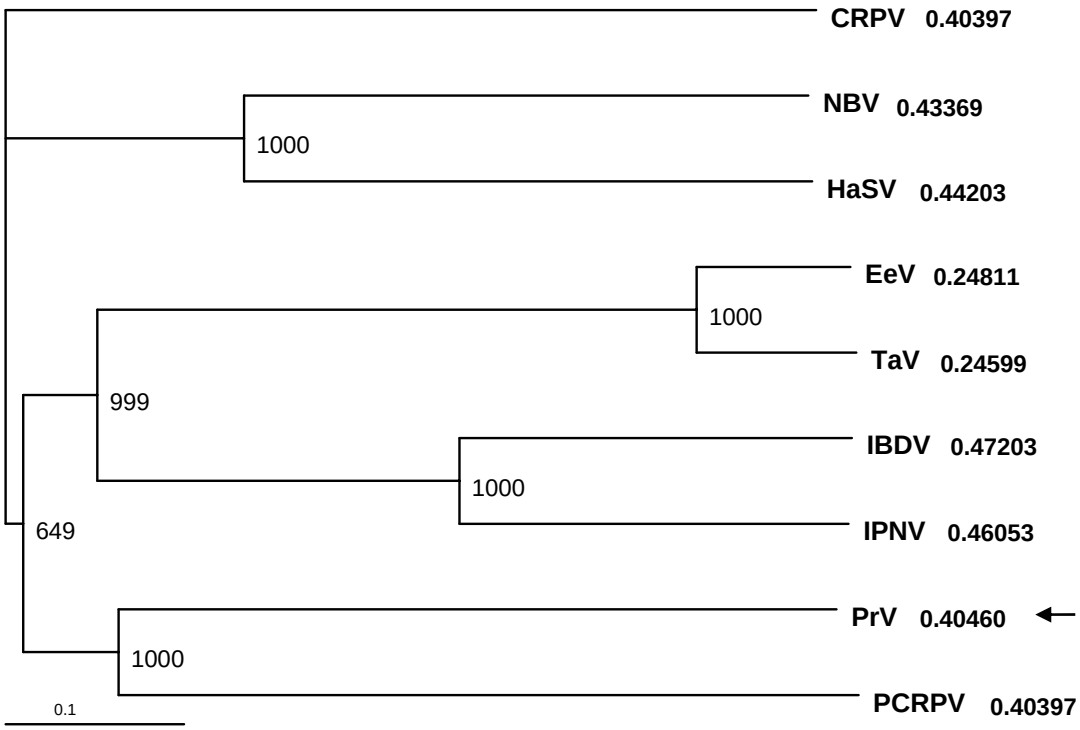

Figure 2.4. Phylogram of aligned RdRp sequences of tetraviruses, two birnaviruses, CrPV and the related PCRPV. The RdRp amino acid sequences of HaSV, NβV, PrV, TaV, EeV, IPNV, IBDV, PCRPV and CrPV were aligned in ClustalW and a phylogam generated using the program, TreeView. The phylogenetic distance of each sequence is represented next to the sequence name and the distance scale in the bottom left corner of the phylogram. Bootsrap values are indicated at the nodes of each branch. The positon of PrV is indicated by the arrow.

Based on this data, it appears that CrPV does indeed serve as a phylogenetic outlier as its position falls outside all smaller branches of the phylogram. As previously predicted (Gorbalenya *et al*., 2002), the RdRp domains of TaV and EeV are phylogenetically closely related and have possibly diverged at the same time from a common ancestor. When aligned with these particular nine sequences, IBDV and IPNV are also closely related, although IBDV is more similar to EeV and TaV, than IPNV

is. Variations with the input data such as gap extensions and gap opening penalties, sequence length and number of sequences submitted produced very similar phylograms, adding confidence to these predictions and adding weight to the suggestion that IBDV might have an evolutionary intermediate and is 'younger' in evolutionary terms than the closely related IPNV (Figure 2.4). Bootstrap values generated for this phylogram support the clustering of each clade with high percentiles of confidence. More information is required to accurately cluster the PrV and PCRPV vs. the permuted RdRp predecessor (which gave rise to EeV, TaV, IBDB and IPNV) since the confidence levels at this branching is only 65 %. The lower bootstrap support value for this particular branch is most likely a result of the inclusion of a linear RdRp motif sequence showing high overall sequence similarity with PrV at the cost of large gap creations in the sequence. This outcome could be the result of the specific sequences chosen in this alignment, since it does not consider the greater genetic variation in the p104 coding sequence and the rest of the PrV genome, nor the possibility of convergent evolution, with particular reference to the PrV/PCRPV relationship. Importantly, this analysis shows that the RdRp regions of PrV and PCRPV are very similar and, relative to the other sequences submitted in this alignment, are 'older' in evolutionary terms than the more recently diverged TaV and EeV RdRp sequences.

2.3.6. Post-translational modification and processing of PrV proteins

Autoproteolytic cleavage

The mechanism of autoproteolytic cleavage of the omegatetravirus capsid protein precursor into the large and small capsid proteins has been well characterized (Agrawal and Johnson, 1995; Gordon *et al*., 1999; Canady *et al*., 2000; Taylor *et al.*, 2002). The event occurs late in the virus lifecycle resulting in the maturation of the provirion into the mature virion following a decrease in the pH of the surrounding environment (Canady *et al*., 2000; Tomasicchio *et al*., 2007). In PrV, the same autoproteolytic event occurs

within the p68 capsid protein to produce the large (p60) and small (p8) capsid proteins, between the amino acids Asn679 and Phe680, resulting in the 60 and 8 kDa major and minor PrV capsid proteins, respectively (Pringle *et al*., 2003). Attempts to increase the relatively low percentage of autoproteolytic cleavage of the PrV capsid precursor protein in purified virus preparations by decreasing the pH of buffered particles did not result any increased cleavage activity (data not shown).

2A-like co-translational processing

Three 2A-like 'cleavage' sites were identified in the PrV genome (F.M. Pringle and L.A. Ball, unpublished data; Luke *et al*., 2008). The first is between amino acids 123 - 131 of the N-terminal of p130; and the second and third between amino acids 46 - 53 and 115 – 122, respectively, of the N-terminal of p81. If these sites were active, the predicted p130 theoretical protein would be processed into 17 and 113 kDa peptides, and if both processing sites within p81 were active the capsid protein precursor would be cleaved into 5.8 (p6), 7.5 (p8) and 68 kDa (p68) proteins. The identification of these co-translational processing signals was based on the minimum conserved signal of the six amino acids DxExNPG↓P, as initially defined in the picornavirus FMDV by Ryan & Drew (1994). Table 2.3 lists the sequence of the 2A-like processing sites found in PrV and other tetraviruses, and their predicted activities *in vitro* (Donnelly *et al.,* 2001b; Luke *et al*., 2008).

Table 2.3. 2A-like processing sequences in tetraviruses

Virus and ORF	Annotation	Sequence	Activity
PrV – p130	PrV-2A$_1$	$_{123}$GDVESNPGP$_{132}$	~99 %
PrV – p81	PrV-2A$_2$	$_{43}$GDIEKNPGP$_{52}$	~94 %
PrV – p81	PrV-2A$_3$	$_{115}$GDVEKNPGP$_{124}$	~99 %
TaV – p82	TaV-2A	$_{148}$GDVEENPGP$_{157}$	~99 %
EeV – p83	EeV-2A	$_{142}$GDVEENPGP$_{151}$	~99 %

Key: Highly conserved nucleotides within the sequence are highlighted. The amino acid number in the coding sequence is noted in subscript for the first and last amino acids in the consensus sequence.

PEST-like sequences and rapidly degraded proteins

The N-terminal 2A-like cleavage products of the capsid protein precursor, p81 (p6 and p8), contain a high percentage of Pro, Glu, Ser and Thr amino acids in a consolidated stretch, referred to as a PEST-like sequence (Rogers *et al.*, 1986). This entire non-structural coding sequence begins at the translational start of p81 and ends at the second 2A-like processing site (Gly123) found within the capsid protein precursor sequence. This peptide stretch (p13) has a predicted molecular weight of 13.4 kDa because the first 2A-like cleavage site (PrV-$2A_2$) in p81 has been shown to exhibit a lowered activity *in vitro* (Luke *et al*., 2008). Thus instead of the p6, p8 and p68 proteins being processed, only a p13 and p68 protein are predicted to result from a single processing event arising from the PrV-$2A_3$ sequence. p13 is located in a similar position on the PrV genome as are the PEST-like sequences found at the N-termini of the capsid precursor of certain tetraviruses (Hanzlik *et al*., 1995; Pringle *et al*., 1999).

The N-terminal sections of the capsid protein precursor of HaSV and PrV, (and to a lesser extent, TaV and NωV) have high percentages of Pro, Glu, Ser and Thr amino acids relative to other amino acids in this peptide stretch (Hanzlik *et al*., 1995; Pringle *et al*., 1999; 2003). It is proposed that peptides containing large numbers of these amino acids are labile and often rapidly degraded by proteolysis (Rogers *et al*., 1986; Reichsteiner and Rogers, 1996). The online program, PESTfind (based on criteria set by Rogers *et al*., 1986) calculates the likelihood of “PEST-like proteins”. Analyses of the PEST-like sequences of DpTV, HaSV, NωV, TaV and PrV assigned PrV the highest score the five sequences (15.87 obtained for amino acids 23 – 34 of p13), and predicted a potential PEST-like sequence from amino acids 25 - 48 of this same peptide stretch, based on its score and amino acid composition. Analyses of analogous peptide stretches of the capsid protein precursor of NωV, and the p17 peptides of HaSV, DpTV and TaV, predicted potential PEST-like sequences in HaSV and DpTV in similar positions to that found in PrV. These PEST-like sequences attained scores of 7.16 and 15.14, respectively while analyses of NωV and TaV

were predicted to contain poor PEST-like sequences, attaining scores of just 1.66 and 0.95, respectively.

According to comparisons of potential PEST-like sequences in other proteins, the scores achieved by the peptide stretches analysed in HaSV, and in particular, PrV, are considered to be significant. A score of 15.87 places the PrV peptide as one of the highest ranking PEST-containing sequences relative to other labile eukaryotic proteins, such as the transcription factor c-fos, calpain large subunits and the heavy subunit of the neurofilament to name a few (Carillo *et al*., 1996). The functions of these PEST-like peptides in tetravirus biology are unknown, although Hanzlik *et al*. (1995) speculated that the PEST-like protein in HaSV (p17) might be a movement protein and play a regulatory role in the lifecycle of tetraviruses.

Phosphorylation of replication proteins

Based on other reports of functionally essential phosphorylation of tombusvirus auxiliary and replicase proteins *in vivo*, (Shapka *et al*., 2005; Jakubiec *et al*., 2006; Jakubiec and Jupin, 2007) the NetPhos program was used to predict regions within the amino acid sequence of the p104 ORF showing a high probability (> 0.95) of being phosphorylated in the eukaryotic cellular environment.

The analysis of possible sites of phosphorylation of the p40/p104 amino acid coding sequence showed two potential regions of phosphorylation – one region on the very N-terminal of the p40/p104 coding sequence, and the other situated over the RdRp domain of p104 (Figure 2.5). It was predicted that a total of thirteen serine residues and one threonine residue in p104 are phosphorylated *in vivo*. p40 alone contained seven of these serine residues predicted to be phosphorylated at probabilities greater than 0.95, with four of these residues concentrated in a region spanning amino acids 6 - 33 of the p40/p104 coding sequence (Figure 2.5).

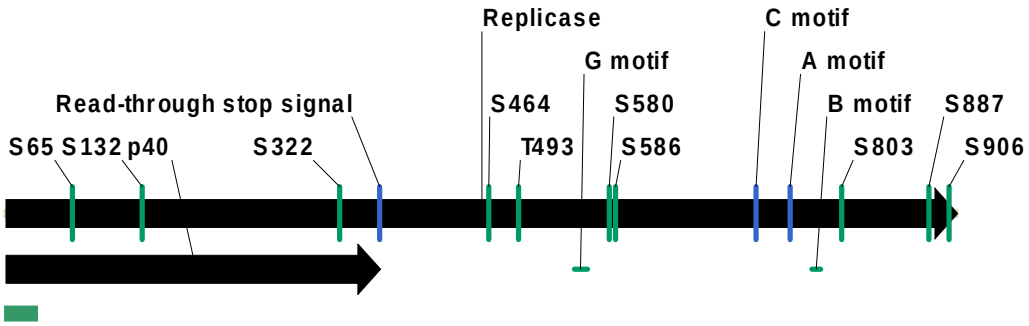

Figure 2.5. Predicted phosphorylation sites for amino acids of the p40/p104 coding sequence. Each site of phosphorylation is annotated by a single amino acid code letter and its amino acid number relative to the p104 coding sequence. The region at the N-terminus of p40 containing seven predicted phosphorylated serine residues is indicated by the horizontal green bar. The A - C and G motifs of the RdRp are labelled on the p104 coding sequence as a reference point.

2.3.7. Identification of potential nuclear import/export signals, RNA binding motifs and subcellular localization and targeting signals on p104

As discussed in Section 1.3.4 of the literature survey, all known RNA-dependent RNA polymerases require host subcellular membranes to anchor themselves and efficiently replicate the viral genome (Buck, 1996; Ahlquist, 2002). Based on this prediction, various online bioinformatic programs and algorithms were used to predict and search for potential sites required for subcellular localization and any potential membrane or organelle-targeting sequences. Analysis of the predicted p104 amino acid sequence using TargetP v. 1.1 revealed no mitochondrial, chlorotic or secretory pathway signal peptide sequences and no predicted subcellular localization. The same result was obtained using the program WoLFpSORT, which predicts and scores potential eukaryotic subcellular localization of peptides and proteins. The program MitoProt II v1.101, which predicts mitochondrial targeting, was used to analyse the p104 coding sequence for the prediction of targeting of this coding sequence, but no such targeting sequences were found.

Next NetNES (v1.1), a program that identifies nuclear export signals was used in the analysis of p104. No significant stretch of amino acids could be

found which would make up a probable nuclear export signal. In case an artefact localization signal, such as a plant peroxisome targeting sequence (PTS) was present, the program 'The PTS1 predictor' was used. No PTS signal and thus, no targeting of p104 to the peroxisome, was predicted using this program.

The presence of potential RNA binding motifs within p104 was investigated. All tombusviruses have an RNA binding motif, termed the RPR motif that is highly conserved amongst members of the family (Panviene *et al*., 2003). The sequence of this arginine/proline rich motif is $Rx^3Px^2Rx^2Rx^2Rx^3P$ (Rajendran and Nagy, 2003) with the number of arginine residues (stated in superscript) in this sequence found to be optimal for tombusvirus genome replication (Panaviene *et al*., 2003). No such RNA binding domain was found in the predicted amino acid sequence of p104. This adds motivation for the experimental testing for any internal replication elements such as RNA or auxiliary recognition elements PrV p104.

2.3.8. Identification of transmembrane helices in p104 and p130

Although initial analyses of p40 and p104 did not reveal any obvious membrane or organelle targeting signal, the prediction of transmembrane helices within the secondary structure of these protein sequences could indirectly signify a subcellular localization as well reinforcing the theory that these replication-related proteins are both intimately associated with a (subcellular) membrane.

Two different online programs were used in the analysis of p40/p104 for the presence of transmembrane helices. The first program, SOSUI engine v. 1.11 (Hirokawa *et al*., 1998), predicted two such transmembrane helices. The first transmembrane helix present in p40/p104 is 23 amino acids in length and spans from amino acids 225 - 247 of p40/p104 (Figure 2.6). The second such transmembrane sequence is 22 amino acids in length and lies in the C-terminal half of the read-through portion of p104 spanning amino acids 644 - 665 of p104. The prediction implies that p40 is

membrane-associated and that the N-terminal half of the protein is cytoplasmic, while the C-terminal section should be located on the other side of the theoretical membrane and is hydrophilic (Figure 2.6).

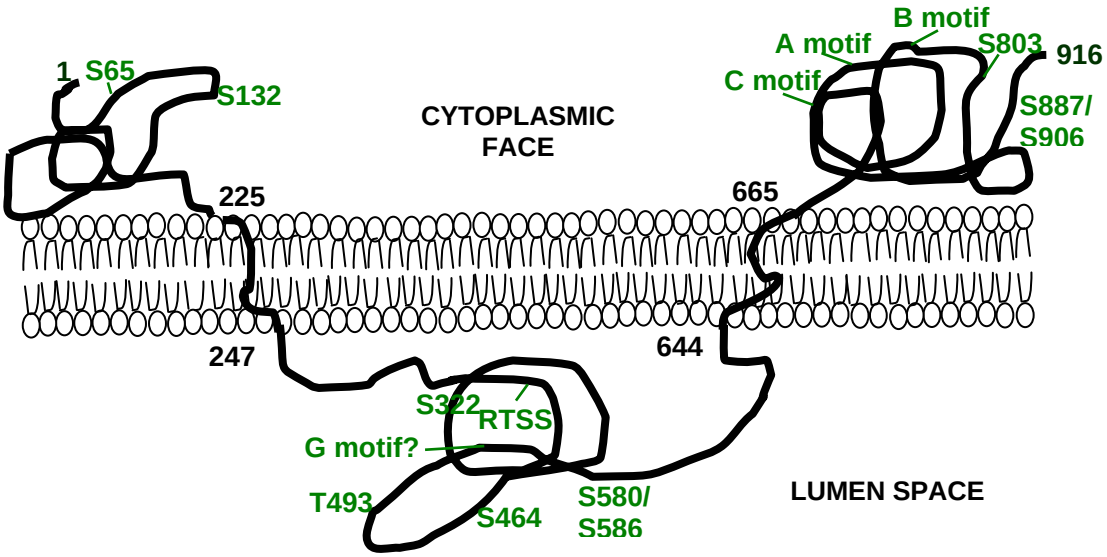

Figure 2.6. Position of the predicted hydrophilic and hydrophobic domains of p40/p104 in relation to probable sites of phosphorylation. The diagram shows the relative position of stretches of the p40/p104 protein relative to a theoretical lipid bilayer. The amino acid stretch of each transmembrane helix is annotated at either side of the lipid bilayer and the length of each stretch is represented to scale. The A - C and G motifs of the RdRp are labelled on the protein sequence as points of reference for the RdRp. The annotation of the phosphorylation sites remains unchanged from Figure 2.5. RTSS refers to the Read Through Stop Signal.

If the N-terminus of p104 membrane-spanning sequences are indeed functional, then this region of p104 will theoretically be membrane-associated. An internal hydrophobic loop from amino acids 248 to 643 is predicted to be situated on the other side of the theoretical membrane and the remaining 259 amino acids (656 - 915 of p104) resurface into the cytoplasm (Figure 2.6). This suggests that a C-terminal third of p40 is membrane-associated while more than 40 % of p104 is membrane associated, although the RdRp domain at the C-terminal is exposed to the cytoplasm as to be expected. The putative G motif would be positioned within the lumen of the theoretical organelle raising questions regarding the annotation of the G motif in this sequence since the motifs of an RdRp interact in close proximity with each other and presumably not across a

lipid bilayer. The predictions of these two transmembrane stretches within p40/p104 were confirmed using the program TMPred (Hoffman and Stoffel, 1993). The scores calculated for these two transmembrane stretches using the TMPred programme were 2068 (a.a 225-247) and 1638 (a.a. 644-655) where a score of over 500 is considered statically relevant. Analysis of p130 using both programs revealed that the protein is predicted to be completely hydrophilic and thus soluble in the cellular environment, containing no transmembrane helices in its amino acid sequence.

2.3.9. Secondary structure of the 5`- and 3`-UTR of the PrV genome

Secondary structure in vRNA has been shown to play essential, varying roles in the replication and recognition of replication-related proteins. Gordon *et al*. (1999) demonstrated that the 3`-UTR region is folded into a clover-leaf tRNA-like structure in both NβV and HaSV, and even went on to suggest that the sequences play a telomeric role and conserve critical information on the terminal sequence by protecting it from exonuclease activity. PrV lacks a tRNA-like cloverleaf secondary structure of its 3`-UTR and does not contain a poly-A sequence (F.M. Pringle and A. Ball pers. comm., 2005). To confirm this apparent lack of significant secondary structure formed by the PrV 3`-UTR, the folding of the nucleotide sequence of this region was analysed using the MFold program. The program was used to predict the secondary viral RNA structure of the 3`-UTR of PrV (nts 6043-6155). Using the default parameters suggested by the program creators found to produce optimal secondary RNA structures, the following structure (Figure 2.7) was predicted to be the most energy-efficient natural form of this region.

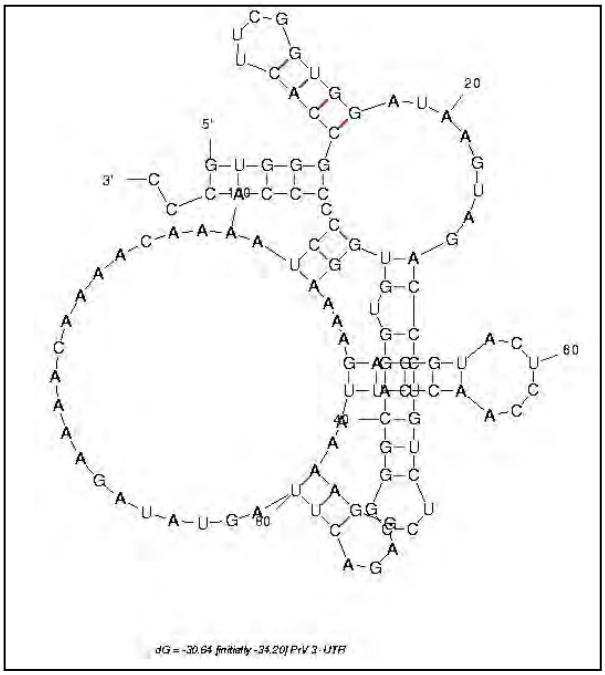

Figure 2.7. The predicted secondary structure of the PrV 3`-UTR. This structure was generated using the MFold program with parameters set at predicting a structure requiring the least amount of energy - a Gibbs free energy (dG) value of -30.64 kcal/mol.

The structure of the 3`-UTR predicted by MFold measured a degree of secondary structure in this particular region. However, these structures were not analogous to the tRNA-like conserved structures of the 3`-UTR present in both RNA1 and RNA2 of HaSV (Gordon *et al*., 1995; Hanzlik *et al*., 1995) or NβV (Gordon *et al.*, 1999). Furthermore the loop regions generated by MFold show no homology to the conserved sequences of HaSV and NωV originally by Gordon *et al*. (1995), nor with the later addition of supporting data within the same region of NβV (Gordon *et al*., 1999). The large loop structures shown in the bottom left (nts 80 – 121 of 3`-UTR sequence), and to a lesser extent, the top right (nts 5 – 116) are not small or structurally stable enough to fold into the tRNA-like structures detected in particular tetraviruses although the two smaller loops (nts 6 - 17 and 57 – 64) could be considered as hairpin-like structures (Figure 2.7).

The PrV genome is unusual in that the first functionally identifiable ORF (p104) starts more than 1000 nts from the start of the genome. Although the p130 ORF is predicted to be translated at the most 5` ORF, its presence and role has yet to confirmed in PrV. The first 1026 nts were

analysed using the MFold program, for secondary structures that might indicate the presence of an IRES. Again, the suggested default parameters were used and the overall secondary structure predicted in this region is displayed in Figure 2.8, panel A.

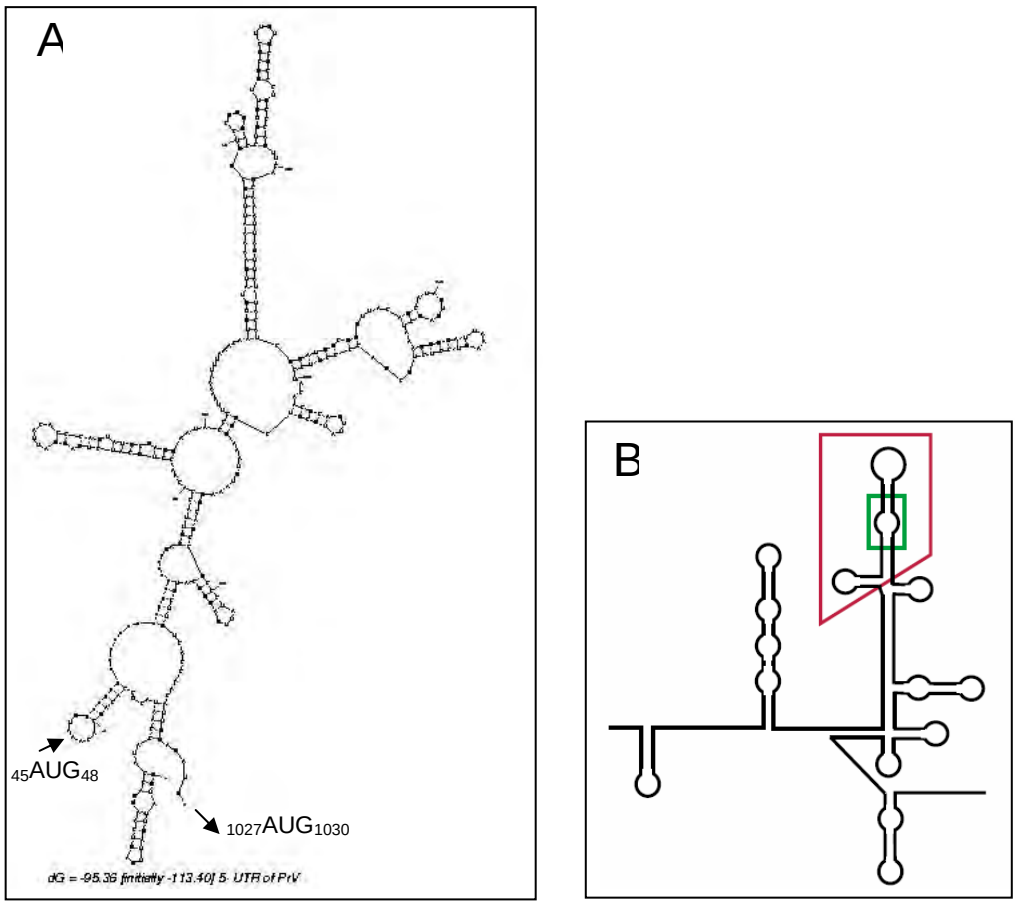

Figure 2.8. IRES secondary structures in the 5`-UTR of (+)ssRNA viruses. (A): Prediction of the secondary structure of the first 1026 nucleotides of the PrV genome This structure was predicted, using the MFold program with parameters set at a structure requiring the least amount of energy - a Gibbs free energy (dG) value of -95.36 kcal/mol. The p104 and p130 translational starts are indicated by arrows leading on from nt 1036 and 44, respectively. (B): A simplified diagram of the HCV IRES secondary structure (taken from Collier *et al.,* 2002). Sequence and secondary structure of the HCV IRES within nucleotides 1 - 362 of HCV genotype 1a. The number and orientation of these stem-loop structures are found in both HCV and Cholera swine fever virus (CSFV) and can be related to other viral IRES structures (Baird *et al.,* 2006). The large box shows the predicted binding site of eIF3, while the smaller rectangle is positioned over a conserved bulge structure that is essential for initiation of translation mediated by an IRES (Collier *et al*., 2002).

Again, this particular stretch of nucleotide sequence exhibited a degree of secondary structure. When compared to the secondary structures often associated with IRES such as the well-characterized IRES of Hepatitis C

Virus (HCV) (Figure 2.8, panel A vs. B), it is evident that this entire stretch of sequence does not contain the finer stem-loop structures and bulges associated with an IRES. Also, the basic pattern of stem-loop structures often modelled from viral IRES sequences is absent. The IRES sequences in members of the *Dicistroviridae* are divergent in both sequence and structure relative to other viral IRES sequences, although functionally important similarities do exist (Sasaki and Nakashima, 1999; reviewed in Sarnow, 2003). Using the MFold program, no obvious IRES-like secondary structure was predicted in the first 1026 nt of the PrV genome. This prediction, due to the length of the PrV 5`UTR has a theoretical level of confidence of approximately 86 %, based on all known base pairing combinations recognized by the MFOLD algorhythm. Further characterization of the viral translation products would need to be performed in order to indirectly analyse the mechanism of translation initiation in this region of the PrV genome and also increase the level of confidence in the prediction of the secondary structure in this region.

2.4. CONCLUSION

Bioinformatic analysis of PrV has uncovered new information regarding sequence homologies, protein processing and modification, RNA structures of the long pre-p104 5-` and 3`-UTR, predicted membrane-association and finally a preliminary phylogenetic analysis of the replication and capsid proteins. Analysis of the non-structural p130 ORF, yielded little information on its predicted function, potential subcellular localization and any post-translational modification. Due to the very short leader sequence before the proposed translational start of this ORF, no IRES signal is predicted to be present at the 5`-UTR. This is unsurprising since one of the two methionine residues with good invertebrate Kozak sequences (Caverner, 1987; Kozak, 1987), present at the start of the p130 coding sequence, are the most likely source of the translational start by the scanning ribosome mechanism.

The replicase (p40/p104)

The sequence of the predicted replicase protein, p104, showed no significant homology with sequences from the database and the RdRp motif displayed a permuted motif arrangement as characterized in TaV and EeV as well as two birnaviruses. This permutation was, in itself, unusual in that the G motif is not as highly conserved, relative to the consensus found in the other permuted domains. The location of this putatively identified motif is more than 163 amino acids upstream of the C motif which differs from the ~128 nt of the other analyzed tetraviruses with permuted RdRp domains, TaV and EeV. Also, the transmembrane helices prediction on p104 place this motif in the luminal cavity making unlikely that is motif is functional since this is dependent on proximal positioning to the other motifs of the RdRp. The significance in the overall tertiary structure and thus activity of the PrV replicase, modelled on the identification of the RdRp domains, remains to be shown and will be encompassed in future research of this protein. Based on the lack of any identifiable guanylyl/methyltransferase or helicase domains, and conservation of the RdRp amino acid sequence with (+)ssRNA viruses belonging to the carmo-like supergroup, the PrV RdRp should be categorized in the carmo-like virus supergroup.

In terms of the predicted subcellular localization of p40 and p104, no obvious subcellular targeting or import/export signal could be detected using available programs. Two transmembrane helices were predicted to occur within p104, one of which is located in p40, suggesting that both might be intrinsically anchored to a subcellular membrane. This characteristic has been experimentally proven for many tombusvirus auxiliary and replicase proteins (Rubino and Russo, 1998; Navarro *et al*., 2004). The predicted hydrophobic region of p104 potentially situated in the lumen of an organelle is quite large and it is surprising that p40 has only one transmembrane helix since two tombusviruses, *Cymbidium ringspot* (CymRSV) and *Carnation Italian ringspot viruses* (CIRV) have two helices (Rubino and Russo, 1998). In the case of CymRSV, only 7 kDa of the

auxiliary protein is anchored into the peroxisomal membrane (Navarro *et al*., 2004). The lack of any obvious subcellular targeting signal has also been found in the analyses of various tombusvirus replication proteins, in particular CymRSV and CIRV (Rubino and Russo, 1998).

Localization of the replication proteins of tombusviruses involves a multistep process with the interaction of the auxiliary proteins, replicase, viral RNA and host proteins before targeting of a pre-replication complex to a membrane (Rajendran and Nagy, 2006; reviewed by Nagy and Pogany, 2006). It is these auxiliary/replicase interaction domains which are also thought to play a role in targeting of Tombuvirus replication proteins to organelles, although a transmembrane sequence within the CNV p33 amino acid sequence is also essential for the targeting of p33 (yet not p92) to the site of replication (Panavas *et al*., 2005a). Due to the complexity of the assembly and targeting of the replication complex, one cannot rule out the multifunctional nature of these domains and thus the overall effect they may have on the functionality and therefore targeting of a pre-replication complex to the site of replication.

Although no RNA binding domains were detected within the p104 sequence, a number of amino acids were predicted to have a high probability of being phosphorylated, a well-characterized post-translational modification in replication-related viral proteins. Two regions within the p104 sequence were predicted to be phosphorylated – the very N-terminus of p40 and the RdRp domain of p104. Further experimental analysis of p40 needs to be performed in order to confirm the phosphorylation of p40 and possibly p104. The two are predicted to act together as replication proteins in the course of PrV infection, although differences in sequence motifs and post-translational modification do exist between p40/p104 and other plant virus replication proteins. It is also predicted that p104 is translated by the standard ribosomal scanning mechanism and not by an IRES signal located in the 5` region upstream of the p104 translational start.

The capsid precursor

Much of the sequence identity of the capsid precursor protein of PrV when compared to other tetraviral proteins remains unchanged from Pringle's initial assessment (Pringle *et al*., 2003). The discovery and sequencing of the DpTV capsid precursor protein does, however, alter these findings on a phylogenetic level. The capsid precursor protein of PrV shares more sequence identity with other omegatetraviruses and is phylogenetically most closely related to NωV and DpTV. Phylogenetically, PrV is positioned between the outlier, CrPV and the omegatetraviruses, ultimately showing more similarity to the omegatetraviruses than the betatetravirus, in terms of its capsid precursor protein sequence.

The analyses reported here have shed light on the potential function of many of the predicted proteins translated by PrV, revealing the interesting evolutionary history of the replicase (p104). These results also confirmed the greater sequence similarity that the capsid precursor protein has, relative to the analogous sequence of omegatetraviruses versus that of other betatetraviruses. In terms of its relatedness to the betatetraviruses, the genome organization of PrV, position and size of its RdRp domain as well as the presence of a third potential ORF suggest a substantial divergence from other members in this genus. The differences between PrV and other members of the betatetraviruses are evident with particular reference to the type member, NβV, as well as EeV and TaV, despite the presence of a permuted RdRp and 2A-like processing sequence at the N-termini of the capsid protein precursor ORFs of these two viruses (Figure 2.9).

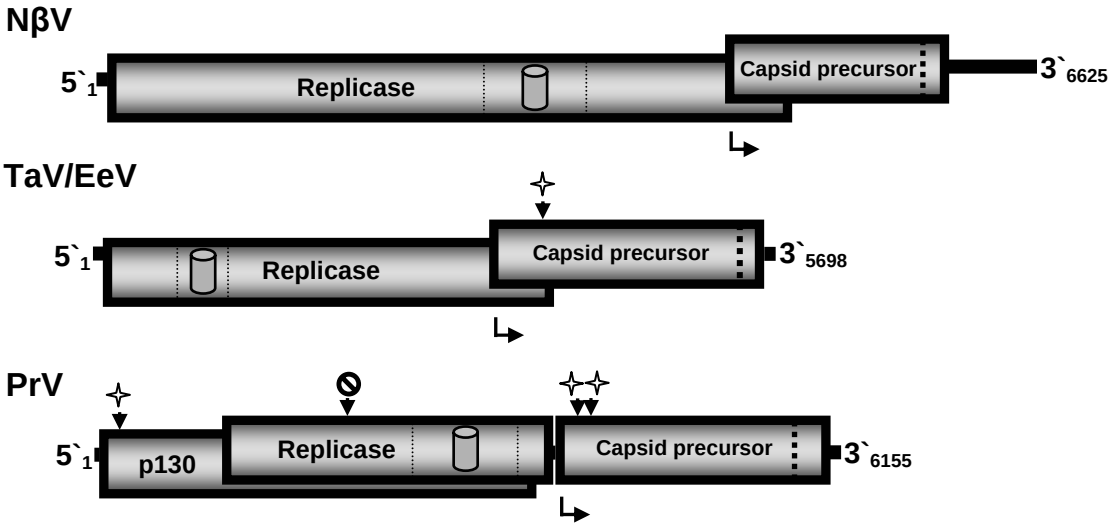

Figure 2.9. Genome organization of the betatetraviruses, NβV, TaV and EeV compared to PrV. The 2A-like processing sites are indicated by stars. The in-frame stop codon within the replicase ORF is indicated by the circle containing the strike-through. The boxed cylinder in the replicase ORF indicates the RdRp domain, drawn to scale. The dashed vertical line in the capsid precursor ORF indicates the site of autoproteolytic cleavage of the capsid precursor protein. The bent arrow indicates the start of subgenomic RNA transcription. The ORFs and UTRs are represented to scale.

The initial classification of PrV into the *Tetraviridae* was based on physical characteristics of the capsid, as well as the significant amino acid homology of the capsid precursor amino acid sequence with that of other tetraviruses (Pringle *et al*., 2003). These results add significant weight to the two hypotheses that PrV a) has a replication mechanism that utilises an auxiliary and full length replicase protein as observed in many (+)ssRNA plant virus families and b) is a unique tetravirus in terms of both sequence homology and replication mechanisms.

CHAPTER 3

Production of antibodies for the detection of PrV translation products

3.1. INTRODUCTION

Bioinformatic analysis of the PrV cDNA sequence identified the presence of three major ORFs in the PrV genome: (1) p130, a large ORF of unknown function, (2) p104, predicted to be the viral replicase based on the presence of a putative RdRp domain and (3) p81, the virus capsid protein precursor. The presence of a read-through stop signal in p104 and 2A-like processing sequences in p130 and p81, suggested that PrV might employ a unique strategy for regulating the expression of viral gene products. In addition, phylogenetic analysis of the replicase coding sequences raised the possibility that the replication of PrV might be different to that of other betatetraviruses and the *Tetraviridae* in general. The first objective of the research described in this thesis was to confirm the annotation of the PrV genome and to investigate the co-translational regulation of viral gene expression focussing on the viral replicase.

This chapter describes the production of anti-p104 polyclonal antiserum using peptides derived from the p104 coding sequence expressed in *Escherichia coli.*

3.2. MATERIALS AND METHODS

3.2.1. General recombinant methods and bacterial strains

Unless otherwise stated, all gene manipulation methods were performed according to Sambrook *et al.* (1989) and protein analysis as described by Laemmli (1970). Crude extracts and purified proteins were quantified by Bradford's assay (Bradford, 1976) to within 99% accuracy relative to a broad range standard linear curve of concentrations of bovine serum

albumin fraction V (BSA) and equimolar quantities of protein analysed by SDS-PAGE and Western analysis. Primer sequences used in the PCR amplification of sequences are listed in Appendix 1. The cloning strain of *Escherichia coli* (*E. coli*) DH5α (Hanahan, 1983) was used for plasmid propagation and strain *E. coli* BL21 (DE3) *(B F–ompT hsdS(rB–mB–) dcm+ Tetr gal (DE3) endA Hte*) (Stratagene) was used for the overexpression of the peptides. The integrity of all PCR amplification products was confirmed by DNA sequencing.

3.2.2. Construction of the 6xhis-p104$_{(644\text{-}915)}$ bacterial expression plasmid pCW21

The bacterial expression plasmid, pCW20 was created by ligating the double stranded oligonucleotide AI11 (Figure 3.1 C) into the pT7-7 derived vector pRD107R, which had been digested with *Xba* I and *Nde* I. pCW20 contained an inducible T7 RNA polymerase promoter, ribosome binding site (RBS) and a 6xhistidine sequence followed by a methionine residue (Figure 3.1 A). This plasmid also contained the coding sequence for the yeast UGA3 gene fused to the 3` end of the 6xhistidine tag.

Amino acids 644 - 915 of the p104 coding sequence were selected for overexpression in the *E. coli* BL21 (DE3) heterologous system. This peptide stretch located at the C-terminus of the p104 protein also contains most of the RdRp domain (amino acids 533 – 825). The coding sequence was amplified from the cDNA clone of the PrV genome (pFLM1) using the primers CW5 and CW6, which introduces a partial *Csp*451 at the 5` end and a *Hin*d III site at the 3` end of the PCR fragment, respectively. Due to the use of a Taq polymerase that created 5` single A nucleotide-overhangs in the amplification of a nucleotide sequence, the *Csp*451 restriction site was reconstituted following ligation of the amplicon into the vector pGEM-T Easy (Promega). The resulting plasmid, pCW19, was digested with *Csp*451 and *Hin*d III and ligated into the pCW20 vector backbone digested with *Csp*451 and *Hin*d III, to create the bacterial expression vector pCW21 (Figure 3.1 B).

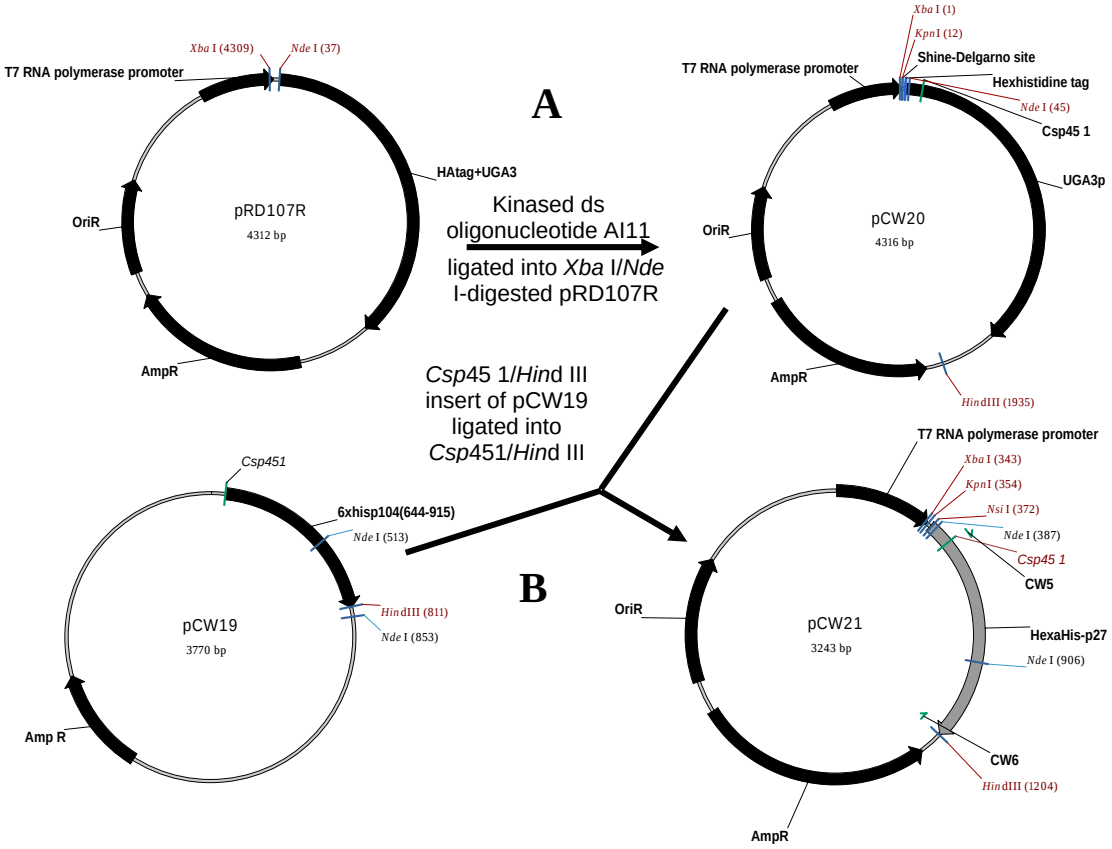

Figure 3.1. Two-step cloning strategy used in the construction of the expression vector, pCW21. Step A: Creation of pCW20. Step B: Creation of pCW21. *OriR*: origin of replication, AmpR*: β-lactamase*, UGA3p: yeast UGA3 coding sequence, p27: $p104_{(644-915)}$. C: Sequence and annotation of the ds oligonucleotide, AI11. RBS: ribosome binding site.

3.2.3. Overexpression and purification of the 6xhis-$p104_{(644-915)}$ peptide (p27)

pCW21 was transformed into chemically competent *E. coli* BL21 (DE3) cells. A volume of 2.5 ml of log phase culture was inoculated into 50 ml Luria-Bertani (LB) broth in the presence of ampicillin (125 µg/ml) and grown with agitation at 37 ºC to an $OD_{600\ nm}$ of 0.4. These cultures were induced with IPTG to a final concentration of 1 mM for 4 hours at 37 ºC, after which

the $OD_{600\ nm}$ of 1 ml of culture was recorded and the cells harvested by centrifugation at 10 000 g for 5 minutes at 4 ºC.

The cells were resuspended in 1 ml of ice-cold phosphate buffered saline (PBS: 140 mM NaCl; 2.7 mM KCl; 10 mM Na_2HPO_4; 1.8 mM KH_2PO_4) and sonicated using a hand-held 3 mm probe (Vibra Cell™) for 20 x 10 second bursts, at 20 % power at 4 ºC. Soluble and insoluble fractions were obtained from these crude lysates by centrifugation of sonicated extracts at 10 000 g for 10 minutes at 4 ºC. All samples were analysed by SDS-PAGE (4% stacking/12 % resolving discontinuous glycine/polyacrylamide gels) and protein bands visualised by Coomassie staining of the resolving gel.

3.2.4. Purification of bacterially overexpressed p27

3.2.4a. Denaturing nickel affinity purification of p27

Denaturing nickel affinity purification using the Qiagen Nickel NTA columns was tested on expressed heterologous p27 from the clarified supernatant of resuspended, sonicated extracts of induced *E. coli* BL21 (DE3) transformed with pCW21 according to the manufacturers' instructions. Protein was eluted into two fractions of 100 µl each from the spin column using elution buffer containing 8 M urea, and then further dialyzed against native purification buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8) using 10 kDa cut-off dialysis tubing.

3.2.4b. Acrylamide gel peptide purification of p27

Due to insolubility of the peptide, a different protein purification protocol (Dr Sharon Braunagel, Department of Biochemistry and Biophysics, Texas A & M University, College Station, Texas, USA. pers. comm., 2005) was used to gel-purify p27 following SDS-PAGE analysis. Ten $OD_{600\ nm}$ units of *E. coli* BL21 (DE3) cellular pellets containing overexpressed protein were heated at 100 ºC in SDS-PAGE sample buffer (10 % glycerol; 125 mM Tris-Cl pH 6.8; 4 % SDS; 10 % β-mercaptoethanol; 0.01 % bromophenol blue) for 5 minutes. Samples were centrifuged at 14 000 g for 10 minutes at

room temperature, the supernatant aspirated from the insoluble pellet and analysed by SDS-PAGE analysis (4% stacking/12 % discontinuous glycine/polyacrylamide gels). The samples were electrophoresed for an appropriate duration of time and the resolving gel incubated in 100 ml of a 100 mM solution of ice cold KCl. After 5 minutes, the localized white bands representing the 27 kDa 6xhis-p104$_{(644-915)}$ was excised from the resolving gel, finely diced and incubated with agitation at 4 °C overnight in 1 X PBS (pH 7.4).

To elute the protein, this supernatant was passed through a 1 ml syringe fitted with a 25 gauge needle and a small glass wool plug inserted into the neck of the syringe to filter out the acrylamide. The resultant protein was further concentrated and the SDS removed by ethanol precipitation as follows: a 1.6 volume of ice-cold ethanol (95 %) was slowly added to the filtered suspension in order to further concentrate the protein and remove the SDS. The mixture was allowed to incubate on ice for 1 hour with occasional mixing, after which the protein was pelleted by centrifugation at 10 000 g for 10 minutes at 4 ºC. The protein pellet was dried and resuspended in PBS, pH 7.4 then analysed by SDS-PAGE and detected by Western analysis using a monoclonal anti-6xhistidine antibody (Roche Applied Science) before quantification by Bradford's assay (Bradford, 1976). For Western analysis, proteins were transferred onto nitrocellulose membrane (GE Healthcare) using the Biorad® Mini Trans-blot electrophoretic transfer cell system according to the manufacture's instructions. The membrane was incubated with a 1 in 2 500 dilution of monoclonal anti-6Xhistidine antibodies prepared in TBST buffer [Tris-buffered saline (150 mM NaCl and 50 mM Tris-Cl pH 7.6)] containing 1 % BSA and 0.1 % Tween-20. The membrane was then probed using goat anti-rabbit alkaline phosphatase-conjugated IgG, followed by addition of substrate according to the manufacturers' specifications (Roche Biosciences) The resultant chemiluminescence was documented by exposure to X-ray film.

3.2.5. Construction of GST-p104$_{(644-915)}$

Difficulties with the solubility of the 6xhis-p104$_{(644-915)}$ peptide prompted the use of an alternative affinity purification system based on the glutathione S transferase (GST) protein. The fusion protein was created by re-amplifying the p27 sequence from the plasmid pCW19 with the oligonucleotides CW7 and CW6 which create a *Bam* HI site and *Hin*d III site on the 5` and 3` of this sequence, respectively. The *Hin*d III site was included in the 3` end of the selected sequence in order to create an in-frame stop codon downstream of the coding sequence. This amplicon was ligated into pGEM-T Easy and the recombinant plasmid (pCW22) sequenced and upon confirmation of the insert sequence, digested with *Bam* HI and *Sal* I to release the p27 coding sequence. The expression vector pGEX-4T-2 (GE Healthcare) (Figure 3.2) was also digested with *Bam* HI and *Sal* I and the backbone ligated with the *Bam* HI/*Sal* I p104$_{(644-915)}$ fragment of pCW22.

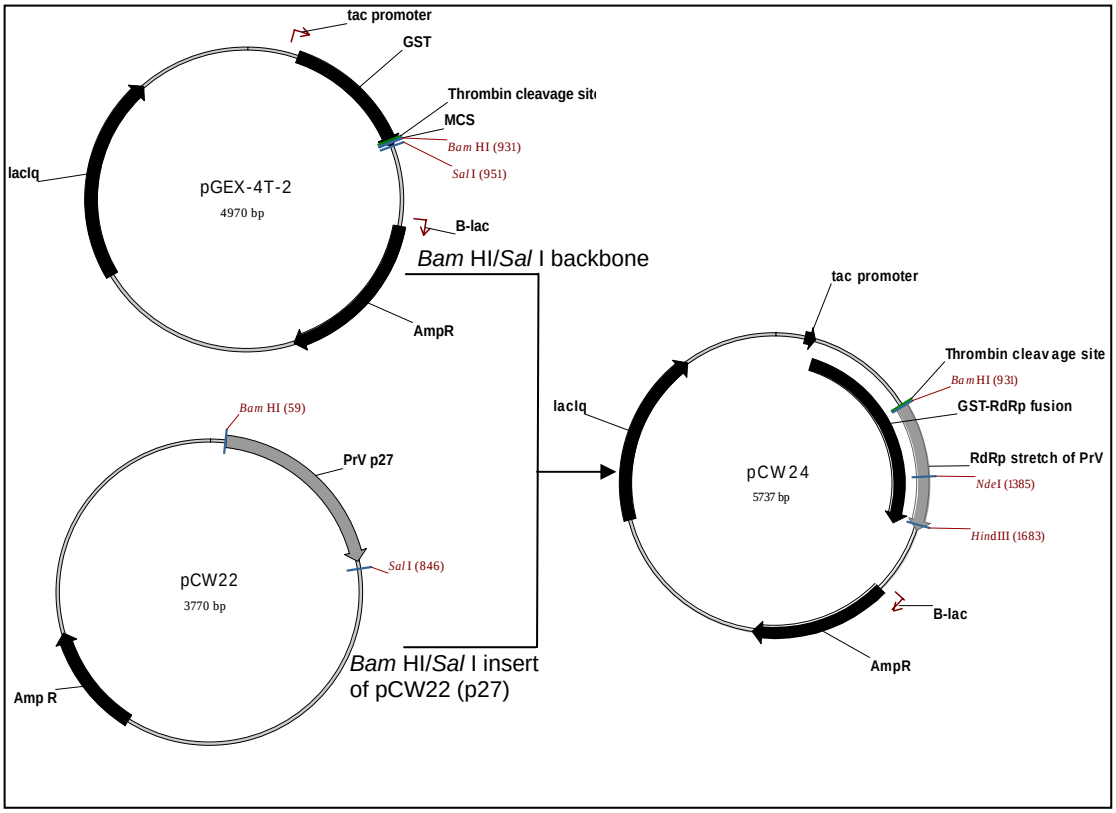

Figure 3.2. Cloning strategy used in the construction of the expression vector, pCW24. OriR: origin of replication, AmpR: β-lactamase, *lacIq*: lactose repressor gene, MCS: multiple cloning site, B-lac: *β-lactamase* promoter, p27/RdRp stretch of PrV: p104$_{(644-915)}$.

This new bacterial expression vector, pCW24, contained the IPTG-inducible hybrid bacterial promoter, *Tac* (tryp-lac hybrid), followed by GST fused to the N-terminal $p104_{(644-915)}$ coding sequence (Figure 3.2). The sequence of the insert in this expression vector was confirmed.

3.2.6. Overexpression and affinity purification of GST-$p104_{(644-915)}$

pGEX-4T-2 (GST alone) and pCW24 (GST-$p104_{(644-915)}$) were transformed into *E. coli* BL21 (DE3) cells and expression of the proteins induced by IPTG, as described in Section 3.2.1. Soluble fractions of induced cell-free extracts were analysed by SDS-PAGE. A 1 ml GSTrap™FF column and the AKTA UPC 900/PJ 920 fast performance liquid chromatography (FPLC) model and fractionator were used for affinity purification of soluble GST and GST-$p104_{(644-915)}$, according to the manufacturer's instructions (GE Healthcare).

3.2.7. Production of anti-$p104_{(644-915)}$ rabbit polyclonal antiserum

A total of 500 µg of purified 6xhis-$p104_{(644-915)}$ protein was used to generate anti-p104 polyclonal rabbit antiserum. Antigen complex preparation and antibody production was performed by Prof. D. U Bellstedt, Department of Biochemistry, University of Stellenbosch, Stellenbosch, Western Cape, South Africa as follows: acid-treated naked bacteria (NB) were prepared from *Salmonella minnesota* R595 as described (Bellstedt *et al.*, 1988). A fine suspension of NB in water (2 mg/ml) was prepared using an homogenizer with a loosely fitting Teflon plunger. A 400 µg amount of the antigen, 6xhis-$p104_{(644-915)}$, in a solution of PBS pH 7.4, was mixed with the NB suspension so that the ratio by dry mass of protein to NB was 1 to 5 (Bellstedt *et al.*, 1987). These mixtures were placed in small pear-shaped flasks and taken to dryness by rotary evaporation. The complexes thus formed were resuspended at a concentration of 240 µg / 0.5 ml PBS (0.15 M, pH 7.2) for immunization purposes.

The 6xhis-p104$_{(644-915)}$-antigen NB complexes were used to immunize rabbits intravenously into the marginal ear vein (125 - 240 μg per immunization). Rabbits were immunized on days 1, 4, 7, followed by booster series on days 14, 18 and 21, and again on days 28, 32 and 35. A 2 ml blood sample was collected on days 1 (as a negative control of all subsequent assays) and 28 (as a test bleed to check antibody titres) followed by a large bleed on day 42 at which time most antigens elicit a strong immune response using this method.

To test the specificity of the antiserum generated against the 6xhis-p104$_{(644-915)}$ peptide (referred to here as anti-p104 antiserum), cell free extracts of the PrV-infected *H. zea* midgut cell line MG8 (MG8); the *H. zea* fat body cell line FB33 (FB33) and *Spodoptera frugiperda* ovarian cell line 9 (*Sf*9) were generated and quantified by Bradford's assay (Bradford, 1976). Details on the maintenance of each cell line are given in Appendix 2. In order to generate cell-free extracts from these cell lines, whole cells were dislodged from the culture flask, pelleted by centrifugation (5 000 g for 2 minutes at room temperature), resuspended in PBS pH 7.4 and disrupted by passing through a 22 gauge needle 20 times. Cell-free extracts were stored in the presence of the protease inhibitor, 100 μM phenylmethylsulphonyl fluoride (PMSF) dissolved in 70 % isopropanol at -80 ºC. Varying amounts of quantified cell-free extracts were analysed by SDS-PAGE (4% stacking/12 % discontinuous glycine/polyacrylamide gels) and the resolved proteins analysed by Western analysis. An optimal dilution of 1 in 150 000 anti-p104 antiserum in 1 % BSA and 0.1 % Tween-20 in TBS buffer (Section 3.2.4b) was used.

To eliminate non-specific antibodies, the antiserum was pre-absorbed with a cell-free extract of uninfected FB33 cells. Approximately 300 μg (in 250 μl) of cell-free extract was mixed with 540 μg (in 250 μl) of anti-p104 antiserum and incubated with agitation, at 4 ºC overnight. The mixture was then centrifuged at 10 000 g for 30 minutes at 4 ºC and the supernatant carefully aspirated into a fresh microfuge tube, discarding the last

approximate 50 µl of antibody-protein precipitate. This pre-absorbed antiserum was used at a dilution of 1 in 50 000 in further Western analyses.

3.2.8. Baculovirus-mediated expression of 6xhis-p104$_{(1-207)}$ peptide in *Spodoptera frugiperda* (*Sf*9) insect cells

A recombinant baculovirus was created for the expression of 6xhistidine-tagged p27 in *Spodoptera frugiperda* ovarian cell line 9 (*Sf*9). The insect/bacterial shuttle vector, pCW30, used for the transposition and subsequent generation of recombinant viruses was constructed using the following multicloning strategy (Figure 3.3): The primers, CW9 and CW10, were created to amplify a 619 nt stretch of the PrV genome (amino acids 1 - 207 of p104) and in the process introduce *Nde*I and *Hin*d III restriction sites on the 5` and 3` of the fragment, respectively. This fragment was ligated into the commercial vector pGEM-T Easy, sequenced and named pCW26. pCW26 was digested with *Nde*I and *Hin*d III and the 619 bp insert ligated into the *Nde*I/*Hin*d III-digested backbone of pCW20. Although the resulting plasmid, pCW27, is a complete bacterial expression vector that could potentially be used in the expression of this PrV coding stretch, it was constructed in order to create a 5`-coding sequence which contained the SD – methionine – 6xhistidine cassette originally introduced into the vector by insertion of dsDNA oligonucleotide, AI11 (Section 3.2.1.). A region of 599 nt (nt 2-661) from pCW27 was amplified using the primers CW11 and CW10. CW10, the complementary primer, was an exact match of the very 3` sequence of the coding stretch in pCW27, whereas the forward primer CW11 introduced a *Bgl*II site into the sequence before the translational start of the optimized cloning cassette. This amplicon was ligated into the commercial vector pGEM-T Easy, sequenced and named pCW28. Finally, the commercial insect/bacterial shuttle vector pFB DUAL, [containing the baculoviral polH and p10 promoters in a bi-directional arrangement (Invitrogen™ Life Technologies)] was digested with *Bam* HI and *Sac*I and the backbone ligated with a *Bgl*II/*Sac*I insert originating from digested pCW28 (Figure 3.3).

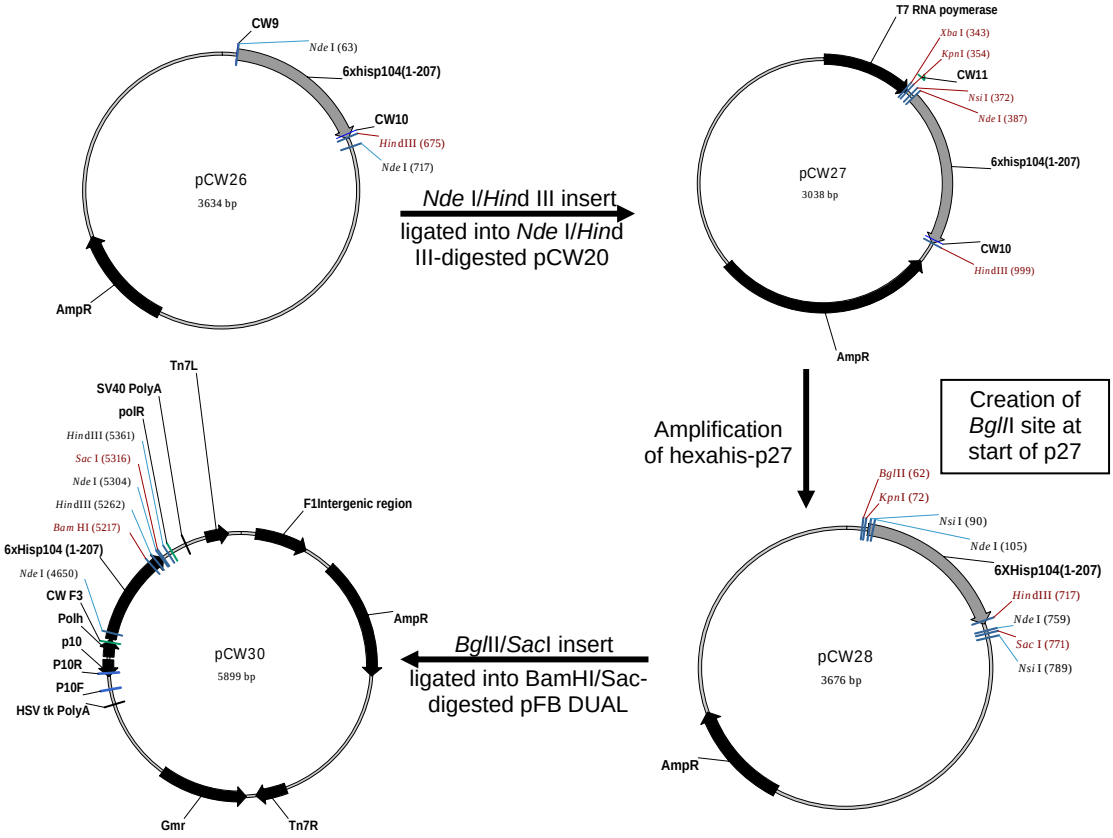

Figure 3.3. Multiple step cloning strategy used in the construction of the insect expression plasmid, pCW30. OriR: origin of replication, AmpR: *β-lactamase,* GentR: gentamicin resistance gene, HSVtk polyA: *Herpes Simplex virus thymidine kinase* poly A sequence, p10: p10 promoter, Polh: polH promoter, SV40 PolyA: *Simian vacuolating virus* 40 poly A sequence, B-lac: *β-lactamase* promoter, p27: $p104_{(644\text{-}915)}$.

The resultant plasmid, pCW30, contained a 6xhistidine-tagged N-terminal 207 amino acids of p104 under the control of the polH promoter (Figure .3.3). The transformation of pCW30 into *E. coli* DH10BAC™ cells and the subsequent isolation of recombinant bacmid DNA was carried out as described in the instruction manual for the Bac-to-Bac® Baculovirus Expression Systems from GibcoBRl (Invitrogen™ Life Technologies). Confirmation of inserted sequence within recombinant bacmid DNA was performed by PCR analysis. The primers, pucF and pucR, were used to PCR amplify and confirm the presence of a 273 bp region in unmodified, transposed bacmid DNA, a 1980 bp fragment of transposed bacmid DNA containing only pFB DUAL and finally, a 2689 bp fragment of bacmid DNA containing transposed pCW30. The amplification reactions were prepared

using 1 μl of template bacmid DNA and *Taq* DNA polymerase (Bioline) following the manufacturer's instructions. The reactions were cycled using an annealing temperature of 55 °C and an extension time of 5.5 minutes over 30 cycles. Recombinant bacmid DNA was transfected into *Sf*9 cells and the recombinant baculovirus passaged four times in *Sf*9 cells following the initial transfection of confirmed, recombinant bacmid DNA.

In order to confirm that these viral passages continued to produce active recombinant baculoviruses containing the transposed pCW30 cassette, PCR analysis was performed on extracted DNA from passage IV. A total volume of 6 ml of high titre recombinant virus was centrifuged at 12 000 g for 15 minutes at 4 °C and the pellet resuspended in 100 μl of virus disruption buffer (10 mM Tris pH 7.6, 10 mM EDTA and 0.25 % SDS). Proteinase K was added to a final concentration of 500 μg/ml and the sample incubated at 37 °C until the resuspended pellet appeared translucent. The resuspended pellet was extracted with equal volumes of phenol, phenol:chloroform:isoamyl alcohol [P.C.I (25:24:1)], and chloroform:isoamyl alcohol [C.I (24:1)]. The first organic phase from the phenol extraction was back-extracted in equal volumes of TE buffer (10 mM Tris-Cl and 1 mM EDTA, pH 8). The baculovirus DNA was precipitated in 205 μl of 3 M sodium acetate (pH 5.2) and 410 μl of ice-cold 96 % ethanol. The DNA was left to precipitate overnight at -20 °C and then centrifuged at 12 000 g for 16 minutes at 4°C. The pellet was washed twice with 200 μl of ice-cold 70% ethanol, resuspended in 25 μl TE buffer and stored at 4 °C.

PCR analysis of extracted recombinant baculovirus DNA was performed using the primers CWF3, which anneals 4583 - 4607 bp and polR, which anneals to 5390 - 5415 bp of pCW30. PCR amplification would produce a 883 bp fragment from the recombinant baculovirus genomic DNA, while control bacmid DNA containing unmodified, transposed pFB DUAL would produce a 172 bp PCR fragment. The PCR reactions were carried out as described in the instruction manual for the Bac-to-Bac® Baculovirus

Expression Systems from GibcoBRl (Invitrogen™ Life Technologies) and O'Reilly *et al*. (1992) using 3 μl of baculovirus DNA and Expand High Fidelity DNA Polymerase (Roche Applied Science). An annealing temperature of 47 °C and extension time of 1 minute (plus an additional, cumulative 5 seconds after each cycle) for 30 cycles was used in the PCR amplification of these inserts.

The virus titre of the supernatant of viral passage 3 stocks containing baculovirus-mediated expression of pCW30 was determined using the tissue culture infectious dose 50 % ($TCID_{50}$) assay described by O' Reilly *et al*. (1992). Eight duplicate wells of a 96-well microtitre plate were each seeded with $2x10^4$ *Sf*9 cells and the plate incubated at room temperature for 1 hour. A ten-fold dilution series of supernatant from virus passage 3 was made up to a final dilution of 10^{-12} in 50 μl of serum-free TC100 medium. Following attachment of the cells to the wells, the 10^{-8} to 10^{-12} dilutions were added to the cells in each of the 8 duplicate wells and incubated at 28 °C for 10 days. After 5 days of incubation, each well was fed with 10 μl of complete TC100 medium and after 10 days each well was scored for cytopathic effects. The $TCID_{50}$ of virus passage 3 of recombinant baculoviruses containing transposed pCW30 was calculated and a T25 (25 cm^3) flask of 100 % confluent *Sf*9 cells infected with recombinant baculoviruses at a multiplicity of infection (MOI) of 5, according to the instruction manual for the Bac-to-Bac® Baculovirus Expression Systems from GibcoBRl (Invitrogen™ Life Technologies). After 5 days of infection, both supernatant and cell lysates (prepared as described in Section 3.2.7) were analysed by SDS-PAGE.

3.3.1. Bacterial expression and purification of a system GST-p104$_{(248-331)}$ peptide

Based on the identification of hydrophilic and antigenic domains in p104 using Kyte & Doolittle hydropathy plots (Kyte and Doolittle, 1982) and the Jameson-Wolf antigenic index (Jameson and Wolf, 1988), a peptide

corresponding to amino acids 248 - 331 of the p40/p104 coding sequence (~10 kDa and referred to as GST-p104$_{(248-331)}$) was selected for overexpression in a bacterial expression system. The hydropathy plots were generated by online submission of the p104 amino acid sequence to the ProtScale program (Gasteiger *et al*., 2005) hosted by the Expert Protein Analysis System (ExPASy), and selecting the Kyte and Doolittle hydropathicity scale. The URI and date on which this online tool was accessed are referenced in Appendix 4.

The nucleotide sequence encoding p104 amino acids 248 – 331 was amplified from the plasmid pFLM1, using the primers CW23 (binding at nts 1756 - 1779 of the PrV cDNA) and CW24 (binding at nts 1986 - 2013), with CW23 introducing a *Bam* HI site at the 5` end of the sequence and CW24, a *Hin*d III and an in-frame stop codon at the 3` end. This amplified product was ligated into the commercial vector, pGEM-T Easy, the insert sequence confirmed and the recombinant plasmid (pCW52) digested with *Bam* HI and *Eco* RI. The *Bam* HI/*Eco* RI fragment containing the p104 fragment was ligated into pGEX-4T-2 (GE Healthcare) and digested with *Bam* HI and *Eco* RI (Figure 3.4.). The resultant plasmid, pCW53 contained the inducible *tac* promoter upstream of a GST-p104$_{(248-331)}$ peptide fusion. Expression of the GST-p104$_{(248-331)}$ fusion peptide was induced in *E. coli* BL21 (DE3) cells transformed with pCW53 as described in Section 3.2.3. Following induction, equivalent amounts of cell-free extracts were analysed by SDS-PAGE and the proteins visualised by Coomassie staining.

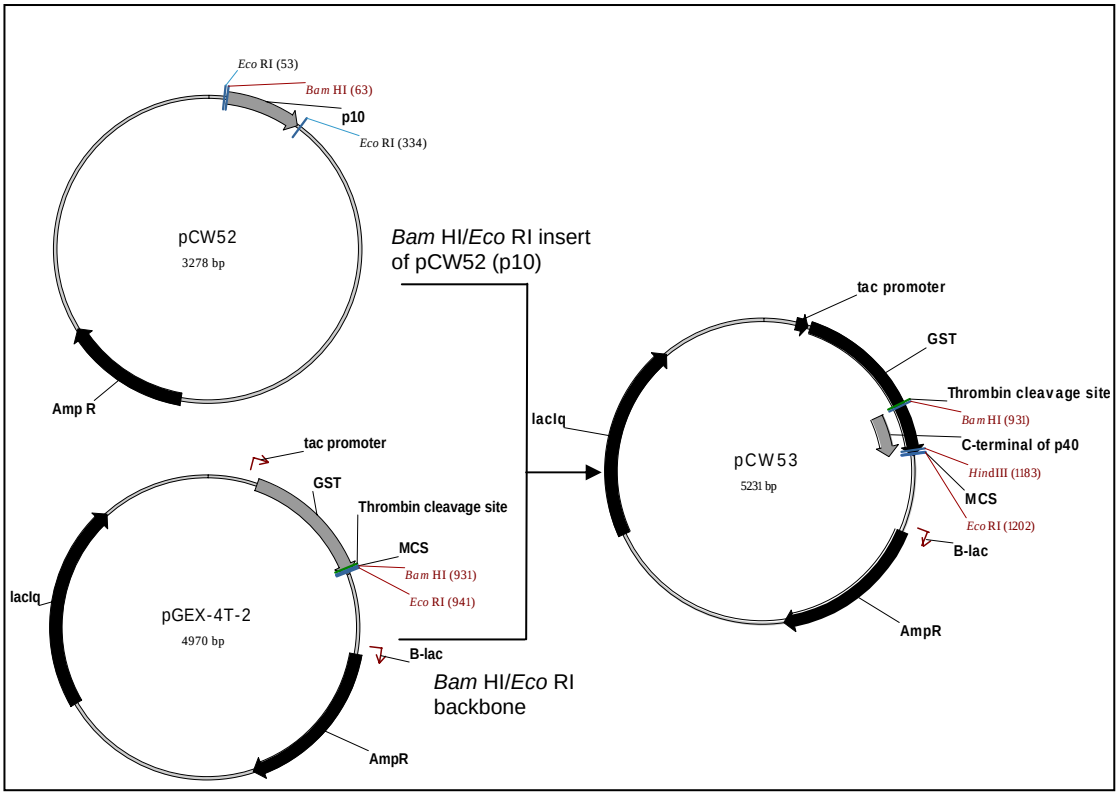

Figure 3.4. Cloning strategy used in the construction of the expression vector, pCW53. OriR: origin of replication, AmpR: *β-lactamase*, tac promoter: hybrid bacterial promoter, *lacIq*: *lactose repressor gene*, MCS: multiple cloning site, B-lac: *β-lactamase* promoter, p10: GST-p104$_{(248\text{-}331)}$.

3.3.2. Purification of the GST-p104$_{(248\text{-}331)}$ peptide fusion protein by affinity purification and FPLC

E. coli BL21 (DE3) cells expressing the GST-p104$_{(248\text{-}331)}$ fusion protein were harvested by centrifugation at 10 000 g for 5 minutes at 4 °C and the pellet resuspended in 1 ml of binding buffer (10 mM DTT in PBS, pH 7.4). The suspension was sonicated using a hand-held 3 mm probe [Vibra Cell™]) for 20 x 10 second bursts at 20 % power at 4 °C, and the cell lysate centrifuged at 10 000 g for 10 minutes at 4 °C. The supernatant (soluble fraction) was filtered through a 0.45 µm filter to remove any remaining whole cells and large/viscous cellular debris. This filtrate was passed through a GSTrap™FF 1 ml column using the AKTA UPC 900/PJ 920 FPLC model and fractionator following the manufacturers' instructions (GE Healthcare). Eluted fractions (500 µl) were collected for SDS-PAGE

analysis followed by consolidation of fractions containing the eluted protein, and the existing buffer exchanged with PBS (pH 7.4) using centrifugal filtration tubes (Amicon Ultra 10 kDa cut-off centrifugal filter devices, Millipore). The pooled purified protein was quantified by Bradford's assay (Bradford, 1976) and approximately 640 µg of this purified protein used to raise rabbit polyclonal anti-p40 antiserum (performed by Prof. D. U Bellstedt, Department of Biochemistry, University of Stellenbosch, Stellenbosch, Western Cape, South Africa as described in Section 3.2.7).

3.3.3. Testing of rabbit polyclonal anti-GST-p104$_{(248-331)}$ antiserum

To test the specificity of the antiserum generated against GST-p104$_{(248-331)}$ fusion protein (referred to here as anti-p40 antiserum), cell free extracts of the MG8, FB33, Baby Hamster Kidney 21 (BHK), HeLa S3 immortalised cervical cancer cells (HeLa) and *Sf*9 cell lines were generated and preserved as described in Section 3.2.7 and quantified by Bradford's assay (Bradford, 1976). 4 µg of each cell-free extract was analysed by SDS-PAGE (4% stacking/12 % discontinuous glycine/polyacrylamide gels) followed by Western analysis using dilutions of the anti-p40 antiserum. An optimal dilution of 1 in 200 000 in 1 % BSA and 0.1 % Tween-20 in TBS buffer was selected for the use of this antiserum in Western analyses.

3.4. RESULTS AND DISCUSSION

Based on the bioinformatic analysis of the p104 coding sequence discussed in Chapter 2, the only enzymatic domain positively identified in this replicase ORF was the RdRp domain. This domain was mapped on the C-terminus of the p104 ORF based on the positions of motifs A - C and G. The RdRp domain was positioned at amino acids 533 – 825 of p104 (Figure 3.5) and as an essential domain in an RNA virus, was predicted to fold correctly following transcription and translation of the p104 coding sequence. Antibodies raised towards this peptide stretch would recognize

the entire p104 protein and therefore, in order to produce antibodies specifically towards the p40 pre-read through N-terminus of this protein, a peptide stretch positioned at amino acids 248 - 331 was also selected for overexpression based on its antigenicity and hydrophilic nature (Figure 3.5).

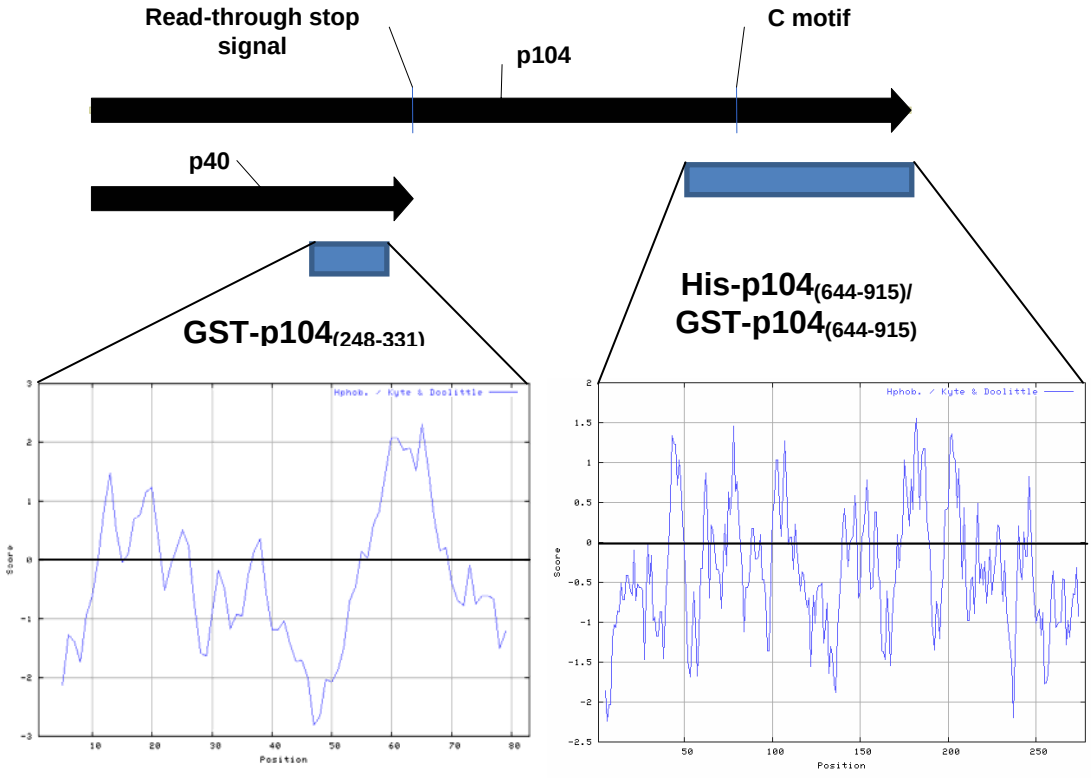

Figure 3.5. Hydropathy plots of the two peptide regions of p104, p104$_{(248-331)}$ and p104$_{(644-915)}$, selected for overexpression in either a bacterial or eukaryotic expression system. The p104$_{(248-331)}$ and p104$_{(644-915)}$ peptide stretches are represented, drawn to scale relative to p104, as blue rectangles and positioned below the pp40/104 coding sequence. The baseline (0 score) is highlighted in each hydropathy plot by a thick horizontal line. The GDD (C) motif is indicated on the p104 coding sequence as a point of reference.

3.4.1. Overexpression in *E. coli* and purification of 6xhis-p104$_{(644-915)}$

With the ultimate aim of generating antibodies towards p104, the expression of the RdRp domain was selected first because this antiserum could be used to identify and detect the essential PrV replicase protein. Initial SDS-PAGE analysis of crude extracts derived from *E.* coli cells

expressing 6xhis-p104$_{(644-915)}$ showed the overexpression of a 27 kDa protein corresponding to the predicted size of the his-tagged peptide (Figure 3.6). However, the peptide appeared to be mostly insoluble since no dominant 27 kDa protein was visible in the soluble fraction (Figure 3.6, lane 5 vs. 6). In an attempt to improve the solubility of the 6xhis-p104$_{(644-915)}$ peptide, the induction temperature was lowered from 37 to 30 ºC and the concentration of IPTG was decreased to 0.5 mM. These measures did not significantly improve the amount of soluble protein present (data not shown).

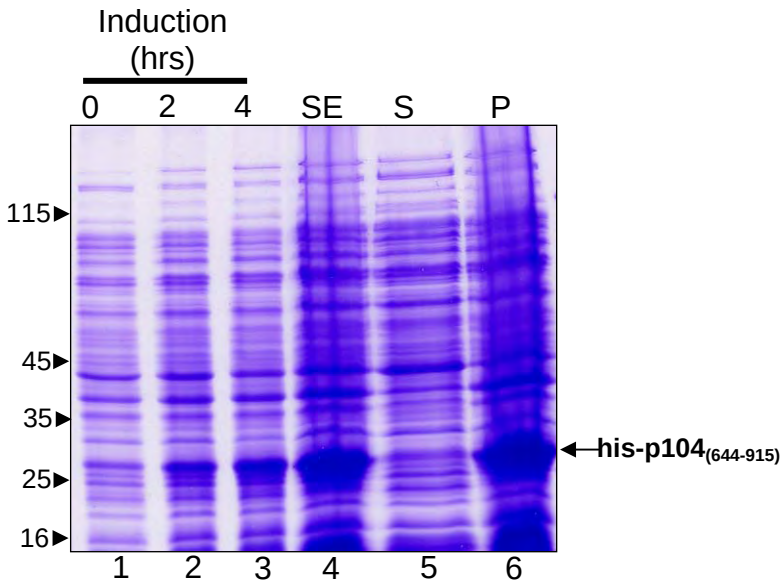

Figure 3.6. Overexpression of the 6xhis-p104$_{(644-915)}$) peptide in *E. coli* BL21 (DE3). SDS-PAGE analysis of protein fractions collected from *E. coli* BL21 (DE3) cells transformed with pCW21. Lanes 1 – 3: extracts derived from cells harvested after 0, 2 and 4 hours' induction with IPTG, respectively. Lane 4: sonicated extract (SE) after 4 hours induction; Lane 5: Soluble (S) fraction; Lane 6: insoluble (P) fraction. The molecular weight standards are indicated on the left of the gel in kDa, while the arrow on the right indicates the p27 expression product, 6xhis-p104$_{(644-915)}$.

Attempts at native purification of 6xhisp104$_{(644-915)}$ were unsuccessful, so it was decided to proceed with denaturing nickel affinity purification of this peptide. Fractions collected during this procedure were collected and analysed by SDS-PAGE (Figure 3.7).

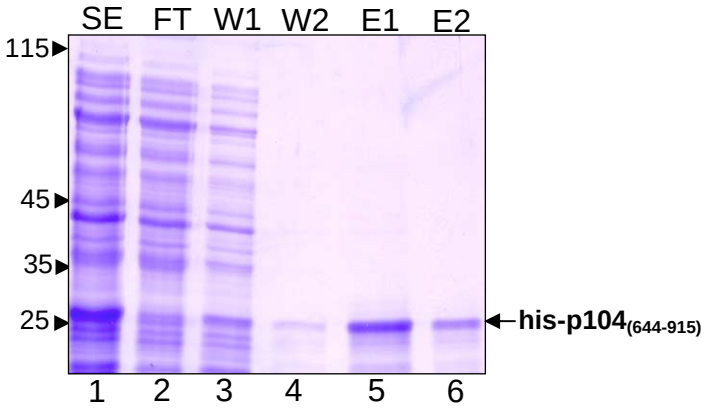

Figure 3.7. **Denaturing nickel affinity purification of 6xhisp104$_{(644-915)}$ peptide expressed in *E. coli* BL21 (DE3).** SDS-PAGE analysis of fractions collected during denaturing nickel affinity purification of 6xhisp104$_{(644-915)}$ from *E. coli* BL21 (DE3) cells transformed with pCW21. Lane 1: sonicated crude extract (SE). Lane 2: flow-through (FT). Lanes 3 and 4: wash 1 and 2 (W1/W2). Lanes 5 and 6: elution 1 and 2 (E1/2). The molecular weight standards are indicated on the left of the gel in kDa, while the arrow on the right indicates the 27 kDa protein corresponding in size to the 6xhisp104$_{(644-915)}$ peptide.

The denaturing affinity purification of 6xhisp104$_{(644-915)}$ from overexpressed *E. coli* BL21 (DE3) (Figure 3.7) was more efficient than the native nickel affinity protocol (data not shown). However, while sufficient quantities of peptide were purified using the denaturing protocol (Figure 3.7 lanes 5 and 6), the purified product only remained soluble in the presence of 8 M urea, making the use of this protein unsuitable in a mammalian immunization schedule. Attempts to re-fold the peptide in a less physiologically toxic buffer by both dialysis and centrifuge column buffer exchange led to the precipitation of the peptide.

3.4.2. Expression in *E. coli* and purification of the GST-p104$_{(644-915)}$ peptide

Since the 6xhisp104$_{(644-915)}$ peptide appeared to be insoluble, it was decided to replace the 6xhist tag with the GST protein in the hope that this would lead to the expression of a soluble fusion peptide. Analysis of crude extracts derived from *E. coli* cells transformed with pCW24 revealed that this peptide was only marginally soluble (Figure 3.8 lane 7), with the majority of the fusion peptide (~50kDa) remaining in the insoluble pellet (Figure 3.8 lane 8).

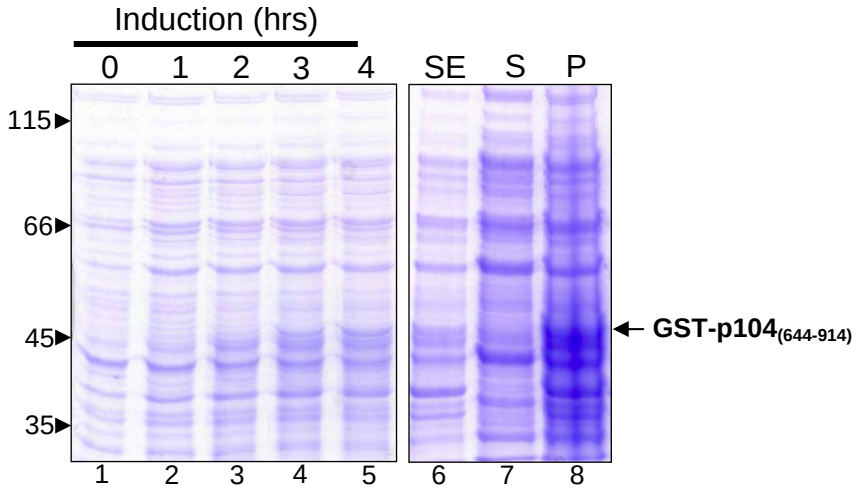

Figure 3.8. Overexpression of GST-p104$_{(644-914)}$ peptide in *E. coli* Bl21 (DE3). SDS-PAGE analysis of protein fractions collected *E. coli* BL21 (DE3) cells transformed with pCW24. Lanes 1 – 5: extracts derived from cells harvested after 0, 1, 2, 3 and 4 hours induction with IPTG, respectively. Lane 6: sonicated extract (SE) after 4 hours induction. Lane 7: soluble supernatant (S) after 4 hours induction. Lane 8: insoluble fraction (P) after 4 hours induction. The molecular weight standards are indicated on the left of the gel in kDa, while the arrow on the right indicates the 50 kDa protein, corresponding to the predicted molecular weight of the GST-p104$_{(644-915)}$ fusion peptide.

3.4.3 Baculovirus-mediated expression of p104 peptides in insect cell culture

With little success in expressing soluble p104$_{(644-915)}$ despite the fusion of the peptide with GST, it was decided to attempt to express the first 204 amino acids of p104 using a recombinant baculovirus expression system. This system was selected because the (over)expression of p40, the N-terminal section of p104, proved unsuccessful in a bacterial heterologous expression system (F. M Pringle, pers. comm., 2005). Recombinant baculoviruses expressing 6xhis-p104$_{(1-207)}$, under the control of the polH baculovirus promoter (pCW30) were used to transfect *Sf*9 cells. SDS-PAGE analysis and Coomassie staining of cell-free extracts obtained from cells infected with the recombinant virus showed no observable expression of the peptide. Furthermore, modified native affinity purification using Nickel-NTA spin columns (Qiagen) and subsequent Western analysis of eluted fractions using monoclonal anti-histidine antibodies did not detect the 6xhisp104$_{(1-207)}$ peptide (data not shown).

3.4.4. Gel purification and solubilisation of the 6xhis-p104$_{(644-915)}$ peptide expressed in *E. coli*.

Isolation of soluble 6xhis p104$_{(644-915)}$ protein by acrylamide gel fragment purification proved to be successful with large amounts of soluble protein isolated (Figure 3.9 panel A). This purified protein presented the 6xhistidine tag sufficiently for detection of this protein by Western analysis, using monoclonal anti-histidine antibodies (Figure 3.9 panel B). This protein was used to raise anti-p104 anti-serum. Mass spectroscopy (MS) analysis was not performed on this peptide due to financial constraints and the undoubted presence of other 27 kDa proteins that would have been co-excised with the overexpressed 6xhis p104(644-915) peptide during the purification procedure.

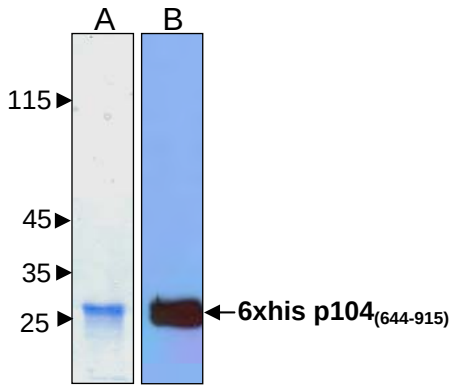

Figure 3.9. Gel purification, solubilization and detection of the 6xhis p104$_{(644-915)}$ peptide expressed in *E. coli* Bl21(DE3). (A) SDS-PAGE analysis of 6xhis p104$_{(644-915)}$ peptide and (B) Western analysis of the purified protein using monoclonal anti-hexahistidine antibodies. The molecular weight standards are indicated on the left of the gel and are represented in kDa, while the arrow on the right indicates a purified, soluble protein migrating at approximately 27 kDa corresponding to the predicted molecular weight of the 6xhis p104$_{(644-915)}$ peptide.

3.4.5. Validation of rabbit anti-p104 polyclonal antiserum

During the optimisation of the heterologous expression of p40 and peptide stretches of p40, the sensitivity and specificity of the rabbit polyclonal anti-p104 antiserum generated in this study was examined. This antiserum was analysed by Western analysis using cell-free extracts prepared from a

number of insect cell lines available at the time of these experiments (Figure 3.10).

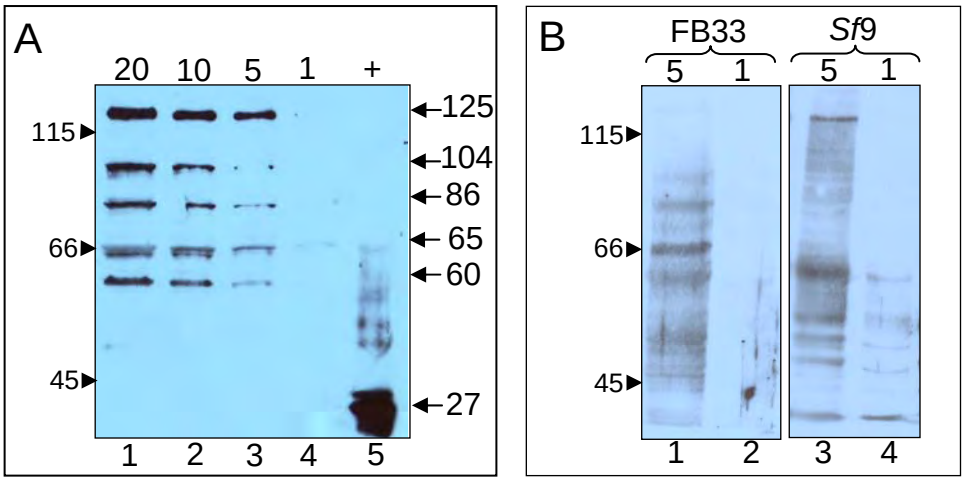

Figure 3.10. Specificity of 6xhis-p104$_{(644-915)}$ antiserum tested with cell-free extracts derived from insect cell lines. (A): Western analysis of varying quantities of MG8 cell-free extracts (lanes 1 to 4) using anti-p104 antiserum at a dilution of 1 in 50 000. Lane 5: 50 ng of purified 6xhis-p104$_{(644-915)}$. (B): Western analysis of varying quantities of FB33 (lanes 1 – 2) and *Sf*9 cell-free extract (lanes 3 - 4) tested with anti-p104 antiserum. The amount of each cell-free extract (in µg) probed are indicated above the relevant lane. The molecular weight standards are indicated on the left of the gel in kDa, while the arrows on the right of the image in panel A represent distinct proteins detected by Western analysis of MG8 extracts.

Although purified 6xhis-p104$_{(644-915)}$ peptide was detected, the antiserum also cross-reacted with numerous cellular protein in all three tested cell lines (Figure 3.10 panel A and B). In particular, five distinctive proteins were detected in the PrV-infected MG8 cell line (125, 104, 86, 65 and 60 kDa) (Figure 3.10 panel A). Only the 104 kDa protein potentially corresponds to the predicted molecular weight of the p104 coding sequence. Importantly, a 40 kDa protein corresponding to p40 was not detected (Figure 3.10 panel A, lanes 1 - 4), as was to be expected since the region chosen for the production of this antiserum is not represented on the p40 protein amino acid sequence.

When the anti-p104 antiserum was also tested against cell-free extracts of FB33 (Figure 3.10 panel B, lanes 1 & 2) and *Sf*9 (Figure 3.10 panel B,

lanes 3- 4) proteins, numerous non-specific proteins were detected. Besides a couple of dominant proteins of various molecular weights depending on the cell line being detected in both cell types, this antiserum also cross-reacted with other cellular proteins of a wide range of molecular weights (Figure 3.10 panel B, lane 1 - 4). Pre-absorption of the anti-p104 anti-serum with FB33 cell-free extract was performed in an attempt to precipitate out antibodies binding non-specifically to *H. zea* cellular proteins.

The Western analysis of the three cell-free extracts (in varying concentrations) was repeated using the pre-absorbed antiserum, but no decrease in the non-specific binding of this antiserum to cellular proteins was observed when tested in all thee of the cell-free extracts (data not shown). The pre-absorption step did not eliminate the non-specific binding of the antiserum to what appears to be at least three or four cellular proteins in the MG8 extract, nor the one or two dominant proteins detected in the FB33 and *Sf*9 cell-free extracts. It is likely that the purified peptide described in Section 3.4.5 did not re-fold during the purification and precipitation steps leading to the generation of anti-serum to co-purified 27 kDa cellular proteins as well as a small proportion of antibodies generated towards correctly re-folded 6xhisp104$_{(644\text{-}915)}$.

3.5.1. Expression of a peptide stretch within the p40 coding sequence in a bacterial heterologous expression system

Since the anti-p104 antiserum generated in this study was non-specific, it was decided to focus on another potentially antigenic region in the N-terminal half of p104, immediately upstream of the read-through stop signal. It was elected to express the p104$_{(248\text{-}331)}$ fragment with GST fused to its N-terminus to increase the likelihood of both the expression and purification of a soluble fusion protein.

The expression plasmid, pCW53, was created for the IPTG-inducible overexpression of a GST-p40 peptide fusion protein. The p40 peptide is an

83 amino acid stretch located at the C-terminal half of the entire coding sequence of p40, and was selected for fusion to GST because this region appeared to be highly antigenic and hydrophilic nature according to hydropathy plots (Figure 3.5) and antigenic indices. Crude extracts derived from *E. coli* BL21 (DE3) cells transformed with pCW53 were induced with IPTG and crude cell extracts analysed by SDS-PAGE (Figure 3.11).

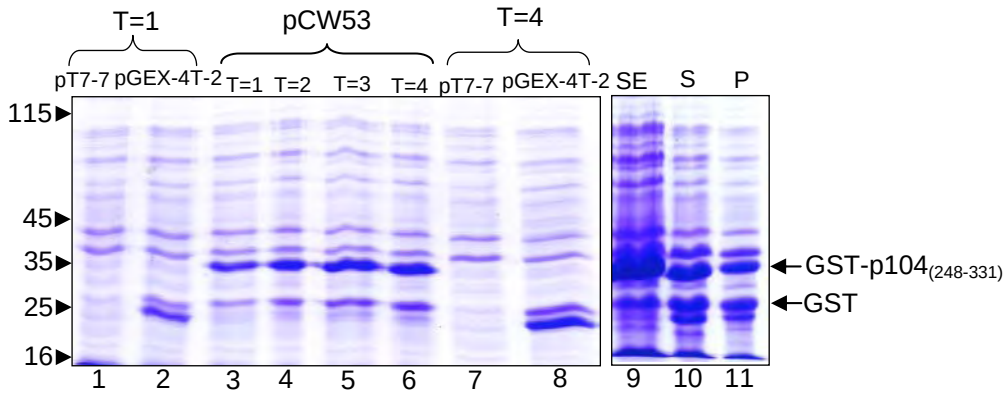

Figure 3.11. Expression and solubility studies on protein extracts from *E. coli* BL21 (DE3) expressing GST-p104$_{(248-331)}$. SDS-PAGE analysis of protein fractions collected from *E. coli* BL21 (DE3) cells transformed with pCW53. Lanes 1 - 2: crude extracts of pT7-7 and pGEX-4T-2 samples 1 hour post induction with IPTG. Lanes 3 – 6: crude extracts of pGEX-4T-2 samples taken 1, 2, 3 and 4 hours post-induction with IPTG. Lanes 7 and 8: crude extracts of pT7-7 and pGEX-4T-2 samples 4 hours post induction with IPTG. Lanes 9: sonicated extract (SE) after 4 hours induction. Lane 10: soluble supernatant (S) after 4 hours' induction. Lane 11: insoluble fraction (P) after 4 hours' induction. The molecular weight standards are indicated on the left of the gel in kDa, while the arrows on the right indicate the expression products, GST-p104$_{(248-331)}$ and GST.

The expression of pCW53-transformed cells, and subsequent SDS-PAGE analysis of samples taken at hourly time intervals during the induction, clearly showed the presence of a 37 kDa protein corresponding to the predicted molecular weight of GST-p104$_{(248-331)}$ (Figure 3.11, lanes 3-6). This protein was not visible in cell-free extracts of *E. coli* BL21 (DE3) transformed with the expression vector pT7-7 (Figure 3.11, lanes 1 and 7), whereas a noticeable 26 kDa band corresponding to the GST protein is visible in cell-free extracts of *E. coli* BL21(DE3) transformed with the GST expression vector, pGEX-4T-2 (Figure 3.11, lanes 2 and 8). A band migrating at the same size as GST is also visible in the protein samples

derived from cells transformed with pCW53 (Figure 3.11, lanes 3 - 6). This is likely a result of C-terminal truncated translation products occurring at the fusion point between GST and the p40 peptide. Solubility studies performed on cell-free extracts expressing the GST-p104$_{(248-331)}$ fusion peptide sampled after four hours post-induction showed that the fusion peptide was soluble (Figure 3.11, lane 9 - 11) and expressed in sufficient quantities for affinity purification (Figure 3.11, lane 10). Following this positive result, native affinity purification of this fusion protein was performed using FPLC and the resultant fractions analysed by SDS-PAGE (Figure 3.12).

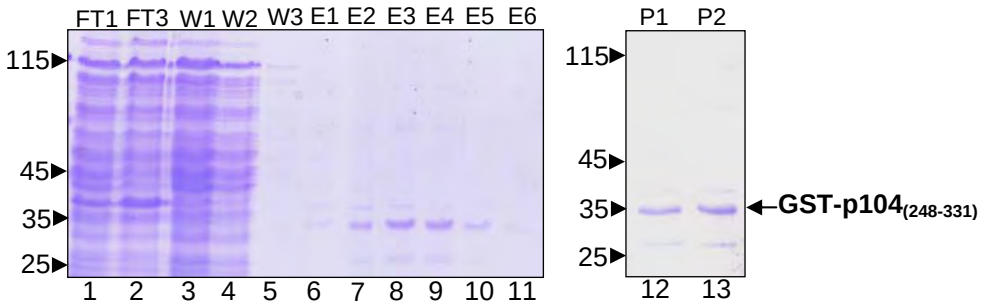

Figure 3.12. Native affinity purification of GST-p104$_{(248-331)}$ fusion peptide. SDS-PAGE analysis of fractions collected following the native affinity chromatography using FPLC. Lanes 1 - 2: Alternate flow-through (FT) fractions collected following addition of the sample and post-void volume release. Lanes 3 – 5: wash fractions 1 – 3 (W 1 - 3) collected following the addition of wash buffer. Lanes 6 – 11: elution fractions (E1 – E6), collected following the addition of reduced glutathione. Lanes 12 – 13: purified, pooled GST-p104$_{(248-331)}$ protein (P1 and P2). The molecular weight standards are indicated on the left of the gel in kDa, while the arrows on the right indicate the expression product, GST-p104$_{(248-331)}$.

FPLC affinity purification of GST-p40 protein resulted in the purification of sufficient quantities of fusion protein purified from soluble supernatant derived from cell-free extracts of *E. coli* BL21 (DE3) cells transformed with pCW53 and induced with IPTG (Figure 3.12). Purified and concentrated fusion protein in PBS buffer was further analyzed by SDS-PAGE (Figure 3.12, lanes 12 and 13). The resultant purified fusion protein was co-eluted and purified with GST and a larger protein of approximately 39 - 40 kDa, but due to its low concentration and the option of being able to remove

contaminating antibodies by pre-absorption, it was decided that the purified protein was sufficiently pure and of a high enough concentration for antiserum production. Since the expression of p40 alone has previously demonstrated to be unsuccessful and considering the overall insoluble nature of p104 peptides expressed without fusion to a tag, thrombin cleavage at the GST-p40 fusion point was not performed post purification in order to maintain the stability and solubility for further studies. MS analysis was not performed on this purified fusion peptide due to financial constraints.

3.5.2. Validation of rabbit anti- p104$_{(248\text{-}331)}$ antiserum

As described in Section 3.4.6, the rabbit anti-p104$_{(248\text{-}331)}$, referred to from here on as anti-p40 polyclonal antiserum was tested for its specificity in Western analysis using cell-free extracts of both insect and mammalian cell lines (Figure 3.13). The degree of non-specific interactions detected by the tested anti-p40 antiserum appeared much less in comparison to the degree of non-specific binding shown by the anti-p104 antibody (compare Figure 3.13 with 3.10).

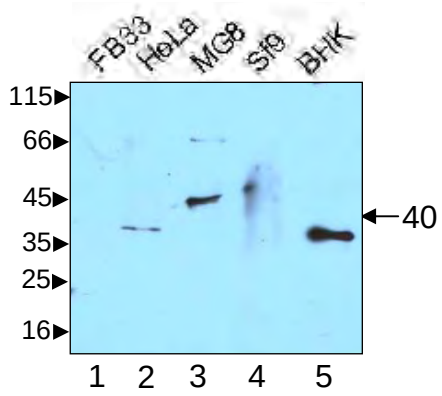

Figure 3.13. Specificity of GST-p104$_{(248\text{-}331)}$ antiserum tested with cell-free extracts derived from insect cell lines. A total of 4 µg of each cell-free extract was probed with a 1 in 200 000 dilution of anti-p40 serum. Lane 1: FB33. Lane 2: HeLa. Lane 3: MG8. Lane 4: *Sf*9. Lane 5: BHK. Molecular weight standards are indicated on the left of the gel, in kDa. The arrow on the right indicates a ~40 kDa protein in the MG8 sample corresponding in size to the replication protein, p40.

The MG8 cell-free extract contained two (suspected viral) proteins which cross-reacted with the anti-p40 antiserum (Figure 3.13, lane 3). These proteins had estimated molecular weights of 42 and 68 kDa and it is likely that the approximately 42 kDa protein represents the hypothesized auxiliary protein, p40 originating from PrV. When the anti-p40 antiserum was tested on resolved proteins of an FB33 extract, no non-specific interactions were detected (Figure 3.13, lane 1) while the HeLa, MG8, *Sf*9 and BHK extracts (Figure 3.13, lanes 2 - 5 respectively) showed cross-reactivity with single cellular proteins. These proteins are most likely cellular GST proteins cross-reacting with the antibodies raised against the GST portion of the fusion protein. Interestingly, the endogenous GST from the FB33 and MG8 cell-free extract was not detected, while the suspected GST in the *Sf*9 cell-free extract was detectable using this antiserum (Figure 3.13 lane 4). Anti-GST antiserum was not available at the time of these experiments to prove this cross-reactivity in the HeLa, *Sf*9 and BHK cell lines. It was concluded from these results that the anti-p40 antiserum would be suitable for use in Western analysis to examine the translation products of PrV.

3.6. CONCLUSION

The expression of sections of p40 and p104 coding sequences were not very successful and when the C-terminal (amino acids 644 - 915) region of p104 was successfully produced in the bacterial heterologous expression system, the protein proved to be insoluble. Although the peptide was eventually solubilized following gel purification, the resulting antibodies raised against this peptide did not appear to cross-react with the native protein expressed in the MG8 cells suggesting that it was likely misfolded. Expression of the N-terminal 207 amino acids of p40/p104 was also unsuccessful in both bacterial and insect heterologous expression systems for reasons unknown, although difficulties in expressing peptides in this region have been observed before (F.M. Pringle, pers. comm., 2005). Figure 3.14 below summarises the peptide stretches of p40/p104 that were

selected for overexpression, the heterologous system used and the outcome of the expression study.

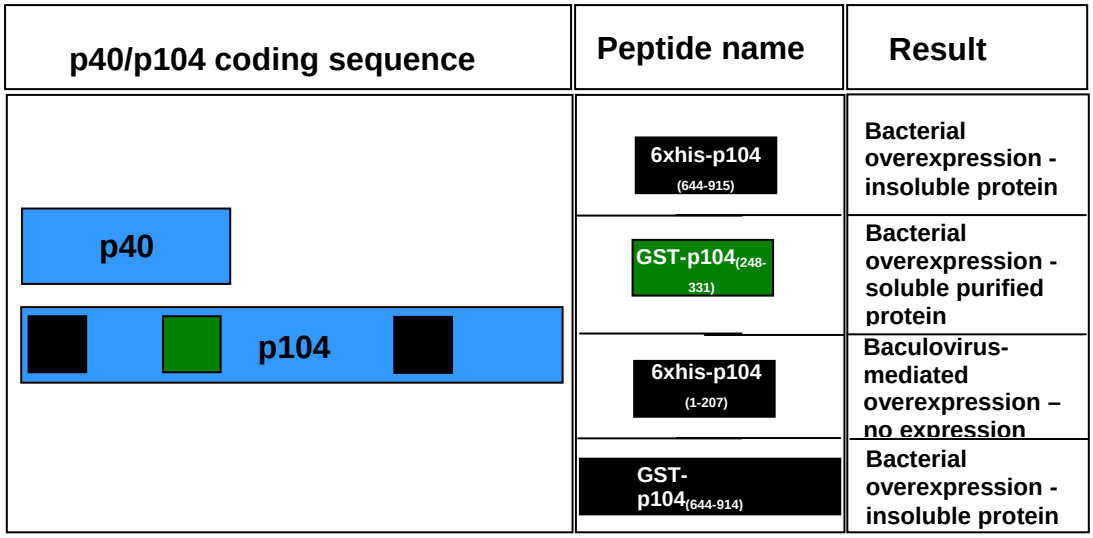

Figure 3.14. Peptide stretches of p40/p104 selected for overexpression in this study. The first column shows the p40/p104 coding sequence drawn to scale. The second column describes the name of each peptide (and its fusion partner) and the third column summarizes the system used and result of expression. The rectangles in the second column are drawn to scale and coloured black if no expression/insoluble protein was observed or green for successful purification.

A hydrophilic and antigenic region at the C-terminal of p40 (fused to the C-terminal of GST) proved to be soluble in the *E. coli* expression system and was used to raise rabbit polyclonal antibodies. This anti-p40 antiserum was proven to be both sensitive and specific, detecting an approximately 40 kDa protein in cell-free extracts of PrV-infected MG8 cells, indicating that these antibodies would be suitable for further studies on the expression of PrV proteins *in vivo.* Future experiments would ideally involve testing of this antiserum against uninfected MG8 cell lines in order to confirm no cross-reactivity between endogenous GST and the antiserum or perhaps GST pull down from the MG8 cell-free extracts followed by Western analysis of clarified extracts using anti-p40 antiserum.

CHAPTER 4
Characterization of the PrV translation products

4.1. INTRODUCTION

PrV was initially identified as a new tetravirus based on its capsid structure, monopartite genome and the amino acid sequence of the capsid protein precursor gene (Pringle *et al*., 2003). However, further analysis of the cDNA sequence has revealed that PrV might not be a typical betatetravirus (Chapter 2). As in the betatetraviruses, the PrV genome encodes a viral replicase (p104) and the capsid precursor gene (p81). However, unlike other members of this family, the 5` end of the genome encodes an additional large ORF (p130) of unknown function. The replicase ORF is unique in that it contains a read-through stop signal, which would result in the translation of two products from the p104 ORF: p40 (the auxiliary replication protein) and p104 (the replicase protein). In addition, the PrV genome encodes three 2A-like 'cleavage' sequences, one at the N-terminus of p130 and two at the N-terminus of p81. The presence of p130 and the co-translational processing signals suggest that PrV might employ a unique strategy for regulating expression of its viral gene products in infected host cells.

The objective of the research described in this chapter was to confirm the annotation of the PrV genome and to uncover more information about the co-translational processing of its gene products, both *in vivo* and *in vitro*.

4.2. METHODS AND MATERIALS

Unless otherwise stated, all gene manipulation methods were performed according to Sambrook *et al.* (1989) and protein analysis as described by Laemmli (1970). Primer sequences used in the construction of the

expression vectors pSTA1 – 5, 7 and 8, are described in Table 4.1 and listed in Appendix 1.

4.2.1. Metabolic labelling of PrV translation products *in vivo*

The methodology used for the *in vivo* labelling of PrV translation products in the PrV-infected MG8 cell line was performed according to Pringle *et al*. (2003). Successful inhibition of cellular transcription by Actinomycin D treatment was confirmed using the following protocol: the ExCell 420 medium from a T25 (25 cm^3) flask containing a near-confluent (80 %) monolayer of MG8 cells was removed and replaced with 6 ml of pre-warmed ExCell 420 medium containing different concentrations (and optimized at 20 µg/ml) of Actinomycin D. The cells were resuspended and aliquoted in 5 X 1 ml quantities into single wells of a 24-well cell culture plate. One millilitre of medium and cells was removed from a single well and collected for further analysis after 6, 12, 25, 36 and 48 hours. The cell suspensions were microfuged at 5 000 g for 2 minutes at room temperature and the pellet resuspended in RIPA buffer (150 mM NaCl, 1 % Nonidet® NP40, 0.5 % sodium deoxycholate (Na-DOC), 0.1 % SDS and 50 mM Tris-Cl, pH 8.0) by passing the cells through a 22 gauge needle four times before being stored in the presence of 100 µM PMSF at -80 ºC. The cell free extracts were quantified by Bradford's assay (Bradford, 1976) in order to ensure equal loading of protein extracts during SDS-PAGE analysis. Cell-free extracts were analysed by SDS-PAGE (4% stacking/12 % resolving glycine/polyacrylamide gels) and subjected to Western analysis using monoclonal antibodies towards VDAC and the PrV capsid protein as described in Chapter 3 (Section 3.2.4b). The nitrocellulose membranes were probed with a 1 in 500 dilution of monoclonal anti-VDAC antibodies (Santa Cruz Biotechnology) and a 1 in 2 000 dilution of polyclonal anti-PrV capsid antiserum, respectively. Antiserum and monoclonal antibodies were diluted accordingly in 1 % BSA and 0.1 % Tween-20 in Tris-buffered saline (150 mM NaCl and 50 mM Tris-Cl pH 7.6).

MG8 cells were allowed to grow to 90 % confluence in complete ExCell 420 medium in a T25 (25 cm^3) flask. The media was removed and replaced with 20 mg/ml of Actinomycin D in media deficient in the amino acid methionine. The cells were incubated at 28 ºC for 30 minutes after which the media was removed and the cells gently rinsed twice with 1 ml of methionine-free medium. Three millilitres of methionine-free medium containing 60 µg of Actinomycin D and 5.55 MBq of [^{35}-S]-methionine (Amersham Biosciences), was added to the rinsed cells and the flask incubated at 28 ºC for 24 hours. Whole cells were then dislodged from the flasks, harvested by centrifugation (5 000 g for 2 minutes at room temperature), resuspended in 400 µl of RIPA buffer and lysed by passing through a 22 gauge needle six times. Cell-free extracts were stored in the presence of 100 µM PMSF at -80 ºC subsequent to quantification by Bradford's assay (Bradford, 1976). A total of 10 µg of cell-free extract was analysed by SDS-PAGE (4% stacking/12 % resolving glycine/polyacrylamide gels), the resolved proteins fixed with 0.25 % Coomassie brilliant blue in 50 % methanol and 10 % acetic acid for 1 hour, destained in 50 % methanol and 10 % acetic acid for 1 hour and finally destained in 7 % methanol and 5 % acetic acid overnight. The fixed gels were dried under vacuum at 80 ºC and exposed to hyperfilm (GE Healthcare) for an appropriate duration of time.

4.2.2. Isolation of tetraviruses from the FB33 cell line

To confirm the uninfected status of the FB33 cell line, a virus purification procedure was performed as described by Pringle *et al*. (2003). The small pellet derived from sucrose cushion ultracentrifugation of FB33 cell-free extract was resuspended in 10 µl of 50 mM sodium phosphate buffer, pH 7.2. 5 µl of this resuspension was adsorbed onto a carbon coated copper grid prepared as described by Dong *et al*. (1998) and examined using a Phillips JEOL 1210 Transmission Electron Microscope.

4.2.3. *In vitro* transcription/translation experiments

4.2.3a. Plasmid construction for *in vitro* transcription/translation

A total of seven expression plasmids were constructed for *in vitro* transcription and translation assays to characterize the PrV ORFs. The vector, pBiEX-3 (Novagen) was digested with *Nco*I, the overhangs from this restriction site filled in using the Klenow fragment of DNA polymerase I and the vector then further digested with *Not* I. Each ORF was PCR amplified from the PrV cDNA (using the plasmid pFLM1), digested with *Not* I and ligated into the expression vector, pBiEX-3 (Figure 4.1).

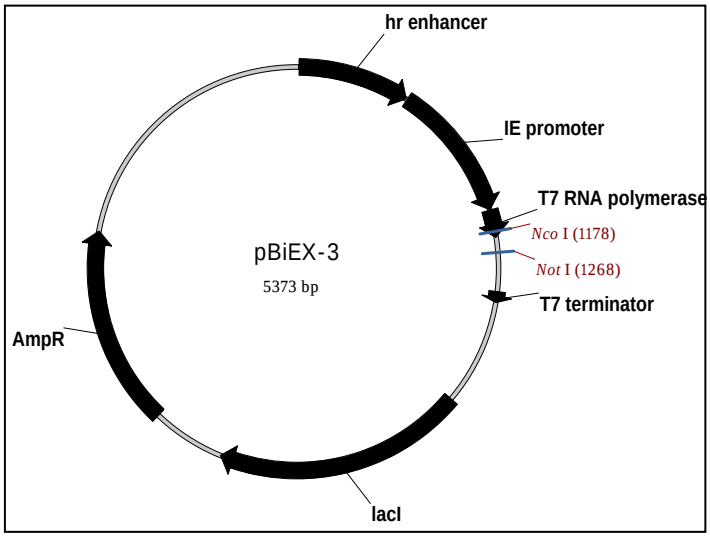

Figure 4.1. Restriction map of the expression plasmid, pBiEX-3. AmpR – *β-lactamase* gene; lacI – *lactose repressor gene*; hr enhancer – baculovirus promoter enhancer element; IE promoter – immediate early *cytomegalovirus* promoter; T7 RNA polymerase - bacterial promoter and T7 terminator – terminator sequence.

The individual plasmids, each containing the unmodified or altered version of a (predicted) PrV coding sequence, are described below (Figure 4.2, Table 4). The p130 ORF was PCR amplified using the primers CW01 and CW02 and ligated into pBiEX-3 to construct pSTA1. pSTA2 was constructed by amplifying a truncated version of the p130 ORF (with the coding sequence starting from an artificial translational start directly downstream of the 2A-like processing site) using the primers CW03 and CW02 and the amplicon ligated into pBiEX-3. pSTA3 was constructed by

amplifying another truncated version of the p130 ORF starting from a second possible translational start of p130 at nt 309 (amino acid 103) and ending at the same stop codon as the extended version of p130 expressed in pSTA1. The primers, CW01b and CW02 were used to amplify this truncated version of the p130 ORF and the amplicon ligated into pBiEX-3.

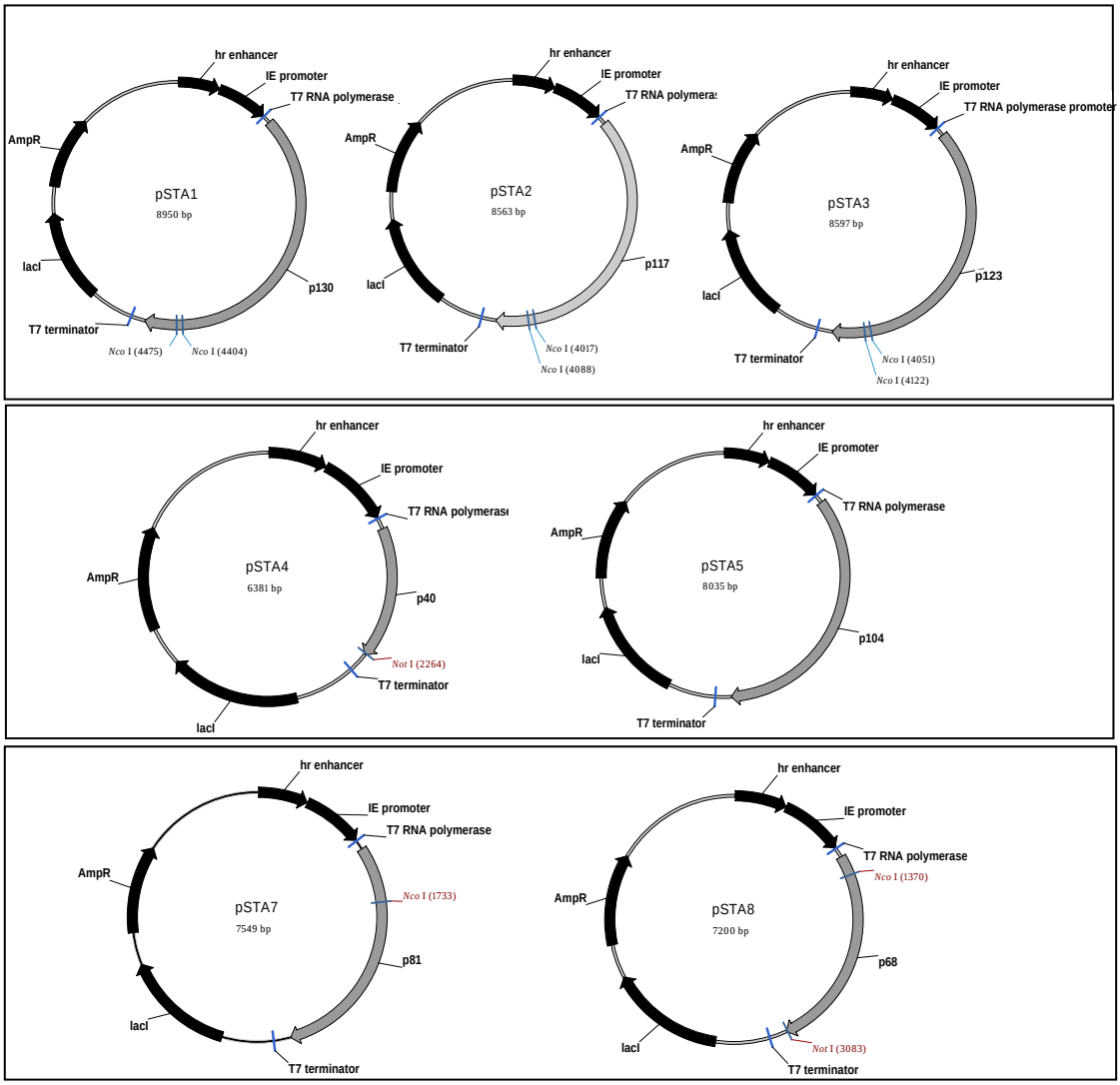

Figure 4.2. Restriction maps of the expression plasmids, pSTA1 – 5, 7 and 8 used for *in vitro* expression of the three major ORFs identified in the PrV genome. pSTA1 contains the coding sequence of p130. pSTA2 contains the coding sequence of a post-2A-like processing signal of the p130 ORF. pSTA3 contains the coding sequence of p130 starting from the second in-frame methionine. pSTA4 contains the p40 coding sequence and pSTA5 contains the entire p104 coding sequence. pSTA7 contains the capsid precursor protein (p81) coding sequence and pSTA8 contains the capsid protein α (p68) coding sequence. AmpR – *β-lactamase* gene, lacI – *lactose repressor gene*, hr enhancer – baculovirus promoter enhancer element, IE promoter – immediate early *cytomegalovirus* promoter, T7 RNA polymerase - bacterial promoter and T7 terminator – terminator sequence.

Table 4.1. Description on the coding sequences cloned into pSTA1 5, 7 and 8, primers used to amplify these coding sequencesp and the molecular weight of the predicted expression product.

Plasmid name	Primer combination used to amplify coding sequence	ORF	Predicted MW of translation products
pSTA1	CW01: 5`GCTGTGAACACTTCAGGTAAAAC'3 CW02: 5`TATATA**GCGGCCGC**TTACACTGCTCCGTTGGCGAAATG`3	p130	117 and 13 kDa
pSTA2	CW03: 5`CCCATCGGCTCTTGGCTTCGAGGA`3 CW02: 5`TATATA**GCGGCCGC**TTACACTGCTCCGTTGGCGAAATG`3	p130 (a.a. 130 – 1221)	117 kDa
pSTA3	CW01b: 5` AAGGAGTCTAATAGTGGTTACG`3 CW02: 5`TATATA**GCGGCCGC**TTACACTGCTCCGTTGGCGAAATG`3	p130(a.a.103-1221	117 and 6 kDa
pSTA4	CW04: 5`TGCGAGCCTTTAAGTTTACGC'3 CW05: 5`TATATA**GCGGCCGC**TTACTTTGTCGGGATTGTTCTTCC`3	p40	40 kDa
pSTA5	CW04: 5`TGCGAGCCTTTAAGTTTACGC'3 CW06: 5`TATATA**GCGGCCGC**TCAATACCCTGACGGCCAGTCAATATC`3	p104	40 kDa and 104 kDa
pSTA7	CW09 : 5`CAGAATTTACCAGTACCCAATG`3 CW10: 5`TATATA**GCGGCCGC**TTACTTGGCTCGTCTAGCGCGG`3	p81	68, 8 and 7 kDa
pSTA8	CW11: 5`CCCCAGATGACCCGCGAGAATAAAG`3 CW10: 5`TATATA**GCGGCCGC***TT*ACTTGGCTCGTCTAGCGCGG`3	p81(a.a.122-754)	68 kDa

Key: The *Not* I site is underlined and in bold and the mutated nucleotides are represented in lower case, in bold.

pSTA4 contains the p40 ORF and a functional stop codon (a.a 1 – 360 of p104) situated directly after the last coding triplet amplified using the primers, CW04 and CW05. A legitimate stop codon (UGA) was introduced at the end of the coding sequence to ensure the termination of translation in the *in vitro* system. pSTA5 was constructed by the amplification of the entire p104 ORF including the read-through stop signal using the primers CW04 and CW06. Finally pSTA7 and 8 were constructed by amplifying the capsid precursor ORF and a truncated version excluding the N-terminal 2A-like cleavage (first 121 amino acids) using the primers CW09 and, CW10 and for the truncated version, using the primers, CW11 and CW10.

4.2.3.b. Coupled *in vitro* transcription/translation using rabbit reticulocyte lysates

Coupled *in vitro* transcription/translation reactions were performed using rabbit reticulocyte lysates (TnT® T7 system, Promega). High quality plasmid DNA used in these reactions was extracted using Qiaprep Spin Miniprep columns (Qiagen). Translation reactions were performed by charging 5 µl *in vitro* rabbit reticulocyte master mix with 100 ng of the plasmids DNA; 0.25 µl double distilled water and 0.5 µl (0.925 MBq) of [^{35}S]-methionine to a final volume of 6.25 µl. This mixture was gently mixed and incubated at 30 ºC for 1.5 hours. 5 µl of each radiolabelled translation reaction was analysed by SDS-PAGE and the labelled proteins detected by autoradiography using a phosphoimager.

Due to the low levels of p104 translation products, immunoprecipitation was used to concentrate the *in vitro* translation products of p104 (and p40), after which these precipitates were probed using anti-p40 antiserum. 100 µl of protein A/G beads (Santa Cruz Biotechnology, Inc.) was rinsed three times with 1 ml of PBS (pH 7.4) and the slurry microfuged at 13, 300 g for 12 seconds to precipitate the beads. 50 µl of anti-p40 antiserum was then added to the washed beads (100 µl) and the mixture incubated with agitation at 4 ºC for 2 hours. The bead/antibody conjugate was then washed four times with 1 ml PBS containing 0.1 % v/v Nonidet® NP40 and

centrifuged at 13 300 g for 20 seconds. In the final wash, BSA was added to a final concentration of 10 mg/ml and the slurry incubated with moderate agitation for 30 minutes at room temperature. The beads were briefly centrifuged and resuspended in 50 µl of PBS (pH 7.4) containing 10 mg/ml BSA. Next, 20 µl of the prepared beads were added to 7 µl of rabbit reticulocyte lysate as well as 1 ml of binding buffer (150 mM NaCl; 20 mM Tris-Cl pH 7.4; 1 mM DTT; 0.05 % v/v Nonidet® NP40 and 10 mg/ml BSA). This mixture was incubated with to incubate at 4 ºC with moderate agitation for 30 minutes. The bead/protein conjugate was microfuged at 13 300 g for 20 seconds, the supernatant carefully removed and the beads washed by the addition of 1 ml of binding buffer. The resuspended bead/protein conjugate was microfuged at 13 300 g for 20 seconds. This was step was repeated a total of five times.

The pelleted beads were resuspended in 10 µl of SDS-PAGE sample buffer (10 % glycerol; 125 mM Tris-Cl pH 6.8; 4 % SDS; 10 % β-mercaptoethanol; 0.01 % bromophenol blue) and analysed by SDS-PAGE followed by Western analysis of the resolved proteins. The anti-PrV capsid antiserum used in this particular Western analysis was a recently generated rabbit polyclonal antiserum generated from approximately 500 µg of PrV purified according to Pringle *et al*. (2003) (Performed by Prof. D. U Bellstedt, Department of Biochemistry, University of Stellenbosch, Stellenbosch, Western Cape, South Africa and described in Section 3.2.7).

4.2.3.c. Coupled *in vitro* transcription/translation using wheat germ extract

Coupled *in vitro* transcription/translation reactions were carried out by adding 12.5 µl of wheat germ extract (Promega); 1 µl TnT® reaction buffer, 0.5 µl TnT® T7 RNA polymerase, 0.5 µl amino acid mix minus methionine, 1.85 MBq of [^{35}-S]-methionine, 0.5 µl *RNAsin*® ribonuclease inhibitor and 500 ng of DNA template, to a final volume of 25 µl. The reactions were gently mixed and incubated at 30 ºC for 90 minutes. 10 µl of each reaction

was analysed by SDS-PAGE (4 % stacking/7.5 % resolving glycine/polyacrylamide gels) and the resolving gels fixed followed by detection of the radiolabelled translation products by autoradiography as described in Sections 4.2.3.b.

4.3. RESULTS AND DISCUSSION

4.3.1. Optimization of [^{35}-S]-methionine metabolic labelling of MG8 cells

The first step towards characterizing the translation products of PrV was to determine whether proteins corresponding to the ORFs identified in the cDNA sequence were present in the persistently infected MG8 cells. Since there was no anti-p130 antiserum available, it was decided to use metabolic labelling with [^{35}-S]-methionine to identify the viral translation products in the MG8 cells. To differentiate between host and viral proteins, the cells were treated with Actinomycin D, which inhibits host transcription and therefore the translation of cellular proteins. Since virus replication is not sensitive to Actinomycin D, one is able to differentially label viral proteins by growing the cells in medium containing [^{35}-S]-methionine.

MG8 cells were treated with varying concentrations of Actinomycin D for up to 48 hours, the rate of cell growth monitored and cell-free extracts assayed to determine the optimal concentration and incubation time to halt cellular transcription and thus translation. Both adhered and lifted cells were collected at each time point, pelleted and lysed into cell-free extracts. Quantification of the total protein of these cell-free extracts by Bradford's assay (Bradford, 1976) indicated that a concentration of 20 µg/ml Actinomycin D efficiently inhibited transcription/translation, without wide-spread cell death, for up to 36 hours. This was attained by the confirmation of a plateau in total protein concentration 36 hours post-treatment.

The voltage-dependent anion channel (VDAC) transporter protein is a ubiquitous eukaryotic protein found in the outer membranes of mitochondria that plays an essential role in many mitochondrion-related signalling pathways and in small molecule or anion transport across the outer membrane (Brown, 1992). This protein has a relatively long half-life and a commercial monoclonal anti-mouse VDAC antibody was able to detect endogenous insect VDAC protein. It was decided to use this protein as a control to confirm Actinomycin D inhibition of cellular transcription in MG8 cells (Figure 4.3 panel A). Anti-PrV capsid antiserum was used to detect capsid protein in the Actinomycin D-treated MG8 cells and to confirm that the presence of the inhibitor did not affect the amount of detectable capsid protein over at least a 24 hour period following commencement of treatment (Figure 4.3 panel B). In both experiments, the total protein was quantified by Bradford's assay (Bradford, 1976) and standardised for each sample in order to ensure equalised loading of total protein.

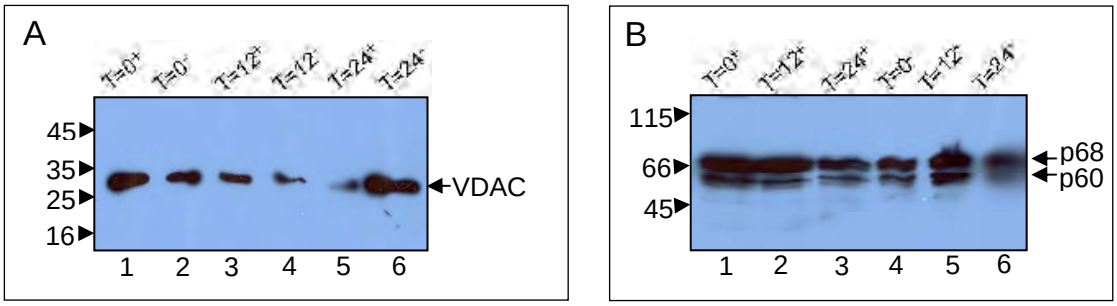

Figure 4.3. Detection of VDAC and PrV VCAP in MG8 cell-free extracts treated with Actinomycin D. (A): Western analysis of MG8 cell-free extracts cultured with (+) or without (-) Actinomycin D and probed with anti-VDAC antibodies. Cells harvested after zero hours (lanes 1, 2); 12 hours (lanes 3, 4) and 24 hours (lanes 5, 6). (B): Western analysis of cell-free extracts probed with anti-PrV capsid antiserum. Lanes 1 – 3: Cells cultured for 0, 12 and 24 hours following treatment with Actinomycin D. Lanes 4 – 6: Cells cultured for 0, 12 and 24 hours in the absence of Actinomycin D. Key: "+" and "-" – cells treated with and without Actinomycin D. The molecular weight standards are indicated on the left of the blot and are represented in kDa, while the arrows on the right either indicate cellular VDAC (30 kDa) or virus capsid proteins (p60 and p68).

Western analysis of the MG8 cell-free extract using the anti-VDAC antibodies resulted in the detection of VDAC protein (30 kDa) in all of the samples taken over a 24 hour period. (Figure 4.3 panel A). Cells grown in

the presence of Actinomycin D showed a significant decrease in the amount of detectable VDAC protein (Figure 4.3 panel A, lanes 1, 3 and 5). This is most likely explained by the increasing number of dying cells (cells began to lift from the bottom of the well after 12 hours). The reduction in VDAC would therefore be expected to result from inhibition of cellular transcription even though the total protein was equilibrated in each sample. Cells growing in the absence of Actinomycin D showed increasing levels of VDAC as is to be expected during normal cellular proliferation (Figure 4.3 panel A, lanes 2, 4 and 6) although it is uncertain why the initial quantity of VDAC detected was lower (Figure 4.3 panel A, lane 2) than that detected in the equivalent time period grown in the presence of Actinomycin D (Figure 4.3 panel A, lane 2 vs. lane 1).

An equivalent amount of PrV capsid protein was present in the cells after 12 hours incubation in the presence or absence of Actinomycin D (Figure 4.3 panel B). The resolution of the chemiluminescence signal in panel B, lane 6 was unfortunately affected by a technical problem during signal capture. Although these results do no conclusively confirm the inhibition of cellular transcription, it was decided to proceed with *in vivo* metabolic labelling experiments since the inhibitory concentration of Actinomycin D was the same that was used by Pringle *et al*. (2003) for *in vivo* labelling of viral RNAs in the same cell line.

4.3.2. Metabolic radiolabelling of PrV translation products in MG8 cells

At least seven radiolabelled proteins were detected in the MG8 cells treated with Actinomycin D (Figure 4.4). The relative molecular weights of these proteins were estimated to be: 128, 104, 85, 68, 63, 61 and 41 kDa. The largest protein of 128 kDa (and also the most dominant in terms of signal detection), corresponds closely in terms of molecular weight to the predicted p130 protein product which contains ten methionine residues in its coding sequence. The 104 kDa protein corresponds in molecular

weight to the replicase protein, p104, with thirty one methionine residues. If these two proteins are the translation products of p130 and p104, then the relative intensity of the two bands suggests that p104 is expressed at a substantially lower rate than p130. A 41 kDa protein potentially represents p40, the potential auxiliary replication protein that contains ten methionine residues in its coding sequence. A comparison of the putative p104 band with that corresponding to p40 suggests that p40 may be expressed at substantially higher levels than p104 (Figure 4.4).

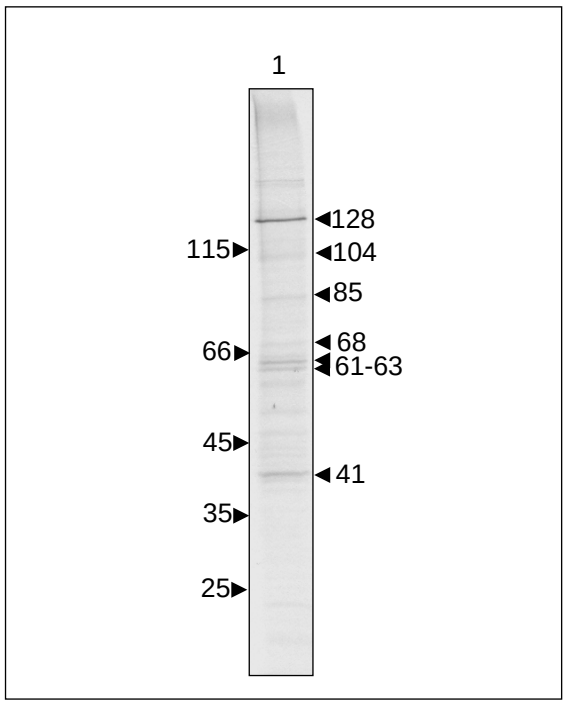

Figure 4.4. Detection of [^{35}S]-methionine-metabolically labelled viral translation products in MG8 cells treated with Actinomycin D. Ten micrograms of total cellular protein of Actinomycin D-treated MG8 cells grown in the presence of [^{35}S]-methionine for 24 hours, was analysed by SDS-PAGE followed by autoradiography. The molecular weight standards are indicated on the left of the autoradiograph and are represented in kDa, while the arrows on the right indicate seven dominant radiolabelled proteins and their molecular weight.

The 85 kDa protein that was detected in this experiment could represent the translated p81 capsid protein precursor translation product, which has a predicted molecular weight of 81 kDa. A putative capsid precursor protein of 68 kDa, which contains 14 methionine residues, was also

detected. Two proteins of approximately 61 and 63 kDa were detected, either of which might represent the cleaved major capsid protein of PrV (p60). The detection of proteins corresponding in size to the capsid protein precursor, p81 and the major capsid protein, p60, was to be expected based on the results of Western analysis by Pringle *et al*. (2003) and in this study (Figure 4.3). No band corresponding to p13, which would be translated at the N-terminus of p81, was detected using this method. This could suggest that this peptide was unstable or marked for proteolysis in this cellular environment as has been documented with peptides containing PEST-like sequences. The predicted ORF of this protein contains four methionine residues. The minor capsid protein (p8), produced after autoproteolytic cleavage during capsid maturation, does not contain any methionine residues and thus would not be detected using this experimental approach, explaining its absence in the autoradiograph.

Since the MG8 cell line is persistently infected and an uninfected cell line is therefore not available, it was not possible to compare the profile of labelled proteins in infected vs. uninfected cells. A similar metabolic labelling experiment could not be conducted using PrV-infected and uninfected FB33 cells because of the low infection rate of this cell line by PrV (Pringle *et al*., 2003). The *Se*1 cell line was also shown to be permissive to PrV infection but at a lower rate of infection (Pringle *et al*., 2003), was not available in the laboratory. Nevertheless, based on the detection of proteins corresponding to within 5 % of the predicted molecular weights of all the identified PrV ORFs, the next objective was to confirm the identity of some of these proteins by Western analysis of the PrV-infected MG8 cells. Although antibodies recognizing the p130 protein were not available during the course of this work, antisera raised against PrV virus particles and p40/p104 were used in the subsequent experiments.

4.3.3. Western analysis of PrV proteins in MG8 cells

Analysis of cell-free extracts probed with anti-capsid antiserum detected the presence of the capsid protein precursor (p68) and the major capsid protein (p60) in MG8 cell-free extracts (Figure 4.5 panel A, lane 1). The minor capsid protein (p8) was not detectable under these conditions because of the percentage resolving gel used in the SDS-PAGE analysis of MG8 cell-free extracts. The faint bands of non-specific proteins between 15 - 30 kDa region (Figure 4.5 panel A, lane 1) were attributed to degraded capsid proteins commonly observed with purified PrV particles and crude preparations containing PrV. The same antiserum detected a faint cross-reacting protein in the FB33 cell-free extract with an approximate molecular weight of 60 kDa, (Figure 4.5 panel A, lane 2), which appears to be a cross-reacting protein from the tissue culture medium used to culture FB33 cells. This cross-reactivity was repeatedly observed in other experiments and was attributed to a protein present in the foetal calf serum that is added to the cell culture medium used for the FB33 cell line. The MG8 cell line does not require foetal calf serum for growth – this cell line cannot survive in FCS-supplemented medium.

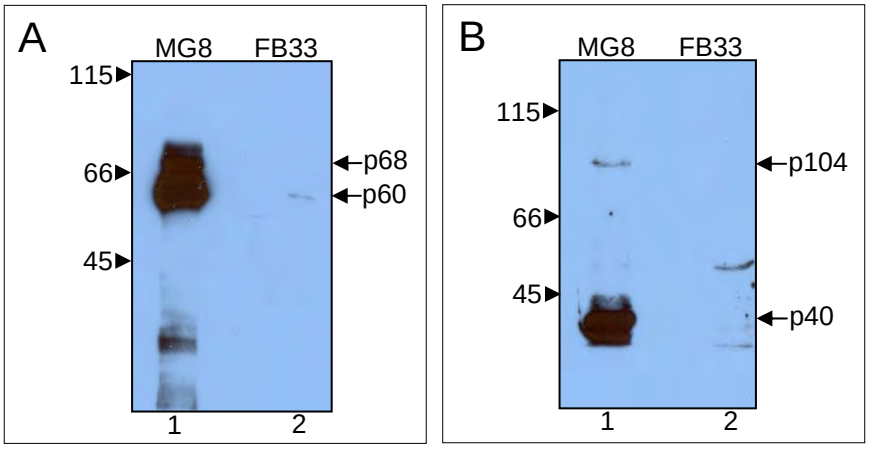

Figure 4.5. Detection of the VCAP and p40/p104 proteins in PrV-infected MG8 cells. (A): Western analysis of 2 µg of MG8 and FB33 cell-free extract using anti-PrV antiserum. The anti-PrV antiserum was diluted 1 in 100 000. (B): Western analysis of 5 µg of MG8 and FB33 cell-free extract using anti-p40 antiserum (diluted 1 in 100 000). Key: MG8 – MG8 cell-free extract; FB33 – FB33 cell-free extract. The molecular weight standards are indicated on the left of the Western blot and are represented in kDa, while the arrows on the right indicate cross-reacting proteins and their respective molecular weights.

Western analysis of MG8 cell-free extracts using anti-p40 antiserum detected a protein migrating at approximately 41 kDa (p40) and a higher molecular weight protein of approximately 100 kDa, which may represent p104 (Figure 4.5 panel B, lane 1). Based upon the intensities of their chemiluminescent signals, it appeared that p40 was present in much higher concentrations than p104. This result is consistent with relative levels of expression of the 41 and 104 kDa proteins observed in metabolic labelling experiments (Figure 4.4). Two cross-reacting proteins migrating at 52 and 35 kDa were detected in the FB33 cell-free extract (Figure 4.5 panel B, lane 2), which appeared to be present in very low concentrations. To eliminate the possibility that these protein bands might represent a tetravirus infection in the FB33 cell line, a tetravirus extraction protocol was performed on the FB33 cells. No virus particles were detected by T.E.M analysis of a sucrose pellet. This result suggested that the FB33 cell line is unlikely to be infected with a tetravirus and that the bands observed in the Western blot were most likely the result of non-specific binding to cellular proteins.

4.3.4. *In vitro* detection of individual PrV translation products

The previous section described the detection of seven radiolabelled proteins *in vivo* that were attributed to the PrV viral translation products of the three ORFs p130, p104 and p81. Western analysis using anti-PrV capsid and anti-p40/p104 antisera identified 4 proteins of 68 and 60 kDa (detected by the anti-PrV capsid antiserum) and 40 and 104 kDa (detected by the anti-40/p104 antiserum). While a protein of approximately 128 kDa was detected in the *in vivo* radiolabelling experiments, the absence of anti-p130 antibodies precluded Western analysis to confirm the presence of this protein in MG8 cells. Thus the data suggested that at least two ORFS, (p81 and p104) in the PrV genome were translated *in vitro*. The N-terminal of p81, which comprises the PEST-like protein p13, was not detected using this form of labelling, possibly due to its unstable nature, as is usually predicted for PEST-containing peptides. Coupled *in vitro*

transcription/translation of constructs expressing each of the three PrV ORFs was conducted to characterize the translation initiation sites, the read-through activity of the stop signal found in p104, as well as the co-translational processing of the three 2A-like cleavage sites in p130 and p81.

Rabbit reticulocyte lysate coupled in vitro transcription/translation of the p130 ORF

Expression vectors were constructed, that would allow for the analysis of the translation of each of the three ORFs identified on the PrV cDNA sequence. Three variants of the p130 theoretical ORF were analysed initially using the rabbit reticulocyte lysate system. The plasmid pSTA1 contained the full-length version of p130, while pSTA2 contained a truncated ORF starting immediately after the 2A-like processing site (a.a. 130 – 1221). The p130 ORF in pSTA3 was also truncated, starting from the second potential translation start at Met103. SDS-PAGE analysis was conducted on the protein extracts originating from lysate charged with each of the three constructs and the [^{35}S]-methionine-labelled translation products were visualized by autoradiography (Figure 4.6).

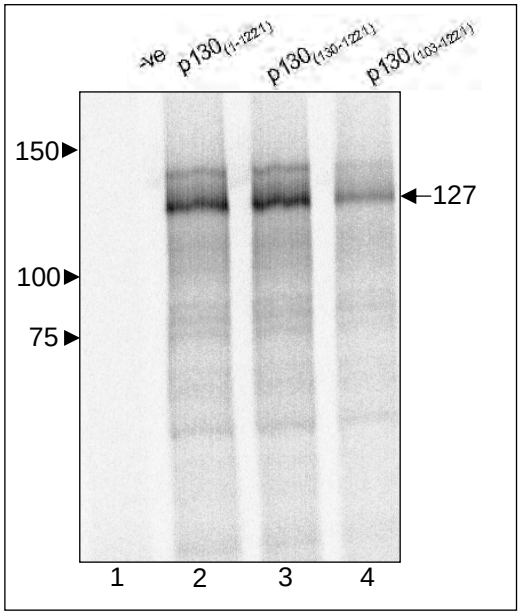

Figure 4.6. Detection of coupled *in vitro* transcription/translation products of constructs expressing p130 in rabbit reticulocyte lysates. Autoradiograph of SDS-PAGE analysis of rabbit reticulocyte lysates charged with pSTA1-pSTA3. Lane 1: control reaction with no input plasmid DNA. Lane 2: pSTA1 (p130). Lane 3: pSTA2 ($p130_{130-1221}$). Lane 4: pSTA3 ($p130_{103-1221}$). The molecular weight standards are indicated on the left of the autoradiograph and are represented in kDa, while the arrows on the right indicate radiolabelled proteins and their molecular weights.

A translation product of 127 kDa was detected in lysates charged with all three coding sequence variants; pSTA1, pSTA2 and pSTA3 (Figure 4.6 lanes 2 - 4). The size of this protein was calculated and confirmed using marker bands as standards. This 127 kDa protein was not present in rabbit reticulocyte lysates that were not charged with plasmid DNA (Figure 4.6 lane 1). The full length version of p130 (pSTA1), predicted to start at the first methionine (which is in an good context for initiation of translation), would be translated into proteins with a predicted size of 117 and 13 kDa protein, if the 2A-like processing site at the N-terminal of this coding sequence was functional. The 2A-like cleavage site did not appear to be functional in this particular *in vitro* translation system (Figure 4.6 lane 2). The truncated version of the coding sequence (pSTA2), translated from a start codon after the 2A-like sequence (PrV-$2A_1$), was predicted to translate a protein of 117 kDa. This was not detected and instead, a protein of 127 kDa was detected as the dominant translation product (Figure 4.6. lane 3). A 10 kDa difference between the theoretical 117 and 127 kDa proteins seen on the autoradiograph should have been detectable

using this method. It is not certain why the translation product of the truncated ORF has a larger than expected molecular weight. This phenomenon was repeatedly recorded in both the wheat germ extracts and rabbit reticulocyte lysates charged with this expression plasmid. SDS-PAGE analysis using high percentage acrylamide ratios were used in the analysis of these translation products with the aim of detected the 17 kDa protein. No labelled proteins were detected using this method.

The possibility that translation of these p130 coding sequences is being initiated at an out of frame methionine relative to the frame of the p130 ORF, resulting in this larger than expected protein product was considered. However, the first eight AUG codons (at least one of which is positioned in one of the two other reading frames) would not be translated into a protein of a significant molecular weight nor do they exhibit better Kozak sequences for the initiation of translation. The coding sequence starting at the second in-frame methionine residue was predicted to be translated into 117 and 6 kDa proteins and yet no detectable difference using this technique could be established between the observed 127 kDa peptide and the other two p130-derived products. The possibility also exists that the p130 frame is translated beyond the first in-frame stop codon, UAA (nts 3661 – 3663) since a methionine residue is the next codon directly after this stop codon. The extension of the p130 coding sequence beyond the first and even a second downstream stop codon only increases the molecular weight of this translation product to 132 and 137 kDa, respectively making this possibility unlikely for the increase of 10 kDa in translation products produced from both pSTA2 and pSTA3.

A protein of 139 kDa was detected at low levels in all three of the *in vitro* translation reactions charged with the p130 constructs, but its significance and relatedness to the p130 coding sequence is unknown as C-terminal extension of this coding sequence is unlikely because of the in-frame stop codons in the adjacent downstream sequence. Another possibility is that the molecular weights of these three products are larger than predicted and that the product translated from the full-length ORF is the larger of the

two potential products resulting from processing at the 2A-like site and thus of a similar molecular weight as the two other truncated translation products. The smaller of the two resultant processing products, with a predicted molecular weight of 17 kDa, might be labile explaining its absence using this method of detection.

Rabbit reticulocyte lysate coupled in vitro transcription/translation of the p104 ORF

The coding sequences of p40 (pSTA4) and p104 (pSTA5) were analysed by radiolabelling of translation products using the rabbit reticulocyte lysate *in vitro* coupled transcription/translation system. SDS-PAGE analysis was performed on immunopreciptated protein extracts originating from lysate charged with pSTA4 and pSTA5, followed by either autoradiography of the dried resolving gel and Western analysis of lysates using anti-p40 antiserum. A 40 kDa protein was expressed from pSTA4 (Figure 4.7 panel A, lane 2) and a corresponding 40 kDa protein also detected by Western analysis of immunoprecipitated translation products using anti-p40 antiserum (Figure 4.7 panel B, lane 1). A 104 kDa protein was the dominant translation product of pSTA5 (containing the native p104 ORF), although a faint 40 kDa band corresponding to p40 was also detected (Figure 4.7 panel A, lane 3). The relative levels of p40 vs. p104 were not equivalent to those seen in the metabolically labelled translation products analysed in MG8 cells. Perhaps this is a consequence of the animal *in vitro* transcription/translation system used. Both the 104 kDa and 40 kDa proteins were detected by Western analysis of immunoprecipitated translation products using anti-p40 antiserum (Figure 4.7 panel B, lane 2).

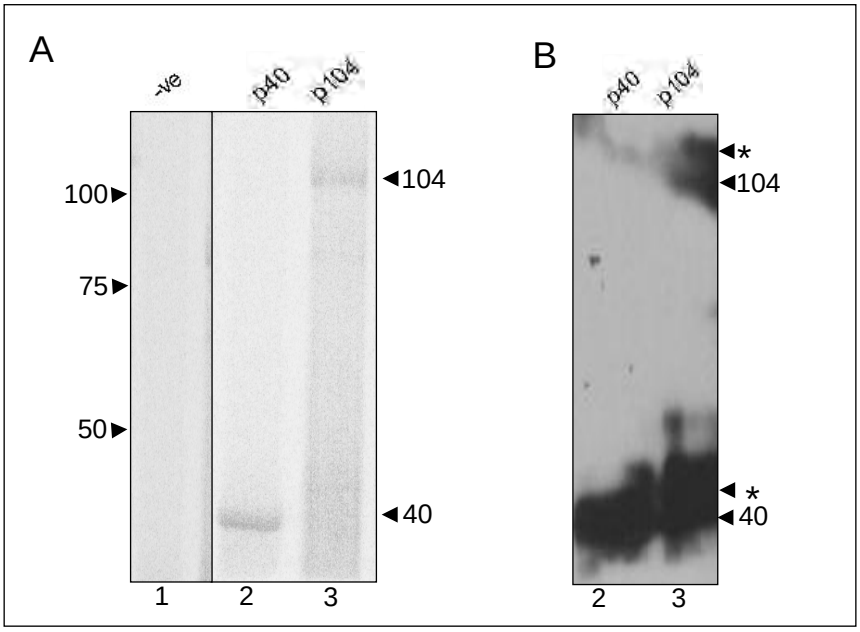

Figure 4.7. Detection of coupled *in vitro* transcription/translation products of constructs expressing p40/p104 in rabbit reticulocyte lysates. (A): Autoradiograph of SDS-PAGE analysis of rabbit reticulocyte lysates charged with pSTA4 and 5. Lane 1: control reaction with no input plasmid DNA (-ve). Lane 2: pSTA4 (p40). Lane 3: pSTA5 (p104). (B): Western analysis of immunoprecipitated protein (anti-p40 antiserum) in lysate charged with pSTA4 (lane 2) and pSTA5 (lane 3) using anti-p40 antiserum (1 in 50 000 dilution). The asterisks (*) indicate cross-reacting rabbit IgG heavy and light chains present as a result of the immunoprecipitation step. The molecular weight standards are indicated on the left of the autoradiograph and Western blot and are represented in kDa, while the arrows on the right indicate the molecular weights of radiolabelled proteins and immunoreactive proteins, respectively.

Since immunoprecipitation of these translation products was required prior to Western analysis (the sensitivity of the anti-p40 serum was not particularly high when probed against pSTA5-charged lysate alone), cross-reacting rabbit immunoglobulin heavy and light chains were also detected. These cross-reacting proteins of approximately 50 and 110 kDa did not interfere with the signals detected for p104 and p40.

Rabbit reticulocyte lysate coupled in vitro transcription/translation of the p81 ORF

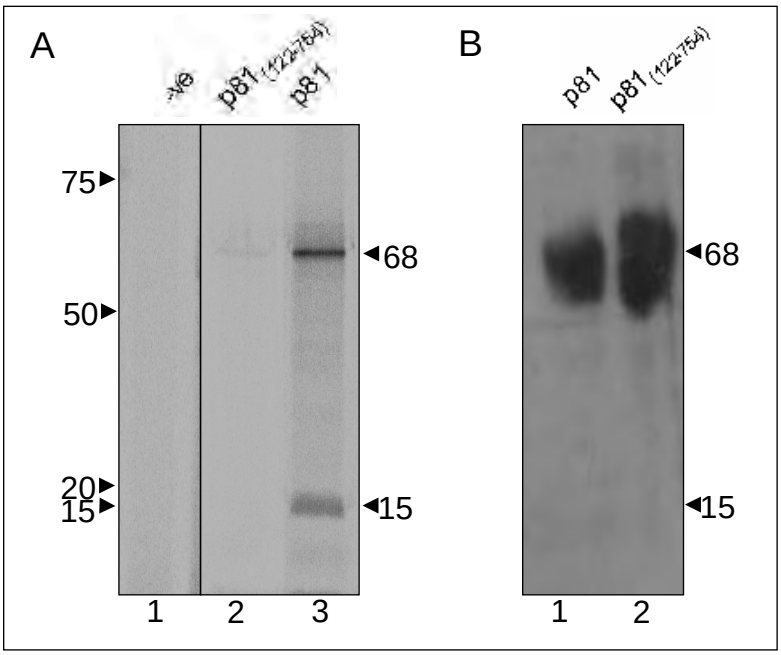

Figure 4.8. Detection of coupled *in vitro* transcription/translation products of constructs expressing p81 in rabbit reticulocyte lysates. (A): Autoradiograph of SDS-PAGE analysis of rabbit reticulocyte lysates charged with pSTA7 and pSTA8. Lane 1: control reaction with no input plasmid DNA (-ve). Lane 2: pSTA8 ($p81_{122-754}$). Lane 3: (p81). (B): Western analysis of lysates charged with pSTA7 and pSTA8 by Western analysis using anti-PrV capsid antiserum. Lane 1: pSTA7 (p81). Lane 2: pSTA7 ($p81_{122-754}$). The molecular weight standards are indicated on the left of the autoradiograph and Western blot and are represented in kDa, while the arrows on the right indicate the molecular weights of radiolabelled and immunoreactive proteins, respectively.

Expression plasmids carrying the coding sequences of the virus capsid protein precursor (p81) (pSTA7) and a truncated version, where the start of translation occurs after the second 2A-like processing (pSTA8 – $p81_{aa121-754}$), were analysed using the rabbit reticulocyte lysate coupled *in vitro* transcription/translation system (Figure 4.8). A 68 kDa translation product corresponding to the processed capsid protein precursor (p68) was detected in lysates charged with both pSTA7 and pSTA8 (Figure 4.8 panel A, lanes 2 & 3 and panel B, lanes 1 & 2). The bands in panel B, although over exposed, are distinct. The reason for this overexposure was to confirm that no 15 kDa protein cross-reacted with the anti-PrV antiserum. The amount of capsid protein originating from the truncated version of the

coding sequence (pSTA8) was much less than that of the equivalent translation product of the full length version of the p81 coding sequence (pSTA7). In addition, a translation product migrating at approximately 15 kDa was detected in the lysates charged with the plasmid expressing the wild-type p81 ORF (Figure 4.8. panel A, lane 3). If the co-translational processing of the p81 capsid precursor ORF occurs, a 13 kDa protein would be produced during translation of p81 in addition to the 68 kDa capsid protein precursor. A single dominant protein of 68 kDa and two dominant proteins of 68 and 15 kDa were detected in lysates charged with the plasmid pSTA8, expressing $p81_{122\text{-}754}$, and pSTA7, expressing p81 (Figure 4.8 panel A, lane 3), respectively. Thus, this result confirms that the second 2A-like cleavage site on the p81 coding sequence (PrV-$2A_3$) is active in the rabbit reticulocyte lysates and that this sequence results in the capsid precursor protein processing mechanism employed by PrV. This co-translational processing also follows the same sequential order and mechanism described in the capsid precursor processing of the betatetravirus, TaV (Pringle *et al*., 2001; Donnelly *et al*., 2001b). The 15 kDa translation product resulting from pSTA7 (Figure 4. 8 panel A, lane 3) is not immunoreactive to anti-PrV capsid anti-serum, suggesting that this peptide may not be present in PrV virion (Figure 4.8 panel B, lane 2). No mature capsid protein (p60) was detected, indicating that no particle maturation had occurred. It is likely that this is due to inefficient assembly *in vitro* of the capsid protein precursor into the provirion in this cell lysate environment which is buffered and will therefore not decrease in pH over the course of infection. It is this decrease in pH (possibly the result of apoptosis that typically results in the particle maturation of tetraviruses in the host cellular environment. Another possibility is that, in the reticulocyte lysate environment, processing of the capsid precursor into mature capsid proteins takes longer than the 1.5 hour time allowed for translation of proteins in this system. This explanation of a sub-optimal protein processing system would also explain the large percentage of p68 present in purified PrV and in the in vivo labelling system.

Wheat germ extract coupled in vitro transcription/translation of PrV translation products

The results of bioinformatic analysis of the PrV replicase suggested a close phylogenetic relationship with the tombus- and umbraviruses (Chapter 2, Sections 2.3.2 and 2.3.5), which are both plant virus families. It was therefore decided to also examine the translation products and co-translational processing of the PrV ORFs in a wheat germ extract coupled *in vitro* transcription/translation system. Plasmid constructs expressing the three ORFs were therefore used to charge a wheat germ extract system, and the [^{35}S]-methionine-labelled translation products were analysed by SDS-PAGE, followed by autoradiography (Figure 4.9).

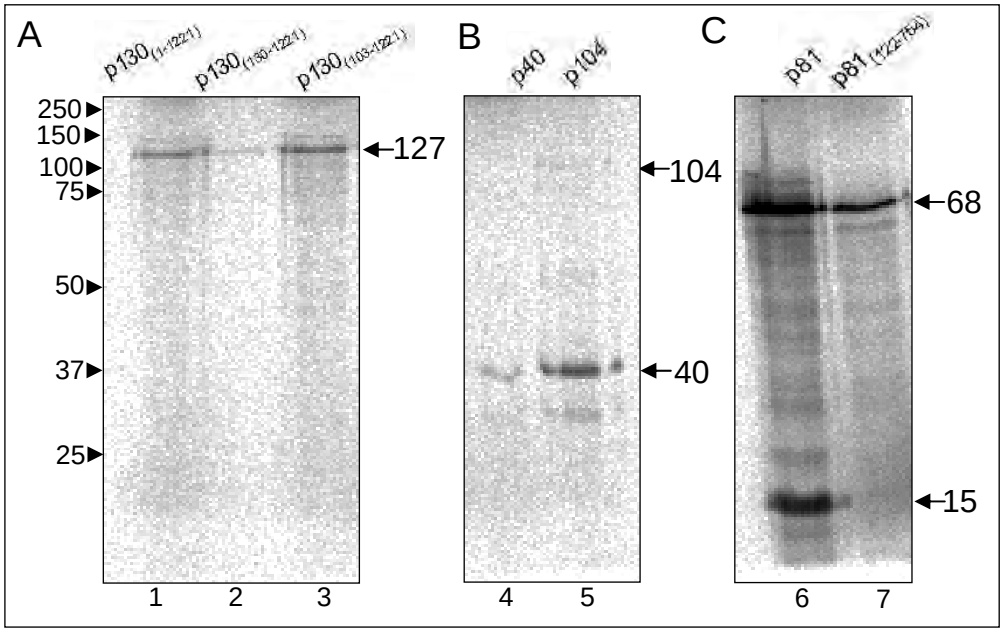

Figure 4.9. Detection of coupled *in vitro* transcription/translation products of constructs expressing p130, p40/p104 and p81 in wheat germ extracts. (A): Autoradiograph of SDS-PAGE analysis of wheat germ extracts charged with pSTA1 – 3. Lane 1: pSTA1 (p130). Lane 2: pSTA2 ($p130_{130-1221}$). Lane 3: pSTA3 ($p130_{103-1221}$) extract charged with pSTA3. (B): Autoradiograph of SDS-PAGE analysis of wheat germ extracts charged with pSTA4 and pSTA5. Lane 4: pSTA4 (p40). Lane 5: pSTA5 (p104). (C): Autoradiograph of SDS-PAGE analysis of wheat germ extracts charged with pSTA7 and 8. Lane 6: pSTA7 (p81). Lane 7: pSTA8 ($p81_{122-754}$). The molecular weight standards are indicated on the left of the autoradiographs and are represented in kDa, while the arrows on the right indicate the molecular weights of radiolabelled proteins.

As was observed in the rabbit reticulocyte system, the three plasmids expressing the p130 coding sequence (pSTA1), a truncated version translated from after the 2A-like processing site (pSTA2) and a third variant where the coding sequence starts at an alterative translation start (pSTA3), produced a radiolabelled protein of approximately 127 kDa (Figure 4.9 panel A, lanes 1 – 3). The absence of a 17 kDa protein from translated pSTA1 and pSTA3 indicated that no co-translational processing of the p130 was occurring at the PrV-$2A_1$ site. This indicated that the 2A-like cleavage site was not active under these *in vitro* experimental conditions.

Wheat germ lysates charged with pSTA4, containing the p40 ORF produced a radiolabelled protein of 40 kDa corresponding to p40 (Figure 4.9 panel B, lane 4). In the case of the constructs expressing the p104 ORF (pSTA5), a 40 kDa and a 104 kDa protein, corresponding to p40 and p104, respectively, were detected in the wheat germ extracts (Figure 4.9 panel B, lane 5). However, unlike the rabbit reticulocyte system, p40 was expressed at significantly higher levels relative to p104 in the wheat germ system suggesting that the suppression of the read-through stop signal was less efficient in the plant system. The levels of p40 relative to p104 in the wheat germ system were similar to those observed in PrV-infected MG8 cells (Figure 4.5 panel B, lane 1). p40 (pSTA4) was translated at high levels in this system and the smaller protein of about 32 kDa detected with this product and other products translated at moderate to high levels (Figure 4. 9 lanes 4 - 7), is most likely an endogenous protein resulting from residual aminoacyl tRNAs - an artefact often observed when using this *in vitro* system. p104 (pSTA5) was translated into a 104 kDa protein and high levels of a 40 kDa protein (Figure 4.9, lane 5).

The two plasmids expressing the capsid precursor ORF, as well as a truncated version starting immediately after the 2A-like cleavage site (pSTA7 and 8), resulted in the detection of a 68 kDa translation product consistent with the size of the capsid protein precursor following 2A-like processing (Figure 4.9 panel C, lanes 1 and 2). A second protein of 15 kDa was detected in wheat germ lysates charged with pSTA7 (expressing

the full p81 ORF), attributed to ‘p13’, the peptide resulting from 2A-like processing of p81 since this protein was not produced by pSTA8. Thus, as observed in the rabbit reticulocyte system, the PrV-$2A_3$ processing sequence appeared to be active, while no PrV-$2A_2$ activity was observed.

4.4. CONCLUSION

In vivo labelling of viral translation products in the persistently PrV-infected MG8 cells detected the presence of seven radiolabelled proteins, of which six corresponded to the predicted molecular weight of the translation products of p130, p104/p40 and p81 (p68 and p60), originating from the predicted ORFs. Western analysis confirmed the presence of proteins corresponding to p40 and p104, indicating that the read-through stop codon identified in the p104 ORF is functional *in vivo*. The capsid protein precursor (p68) and major capsid protein (p60) were also detected, confirming the results of Pringle *et al.* (2003). However, the three peptides predicted to result from 2A-like processing at the N-terminus of p81 to produce p68 were not detected and instead a single small 15 kDa protein and the capsid precursor protein of 68 kDa were observed in both systems.

The expression of p130 in virus-infected cells remains to be confirmed and is dependent on the availability of p130-specific antibodies. Also, a future experiment should involve the isolation of (over)expressed p130 followed by the MS analysis of this peptide to confirm its amino acid sequence and thus translational origins. The results of the *in vitro* coupled transcription/translation systems showed that p130, p104 and p81 were translated in both animal and plant systems. In addition, it was demonstrated that the read-through stop codon present in p104 was functional, since the presence of p104 and p40 was detected. However, it appeared that the read-through activity of this stop codon was different in the two translation systems, with more p104 produced in the rabbit reticulocyte system. The levels of p40 relative to p104 in the wheat germ translation system corresponded to the levels of both proteins detected in

PrV-infected cells, supporting the bioinformatic data that the read-through signal could have originated from a plant virus. Only one of the three 2A-like sequences, PrV-$2A_3$, appears to be active *in vitro.* This contradicts the data of Luke *et al*. (2008), who reported that both PrV-$2A_1$ and PrV-$2A_2$ were active *in vitro*, although not tested in the context of the wild-type surrounding coding sequence on either side of these processing signals. An interesting experiment to perform would be to measure the activities of these processing sites using a reporter system in the context of the wild-type sequence.

CHAPTER 5

Establishing a system for studying the subcellular distribution of p40 and viral RNA in MG8 cells

5.1. INTRODUCTION

The replicase proteins of (+)ssRNA viruses are known to encode at least an RNA-dependent RNA polymerase catalytic core. This domain associates itself with a subcellular membrane within the host cell and makes up a critical component of the viral replication complex (Ahlquist *et al*., 2003). This association with a subcellular membrane that usually originates from an organelle, is thought to be important in: (1) physically shielding the newly produced viral RNA from host defences such as RNAi (Jonczyk *et al*., 2007), (2) providing a physical anchor for the replication complex to provide the essential membrane lipids required for polymerization and (3) as a means of concentrating these components into small areas within the host cellular environment (Lee *et al*., 2006). The subcellular localization and thus the variety of organelles used for this purpose vary substantially, even between viruses of the same genus (Ahlquist *et al*., 2003).

Another important morphological structure, which has been observed at the site of some replicating viruses, is the aggresome (Wileman, 2006; Netherton *et al*., 2008). Aggresomes act as a component-concentrating subcellular structure and, although well-characterized in terms of their role in the replication of large dsDNA viruses (Wileman, 2006), little is known about their biological importance and frequency of occurrence during the infection of a cell by certain ssRNA viruses. The aggresome is surrounded by filaments of vimentin and these complexes have been shown to recruit chaperones, proteasomes and mitochondria, most likely to facilitate as a site for protein folding and degradation (Wileman, 2006). The default protein degradation pathway in these structures must be altered in order to

keep the replication complex intact (Wileman, 2006). Pausing this cellular pathway might entail halting the entry of autophagosomes into the aggresome and/or subsequent fusion of the complex with lysosomes (reviewed in Wileman, 2006).

The presence of either structurally altered subcellular organelles or aggresomes during virus infection are two key subcellular structures indicating the presence of replicating viral RNA (Netherton *et al*., 2008). Information on these structures, as well as knowledge on the site of replication of related viruses, is used in predicting this site within a virus-infected host cell (reviewed in Netherton *et al*., 2008). Data analyzed and discussed in Chapters 2 and 4 has revealed that PrV may employ a unique replication strategy compared to other members of the *Tetraviridae* family. Like the tombusviruses, PrV might be utilizing an auxiliary protein and a full- length replicase protein for replication of its viral genome. Based on prior reports of the co-localization of the auxiliary and replicase proteins at the site of tombusvirus replication (Scholthof *et al*., 1995; Weber-Lotfi *et al*., 2002; Pantaleo *et al*., 2004), it is hypothesized that PrV p40 recruits and then co-localizes with other p40 molecules, p104, specific host proteins and viral RNA to a membrane-associated subcellular site of replication within the PrV-infected cells.

Based on this hypothesis, this chapter serves to investigate the subcellular distribution of p40 both in transiently expressing cells and within the PrV-infected MG8 cell line. Fluorescence microscopy and biochemical fractionation techniques were developed to investigate the distribution of transiently expressed p40 fused to enhanced green fluorescent protein (eGFP) in *Sf*9 cells and ‘endogenous’ p40 (and p104) in persistently infected *H. zea* MG8 cells. The distribution of viral RNA in the infected MG8 cell line was also examined and compared with the subcellular distribution of p40/p104.

5.2. MATERIALS AND METHODS

Unless otherwise stated, all gene manipulation methods were performed according to Sambrook *et al.* (1989) and protein analysis as described by Laemmli (1970). The cloning strain of *E. coli* (DH5α) (Hanahan, 1983) was used to propagate and clone all plasmids described below. Primers sequences are listed in Appendix 1 and additional montages of confocal images are shown in Appendix 7.

5.2.1. Expression plasmid design and construction

The plasmid pA1 (generously supplied by Prof. Vernon Ward, Department of Microbiology and Immunology, School of Medical Sciences, University of Otago, Dunedin, New Zealand) contains the kanamycin resistance and ampicillin resistance genes, the SV40 origin of replication and the baculoviral *AcMNPV* polH promoter. The coding sequence of eGFP was cloned into this vector by digesting pA1 and the commercial mammalian expression vector, pEGFP-N1 (Clontech), with *Hin*d III and *Not* I. The 779 bp *Hin*d III/*Not* I fragment containing the eGFP ORF from pEGFP-N1 was ligated into the *Hin*d III/*Not* I backbone of pA1. The resultant construct, pCW31, contains the eGFP coding sequence immediately downstream of the *AcMNPV* polH promoter (Figure 5.1).

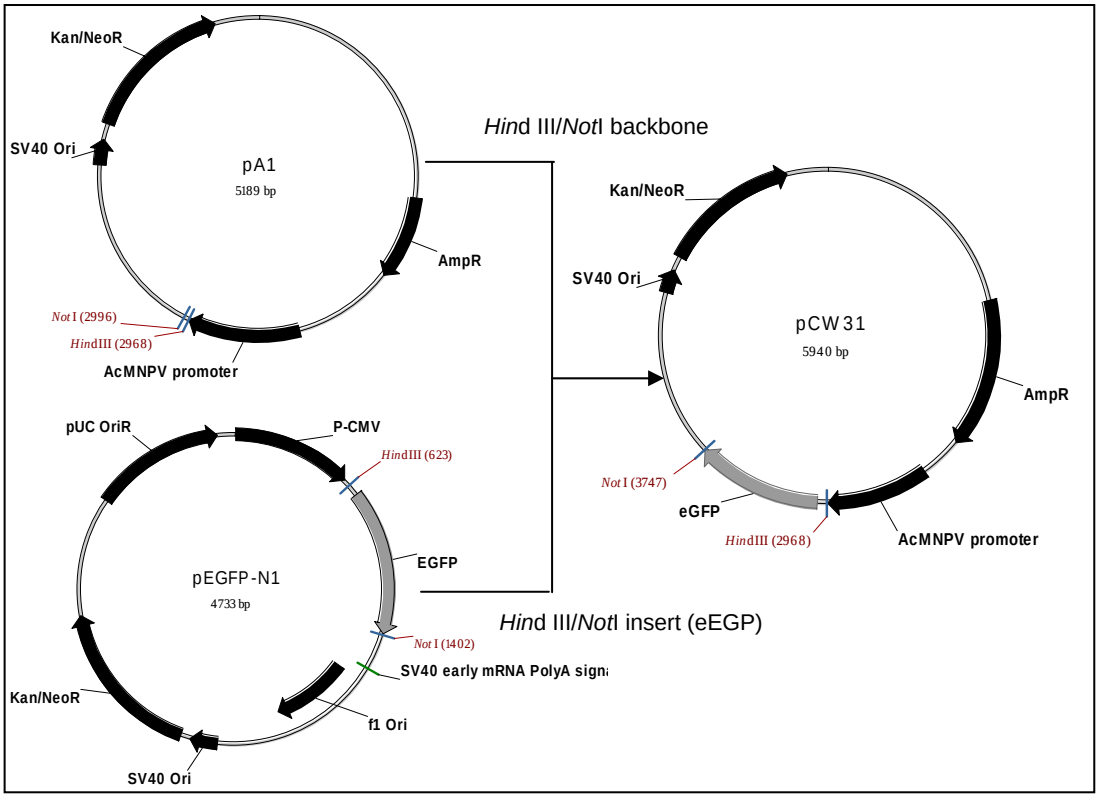

Figure 5.1. Cloning strategy used in the construction of the expression plasmid, pCW31. AmpR: *β-lactamase gene*; pUC oriR: bacterial origin of replication; SV40 Ori: *Simian vacuolating virus* 40 origin of replication; SV40 poly A signal: *Simian vacuolating virus 40* poly-A sequence; Kan/NeoR: *kanamycin/neomycin phosphotransferase*, P-CMV: *cytomegalovirus* promoter; f1 Ori: F1 bacteriophage origin or replication.

The p40 coding sequence was amplified from pFLM1 (containing the entire PrV genome) using the primers CW9 (which introduces a *Nde* I site at the 5`-end of the p40 coding sequence) and CW13 (which creates a *Bcl* I site in place of the read-through stop codon). The 1085 bp amplicon was ligated into the commercial vector pGEM-T Easy (Promega) to produce pCW35. pCW35 was then digested with *Sac* II and *Bcl* I and the insert containing the p40 coding sequence minus the UAG stop codon was ligated into the *Sac* II/*Bam* HI-digested backbone of pCW31 (Figure 5.2). The final construct, pCW39, is therefore an insect expression vector driving the expression of p40-eGFP under the control of the *AcMNPV* promoter.

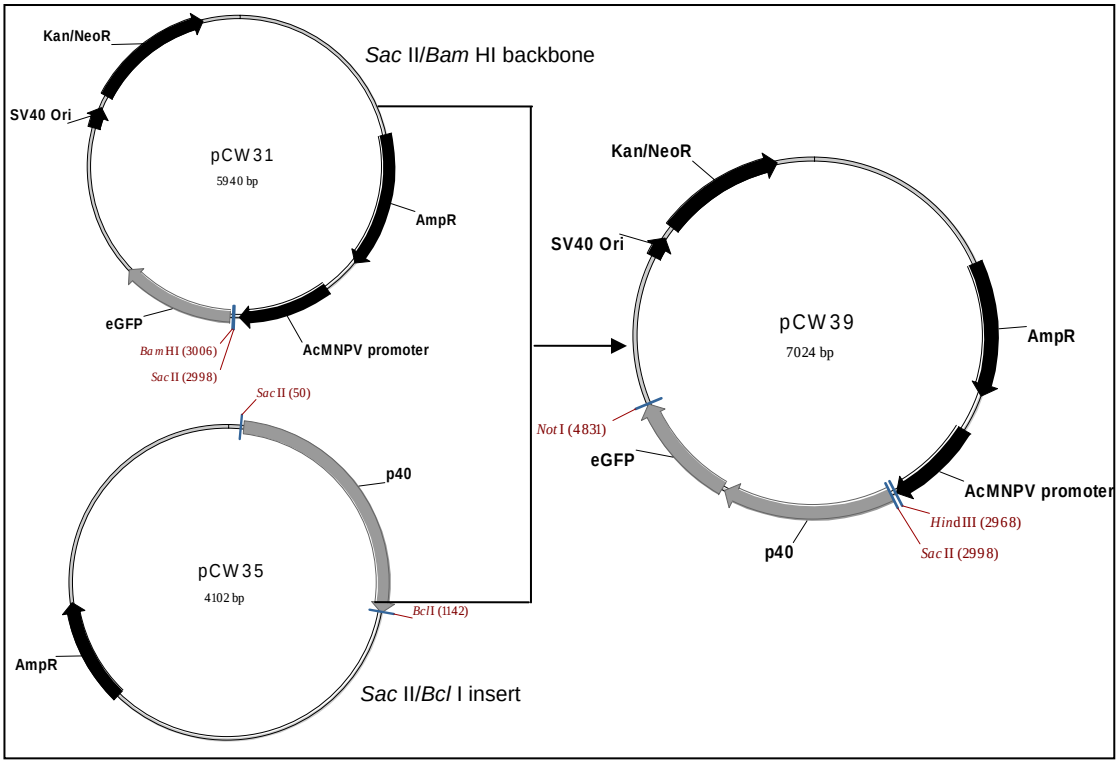

Figure 5.2. Cloning strategy used in the construction of the expression plasmid, pCW39. AmpR: *β-lactamase gene*; pUC oriR: bacterial origin of replication; SV40 Ori: *Simian vacuolating virus* 40 origin of replication; SV40 poly-A signal: *Simian vacuolating virus* 40 poly-A sequence; Kan/NeoR: *kanamycin/neomycin phosphotransferase*.

An expression vector, expressing the p40-eGFP fusion protein under the control of an enhanced promoter, was derived from the expression vector phrA1GFP (generously supplied by Vernon Ward, Department of Microbiology and Immunology, School of Medical Sciences, University of Otago, New Zealand), which carries an enhancer-promoter fusion – *hr/AcMNPV*. The enhancer element comprises nine small 150-800 bp of repeated sequences derived from the genome of the baculovirus, AcMNPV. These nine repeats, referred to as homologous regions (*hrs*) (Ayers *et al*., 1994), have each been found to act as early transcriptional enhancers (Guarino *et al*., 1986; Guarino and Summers, 1986). These repeats have been engineered into insect expression promoters due to the marked increase these enhancers confer on insect virus promoter activity (Lo *et al*., 2002). The enhanced baculovirus promoter (*hr/AcMNPV*) was PCR amplified from phrA1GFP using primers CW19 (introducing an *Aat* II restriction site at the start of the enhanced promoter sequence) and CW20 (introducing a *Hin*d III restriction site at the 3`-end of the enhanced

promoter sequence). The 1211 bp amplification product was directly ligated into pGEM-T Easy vector and the resultant plasmid (pCW44) sequenced. Amplicons that had ligated into pGEM-T Easy vector in the correct orientation (pCW44R) were used in the subsequent sub-cloning experiments, whereby the *AcMNPV* promoter in pCW39 was substituted for the enhanced version, hr/AcMNPV. The expression plasmid, pCW55, was constructed by digesting pCW44R with *Aat* II and *Sac* II, and the resultant 1222 bp fragment containing the enhanced promoter ligated into the *Aat* II/*Sac* II backbone of pCW39 (Figure 5.3). The plasmid, pCW55, contained the hr/AcMNPV promoter driving the expression of p40 fused to the N-terminus of eGFP.

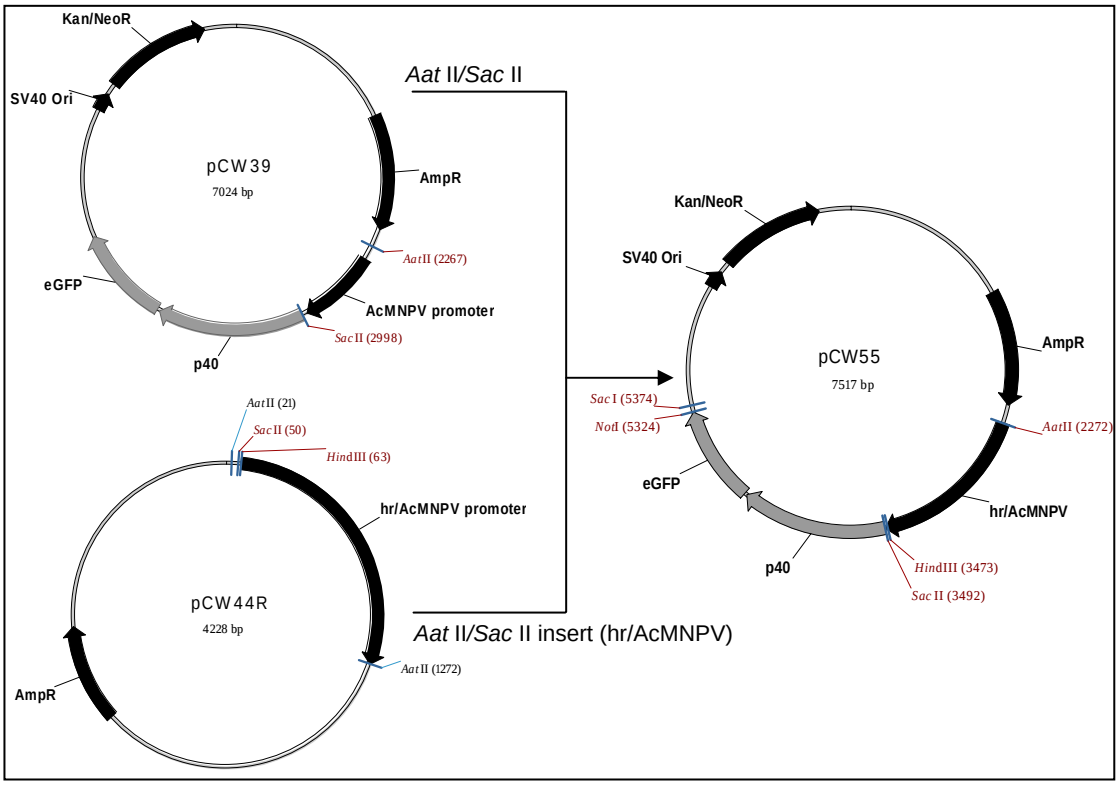

Figure 5.3. Cloning strategy used in the construction of the expression construct, pCW55. AmpR: *β-lactamase gene*; pUC oriR: bacterial origin of replication; SV40 Ori: *Simian vacuolating virus* 40 origin of replication; SV40 poly-A signal: *Simian vacuolating virus* 40 poly A sequence; Kan/NeoR: *kanamycin/neomycin phosphotransferase*.

In order to construct an expression construct with the p40 coding sequence fused to the C-terminus of eGFP, the *Aat* II/*Sac* II fragment of pCW44R, containing the enhanced promoter, was ligated into the *Aat* II/*Sac* II-digested backbone of pCW31. The resultant insect expression plasmid, pCW56, contained the enhanced promoter (*hr/AcMNPV*) driving the expression of eGFP (Figure 5.4).

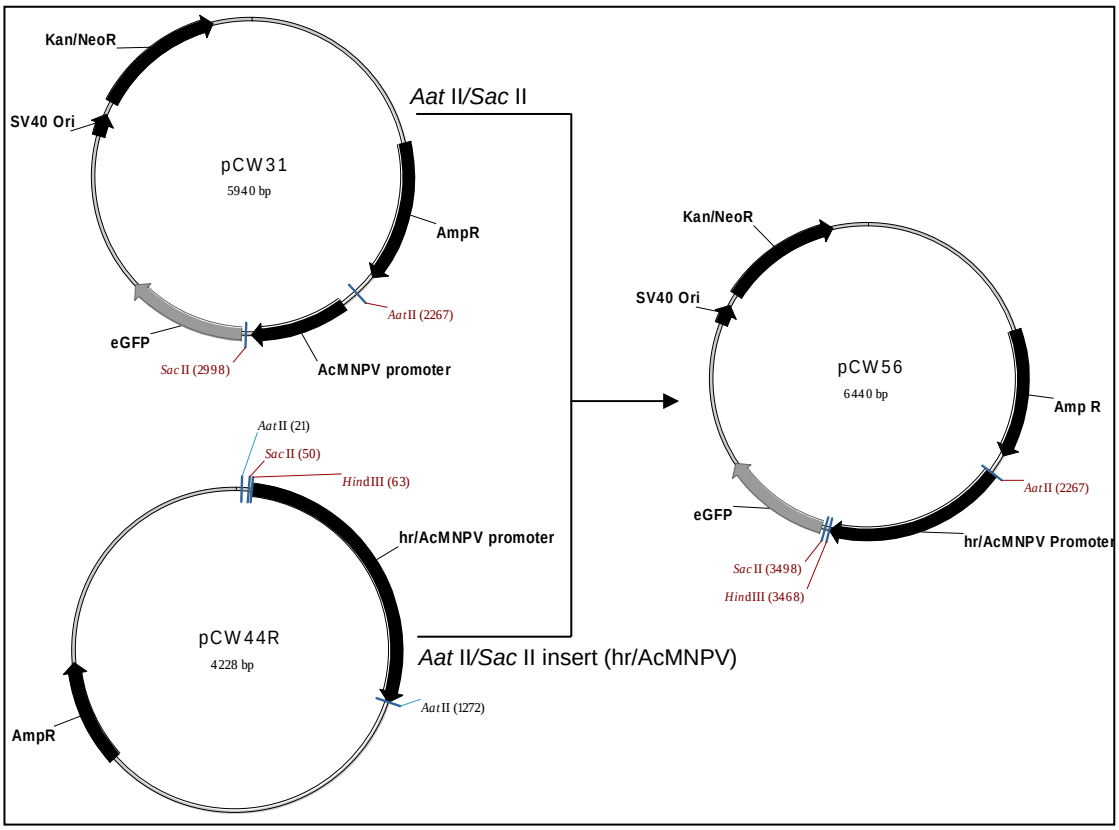

Figure 5.4. Cloning strategy used in the construction of the expression plasmid, pCW56. AmpR: *β-lactamase gene*; pUC oriR: bacterial origin of replication; SV40 Ori: *Simian vacuolating virus* 40 origin of replication; SV40 poly-A signal: *Simian vacuolating virus* 40 poly A sequence; Kan/NeoR: *kanamycin/neomycin phosphotransferase*.

The commercial mammalian expression vector, pEGFP-C1 (Clontech), was linearized with *Eco* RI. pCW35 was digested with *Eco* RI, and a 1103 bp fragment (containing the p40 coding sequence) was ligated into the linearized pEGFP-C1 to construct pCW54 – a mammalian expression vector driving the expression of the eGFP-p40 fusion protein. Using the existing eGFP-p40 fusion within the plasmid pCW54, this fragment was excised by digestion of pCW54 with *Age*I/*Nsi* I and the 2832 bp fragment

ligated into the *Age*I/*Nsi* I backbone of pCW56, thus substituting the eGFP in pCW56 with the eGFP-p40 fusion removed from the vector, pCW54. The resultant insect expression plasmid, pCW57, contained the enhanced promoter immediately upstream of the eGFP-p40 fusion (Figure 5.5).

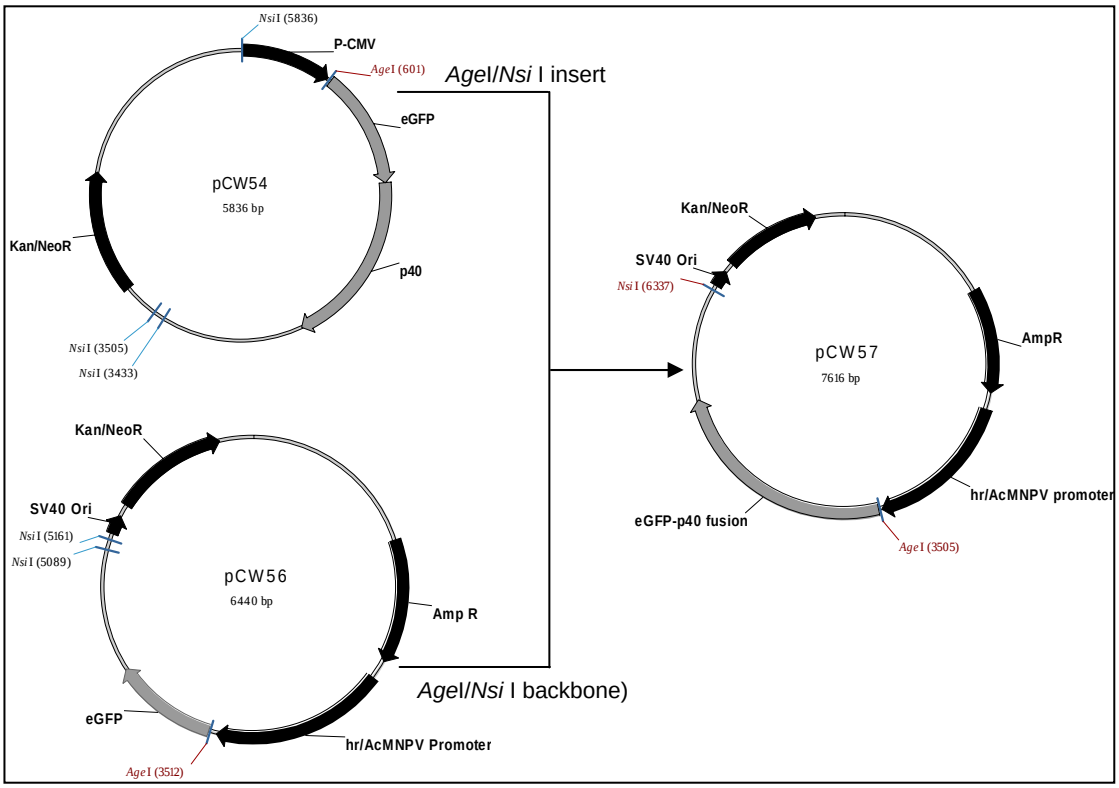

Figure 5.5. Cloning strategy used in the construction of the expression construct, pCW57. AmpR*: β-lactamase gene*; pUC oriR: bacterial origin of replication; SV40 Ori: *Simian vacuolating virus* 40 origin of replication; SV40 poly A signal: *Simian vacuolating virus* 40 poly A sequence; Kan/NeoR: *kanamycin/neomycin phosphotransferase*, P-CMV: *cytomegalovirus* promoter.

The leader sequences, invertebrate Kozak sequences (Caverner, 1987, Sano *et al*., 2002), of each expression vector as well as the correct stop codon at the end of each fusion protein were confirmed by sequence analysis. Due to the presence of the promoter enhancer element in the two expression vectors, pCW55 and pCW57, these two plasmids were selected for the expression of large quantities of p40-eGFP and eGFP-40 protein, respectively, in transfected insect cells.

5.2.2. Transfection of insect cell lines

*Sf*9 monolayers were transfected using a method obtained from Prof, Vernon Ward (Department of Microbiology and Immunology, School of Medical Sciences, University of Otago, Dunedin, New Zealand, pers. comm., 2006). Single glass coverslips were added to single wells of a 24-well cell culture plate. 1 ml of complete TC100 medium was added to the coverslip followed by the addition of approximately 1.8 X 10^5 *Sf*9 cells resuspended in fresh, complete medium. The plate was gently agitated in order to disperse the cells evenly on the coverslip and then incubated at 28 ºC until the cell monolayer reached 50 % confluence. 1 µg of plasmid DNA (prepared using Qiaprep Spin Miniprep columns by Qiagen®) was added along with 4 µl of Fugene® HD (Roche Applied Science) to serum-free medium to make up a final volume of 100 µl. This transfection mixture was incubated at room temperature for 45 minutes after which it was added drop-wise to the *Sf*9 cells growing onto the coverslip. The transfection mixture was left at 28 ºC for 4 hours, after which a further 1 ml of complete medium was added and the cells incubated at 28 ºC for 3 days or until each monolayer had reached complete confluence. The medium was removed from the well and rinsed once with 400 µl of phosphate-buffered saline (PBS) (pH 7.4). To fix the cells on the coverslips, 400 µl of paraformaldehyde (4 % w/v in PBS pH 7.4) was added to each well and incubated at room temperature for 20 minutes. Each coverslip was rinsed twice in 400 µl of PBS for 10 minutes at room temperature, with gentle agitation. Following the first two washes in PBS, the nuclear stain, 4', 6-diamidino-2-phenylindole (DAPI), was added to a final concentration of 5 µg/ml in 400 µl of PBS for 5 minutes and the coverslip rinsed once more in PBS. Finally, the coverslips were rinsed in triple-distilled water, the reverse side of the slip dried and the side containing the cells mounted face down into ~2.5 µl of mounting medium (Dako Cytomation) and placed in the centre of a glass slide.

5.2.3. Biochemical fractionation to determine subcellular distribution of p40 in insect cell lines

The following protocol, which is based on a modified protocol obtained from Prof. Tom Wileman (School of Medicine, Health Policy and Practise, University of East Anglia, Norwich, England), was used to obtain subcellular fractions of insect cells transiently expressing p40-eGFP fusions. The protocol was validated using the mitochondrial membrane protein, VDAC, (a known membrane-associated 30 kDa protein) as a marker in subsequent Western analyses. Western analysis of VDAC protein using anti-VDAC antibodies (Santa Cruz Biotechnology) detected the 30 kDa mitochondrial protein in the post-nuclear and membrane fractions.

To fractionate membrane-associated proteins from insoluble protein and aggregates, a 100 % confluent monolayer of transfected *Sf*9 cells in a 24-well cell culture plate, transfected with pCW55 or pCW57, was resuspended in the existing growth medium. The cells were pelleted by centrifugation at 5 000 g for 2 minutes at room temperature and the cellular pellet resuspended in 1 ml of chilled buffer (50 mM Tris-Cl, pH 7.4; 1 mM EDTA; 8 % (w/v) sucrose and one Complete™ EDTA-free protease inhibitor cocktail tablet per 50 ml of buffer (Roche Applied Science). The suspended cells were disrupted by passing through a 25 gauge needle twenty times. These lysates were then centrifuged at 3 300 g for 2 minutes at 4 ºC, the supernatant transferred to a fresh microfuge tube and the pellet re-extracted as described above. The pooled supernatant, referred to as the post-nuclear fraction due to the absence of the nuclei, was further fractionated to obtain a solubilized membrane fraction and an aggregates/insoluble protein fraction.

The post-nuclear fraction was centrifuged at 17 000 g for 20 minutes at 4 ºC and the resultant pellet resuspended in 500 µl of chilled IPB buffer (see Appendix 6) containing 2 % v/v Triton X-100. The resuspended extract was centrifuged at 17 000 g for 20 minutes at 4 ºC and the supernatant

(solubilized membrane fraction) transferred to a new microfuge tube. The pellet was resuspended in 500 µl of chilled IPB buffer containing 2 % v/v Triton X-100, centrifuged at 17 000 g for 20 minutes at 4 ºC and the supernatant added to the solubilized membrane fraction. The pellet was resuspended (and solubilized for further analysis) in 500 µl of 50 mM HEPES buffer pH 7.4 containing 1 % v/v SDS. The protein concentration in these fractions was quantified by Bradford's assay (Bradford, 1976), then concentrated by ethanol precipitation, if necessary (Chapter 3, Section 3.2.2) and resolved by SDS-PAGE followed by Western analysis. A total of 2 µg from each fraction was electrophoresed, in duplicate, in 12 % glycine/acrylamide resolving gels, transferred to nitrocellulose membrane and subjected to Western analysis using polyclonal anti-p40 antiserum as described in Chapter 3, Section 3.2.2.

The MG8 cell line was subjected to the same subcellular fractionation protocol described above. Additionally, total membrane and cytoplasmic fractions were obtained by halving the post-nuclear fraction and subjecting one half to the procedure described above. The remaining half was fractionated as follows: the post-nuclear fraction was centrifuged at 17 000 g for 20 minutes at 4 ºC, the supernatant made up to 1 ml in 50 mM HEPES, pH 7.4 and the pellet resuspended in 1 ml of HEPES buffer. Both supernatant and pellet were centrifuged at 17 000 g for 20 minutes at 4 ºC, the supernatants pooled and made up to 4 ml in HEPES buffer followed by a final centrifugation at 17 000 g for 20 minutes at 4 ºC. The supernatant (cytoplasmic fraction) was transferred to a fresh microfuge tube while the resultant pellet (membrane fraction) was resuspended in 1 X IPB containing 2 % Triton X-100 (v/v). The protein concentration in subcellular fractions of MG8 cells was quantified and analysed as above and subjected to Western analyses using polyclonal anti-PrV and anti-p40 antisera as described in Chapter 4, Section 4.2.2.

5.2.4. Bromouridine labelling of viral RNA in the MG8 cells

The protocol for bromouridine labelling of viral RNA was carried out according to a protocol modified from Wansink *et al*., 1993 and Dr. D. J Miller (Department of Medicine and the Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbour, Michigan, USA, pers. comm., 2006). Single glass coverslips were added to wells of a 24-well cell culture plate. Approximately 5 X 10^5 MG8 cells resuspended in fresh, complete ExCell 420 medium were added to the coverslip to obtain 80 – 90 % confluence. The plate was gently agitated in order to disperse the cells evenly on the coverslip. Once settled, the cells were treated with a 20 µg/ml concentration of Actinomycin D as described in Chapter 4, Section 4.2.1, with the exception that complete medium was used during the addition of Actinomycin D. The labelling/transfection mixture was made up as follows: 5.1 µl of Fugene® HD (Roche Applied Science); 8.5 µl of 5` bromouridine 5` triphosphate sodium salt (BrUTP) (10 mM final concentration); 3.4 µl of Actinomycin D (20 µg/ml final concentration); and 68 µl of complete medium (to a final volume of 85 µl), gently mixed and allowed to equilibrate to room temperature for 15 minutes. Following the pre-treatment with Actinomycin D and equilibration of the labelling/transfection mixture, the existing medium over the monolayer of cells was removed and 85 µl of equilibrated mixture added drop-wise to each well. Following the observation by Pringle *et al*. (2003) that sufficient RNA synthesis was detected after 3 hours of [^{3}H]-uridine metabolic radiolabelling, MG8 cells were incubated in the presence of 20 µg/ml Actinomycin D for 3 hours at 28 ºC. After bromouridine incorporation, the MG8 cells were fixed as described above and subjected to immunofluorescence as described below.

5.2.5. Fluorescence microscopy

Visualization of p40-eGFP in transfected Sf9 cells

The expression plasmids, pCW55 and pCW57, were used to individually transfect *Sf*9 cells. After an appropriate period of time, the cells were fixed, stained with DAPI, mounted onto a glass slide and imaged by confocal microscopy. In all experiments, a Zeiss LSM 5-Meta inverted confocal laser scanning microscope (Carl Zeiss Optics®) was used to view transfected or labelled cells under a 63 X oil objective and images captured directly by the detector system. A dual track sequential line by line mode was set to scan and image 1 µm optical slices of cells at a resolution of 1768 x 1768 pixels unless otherwise stated. The helium/neon and argon lasers were set at excitation wavelengths of 488 (GFP) and 405 (DAPI) nm and attenuated to 2 - 10 % and 2 % laser power, respectively, during the imaging of cells in Figures 5.2 and 5.3 below. Two filter channels were set to detect wavelength emissions at 420 nm and 505 - 530 nm in order to visualize the eGFP and DAPI staining, respectively. All images were analyzed using the Zeiss LSM 5 Image Browser software program v 4.2.0.

Immunofluorescence and fixing of bromouridine- labelled MG8 cells

Following labelling, the medium was removed from each well containing a cell monolayer and rinsed once with 400 µl of PBS (pH 7.4). To fix the cells, 400 µl of paraformaldehyde (4 % w/v in PBS pH 7.4) was added to each well and incubated at room temperature for 20 minutes. Each well was rinsed once with 400 µl of PBS and the cells equilibrated in 400 µl of permeabilization buffer (1 % Triton X-100; 10 % sucrose; 5 % goat serum in PBS) for 20 minutes at room temperature, with gentle agitation. A 1 in 50 dilution of mouse monoclonal anti-BrUTP primary antibody (Sigma®) in a total of 400 µl of permeabilization buffer was added to each well and incubated for 90 minutes at room temperature with gentle agitation.

Following incubation in primary antibody, each well was rinsed three times in 400 µl of PBS-Tween (1 % v/v Tween-20) for 10 minutes at room temperature with gentle agitation. Next, the secondary antibody, anti-mouse antibody conjugated to Texas Red (sulphur rhodamine) (Invitrogen Molecular probes™) was diluted in 400 µl of permeabilization buffer, added to each well and incubated for 30 minutes at room temperature with gentle agitation. Each well was then rinsed three times in 400 µl of PBS-Tween for 10 minutes at room temperature, with gentle agitation. In the second of the three washes, DAPI was added to a final concentration of 5 µg/ml. The coverslips were rinsed in dddH_20 and mounted on glass slides as described in Section 5.2.2. The primary and secondary antibody dilutions were optimized and the fixed cells visualized as described in Section 5.2.5 using confocal laser scanning microscopy. The fluorophore, Texas Red, was excited using a helium/neon laser attenuated to 30-40 % and the fluorescence subsequently detected at 546 nm. The Texas Red fluorophore combined with the filter sets and lasers did not produce an optimal signal output but the use of this particular secondary antibody conjugate was unavoidable at the time of experimentation.

5.2.6. Immunofluorescence of p40 (and p104) in MG8 cells.

Immunofluorescence was used for the detection of p40/p104 in the MG8 cell line as described in Section 5.2.5. Primary rabbit anti-p40 antiserum (Chapter 3) and secondary antibody [goat anti-rabbit antiserum conjugated to fluorescein isothiocyanate (FITC) (Zymax™ by Zymed] dilutions were optimized and fixed cells visualized as described in Section 5.3.3, using confocal laser scanning microscopy. FITC was excited using argon laser attenuated to 1 - 10 % and the fluorescence subsequently detected at a wavelength of 488 nm. The MG8 cells were directly probed with anti-p40 antiserum (1 in 100 dilution) followed by the addition of the secondary antibody, goat anti-rabbit FITC conjugate (1 in 400 dilution).

Co-localization of PrV p40 and vRNA in MG8 cells

Due to a high degree of both green/yellow auto-fluorescence as well as low levels of p40 production in all the MG8 cells, the experiments below were optimized in order to gain preliminary evidence of vRNA and p40/p104, as well as other potentially cross-reacting cellular proteins co-localizing in these cells. Co-localization of the auxiliary protein, p40 (and p104 due to the use of the anti-p40 antiserum) and PrV vRNA (bromouridine labelling) in persistently infected MG8 cells required a dual immunofluorescence labelling of these two molecules. This was achieved following BrUTP incorporation in the presence of Actinomycin D so as to eliminate cellular RNA background signal. Accordingly, monolayers of MG8 cells were subjected to co-localization experiments by using both anti-p40/anti-rabbit FITC and anti-BrUTP/anti-mouse Texas Red primary and secondary antiserum/conjugate combinations in order to detect immunofluorescence as described in Sections 5.2.4 and 5.2.5.

Cells exhibiting typical p40 expression were identified and selected for imaging using filter combinations set to detect Texas Red, FITC and DAPI emissions. The lasers and settings remained unaltered from those described in Sections 5.2.4 and 5.2.5 and candidate cells (those with both strong Texas Red and FITC signals) were selected for imaging, followed by the FITC and finally DAPI channel to minimize the photobleaching effect of the shorter wavelengths on the specimen. The fluorescence signal from the Texas Red was digitally summed up four times so as to minimize background signal otherwise created by alternatively increasing the laser power in order to equalize this signal to that emitted by the FITC fluorophore. Three images of each candidate cell were independently acquired using separate channels, digitally superimposed and co-localization of the FITC and Texas Red signals statistically evaluated using the Zeiss LSM image browser software package.

5.3. RESULTS AND DISCUSSION

5.3.1. Visualization of *Sf*9 cells transfected with eGFP-p40 and p40-eGFP fusion proteins

The optimization of transfection of *Sf*9 cells was done using the plasmid pCW56, expressing just eGFP, as a measure of the transfection efficiency. The average transfection efficiency was measured at 25 %. Next, the plasmids pCW55 and pCW57, expressing both the N-terminally fused p40 to eGFP and C-terminally fused p40 to eGFP respectively, together with pCW56 (expressing only eGFP) were transfected into monolayers of *Sf*9 cells grown to confluence on coverslips. Two representative images of *Sf*9 cells transfected with pCW56 (eGFP), pCW55 (p40-eGFP) or pCW57 (eGDP-p40) are depicted in Figure 5.6 and additional images can be found in Appendix 7, Figures A.1 and A.2.

The expression of eGFP under the control of the *hr/AcMNPV* promoter in transfected *Sf*9 cells resulted in the successful overexpression of eGFP in these cells (Figure 5.6 panels A - B). Transfection efficiencies of 25 % were achieved in *Sf*9 cells transfected with pCW56. As expected, the general distribution of eGFP observed in transfected *Sf*9 cells was diffuse with no particular pattern of accumulation or specific localization in either the nucleus or cytoplasm.

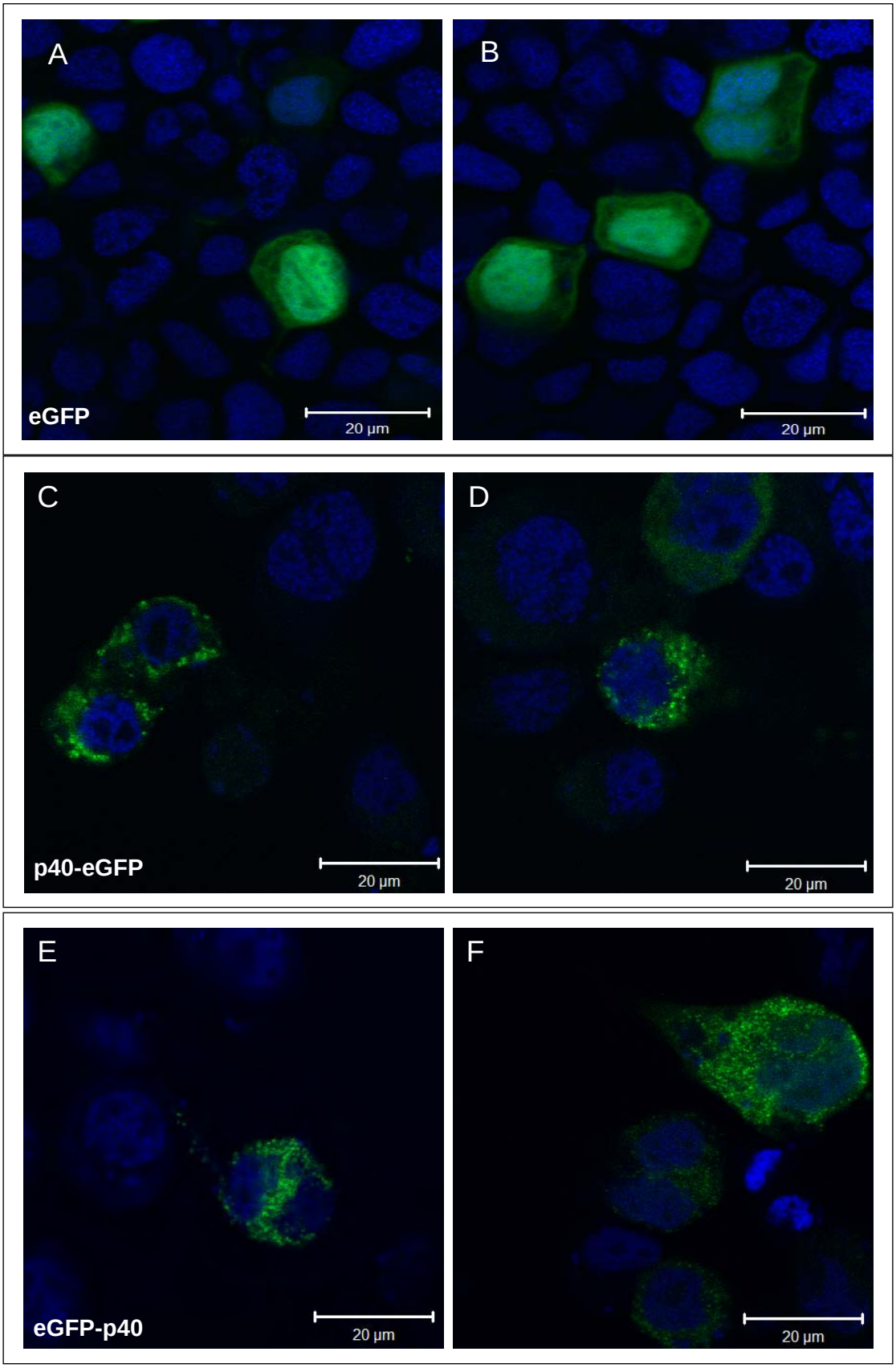

Figure 5.6. Distribution of eGFP and eGFP/p40 fusion proteins in transfected *Sf*9 cells. A-B: Distribution of eGFP in *Sf9* cells transfected with pCW56. C - D: Distribution of p40 fused to the N-terminus of eGFP in *Sf9* transfected with pCW55. E - F: Distribution of p40 fused to the C-terminus of eGFP in *Sf9* transfected with pCW57. Nuclei were stained with DAPI. Images A - F acquired at 2 X scanning zoom. Scale bars represent 20 µm

Next, *Sf*9 cells were transfected with pCW55 and pCW57 in order to assess if: a) the p40 (fused to eGFP) showed any pattern of subcellular distribution within transfected insect cells, b) if the position of the eGFP coding sequence fused at either the N- or C-terminus of p40 altered any observed pattern of distribution of p40, and lastly c) to observe the general pattern of p40 distribution, aggregation and/or oligomerization in the context of an insect intracellular environment. Both p40-eGFP and eGFP-p40 fusion proteins appear to have a noticeably less diffuse, cytoplasmic distribution pattern as compared with eGFP (Figure 5.6 panels C - D and E - F, respectively, as compared with panels A - B). The fusion proteins were excluded from the nucleus suggesting the absence of a functional nuclear import signal. The intracellular punctate structures observed in cells expressing the eGFP-p40 fusion proteins showed no obvious co-localization with organelles although this remains to be confirmed by co-localisation with organellular markers. However, these structures were larger and thus emitted a stronger signal closer to the nuclei of transfected cells, indicating the accumulation or targeting of p40 and its fusion partner in this location.

This observation could be an artefact of the protein post-translational modification pathway, with many of these structures still in transit within the ER adjacent to the nucleus. These structures could also represent accumulated membranes derived from organelles, which have been rearranged in the presence of p40 such as those observed in many picornaviral infections (reviewed by Netherton *et al*., 2008). Alternatively, they could be composed of membranes contributed by more than one type of organelle. Based on the general volumes of ER within the average cell, they could include a large percentage of ER and located closer to the nucleus. Also, these apparent aggregations of eGFP could be an artefact of the fixation method used on these samples, but live cells transfected and imaged under the same conditions as above and viewed by confocal laser scanning microscopy were still found to contain these distinct punctate structures of eGFP so this is unlikely. Ultimately, co-localization with

cellular organelles cannot be ruled and still needs to be investigated based on the unusual distribution pattern observed in these experiments.

No obvious differences in the rate of expression or overall distribution were observed between the p40-eGFP (Figure 5.6 panels C - D) and eGFP-p40 (Figure 5.6 panels E - F). In a significant number of transfected cells, both versions of the fusion protein appear to localize in a slightly polarized fashion within the cell (See additional images in Appendix 7, Figures A.1 and A.2). Upon closer examination of magnified cells expressing the two forms of the fusion protein, one could contend that the p40-eGFP fusion protein (Figure 5.6, panels C - D) produces larger punctate structures which, due to their size and overall fluorescence, were more noticeable relative to the disseminated, more diffuse pattern observed by eGFP-p40 fusion protein expressed in these cells (Figure 5.6 panels E - F). These punctate structures varied in size and to an extent, also in shape, with some structures being more oblong than others (Figure 5.6 panels C - F).

5.3.2. Biochemical fractionation of *Sf*9 cells transfected with eGFP-p40 and p40-eGFP fusion proteins

The next objective was to determine whether the punctate structures observed in transfected cells were membrane-associated complexes formed in the presence of tagged p40, or if the structures observed were an artefact of aggregating protein within the cell. To answer this question, transfected cells were fractionated into membrane-associated and insoluble protein/ aggregate fractions and these sub-cellular components subjected to Western analysis to determine the presence of the eGFP-p40 fusion protein.

The subcellular fractionation procedure was validated using the membrane-associated VDAC protein as an experimental control. Monolayers of *Sf*9 cells transfected with pCW55 (p40-eGFP) and pCW57 (eGFP-p40) were subjected to the same subcellular fractionation protocol. The resultant subcellular fractions were analyzed in duplicate by SDS-PAGE followed by

Western analyses. One set of resolved, transferred proteins was probed with monoclonal anti-VDAC antibodies, and the other with polyclonal anti-p40 antiserum (Figure 5.7).

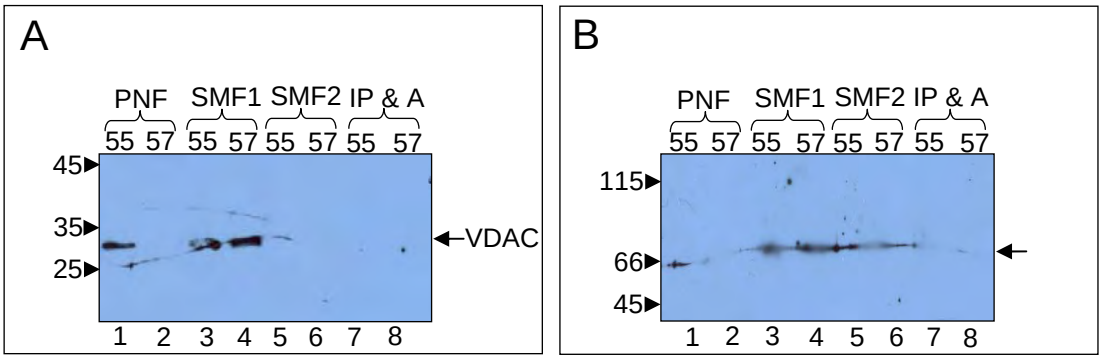

Figure 5.7. Distribution of p40-eGFP fusion proteins in subcellular fractions of *Sf*9 cells. (A): Western analysis of subcellular fractions probed with anti-VDAC antibodies (1 in 500). (B): Western analysis of subcellular fractions probed with anti-p40 antiserum (1 in 100 000). Key: PNF – post-nuclear fraction; SMF – soluble membrane fraction; IP & A – insoluble protein and aggregates; 55 – pCW55; 57 – pCW57. The molecular weight standards are indicated on the left of the blots and are represented in kDa, while the arrows on the right either indicate cellular VDAC (30 kDa) or fusion protein (~68 kDa).

VDAC was detected in the post-nuclear and soluble membrane fractions (Figure 5.7 panel A, lanes 1 - 4). No carry-over of VDAC protein was evident in either the second soluble membrane wash or the insoluble protein and aggregate fraction (Figure 5.7 panel A, lanes 5 - 8). No VDAC protein was detected in the second soluble membrane wash because of the small amounts of total protein available for fractionation from a single monolayer originating from a 24-well culture plate well. This suggested that the biochemical fractionation protocol had successfully separated the soluble membrane-associated proteins from cytosolic and aggregated protein fractions.

The 68 kDa protein, corresponding to p40-eGFP and eGFP-p40 was detected in both the post-nuclear and soluble membrane fractions (Figure 5.7 panel B, lanes 1 - 4). The reasons for the lack of immunoreactive protein in the post-nuclear fraction of pCW57 (Figure 5.7 panel B, lane 2) is unclear although the amount of fusion protein fractionated from pCW57 was, in general lower, than that observed for pCW55 suggesting that this

particular fraction might have been over-diluted. This would explain the weak immunoreactive signal at this particular time exposure. Also, a small amount of the fusion protein was detected in the insoluble protein and aggregate fraction (Figure 5.7 panel A, lanes 7 - 8). Considering the high levels of heterologous protein expression observed in this system, it is not surprising that a (small) proportion of p68 was insoluble and/or aggregating possibly because of its intracellular concentration.

Based on the efficient fractionation of a marker cellular protein (VDAC) from *Sf*9 cells (Figure 5.7 panel A), the procedure was repeated and the fractionated fusion proteins subjected to SDS-PAGE and Western analysis using anti-p40 antiserum (Figure 5.7 panel B). The analysis revealed that both p40-eGFP (pCW55) and eGFP-p40 (pCW57) fusion proteins were membrane-associated (Figure 5.7 panel B, lanes 3 - 6), with a small percentage detectable in the insoluble fraction (Figure 5.7 panel B, lanes 7 – 8).

Initially, one would predict that the punctate structures observed within *Sf*9 cells transfected with both pCW55 (Figure 5.6, panels C - D) and pCW57 (Figure 5.6, panels E - F) may possibly represent aggregating protein especially since these foci showed no pattern of distribution similar to any network of organelles or central nuclear structure. Further biochemical analysis of these proteins suggested that the majority of this heterologous protein is membrane-associated (Figure 5.7, panel B). Due to a lack of antibodies generated towards mammalian organelles that sufficiently cross-react with insect organelles or stains available for the subcellular co-localization of p40/p104 to a particular intracellular location, it was decided to direct further studies towards co-localization of vRNA and p40/p104 in the PrV-infected MG8 cell line. This approach was decided upon since the exact role or contribution, if any, of eGFP in the formation of the punctate structures observed in *Sf*9 cells transiently expressing the eGFP-p40 fusions is unknown. Also, the availability of the MG8 cell line and anti-p40 antiserum for the direct analysis of p40 would eliminate potential artefacts created by the transient expression of eGFP fusion proteins.

5.3.3. Biochemical fractionation and Western analysis of the MG8 cell line

The aim of these experiments was to determine whether similar patterns of distribution of p40 could be observed in PrV-infected MG8 cells. This would also eliminate any artefacts due to the presence of eGFP. Preliminary data on the localization and solubility of transiently expressed p40-eGFP fusions suggested that the eGFP tag did not significantly change any inherent localization of the protein when fused to either the N- or C-termini of the coding sequence. Although mainly membrane-soluble, the foci imaged during expression of these protein within *Sf*9 cells could represent aggregating protein within the subcellular environment. To compare the behaviour of p40 in PrV-infected cells and within the host cell environment, MG8 cells were subjected to subcellular fractionation. Based on the efficient subcellular fractionation of VDAC protein within MG8 cells (Figure 5.8 panel A, lanes 3 and 4), subcellular fractions from this procedure were analyzed by SDS-PAGE followed by Western analyses using monoclonal anti-VDAC antibodies and polyclonal anti-p40 and anti-PrV antisera (Figure 5.8 panels B and C).

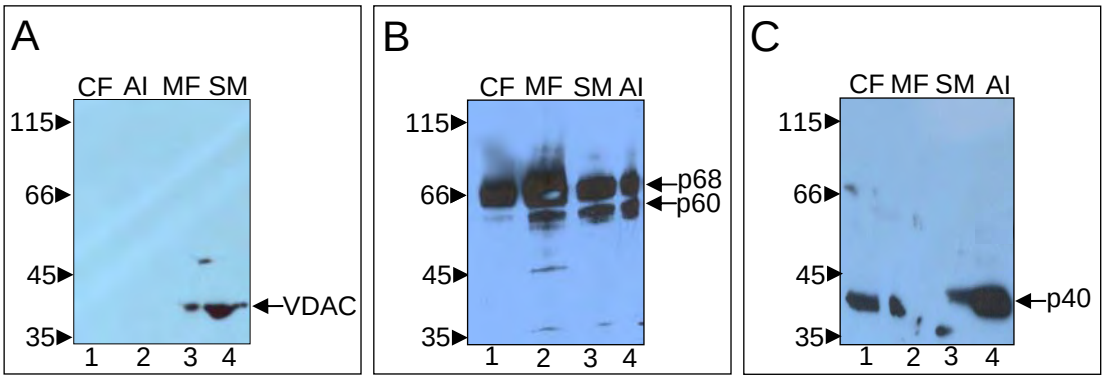

Figure 5.8. Subcellular localization of VDAC, p40 and VCap proteins in PrV-infected MG8 cells as determined by biochemical fractionation. (A): Western analysis of soluble and insoluble fractions of MG8 cells, using anti-VDAC antibodies. (B): Western analysis of soluble and insoluble fractions of MG8 cells, using anti-PrV antiserum. (C): Western analysis of soluble and insoluble fractions of MG8 cells, using anti-p40 antiserum. Key: CF – cytoplasmic fraction; MF – membrane fraction; SM – soluble membrane fraction; AI – aggregate and insoluble protein fraction. The molecular weight standards are indicated on the left of the blots and are represented in kDa while the arrows on the right either indicate PrV capsid precursor and major capsid protein (p68/p60) or p40.

The PrV virus capsid proteins were shown to be associated with both the membrane fraction and cytoplasmic fraction based on the detection of the capsid precursor and major capsid protein in both of these fractions (Figure 5.8 panel B, lane 1 - 2). A large proportion of the capsid precursor and major capsid protein were shown to be soluble but a lesser amount of these proteins were detected in the insoluble protein/aggregate fraction (Figure 5.8 panel B, lane 1-3 vs. lane 4). Interestingly, the majority of the viral protein detected in the cytoplasmic fraction is capsid protein α (p68) (Figure 5.8 panel B, lane 1) suggesting that PrV might be translated in the cytoplasm but is assembled and matures in membrane inclusions. This theory is supported by the imaging of large membrane-bound vesicles in infected MG8 and FB33 cells harbouring PrV virions, whereas in the regions outside of these vesicles far fewer virus particles were observed (Pringle *et al*., 2003).

The distribution of p40 within various subcellular fractions of MG8 cells was determined by Western analysis using anti-p40 antiserum. Although some p40 was detected in the soluble fraction, a substantial percentage of this protein was insoluble or in aggregates (Figure 5.8 panel C, lanes 1 - 3 vs. lane 4). p40 was clearly detectable in the membrane fraction (Figure 5.8 panel C, lane 2), but was present in much lower concentrations in the soluble membrane fraction (Figure 5.8 panel C, lane 3).

No p104 was detected, possibly because only 2 µg of total protein could be analyzed in each fraction, the amount of p104 was probably below the detection limit of the antiserum used. Although the replicase (p104) should theoretically be isolated with the membrane faction, more protein needed to be analyzed in order to detect p104 using this protocol. Since p40 was present in abundance, it was decided to focus on this protein in further experiments, as a ‘marker’ in the subcellular localization of viral replication complexes and co-localization of p40 and vRNA in MG8 cells.

5.3.4. Bromouridine labelling of viral RNA in the MG8 cell line

It was decided to examine the subcellular distribution of vRNA and virus replication complexes in PrV-infecting MG8 cells. PrV vRNA was initially labelled in the presence of 5 µg/ml or 20 µg/ml of Actinomycin D, in order to distinguish between MG8 cells transcribing both viral and cellular RNA, from those transcribing just vRNA (in the presence of the inhibiting concentration of 20 µg/ml Actinomycin D). Control experiments were performed by excluding either the primary or secondary antibodies from the immunofluorescence procedure described above. Cells were then imaged so as to confirm that the signal detected was not the result of non-specific binding by either antibody or due to inherent auto-fluorescence from the MG8 cells. In these cells, RNA was seen to be distributed at high levels throughout the cell including the nucleus (see white arrows in Figure 5.9 panels A and B). However, transcription was inhibited in the majority of the cells, and replicating vRNA was observed to take on a distinct distribution pattern and at low levels compared to uninhibited cellular transcription (Figure 5.9 panel C - D). An optimal dilution of primary antibody, mouse anti-BrUTP, was determined to be 1 in 50 and the dilution of the secondary antibody, goat anti-mouse antibody conjugated to Texas Red (sulphur rhodamine), was 1 in 750. Additional representative images of BrUTP-labelled vRNA detected by immunofluorescence can be seem in Appendix 7, Figure A.3.

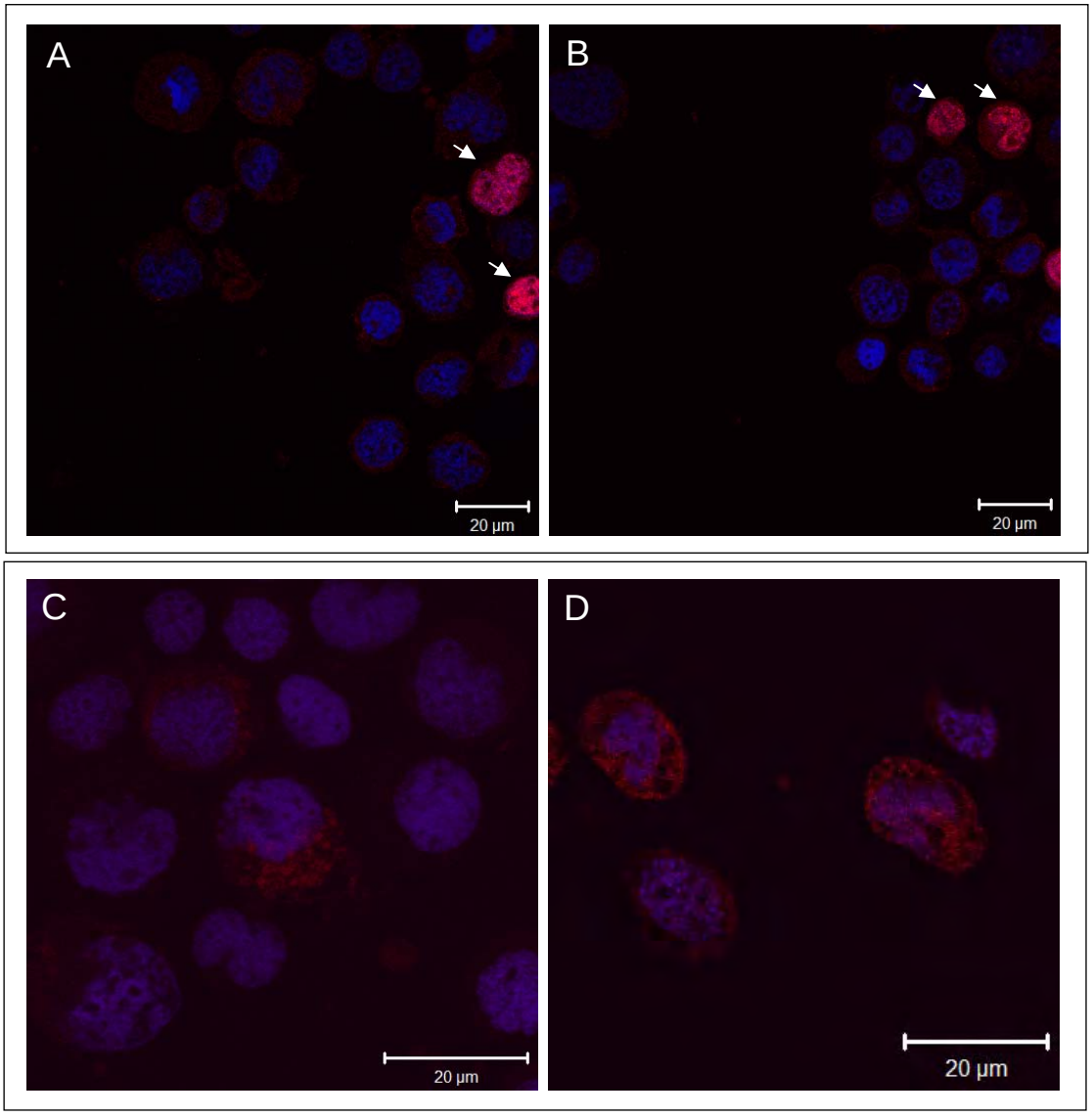

Figure 5.9. Immunofluorescence detection of viral RNA accumulation and distribution within MG8 cells. BrUTP was incorporated into PrV vRNA in the presence of 5 mg/ml Actinomycin D (A – B) or 20 mg/ml Actinomycin D (C – D) for 3 hours. BrUTP-incorporated vRNA was detected by immunofluorescence using anti-BrUTP and goat anti-mouse antibody conjugated to Texas Red. Cells were imaged by confocal laser scanning microscopy. Nuclei were stained with DAPI. Images A and B were acquired at 1 X scanning zoom and C and D, at 2 X scanning zoom. Scale bars represent 20 µm.

The cells were stained with this nuclear marker so as to identify all the cells on the fixed coverslip and not limiting the analysis of cells just emitting fluorescence from the Texas Red fluorophore. The visualization of vRNA by immunofluorescence resulted in a number of noteworthy observations in the infected MG8 cells. Firstly, not all of the DAPI-stained MG8 cells were harbouring active RNA replication complexes. Secondly, the detected viral RNA levels appeared to quite attenuated, emitting low levels of fluorescence (compare indicated cells in panels A and B with the cells in

panels C – D of Figure 5.9), which is to be expected since PrV is a persistent infection in the MG8 cell lines and exhibits no obvious cytopathic effect (CPE), possibly indicating low levels of replication. Thirdly, viral RNA appears to be transcribed in a polarized fashion within the cytoplasm of MG8 cells, forming variably sized punctate structures within these polarized regions of viral transcription. The levels of vRNA synthesis based on the signal attained in these experiments varied between cells, and certain cells harbouring transcribing viral RNA showed brighter zones of transcription on the cell periphery and more dominantly in perinuclear clusters.

The finding that not all of the MG8 cells subjected to immunofluorescence were supporting active viral replication complexes was not unexpected when one considers a) the small quantities of PrV purified from MG8 cell monolayers (observed in this study and by Pringle *et al.*, 2003) b) the lack of any observable CPE or other side-effect on the MG8 cells as a result of the persistent tetravirus infection (observed in this study and by Pringle *et al.*, 2003), and c) the lack of cell typing and other histopathological experiments on the MG8 cell line has not ruled out the possibility that more than one cell type is present within this particular cell line. Lepidopteron midgut tissue is made up of four distinguishable cell types: columnar, goblet, basal regenerative and endocrine cells (Billingsley and Lehane, 1996). Current knowledge of tetraviral infection only reports whole larval midgut tissue as permissive to these viruses with no extensive reports on the susceptibility of these individual cell types to tetraviral infection (Grace and Mercer, 1965; Hanzlik and Gordon, 1997; Bawden *et al*., 1999; 2002).

The general distribution of the viral RNA appeared to be polarized to varying degrees in MG8 cells supporting active replication complexes (Figure 5.9). The punctate structures forming in these polarized regions were variable in size and general location, based on observations in a sample set of cells. Furthermore, the punctate, polarized distribution bore similarities to transiently expressed p40-eGFP fusions in *Sf*9 cells (see Figure 5.6 panels C - F), in particular with respect to the variable aggregate-like formation and the polar clustering of these foci in the cell.

Immunofluorescence detection of p40 (and p104) in PrV-infected MG8 cells

MG8 cells were subjected to the same procedure described in the immunofluorescence-based detection of labelled RNA, but without the use of the inhibitor of cellular transcription, Actinomycin D. The primary and secondary antibody combination used to detect p40/p104 was anti-p40 antiserum and anti-rabbit conjugated to FITC, respectively. Negative controls involving the exclusion of either the primary or secondary antibody were performed in parallel to these experiments to confirm that the distribution pattern observed here was not a general effect from the secondary antibody nor autofluorescence. Following this, MG8 cells were imaged by scanning confocal microscopy (Figure 5.10).

Following the immunofluorescence of p40 within MG8 cells, imaging of cells expressing p40 showed that this protein was readily detectable in 24 % of the general cell population. Since the antibodies used here were directed against a peptide region also present on p104, it is likely that some fluorescence observed might be due to the presence of the read-through version.

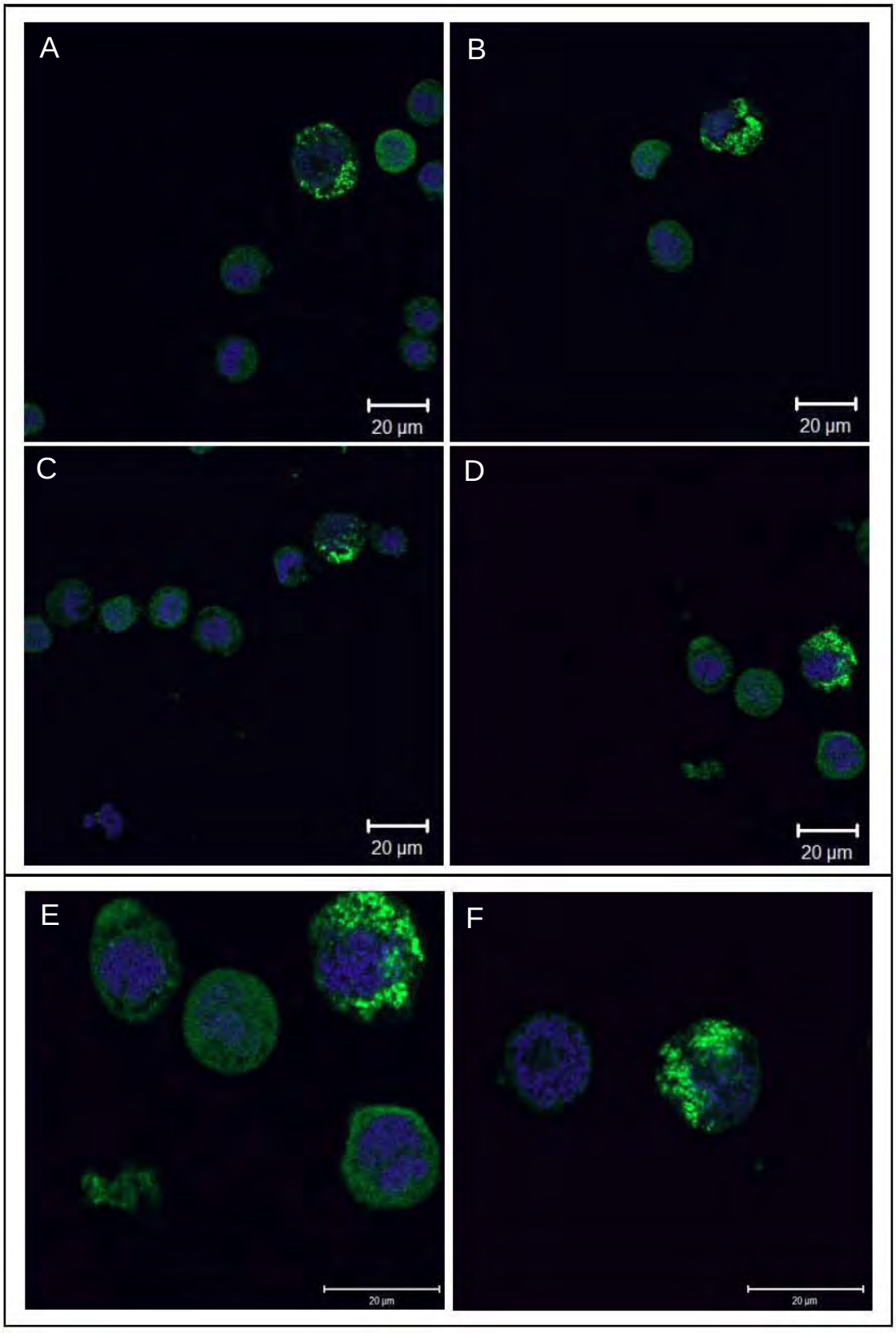

Figure 5.10. Immunofluorescence detection of PrV p40 (and p104) accumulation and distribution in MG8 cells. The predicted quantitatively dominant p40 auxiliary protein was detected by immunofluorescence using rabbit anti-p40 and goat anti-rabbit antibody conjugated to FITC. Nuclei were stained with DAPI. Cells were imaged by confocal laser scanning microscopy. Images A – D) were acquired at 1 X scanning zoom and E – F, at 2 X scanning zoom. Scale bars represent 20 µm.

Once again, not all cell types appeared to be expressing p40/p104 (Figure 5.10 panels A - F). The weak signal, seen in the remainder of the MG8 cells (Figure 5.10), was investigated and shown not to be a result of cellular auto-fluorescence or background signal from the secondary antibody conjugate, and thus may potentially represent cells expressing p40. Here, p40 expression is probably occurring at low levels and the protein is not aggregating or causing an aggregating effect in the cell. The phenomenon of only ~ 25 % of the imaged cells were expressing p40 (adjudged by the bright foci as opposed to weak diffuse expression) could be a result of not all of the four midgut cell types being permissive to PrV infection, or perhaps a defective virus replication cycle whereby only a portion of (permissive) cells are infected. The fluorescence detected in the above experiment again shows the formation of punctate-like structures clustering in a polarized fashion within p40-expressing cells. These structures show both variably sized 'aggregates' as well as shapes of these individual foci (Figure 5.10).

Importantly, the structures observed here were similar in localization and general morphological structure to transiently expressed p40 fused to eGFP. In MG8 cells, as compared with *Sf*9 cells, the general size and structure of these fluorescent foci were similar (Compare Figure 5.10 with Figure 5.6).

5.3.5. Co-localization of PrV p40 and vRNA in MG8 cells

The similarity in appearance (although not intensity) of these p40/p104-related structures and the polarized foci associated with viral RNA in MG8 cells (Figure 5.9) warranted further investigation. To determine whether p40/p104 is associated with the PrV replication complex *in vivo*, MG8 cells were cultured in the presence of Actinomycin D and vRNA labelled with BrUTP. The presence of vRNA and p40/p104 was then detected using anti-p40/anti-rabbit conjugated to FITC, and anti-BrUTP/anti-mouse conjugated to Texas Red antibody combinations with minor alterations to

the existing protocol (Section 5.2.6). Figures 5.11 and 5.12 represent two examples of cells where co-localization of the two fluorophore signals was detected. Appendix 7, Figure A.4 contains a further montage of representative co-localized fluorophore signals.

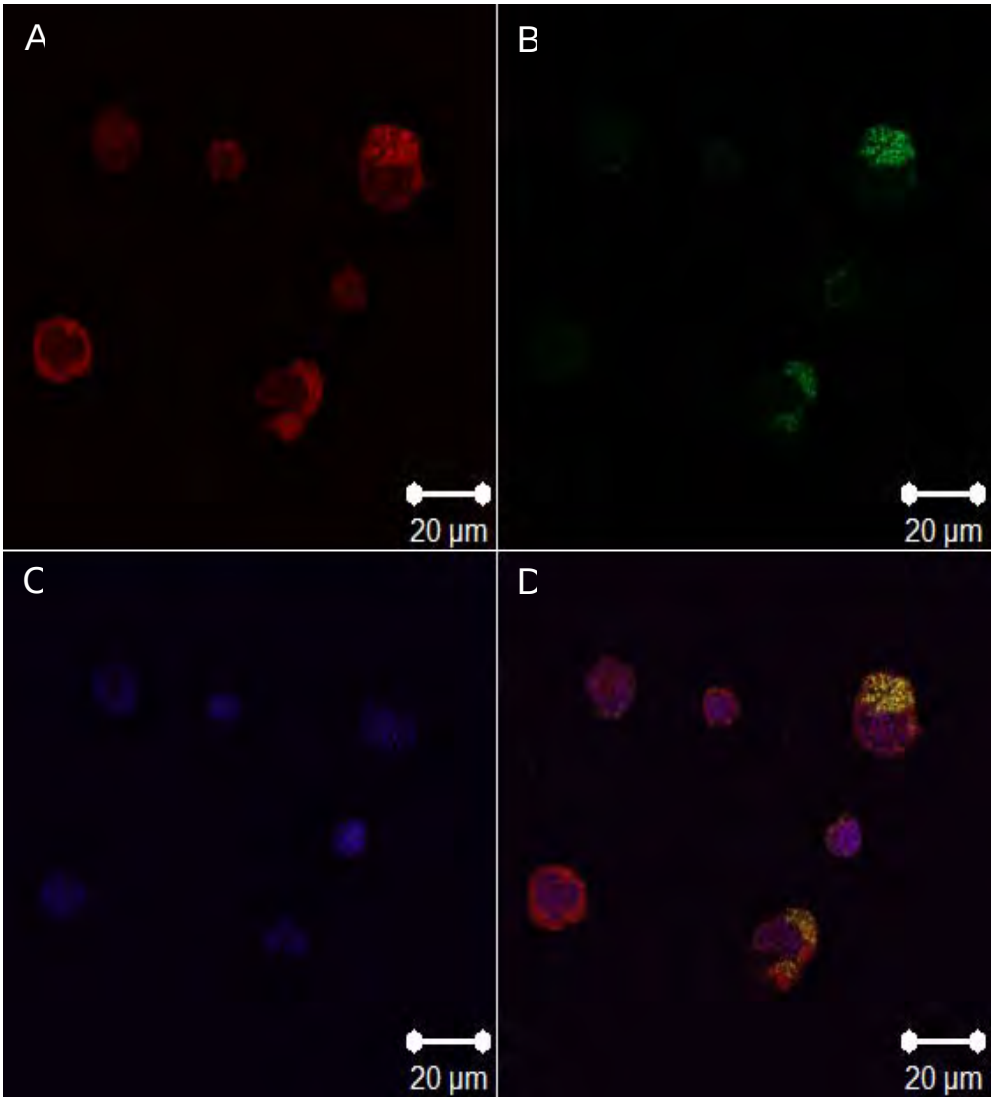

Figure 5.11. Co-localization of p40 and viral RNA in PrV-infected MG8 cells. BrUTP was incorporated into PrV vRNA in the presence of Actinomycin D for 3 hours. The BrUTP-incorporated vRNA was detected by immunofluorescence using anti-BrUTP/goat anti-mouse antibody conjugated to Texas Red. p40 was detected by immunofluorescence using rabbit anti-p40/goat anti-rabbit antibody conjugated to FITC. Cells were imaged by confocal laser scanning microscopy. A: fluorescent signal from Texas Red indicating vRNA. The Texas Red signal was digitally summed up four times during acquisition of each image; B: fluorescent signal from FITC indicating p40 (and p104); C: fluorescent signal from DAPI stain indicating the nucleus; D: merged image of all three fluorescence emissions. The images were acquired at 1 X scanning zoom. Scale bars represent 20 µm.

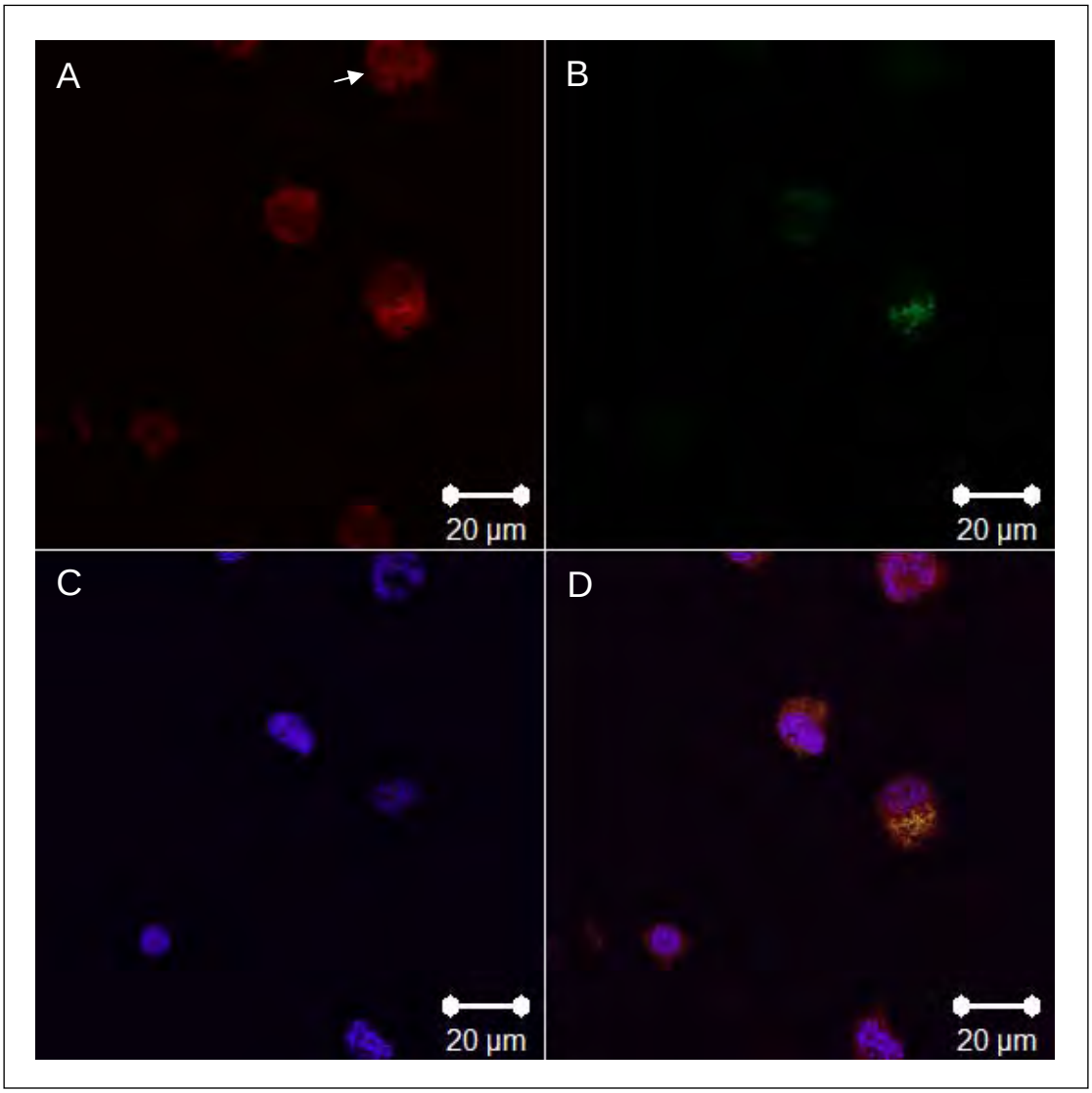

Figure 5.12. Co-localization of p40 and viral RNA in PrV-infected cells. BrUTP was incorporated into PrV vRNA in the presence of Actinomycin D for 3 hours. The BrUTP-incorporated vRNA was detected by immunofluorescence using anti-BrUTP/goat anti-mouse antibody conjugated to Texas Red. p40 auxiliary protein was detected by immunofluorescence using rabbit anti-p40/goat anti-rabbit antibody conjugated to FITC. Cells were imaged by confocal laser scanning microscopy. A: fluorescent signal from Texas Red indicating vRNA. The Texas Red signal was digitally summed up four times during acquisition of each image; B: fluorescent signal from FITC indicating p40 (and p104); C: fluorescent signal from DAPI stain indicating the nucleus; D: merged image of all three fluorescence emissions. The images were acquired at 1 X scanning zoom. Scale bars represent 20 µm.

The co-localization of p40 and vRNA was clearly evident in the confocal laser scanning images of dual-labelled vRNA and p40 (Figure 5.11 and 5.12). Again, a polarized distribution of both fluorescent signals was observed in MG8 cells expressing both p40 and exhibiting replicating vRNA (Figure 5.11 and 5.12). As observed with the immunofluorescent detection

of p40/p104 alone (Figure 5.10) and BrUTP-labelled vRNA alone, not all of the MG8 cells were expressing either vRNA or p40/p104 however the absence of one of these molecules within a cell did not exclude the presence of the other, i.e. vRNA might be labelled in a cell where no p40/p104 was detectable above background levels and vice versa (Figure 5.12 panel A, indicated by arrow).

Analysis of this phenomenon revealed that, of the seventy four MG8 cells imaged in this particular experiment: p40/p104 was detected in eighteen cells (24 %); replicating vRNA was detected in twenty nine cells (39 %) and fifteen cells (20%) were observed to be expressing both p40/p104 and vRNA, suggesting co-localization of the two signals. Cells were marked as positive or negative for either fluorophore based on the strong, punctate distribution of both molecules previously observed (Figure 5.9 and 5.10) and not the weak, diffuse signal.

The extent of the co-localization of five randomly selected MG8 cells exhibiting co-localization was digitally analyzed in terms of pixel distribution and overall percentage area of co-localization. These cells all showed grouping of pixels from both Texas Red and FITC emissions in a co-localized pattern based on histogram analysis followed by calculation of the percentage area of this co-localization. The average percentage area exhibiting co-localization in these five representative cells was measured at 19 %. However, this value is based on a two dimensional area which was imaged because of its significant co-localization and is thus not representative of the three dimensional area of co-localization within a larger group of MG8 cells exhibiting varying degrees of co-localization.

This analysis represents only a preliminary investigation and could over-represent the number of cells exhibiting co-localized p40 and vRNA, because these cells were being actively selected during the imaging process. Also, the number of vRNA-harbouring cells could be over-represented because of the need to employ digital summing up of the fluorescence signal during acquisition, thereby potentially leading to false

positives. More cells should be imaged in a completely unbiased fashion in order to acquire a more representative sample for statistical analysis. Based on this data the following two conclusions can be drawn: (a) p40/p104 and vRNA do co-localize in a polarized, punctate fashion within the cell at a frequency of 20 % relative to the general population of MG8 cells, and (b) a significantly larger proportion of MG8 cells support actively replicating vRNA (39 %) when compared to those which appear to be expressing p40/p104 and not supporting vRNA replication (24 %).

5.4. CONCLUSION

Following on from the analysis of the PrV replication proteins, p40 and p104 discussed in Chapter 4, this chapter provides insight into the presence and distribution of p40 (and probably p104) in PrV-infected MG8 cells. Initially, a transient expression system was used to detect eGFP-tagged p40 fusion proteins in *Sf*9 cells, serving as a system of studying solely p40. The results suggested that p40-eGFP fusions exhibit an unusual punctate distribution pattern in the cytoplasms of *Sf*9 cells and suggest that these fusion proteins may be mostly membrane associated. These data suggested that p40 might actively induce and associate with membranous structures originating from the transfected cells into atypical patterns of diffusion. However, this system has several potential flaws: firstly, the inclusion of the eGFP tag to the fusion protein adds substantial molecular weight to the fusions which could impact on the mobility and distribution of these fusions in the cells. Also the transient expression was controlled by an enhanced viral promoter which could have produced large amounts of insoluble and/or aggregating proteins, thus creating an artefact during the imaging of transiently expressed *Sf*9 cells. Another potential drawback to the transient expression system would be the absence of p104, expression which is likely to play a role in replication complex formation and morphology albeit in smaller amounts compared to an auxiliary protein.

Given the potential problems with the transient expression of eGFP fusion proteins, it was decided to focus on 'endogenous' p40 and p104 in PrV-infected MG8 cells. These immunofluorescence studies revealed a similar distribution pattern that was often polarized to one side of the cell. Not all of the imaged MG8 cells exhibited this pattern of intracellular p40/p104 accumulation and some were shown to emit a very weak signal corresponding to low levels of diffuse p40/p104 expression. This could be attributed to a mixed cell line where only certain cell types within the midgut tissue are permissive to PrV infection or alternatively, these cells exhibiting little to no p40 expression could represent various stages of replication complex formation and p40 expression. Since the antibody used for these studies is unable to distinguish between p40 and p104, it is possible that these cells may be in fact expressing p104, rather than p40. It would be important to use antibodies that would be able to detect only p104 (raised against the C-terminus of this protein) to eliminate this possibility.

Immunofluorescence studies on the distribution of BrUTP-labelled vRNA revealed a distribution and accumulation of vRNA similar to that observed in the detection of p40. Polarized accumulations of vRNA could be observed as well as 'hot spots' of signal at the cell membrane interface most likely representing accumulation of virion particles at the cell membrane as observed by Pringle *et al*. (2003). Again, not all MG8 cells appeared to support actively replicating RNA although it was interesting that more cells on average exhibited actively replicating vRNA compared to cells expressing p40. No particular pattern could be observed between cells expressing either p40 or containing actively replicating vRNA and cells undergoing mitosis, cell lysis or membrane blebbing (an indicator of apoptosis). Phase contrast microscopy and the analysis of various unfixed cell structures would need to be performed in order to confirm this observation.

Upon dual immunofluorescence detection of both p40 and vRNA in the MG8 cells, several cells exhibiting co-localization were observed. Although the vRNA fluorescent signal was weak due to the combination of

conjugated secondary antibody and detector parameters as well as a potentially low level of vRNA replication, compounding of and merging of this amplified signal with that from p40 resulted in the detection of co-localized punctate and polarized structures within these cells. Biochemical fractionation of MG8 cells revealed that p40 is largely insoluble and/or associated with aggregates, although a portion of this protein is membrane-associated. It is possible that the structures observed to form were the result of virus-induced aggresomes that act as sites of replication mediated by at least p40 and, most likely, p104. Co-localization experiments using markers specific for organelle-derived membranes as well as markers specific to autophagy and aggresomes (such as anti-LC3 antiserum) are required for further analysis into the site and nature of PrV replication.

CHAPTER 6
General discussion and future avenues of research

Providence virus, discovered by Pringle *et al*. (2003), was the first tetravirus able to infect cultured insect cell lines. Apart from this novel characteristic, PrV was found to be a member of the genus *Betatetravirus*, based on its monopartite genome, outer capsid appearance and moderate capsid sequence conservation with other betatetraviruses. Unpublished data generated by Dr. Fiona Pringle and Prof. Andrew Ball at the University of Alabama at Birmingham, shortly after their 2003 publication detailing the discovery and characterization of PrV, began to suggest an atypical replicase gene sequence, organization and gene conservation. These unique features and the availability of a tetravirus that infects insect cell lines lead to the conceptualization of the research described in this thesis. The overall aim of this study was to characterize the translation products of PrV and to set up an experimental system for studying virus replication in PrV-infected MG8 cells.

6.1. PrV has a unique set of replication proteins and might follow a replication strategy similar to the *Tombus*- and *Umbraviridae* families

The bioinformatic analysis of the coding sequences, protein products, phylogenetics and possible post-translational modifications or processing mechanisms predicted in the PrV genome resulted in the accumulation of data supporting the initial theory on the divergent genome organization and protein processing of PrV. The presence of a highly conserved read-through stop signal, significant sequence conservation and phylogenetic relationships with other tombus- and umbraviruses suggested that the putative PrV replicase protein might have originated, or evolved in a convergent manner from a (+)ssRNA plant virus. The in-frame group I

read-through stop signal is a recoding mechanism found mostly in plant viruses and mammalian retroviruses (Beier and Grimm, 2001, Harrell *et al.*, 2002). Based on its position in the replicase ORF, this stop signal is likely to function in a similar fashion to those found in the replicase ORFs of exclusively plant viruses and was thus predicted to function as a recoding mechanism, resulting in the translation of asymmetric quantities of auxiliary versus replicase protein.

In PrV, the protein sequence preceding the stop signal (p40) may act as an auxiliary protein binding to other auxiliary, host proteins and the read-though product, the replicase (p104), to form the basis of a replication complex. Given the relatedness of PrV to carmo-like plant viruses, bioinformatic analysis was focused on uncovering characteristics of the PrV genome that are known to exist in the well-studied tombusviruses. Further analyses of the replicase ORF and its predicted protein product, p104, revealed a carmo-like replicase based on the presence of only the essential catalytic RdRp domain situated at the C-terminus of p104. Furthermore, this RdRp domain showed significant protein sequence conservation with other tombus- and umbraviruses despite its unusual permuted RdRp motif arrangement - a feature not associated with carmo-like viruses. Both sequence and spatial orientation of the motifs within the PrV RdRp domain appear to be divergent from the permuted domains of two other betatetraviruses, EeV and NβV. The exact structural and catalytic alterations that might occur because of these differences should be experimentally determined since homology modelling will prove difficult, based on the lack of crystallographic data on related RdRp domains. Furthermore, a measure of the fidelity and catalytic activity of the PrV RdRp would provide indirect evidence of its evolutionary stability and overall rate of vRNA polymerization in its permissive host, the MG8 cell line.

Other similarities also exist between the replicases of PrV and tombusviruses. Numerous potential sites of phosphorylation, a functional modification often associated with replication-related proteins and

demonstrated in both the auxiliary and replicase proteins of tombusviruses (Shapka *et al*., 2005) are predicted to be present in the coding sequences of both p40 and p104. The prediction of two hydrophobic helices further substantiates the hypothesis that p40 and p104 could be membrane-associated proteins, forming part of a replication complex associated, and most likely anchored to a subcellular membrane by these transmembrane domains. The presence of these two helices provides further evidence of the similarities between the p40/p104 PrV replication proteins and the auxiliary/replicase proteins of the tombusviruses. The helix-to-helix stretch in the tombusvirus, CymRSV, has been found to be essential for both the anchoring of p33 as well as the targeting of this protein to the peroxisome (Navarro *et al*., 2004). If such an anchoring event did occur in p40, it would be as a result of a single transmembrane helix, although p104 is predicted to contain two helices that could result in the insertion of more than 400 amino acids into a theoretical subcellular lumen.

The availability of polyclonal antiserum raised against a peptide stretch of p40, thus recognizing both the predicted (and correctly folded) p40 and p104 replication proteins resulted in experimental proof of the existence of an auxiliary (p40) and replicase (p104) protein combination, detected in the persistently infected MG8 cells. Western analysis of PrV-infected MG8 cell-free extracts confirmed the presence of p40 and p104 and showed that they were expressed in an asymmetric ratio similar to that of tombusviruses (Skuzeski *et al.*, 1991; Panaviene *et al*., 2003). *In vivo* metabolic radiolabelling of PrV proteins resulted in the detection of a 40 and a 104 kDa protein. *In vivo* metabolic labelling of PrV in the MG8 cells also led to the detection of a further five proteins potentially translated from the PrV genome. Two of these (p68 and p60) were identified by Western analysis as the viral capsid protein precursor and mature capsid protein. A 128 kDa protein, proposed to be the translation product of p130, was also detected in metabolically labelled MG8 cells and the transcription/translation of p130 was confirmed *in vitro*. The function of p130, unique to PrV, and not present in the genome of any other sequenced tetraviruses, remains unknown.

In the rabbit reticulocyte coupled *in vitro* transcription/translation system, the p104 ORF was translated into a 40 and 104 kDa protein, although not in the asymmetric ratios expected to result from the in-frame stop signal and was observed by Western analysis of PrV-infected MG8 cells using anti-p40 antiserum. Interestingly, the ratio of p40 to p104 resulting from translation of the p104 ORF in the wheat germ system resulted in substantially higher levels of p40 as compared with p104. This provided support for the proposed plant virus origin of the PrV replicase and the conserved read-through stop signal. Importantly, these results showed both the presence of two replication proteins as well as demonstrating the functionality of the read-through stop signal in two coupled *in vitro* transcription/translation systems. The functionality of the PrV read-through stop signal needs to be tested in a more analogous (invertebrate) system since the *in vitro* analyses of viral read-through stop signals can lead to ambiguous results (Sasaki *et al*., 1998).

6.2. Establishing a system for studying the subcellular distribution of p40 and viral RNA in MG8 cells

With both the presence and origins of p40/p104 established, the next objective was to investigate the presence and distribution of both viral RNA and PrV replication proteins in PrV-infected MG8 cells. Since p40 was shown to be translated in greater quantities than p104 and an antiserum was available for the detection of both p40 and p104, it was elected to study the presence and distribution of this protein and BrUTP-labelled vRNA in PrV-infected MG8 cell line. The results of these experiments revealed the presence of punctate, polarized distribution patterns of p40 (and possibly p104) as well as newly synthesized vRNA in a proportion of MG8 cells. Although not every cell supported either replicating vRNA or p40, the co-localization of these two molecules was evident in a fraction of cells. Even though p40 was shown to be associated with the membrane fraction of biochemically fractionated cell-free extracts, it was most strongly associated with the insoluble protein and aggregates fraction of MG8 cells.

The question that arises from this result is whether the punctate structures observed in the immunofluorescent detection using anti-p40/p104 antibodies represent aggregated proteins (or perhaps viral-induced aggresomes) or whether they are the result of membrane rearrangements induced by PrV infection and replication. Future work would involve attempting to localize p40/p104 and vRNA to an autophagosome or a subcellular membrane. Such analyses would also lead to a better understanding of the location and mechanism of replication employed by PrV in the MG8 cell line. The availability of a p104-specific antiserum would allow for the confirmation of the presence and distribution of only p104 in relation to p40 and vRNA in PrV-infected MG8 cells.

Although very little is known about the replication strategies of other tetraviruses, inherent differences in replicase sequence and motif arrangement have already been established in TaV and EeV (Gorbalenya *et al*., 2002), and from this study, also in PrV. Based on the data summarized above, PrV has been demonstrated to translate proteins both with proposed functions and in a mechanism similar to other (+)ssRNA plant viruses and instead of other members of the *Tetraviridae* family.

6.3. Co- and post-translational processing mechanisms of PrV

During this study, it became increasingly evident that PrV is an unusual tetravirus, not only in terms of its divergent replication strategy but also because of the number of co- and post-translational processing mechanisms identified in the protein coding sequences of the three major ORFs of this virus. Three major mechanisms exist, firstly the PrV read-through stop signal is only the second reported occurrence of a functional signal found in a virus infecting exclusively invertebrate hosts. Such a signal has been discovered in the dicistrovirus, PSIV, although this stop signal was shown to be inactive *in vitro* and the capsid protein precursor translated by the re-initiation of translation from an IRES signal in this region and not a leaky read-through mechanism (Sasaki *et al.*, 1998;

Sasaki and Nakashima *et al*., 1999). The regulation of polymerase expression is widely observed in both (+)RNA viruses and is especially well-documented in retroviruses (Shehu-Xilaga *et al*., 2001). Although the mechanism of polymerase expression control in these lineages of viruses occur by very different means, the resultant polymerase attenuation and accessory protein overexpression are essential for effective virus assembly, release and infectivity (reviewed in Ahlquist, 2006). It is interesting to speculate on the proposed multifunctional nature of the ribosome coding mechanism in PrV. Does this mechanism result in attenuated levels of replication or does it play a more active role in the divergent replication mechanism of this unique virus?

A second co-translational processing mechanism is the presence of a 2A-like processing sequence at the N-terminus of the capsid protein precursor coding sequence, which is also present in the betatetraviruses, TaV (Pringle *et al*., 1999) and EeV (Donnelly *et al*., 2001b). In TaV, this processing sequence results in the processing of the translated capsid protein precursor into a smaller 17 kDa protein and the capsid precursor (Pringle *et al.*, 1999; 2001). The capsid precursor is then further cleaved during particle maturation by thirdly, an autoproteolytic site at the C-terminus of the precursor, resulting in the major and minor capsid proteins. 2A-like sequences are found in analogous positions in both EeV and PrV (PrV-$2A_3$). PrV has a second 2A-like processing sequence (PrV-$2A_2$) in the p13 sequence further upstream of the PrV-$2A_3$ sequence, although this sequence has been shown to exhibit a lowered activity in an *in vitro* reporter system (Luke *et al*., 2008). PrV-$2A_2$ was not active *in vitro* in the context of the wild-type surrounding sequence reported during this study and, since neither p13 nor the predicted products of PrV-$2A_2$ processing have been detected *in vivo*, the function of this sequence in this context is unknown. In this study, PrV-$2A_3$ was shown to be active upon *in vitro* transcription/translation of the p81 coding sequence, suggesting that PrV undergoes a similar sequence of processing and cleavage events in the production of the major and minor capsid proteins. A third 2A-like processing sequence, PrV-$2A_1$, was identified in the N-terminal region of

the theoretical p130 protein and this sequence was shown to be inactive *in vitro* although interpretation of these data were complicated by the ambiguous translation products produced from the p130 coding sequences and two truncations thereof.

Phylogenetic analysis of these three 2A-like sequences relative to analogous sequences in TaV, EeV and certain members of the *Dicistroviridae*, predicted a clade of four out of the five betatetravirus sequences excluding PrV-$2A_2$ (Luke *et al*., 2008). The exact role and functionality, if any, of PrV-$2A_1$ and PrV-$2A_2$ in the *in vivo* processing of p130 and p81 respectively, need to be examined further.

6.4. Is PrV an attenuated or latent infection in the MG8 cell line or are only a proportion of the cells or cell types infected?

A survey on publications in the last 5 years reporting the discovery of latent viruses shows that a number of (typically (+)ssRNA carmo-like) viruses have been discovered by researchers and reported in various peer-reviewed publications or by less formal means (Ohki *et al*., 2006; Tzanetakis and Martin, 2008; Vaira *et al*., 2008). Many of these viruses were unintentionally discovered, yet when one considers the number of latent plant and animal viruses that potentially exist, a fascinating account of virus replication and genome processing mechanisms would surely be revealed. A number of facts and observations discovered during the course of this work make the possibility of PrV being a latent or highly attenuated virus plausible. These include: (1) the very low levels of virus particle recovery from infected MG8 cells despite attempts to optimize this process (supported by Pringle *et al*., 2003), (2) the steady growth rate of the MG8 cell line in medium excluding foetal calf serum (FCS) and the marked decrease of cell growth leading to eventual cell death in the presence of this supplement, (3) a lack of replicating vRNA structures/accumulated p40 protein in a large percentage of cells in the MG8 cell line, (4) the presence of a read-through stop signal in the

replicase of an animal virus, and (5) large amounts of unprocessed capsid precursor protein (p81) detected in radiolabelled viral translation products in MG8 cells by Western analysis of purified particles and MG8 cell-free extract and also evident by the lack of autoproteolytic processing of this protein *in vitro.* It is not common for a differentiated cell line to proliferate over the long term in the absence of growth supplements usually supplied in the form of FCS. FCS provides essential growth and mitogen proteins in addition to the basal cell medium and exclusion of FCS has been shown to induce cell differentiation in embryonic and/or stem cell lines leading to cell mitosis. In the case of cell lines harbouring latent retroviruses, shifting of the cell cycle into mitosis can then further lead to the induction of latent virus back into the active form. An essential future experiment will be the cell typing of the MG8 line so that the presence of any undifferentiated cell types potentially supporting latent PrV can be investigated.

Alternatively, it could also be argued, that PrV is not an attenuated or latent virus. PrV has been shown to be able to infect at least two other insect cell lines. Although the labelling of both viral RNA and p40 in MG8 cells could be an indication of an attenuated rate of replication based on an average of 20 % of examined cells exhibiting co-localization of these two molecules, it could also be the likely result of a mixed cell line. When one considers the size of midgut tissue in early instar lepidopteran larvae, it is quite possible that this cell line is composed of all four cell types making up midgut tissue and perhaps even peripheral cell types. A mixed cell line might contain cell types not permissive to tetraviral infection explaining the low numbers of cells supporting active vRNA replication and/or p40 expression and not as a result of attenuated replication in a single cell type. A quantitative study of the cell populations in HaSV-infected midgut tissue of *H. armigera* larvae showed that the endocrine cells of virus-challenged midgut tissue were excluded during infection by HaSV (Brooks *et al*., 2002). As mentioned earlier, further cell sorting and sub-culturing experiments would need to be performed on this particular cell line to confirm its mixed cell type status and also to separate out the proposed different cell types that could potentially be supporting actively replicating PrV vRNA. The relatively

large amount of capsid precursor detected in both purified PrV virions and in cell-free extracts of MG8 cells was often observed in this study and, although not evident by Coomassie staining of PrV proteins resolved by SDS-PAGE, comparative levels of the precursor protein was detected by Western analysis of FB33 cells infected with PrV, 48 hours post-infection (Pringle *et al*., 2003). The degree of autoproteolytic cleavage from p68 into p60 and p8 might not be as efficient in PrV as observed in other tetraviruses (Hanzlik *et al*., 1993; Pringle *et al*., 1999; Yi *et al*., 2005; Walter *et al*., 2008).

What role, if any, does p130 play in the lifecycle of PrV? This proposed non-structural protein is the first ORF in the PrV genome and positioned in a good context for initiation of translation. Could p130 play a role in the attenuation of PrV replication and infection at a translational, RNA or protein level or would the apparent low levels of infection and replication of vRNA be a result of a read-through stop signal amounting to low levels of replicase protein? Again, additional analyses on the transcription, translation and the functions of p104 and p130 need to be addressed further in order to begin to answer some of these questions.

6.5. CONCLUSIONS

With the first feasible opportunity to study tetravirus replication possible since the discovery of the unique tetravirus, PrV, the overall aims of this work were fulfilled. The genome organization and protein processing of PrV have been both bioinformatically and experimentally investigated. A hypothesis on the replication strategy of PrV has also been put forward following the discovery of two replication proteins and a recoding mechanism controlling their expression. Based on data generated in this work, it is likely that PrV is not a tetravirus and it is suggested that this virus undergoes reclassification at the family level.

APPENDICES

APPENDIX 1 – PRIMERS USED IN THIS STUDY

Table A.1. Name, sequence and description of primers used in this study.

Name	Sequence	Description
AI11	**5`CTAG**AGGTACCAAGGAGAAAAAAATGCATCATCATCACCAT**CA**`3 3` TCCATGGTTCCTCTTTTTTTACGTAGTAGTAGTGGTA**GTAT**`5	ds oligonucleotides *Xba*I/*Nde*I - pCW20
CW5	5`**CGAA**CCTTGCTACCACGAACGAAAACAGC`3	*Csp*45I - pCW21
CW6	5`**AAGCTT**CAATACCCTGACGGCCAGTCAATATCGG`3	*Hin*d III - pCW21, pCW22
CW7	5`**GGATCC**AACCTTGCTACCACGAACG`3	*Bam* HI - pCW22
CW9	5`**CATATG**TGCGAGCCTTTAAGTTTACGC`3	*Nde*I - pCW26,27, pCW35
CW10	5`**AAGCTT**CCGCCGGTCCCATT`3	*Hin*dIII - pCW26, 27
CW11	5`**GATCT**GGTACCAAGGAGAAAAAAATGC`3	*Bgl*II - pCW28
CWF3	5`CATACCGTCCCACCATCGGGCGCGG`3	pFB Dual, pCW30
polR	5`ACCTCTACAAATGTGGTATGGCTGAT`3	pFB Dual, pCW30
pUCF	5`CGCCAGGGTTTTCCCAGTCACGAC`3	pFB Dual, pCW30
pUCR	5`TCACACAGGAAACAGCTATGAC`3	pFB Dual, pCW30
CW23	5`**GGATCC**CTACTGGAGGAGGGGGAG'3	*Bam* HI - pCW52, pCW53
CW24	5`**AAGCTT**TTGGTGGTCTATCAGAGACTCG'3	*Hin*d III - pCW52
CW01	5`GCTGTGAACACTTCAGGTAAAAC '3	pSTA1
CW02	5`TATATA**GCGGCCGC***TTA*CACTGCTCCGTTGGCGAAATG`3	*Not*I - pSTA1-3
CW03	5`CCCATCGGCTCTTGGCTTCGAGGA`3	pSTA2
$CW01_b$	5` AAGGAGTCTAATAGTGGTTACG`3	pSTA3
CW04	5`TGCGAGCCTTTAAGTTTACGC'3	pSTA4, pSTA5
CW05	5`TATATA**GCGGCCGC***TTA*CTTTGTCGGGATTGTTCTTCC`3	*Not*I - pSTA4
CW06	5`TATATA**GCGGCCGC***TCA*ATACCCTGACGGCCAGTCAATATC`3	*Not*I - pSTA5
CW09	5`CAGAATTTACCAGTACCCAATG`3	pSTA7
CW10	5`TATATA**GCGGCCGC***TTA*CTTGGCTCGTCTAGCGCGG`3	*Not*I - pSTA7
CW11	5`CCCCAGATGACCCGCGAGAATAAAG`3	pSTA7
CW13	5`**TGATCA**TTTGTCGGGATTGTTC`3	pCW35
CW19	5`**GACGTC**GCTCACTCATTAGGCACC`3	*Aat*II - pCW44
CW20	5`**AAGCTT**GGTTGTTCACGATCTTGTCG`3	*Hin*d III - pCW44

Key: Complete restriction sites are in bold and underlines. Partial restriction sites are in bold. Stop codons are in italics. Mutated nucleotides are represented in lower case and are in bold. 5`: five prime direction of sequence. 3`: three prime direction of sequence

APPENDIX 2 – DESCRIPTION AND MAINTENANCE OF CELL LINES USED IN THIS STUDY

H. zea fat body BCIRL-HZ-FB33 (referred to in text as FB33) (Kariuki *et al*., 2000) and *Spodoptera frugiperda* ovarian cell line 9 CRL 1711 (referred to as *Sf*9 cells) (Vaughn *et al*., 1977) were grown at 28 °C in TC100 medium containing 1 % penicillin/streptomycin and 5 % heat-inactivated foetal calf serum (FCS). *H. zea* midgut cell line BCIRL-HZ-MG8 (Pringle *et al*., 2003) (referred to in text as MG8) were grown at 28 °C in Excell 420 complete medium containing 1 % penicillin and streptomycin. Attempts to proliferate this cell line in complete medium supplemented with FCS were unsuccessful and resulted in cell death after two passages in the presence of FCS.

Baby Hamster (Syrian) Kidney cell line IZSBS BS CL 21 (Macpherson and Stoker, 1962) (referred to as BHK 21 cells in text) and the HeLa S3 immortalized human cervical cancer cell line ECACC 93021013 (Gey *et al*., 1952) (referred to as HeLa cells in text) were grown in the presence of 10 % CO_2 at 37 °C in Dulbecco's Modified Eagle Medium (DMEM) containing 100 U/ml penicillin, 25 µg/ml Amphotericin B and 10 % heat-inactivated FCS.

APPENDIX 3 – SEQUENCING OF THE p104 ORF OF PrV

The PrV p104 ORF in the plasmid pFLM1, was sequenced using the dideoxy chain termination principal and the products resolved by automated gel separation using an ABI sequencer (v. 3100) (Rhodes University DNA Sequencing Unit). The chromatograms were visualized using the Chromas v.2.1 program. Primers used in the sequencing reactions of p104 within pFLM1 were designed to be complimentary to the theoretical sequence provided by Prof. L Andrew Ball (Department of Microbiology, University of Alabama at Birmingham, Alabama, USA). Nucleotide sequences were analyzed using the Vector NTi10 sequence analysis and data management software.

The plasmid pFLM1 (generously donated by Prof. L. Andrew. Ball), contained the entire 6155 nt of cDNA sequence of the PrV genome cloned into the vector pTVT7R. Analysis of the replicase sequence (completed in triplicate) resulted in the detection of the following substitutions when compared to the original p104 sequence: a substitution at nt 2466 (T - C base change) did not result in an amino acid substitution. Three amino acid substitutions were detected in the p104 coding sequence: 1496 nt (A - G) resulting in a glutamine to an arginine substitution; 1499 nt (G - A) resulting in an alanine to a threonine substitution and finally, 1553 nt (A - G) resulting in a histidine to an arginine substitution. These substitutions are not located within the RdRp domain of the replicase protein and therefore should not impact on the genome replication ability of PrV.

APPENDIX 4 – UNIFORM RESOURCE IDENTIFIERS (URI) AND ACCESS DATES OF ONLINE BIOINFORMATIC PROGRAMS AND SERVERS

Table A.2. Name, address and date of access of bioinformatic programs used in this study

Name of program	URL	Date accessed
ClustalW	http://www.ebi.ac.uk/Tools/clustalw2/index.html	28-29/09/2008
MFold	http://mfold.bioinfo.rpi.edu/cgi-bin/rna-form1.cgi	29/09/2008
MitoProt II v1.101	http://ihg2.helmholtz-muenchen.de/ihg/mitoprot.html	26/09/2008
NCBI – BLAST	http://blast.ncbi.nlm.nih.gov/	26-27/09/2008
netNES 1.1	http://www.cbs.dtu.dk/services/NetNES/	26/09/2008
NetPhos 2.0	http://www.cbs.dtu.dk/services/NetPhos/	26/09/2008
PESTfind	https://emb1.bcc.univie.ac.at/toolbox/pestfind/pestfind-analysis-webtool.htm	29/09/2008
ProtScale	http://ca.expasy.org/tools/protscale.html	09/11/2008
SOSUI	http://bp.nuap.nagoya-u.ac.jp/sosui/sosui_submit.html	26/09/208
TargetP 1.1	http://www.cbs.dtu.dk/services/TargetP/	26/09/2008
The PTS1 Predictor	http://mendel.imp.ac.at/mendeljsp/sat/pts1/PTS1predictor.jsp	26/09/2008
TMpred	http://www.ch.embnet.org/software/TMPRED_form.html	27/09/2008
WoLFpSORT	http://wolfpsort.org/	26/09/208

APPENDIX 5 – ACCESSION NUMBERS OF NUCLEOTIDE AND AMINO ACID SEQUENCES USED IN THIS STUDY

Table A.3. Sequence names and accession numbers of amino acid sequences used this study

Sequence name	Accession number
Nudaurelia capensis ω virus p70	AAY54604
Nudaurelia capensis β virus p67 and RdRp	AAC97510 and AAC97509
Dendrolimus punctatus tetravirus p71 and p17	YP0250956 and YP025095
Helicoverpa armigera stunt virus p71 and RdRp	AAC37885 and NP049235
Euprosterna elasaea virus p83	NP573542
Thosea asigna virus p83	AAC 97195
Providence virus p81	AAO 73881
Infectious pancreatic necrosis virus RdRp	BAA 05533
Infectious bursal disease virus RdRp	NP 690839
Cricket paralysis virus structural polyprotein and RdRp	NP647482 and AAA 42889
Pelargonium chlorotic ring pattern virus replicase	YP05295

APPENDIX 6 – MEMBRANE-STABILISING IPB BUFFER (CHAPTER 5)

10 X IPB
24.2 g Tris base
36 g NaCl
8 ml 0.5 M EDTA (pH 8.0)

Made up to 500 ml in double-distilled water after adjusting to pH 7.8 with HCl and stored at room temperature.

1 X IPB
50 ml 10 X IPB
1 g iodoacetamide
1 tablet of Complete™ EDTA-free protease inhibitor cocktail (Roche Applied Science)
5 ml x 100 PMSF (100 mM stock in 70 % ethanol)
17 ml Brij-30

Made up to 500 ml in double-distilled water on the day of each subcellular fractionation and stored at 4 ºC.

APPENDIX 7 – ADDITIONAL CONFOCAL MICROSCOPE IMAGES OF *Sf*9 CELLS TRANSFECTED WITH pCW55 AND pCW57

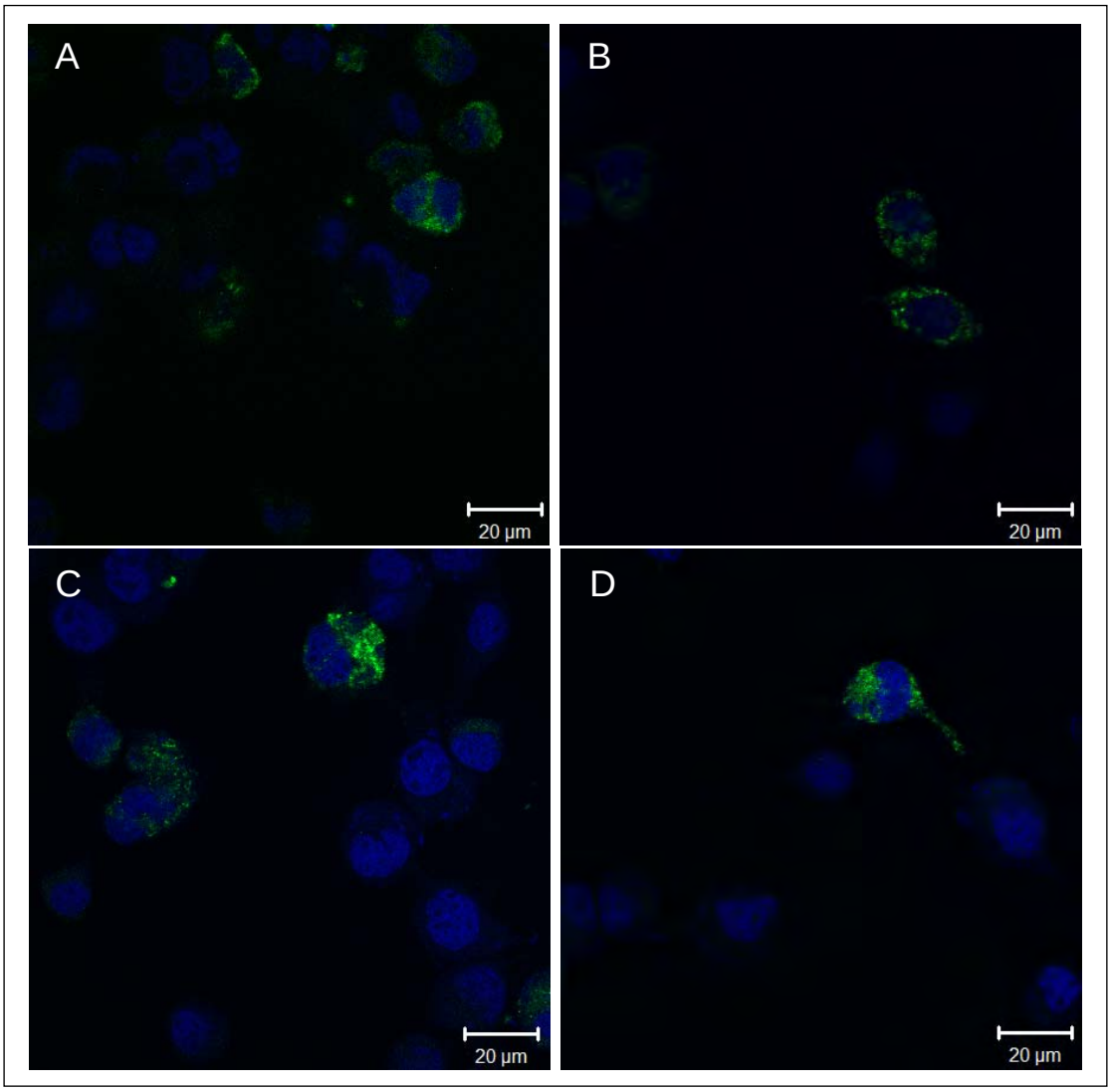

Figure A.1. Distribution of eGFP/p40 fusion proteins in *Sf*9 cells transfected with pCW55. A - D: Punctate, cytoplasmic distribution of p40 fused to the N-terminus of eGFP in *Sf9* cells transfected with pCW55. Images acquired at 1 X scanning zoom. Scale bars represent 20 µm.

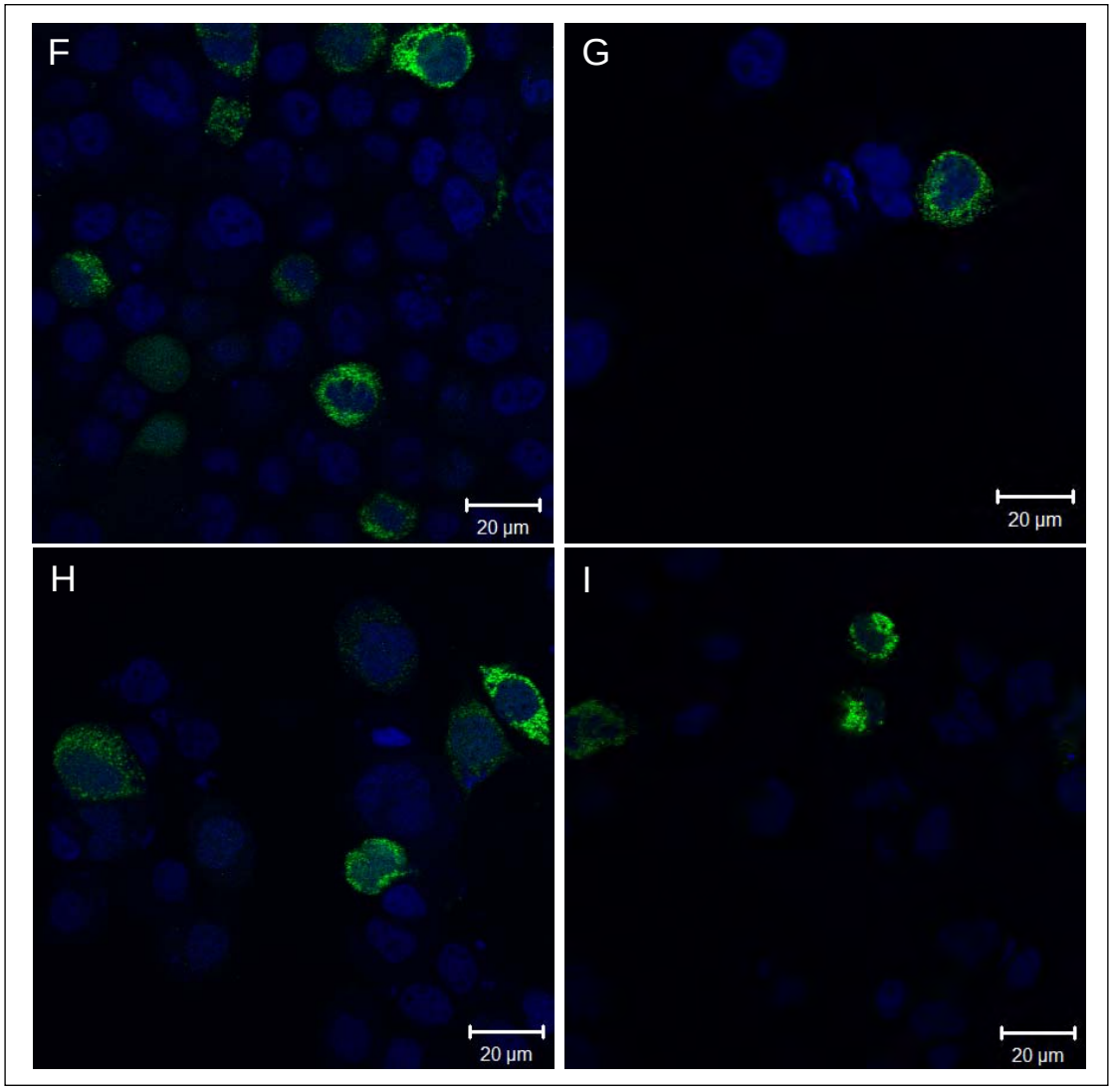

Figure A.2. Distribution of eGFP/p40 fusion proteins in *Sf*9 cells transfected with pCW57. F - I: Punctate cytoplasmic distribution of p40 fused to the C-terminal of eGFP in *Sf9* cells transfected with pCW57. Images acquired at 1 X scanning zoom. Scale bars represent 20 µm.

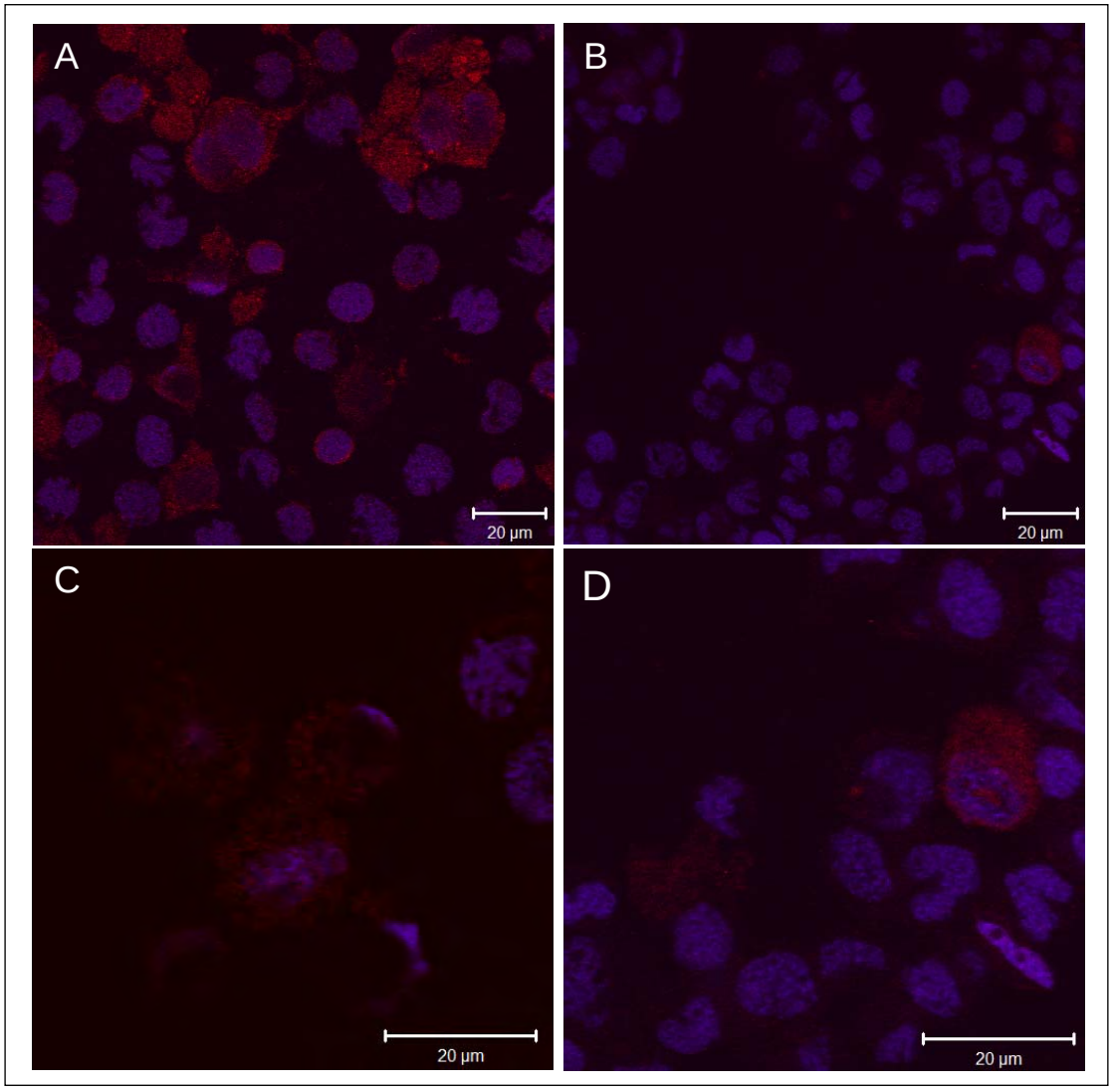

Figure A.3. Immunofluorescence detection of viral RNA accumulation and distribution in PrV-infected MG8 cells. A - B: BrUTP was incorporated into PrV vRNA in the presence of 20 mg/ml Actinomycin D for 3 hours. BrUTP-incorporated vRNA was detected by immunofluorescence using anti-BrUTP and goat anti-mouse antibody conjugated to Texas Red. Cells were imaged by confocal laser scanning microscopy. A - B were acquired at 1 X scanning zoom and C - D, at 2 X scanning zoom. Scale bars represent 20 µm.

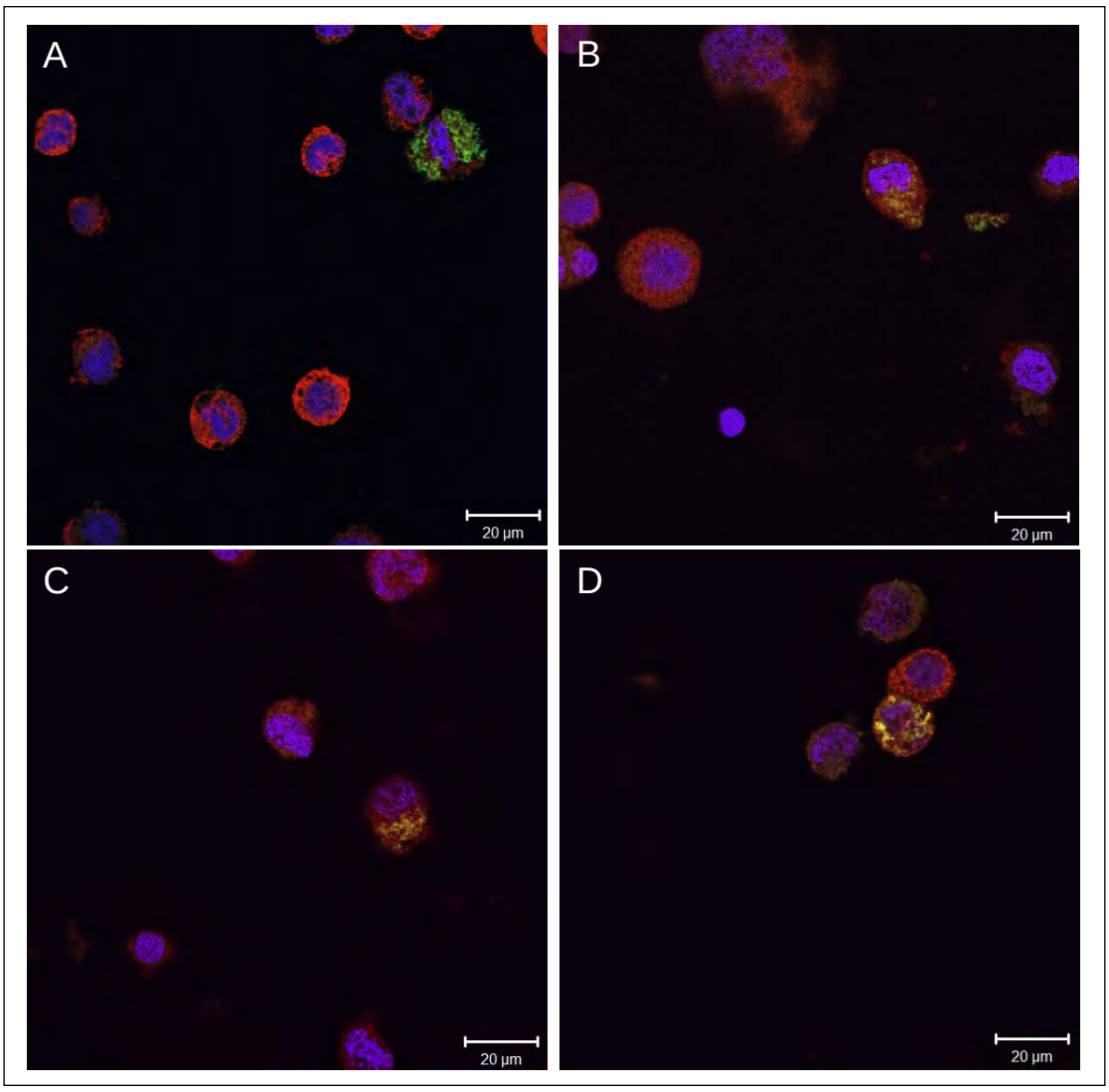

Figure A.4. Co-localization of p40 and viral RNA in PrV-infected MG8 cells. A - D: BrUTP was incorporated into PrV vRNA in the presence of Actinomycin D for 3 hours. The BrUTP-incorporated vRNA was detected by immunofluorescence using anti-BrUTP and goat anti-mouse antibody conjugated to Texas Red. p40 auxiliary protein was detected by immunofluorescence using rabbit anti-p40 and goat anti-rabbit antibody conjugated to FITC. Cells were imaged by confocal laser scanning microscopy. The Texas Red signal was digitally summed up four times during acquisition of each image. Images were captured at 1 X scanning zoom. Scale bars represent 20 µm.

REFERENCES

Agrawal, D.K. and Johnson, J.E. 1992. Sequence analysis of the capsid protein of *Nudaurelia capensis* ω virus, an insect virus with *T*=4 icosahedral symmetry. *Virol*. **190**: 806-814.

Agrawal, D.K. and Johnson, J.E. 1995. Assembly of the *T*=4 *Nudaurelia capensis ω* virus capsid protein, post-translational cleavage, and specific encapsidation of its mRNA in a baculovirus expression system. *Virol*. **207**: 89-97.

Agol, V.I., Paul, A.I. and Wimmer, E. 1999. Paradoxes of the replication of picornaviral genomes. *Virus Res*. **62**: 129-147.

Ahlquist, P. 2002. RNA-dependent RNA polymerases, viruses, and RNA silencing. *Science*. **296**: 1270-1273.

Ahlquist, P. 2006. Parallels among positive-strand RNA viruses, reverse-transcribing viruses and double-stranded RNA viruses. *Nat. Rev. Micro*. **4**: 371-382.

Ahlquist, P., Noueiry, A.O., Lee, W-M., Kushner, D.B and Dye, B.T. 2003. Host factors in positive-strand RNA virus genome replication. *J. Virol*. **77**: 8181-8186.

Altschul, S.F., Madden, T.L., Shäffer, A.A., Zhang, J., Zhang, Z., Miller, W and Lipman, D.J. 1997. Gapped-BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. **25**: 3389-3402.

Aminev, A.G., Amineva, S.P. and Palmenberg, A.C. 2003. *Encephalomyocarditis virus* (EMCV) proteins 2A and 3BCD localize to nuclei and inhibit cellular mRNA transcription but not rRNA transcription. *Virus Res*. **95**: 59-73.

Ayers, M.D., Howard, S.C., Kuzio, J., Lopez-Ferber, M. and Possee, R.D. 1994. The complete DNA sequence of *Autographa californica* multiple-nucleocapsid nucleopolyhedrovirus. *Virol*. **202**: 586-605.

Baca, A.M. and Hol, W.G.J. 2000. Overcoming codon bias: A method for high-level overexpression of *Plasmodium* and AT-rich parasite genes in *Escherichia coli*. *Inter. J. Parasit.* **30**: 113-118.

Baird, S.D., Turcotte, M. and Korneluk, R.G. 2006. Searching for IRES. *RNA*. **12**: 1-31.

Baltimore, D. 1971. Expression of animal virus genomes. *Bacteriol. Rev*. **35**: 235-241.

Batten, J.S., Turina, M and Scholthof, K-B, G. 2006. *Panicovirus* accumulation is governed by two membrane-associated proteins with a newly identified conserved motif that contributes to pathogenicity. *Virol. J*. **3**: 1-12.

Bawden, A.L., Gordon, K.H.J and Hanzlik, T.N. 1999. The specificity of *Helicoverpa armigera Stunt Virus* infectivity*. J. Inver. Pathol*. **74**: 156-163.

Bedard, K.M. and Semler, B.L. 2004. Regulation of picornavirus gene expression. *Microbes. Infect*. **6**: 702-713.

Beier, H. and Grimm, M. 2001. Misreading of termination codons in eukaryotes by natural nonsense suppressor tRNAs. *Nucleic. Acids. Res*. **23**: 4767-4782.

Bellstedt, D.U., Human, P.A., Rowland, G.F. and Van der Merwe, K.J. 1987. Acid-treated, naked bacteria as immune carriers for protein antigens. *J. Immunol. Meth.* **98**: 249-255.

Bellstedt, D.U., Van der Merwe, K.J. and Galanos, C. (1988). Immune carrier properties of acid-treated *Salmonella minnesota* R595 bacteria. The immune response to TNP-bacterial conjugates in rabbits and mice. *J. Immunol. Met.* **108**: 245-254.

Belsham, G.J and Sonenberg, N. 2000. Picornavirus RNA translation: roles for cellular proteins. *Trends. Microbiol*. **8**: 330-335.

Billingsley, P.F. and Lehane, M.J. 1996. In: Structure and ultrastructure of the insect midgut. Lehane, M.J., Billingsley, P.F. (Eds). *In* Biology of the insect midgut. Chapman and Hall, London, pp 3-29.

Bradford, M.M. 1976. A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**: 248-254.

Brooks, E.M., Gordon, K.H.J., Dorrian, S.J., Hines, E.R. and Hanzlik, T.N. 2002. Infection of its lepidopteron host by *Helicoverpa armigera stunt virus* (*Tetraviridae*). *J. Invert. Pathol.* **80**: 97-111.

Brown, G.C. 1992. Control of respiration and ATP synthesis in mammalian mitochondria and cells. *Biochem J*. **284**: 1-13.

Buck, K.W. 1996. Comparison of the replication of positive-stranded RNA viruses of plants and animals. *Adv. Virus. Res*. **47**: 159-251.

Canady, M.A., Tihava, 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). *J. Mol. Biol*. **299**: 573-584.

Carillo, S., Pariat, M., Steff, A-M., Jariel-Encontre, I., Poulat, F., Berta, P. and Piechaczyk, M. 1996. PEST motifs are not required for calpain-mediated proteolysis of c-fos protein. *Biochem. J.* **313**: 229-235.

Caverner, D.R. 1987. Comparison of the consensus sequence flanking translation start sites Drosophila and invertebrates. *Nucleic. Acids. Res*. **15**: 1353-1361.

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

Chen, J. and Ahlquist, P. 2000. Brome Mosaic virus polymerase-like protein 2a is directed to the endoplasmic reticulum by helicase-like viral protein 1a. *J. Virol*. **74**: 4310-4318.

Christian, P., Carstens, E., Domier, L., Johnson, K., Nakashima, N., Scotti, P. and van der Wilk, F. 1999. Cricket paralysis-like viruses. *In* "Virus taxonomy. Seventh Report on the International Committee on the Taxonomy of Viruses" M.H. V Van Regenmortel, C. M. Fauquet D.H.L. Bishop, C. H. Calisher, E.B., Carsten, M.K. Estes, S.M. Lemon, Maniloff, J., Mayo, M.A., McGeoch, D.J., Pringle, C.R. and Wickner, R.B. (eds). Academic Press, San Diego, CA. pp 678-683.

Collier, A.J., Gallego, J., Klinck, R., Cole, P.T., Harris, S.J., Harrison, G.P., Aboula-ela, F., Varani, G. and Walker, S. 2002. A conserved RNA structure within the HCV IRES eIF3-binding site. *Nature Struct. Biol.* **9**: 375-380.

Craigen, W.J., Cook, R.G., Stockwell, P.A. and Tate, W.P. 1985. Bacterial peptide chain release factors: conserved primary structure and possible frameshift regulation of release factor 1. *Proc. Natl. Acad. Sci*. **82**: 3616-3620.

Dalphine, M.E., Brown, C.M., Stockwell, P.A. and Tate, W.P. 1997. The translation signal database, TransTerm: more organisms, complete genomes. *Nucleic Acids Res*. **25**: 246-247.

de Felipe, P., Luke, G.A., Hughes, L.E., Gani, D., Halpin, C. and Ryan, M.D. 2006. *E unum pluribus*: multiple proteins from a self-processing polyprotein. *Trends. Biotechnol*. **24**: 68-75.

Devaney, M.A., Vakharia, V.N., Lloyd, R.E., Ehrenfeld, E. and Grubman, M.J. 1988. Leader protein of foot-and-mouth disease virus is required

for cleavage of the p220 component of the cap-binding protein complex. *J. Virol.* **62**: 4407-4409.

Dong, X.F., Natarajan, P., Tihova, M., Johnson, J.E. and Schneemann, A. 1998. Particle polymorphism caused by deletion of a peptide molecular switch in a quasi-equivalence virus. *J. Virol*. **72**: 6024-6033.

Donnelly, M.L.L., Luke, G., Mehrotra, A., Li, X., Hughes, L.E., Gani, D. and Ryan, M.D. 2001a. Analysis of the apthovirus 2A/2B polyprotein 'cleavage mechanism' indicates not a proteolytic reaction but a novel translational effect: a putative ribosomal 'skip'. *J. Gen. Virol*. **82**: 1013-1025.

Donnelly, M.L.L., Hughes, L.E., Luke, G., Mendoza, H., ten Dam, E., Gani, D. and Ryan, M.D. 2001b. The 'cleavage' activities of foot-and-mouth disease virus 2A site-directed mutants and naturally occurring '2A-like' sequences. *J. Gen. Virol.* **82**: 1027-1041.

Dreher, T.W. and Miller, W.A. 2006. Translational control in positive strand RNA plant viruses. *Virol*. **344**: 185-197.

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). *Arch. Virol*. **150**: 2397-2402.

Dye, B.T., Miller, D.J., and Ahlquist, P. 2005. *In vivo* self-interaction of nodavirus RNA replicase protein A revealed by fluorescence energy transfer. *J. Virol*. **79**: 8909-8919.

Egger, D., Teterina, N., Ehrenfeld, E. and Bienz, K. 2000. Formation of the Poliovirus replication complex requires coupled viral translation, vesicle production and viral RNA synthesis. *J. Virol*. **74**: 6570-6580.

Egger, D. and Bienz, K. 2005. Intracellular location and translocation of silent and active poliovirus replication complexes. *J. Gen. Virol*. **86**: 707-718.

Fauquet C.M., Mayo M.A., Maniloff J., Desselberger, U. and Ball L.A. (eds),. 2005. Ball L.A., Hendry, D.A., Johnson, J.E., Rueckert, R.R., Scotti, P.D. 2005. Family *Tetraviridae*. pp 873-883. Family *Nodaviridae*. pp 884-887. Christian, P., Carstens, E., Domier, L., Johnson, J., Johnson, K., Nakashima, N., Scotti, P. and van der Wilk, F. (eds.). In: Virus taxonomy. Eighth Report of the International Committee on Taxonomy of Viruses, Elsevier Academic Press, San Diego and London

Feng, Y.X., Yuan, H., Rein, A. and Levin, J.G. 1992. Bipartite signal for read-through suppression in murine leukaemia virus mRNA: an eight-nucleotide purine-rich sequence immediately down-stream of the gag

termination codon followed by a RNA pseudoknot. *J. Virol*. **66**: 5127-5132.

Ferrer-Orta, C., Arias, A., Perez-Luque, R., Escarmis, C., Domingo, E. and Verdaguer, N. 2004. Structure of Food-and-Mouth Disease Virus RNA-dependent RNA polymerase and its complex with a template-primer RNA. *J. Biol. Chem*. **279**: 47212-47221.

Ferrer-Orta, C., Arias, A., Escarmis, C. and Verdaguer, N. 2006. A comparison of viral RNA-dependent RNA polymerases. *Curr. Opin. Struc. Biol*. **16**: 27-34.

Fields, B.N., Knipe, D.M., Howley, P.M., Chanock, J., Melnick, L., Monath, T.P., Roizman, B. and Straus, S.E. 1996. In: Fundamental Virology, Third Edition. Lippincott-Raven. Philadelphia. pp 401-402.

Fujiyuki, T., Takeuchi, H., Ono, M., Ohka, S., Sasaki, T., Nomoto, A. and Kubo, T. 2004. Novel insect picorna-like virus identified in the brains of aggressive worker honeybees. *J. Virol*. **78**: 1093-1100.

Gallagher, T.M. and Rueckert, R.R. 1988. Assembly-dependent maturation cleavage in provirions of a small icosahedral insect ribovirus. *J. Virol*. **62**: 3399-3406.

Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D. and Bairoch, A. 2005. In: The Proteomics Protocols Handbook. John M. Walker (ed). Protein Identification and Analysis Tools on the ExPASy Server. Humana Press inc. Totowa, NJ. pp. 571-607.

Gey, G.O., Coffman, W.D. and Kubicek, M.T. 1952. Tissue culture studies of the proliferative capacity of cervical carcinoma and normal epithelium. *Cancer. Res.* **12**: 264-265

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

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

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

Gordon, K.H.J., Williams, M.R., Baker, J.S., Gibson, J.N., Bawden, A.L., Millgate, A.G., Larkin, J. and Hanzlik, T.N. 2001. Replication-

independent assembly of an insect virus (*Tetraviridae)* in plant cells. *Virol*. **288**: 36-50.

Grace, T.D.C. and Mercer, E.H. 1965. A new virus of the *Saturniid Antheraea eucalypti* Scott. *J. Inverter. Pathol*. **7**: 241-244.

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

Guarino, L.A. and Summers. M.D. 1986. Interspersed homologous DNA of *Autographa californica* Nuclear Polyhedrosis virus enhances delayed-early gene expression. *J. Virol.* **60**: 215-223.

Guarino, L.A., Gonzalez, M.A. and Summers, M.D. 1986. Complete sequence and enhancer function of the homologous DNA regions of *Autographa californica* Nuclear Polyhedrosis virus. *J. Virol.* **60**: 224-229.

Hanahan, D. 1983. Studies on transformation of *Escherichia coli* with plasmids. *J. Molec. Biol.* **166**: 557.

Hansen, J.L., Long, A.M. and Schultz, S.C. 1997. Structure of the RNA-dependent RNA polymerase of poliovirus. *Structure*. **5**: 1109-1122.

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*. *J. Gen*. *Virol*. **74**: 1805-1810.

Hanzlik, T. N., Dorrian, S .J., Johnson, K. N., Brooks, E .M. and Gordon, K .H. 1995. Sequence of RNA 2 of the *Helicoverpa armigera stunt virus* (*Tetraviridae*) and bacterial expression of its genes. *J. Gen. Virol*. **76**: 799-811.

Hanzlik, T.N. and Gordon, K.H.J. 1997. The *Tetraviridae. Adv. Virus. Res*. **48**: 101-168.

Harrell, L., Melcher, U. and Atkins, J.F. 2002. Predominance of six different hexanucleotide recoding signals 3` of read-through stop codons. *Nucleic Acids. Res*. **30**: 2011-2017.

Helgstrand, C., Munshi, S., Johnson, J.E. and Liljas, L. 2004. The refined structure of *Nudaurelia capensis* ω *virus* reveals control elements for a *T*=4 capsid maturation. *Virol*. **318**: 192-203.

Hendry D.A., Becker, M.F. and van Regenmortel, M.V.H. 1968. A non-inclusion virus of the pine emperor moth *Nudaurelia cytherea* Stoll. *S. Afr. Med. J*. **42**: 117.

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

Hendry, D.A. and Agrawal, D.K. 1994. Tetraviruses. In: Encyclopedia of Virology. Second Edition. A. Granoff and R.G. Webster (eds). Academic Press, San Diego, CA. pp 1416-1422.

Heras, S.R., Thomas, M.C., García-Canadas, M., de Felipe, P., García-Pérez, J.L., Ryan, M.D. and López, M.C. 2006. L1Tc non-LTR retrotransposons from *Trypanosoma cruzi* contain a functional viral-like self-cleaving 2A sequence in frame with the active proteins they encode. *Cell. Mol. Life. Sci*. **63**: 1449-1460

Hirokawa, T., Boon-Chieng, S. and Mitaku, S. 1998. SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics.* **14**. 378-379.

Hoffman, K. and Stoffel, W. 1993. TMbase - A database of membrane spanning proteins segments. *Biol. Chem. Hoppe-Seyler.* **374**: 166.

Isawa, H., Asano, S., Sahara, K., Iizuka, T., and Bando, H. 1998. Analysis of genetic information of an insect picorna-like virus, infectious flacherie virus of silkworm. Evidence for evolutionary relationships among insect, mammalian and plant picorna (-like) viruses. *Arch. Virol*. **143**: 127-143.

Jakubiec, A., Tournier, V., Drugeon, G., Pflieger, S., Camborde, L., Vinh, J., Hericourt, G., Redeker, V. and Jupin, I. 2006. Phosphorylation of viral RNA-dependent RNA polymerase and its role in replication of a plus strand RNA virus. *J. Biol. Chem*. **281**: 21236-21249.

Jakubiec, A. and Jupin, I. 2007. Regulation of positive-strand RNA virus replication: the emerging role of phosphorylation. *Virus. Res*. **129**: 73-79.

Jameson , B.A and Wolf, H. 1988. The antigenic index: a novel algorithm for predicting antigenic determinants. *Comput. Appl. Biosci*. **4**: 181-186.

Johnson, J.E., Munshi, S, Liljas, L., Agrawal, D., Olson, N.H., Reddy, V., Fisher, A., McKinney, B., Schmidt, T. and Baker, T.S. 1994. Comparative studies of *T*=3 and *T*=4 icosahedral RNA insect viruses. *Arch. Virol*. **9**: 497-512.

Johnson, K.N., Johnson, K.L., Dasgupta, R., Gratsch, T. and Ball, L.A. 2001. Comparisons among the larger genome segments of six nodaviruses and their encoded RNA replicases. *J. Gen. Virol.* **82**: 1855-1866.

Jonczyk, M., Pathak, K.B., Sharma, M. and Nagy, P.D. 2007. Exploiting alternative subcellular location for replication: Tombusvirus replication switches to the endoplasmic reticulum in the absence of peroxisomes. *Virol*. **362**: 320-330.

Juckes, I.R.M. 1974. Purification and characterization of viruses of the pine emperor moth, *Nudaurelia cytherea capensis* (Fabr.). PhD thesis. University of Cape Town.

Kampmueller, K.M. and Miller, D.J. 2005. The cellular chaperone heat shock protein 90 facilitates Flock House Virus RNA replication in *Drosophila* cells. *J. Virol*. **79**: 6827-6837.

Kariuki, C.W., McIntosh, A.H. and Goodman, C.L. 2000. In vitro host range studies with a new baculovirus isolate from the diamondback moth *Plutella xylostella* (L.) (Plutellidae: Lepidoptera). *In. Vitro. Cell Dev. Biol. Anim.* **36**: 271-276.

Knox, C., Moffat, K., Ali, S., Ryan, M.D. and Wileman, T. 2005. Foot-and-mouth disease virus replication sites from next to the nucleus and close to the Golgi apparatus, but exclude maker proteins associated with host membrane compartments. *J. Gen. Virol*. **86**: 687-696.

Koonin, E.V. 1991. The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses. *J. Gen. Virol.* **72**: 2197-2206.

Koonin, E.V. and Dolja, V.V. 1993. Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequence. *Crit. Rev. Biochem. Mol. Biol.* **28**: 375-430.

Korneeva, V.S. and Cameron, C.E. 2007. Structure-function relationships of the viral RNA-dependent RNA polymerase. *J. Biol. Chem*. **282**: 16135-16145.

Kost, T.A., Condreay, J.P. and Jarvis, D.L. 2005. Baculoviruses as versatile vectors for protein expression in insect and mammalian cells. *Nat. Biotechnol*. **23**. 567-575.

Kozak, M. 1987. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic. Acids. Res.* **15**: 8125-8148.

Kushner, D.B., Lindenbach, B.D., Grdzelishvili, V.Z., Noueiry, A.O., Paul, S.M. and Ahlquist, P. 2003. Systematic, genome-wide identification of host genes affecting replication of a positive-strand RNA virus. *Proc. Natl. Acad. Sci*. **100**: 15764-15769.

Kyte, J. and Doolittle, R.F. 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**: 105-132.

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

Larkin M.A., Blackshields, G., Brown N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics.* **23**: 2947-2948.

Lee, Y.F., Nomoto, A., Detjen, B.M. and Wimmer, E. 1977. A protein covalently linked to poliovirus genome RNA. *Proc. Natl. Acad. Sci*. **74**: 59-63.

Lee, H., Liu, Y., Mejia, E., Paul, A.V. and Wimmer, E. 2006. The C-terminal hydrophobic domain of Hepatitis C Virus RNA polymerase NS5B can be replaced with a heterologous domain of Poliovirus protein 3A. *J. Virol*. **80**: 11343-11354.

Lin, H-X., Wei, X. and White, K.A. 2007. A multicomponent RNA-based control system regulates subgenomic mRNA transcription in a tombusvirus*. J. Virol*. **81**: 2429-2439.

Lo, H-R., Chou, C-C., Wu, T-Y., Yuen, J. P-Y. and Chao, Y-C. 2002. Novel baculovirus elements strongly stimulate activities of exogenous and endogenous promoters. *J. Biol. Chem*. **277**: 5256-5264.

Luke, G.A., de Felipe, P., Lukashev, A., Kallioninen, S.E., Bruno, E.A. and Ryan, M.D. 2008. Occurrence, function and evolutionary origins of '2A-like' sequences in virus genomes. *J. Gen. Virol*. **89**: 1036-1042.

Mackey, A.J., Haystead, T.A. and Pearson, W.R. 2002. Getting more from less: algorithms for rapid protein identification with multiple short peptide sequences. *Mol. Cel. Proteom.* **1**: 139-147.

Macpherson, I. and Stoker, M. 1962. Polyoma transformation of hamster cell clones - an investigation of genetic factors affecting cell competence. *Virol.* **16**: 147-151.

Marcotte, L.L., Wass, A.B., Gohara, D.W., Pathak, H.B, Arnold, J.J., Filman, D.J., Cameron, C.E. and Hogle, J.M. 2007. Crystal structure of Poliovirus 3CD protein: virally encoded protease and precursor to the RNA-dependent RNA polymerase. *J. Virol*. **81**: 3583-3596.

Martínez-Salas, E., Pacheco, A., Serrano, P. and Fernandez, P. 2008. New insights into internal ribosome entry site elements relevant for gene expression. *J. Gen. Virol.* **89**: 611-626.

Masoumi, A., Hanzlik, T.N., and Christian, P.D. 2003. Functionality of the 5`- and intergenic IRES elements of *cricket paralysis virus* in a range of insect cell lines, and its relationships with viral activities. *Virus. Res*. **94**: 113-120.

Mathews, D.H. Sabina, J., Zuker, M. and Turner, D.H. 1999. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.* **288**: 911-940.

Mayo, A.M. 2002. Virus taxonomy – Houston 2002. *Arch. Virol*. **147**: 1071-1076.

McCarthy, J.E.G. 2000. Translation initiation: Insect virus RNAs rewrite the rule book. *Curr. Biol*. **10**: 715-717.

Moffat, K., Howell, G., Knox, C., Belsham, G., Monaghan, P., Ryan, M.D. and Wileman, T. 2005. Effects of Foot-and-Mouth Disease Virus nonstructural proteins on the structure and function of the early secretory pathway: 2BC but not 3A blocks endoplasmic reticulum-to-Golgi transport. *J. Virol*. **79**: 4382-4395.

Monkewich, S., Lin, H-X., Fabian, M.R., Xu, W., Na., H., Ray, D., Chernysheva, O.A., Nagy, P.D. and White, K.A. 2005. The p92 polymerase coding region contains an internal element required at an early step in tombusvirus genome replication. *J. Virol*. **79**: 4848-4858.

Moore, N.F., Reavy, B. and King, L.A. 1985. General characteristics, gene organization and expression of small RNA virus of insects. *J. Gen. Virol*. **66**: 647-659.

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*. *Intervirol*. **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. *J. Mol. Biol*. **261**: 1-10.

Nagy, P.D. and Pogany, J. 2006. Yeast as a model host to dissect functions of viral and host factors in tombusvirus replication. *Virol.* **344**: 211-220.

Navarro, B., Rubino, L. and Russo, M. 2004. Expression of the *Cymbidium ringspot virus* 33-kilodalton protein in *Saccharomyces cerevisiae* and molecular dissection of the peroxisomal targeting sequence. *J. Virol*. **78**: 4744-4752.

Netherton, C., Moffat, K., Brooks, E. and Wileman, T. 2008. In: A guide to viral inclusions, membrane rearrangements, factories, and virioplasm produced during virus replication. *Adv. Virus. Res*. **70**: 101-182.

Nishikori, M., Dohi, K., Mori, M., Meshi, T., Naito, S. and Ishikawa, M. 2006. Membrane-bound Tomato Mosaic virus replication proteins participate in RNA synthesis and are associated with host proteins in a pattern distinct from those that are not membrane bound. *J. Virol*. **80**: 8459-8468.

Novak, J.E. and Kirkegaard, K. 1994. Coupling between genome translation and replication in an RNA virus. *Gene Dev*. **8**: 1726-1737.

Ollis, D.L., Brick, P., Hamlin, R., Xuong, N.G. and Steitz, T.A. 1985. Structure of large fragment of *Escherichia coli* DNA polymerase I complexed with dTNP. *Nature*. **313**: 762-766.

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

Ohki, T., Uematsu, K., Nakayama, Y., Lesemann, D.-E., Honda, Y., Tsuda, S. and Fujisawa,I. 2006. Characterization of Grapevine Algerian latent virus isolated from nipplefruit (*Solanum mammosum*) in Japan. *J. Gen. Plant. Pathol*. **72**: 119-122.

O' Reilly, D.R., Miller, L.K. and Luckow, V.A. 1992. Baculovirus expression vectors – a laboratory manual. New York, D.H Freeman and Company. pp 1-10.

Page, R.D.M. 1996. TREEVIEW: An application to display phylogenetic trees on personal computers. *Comput. Applic. Biosci.* **12**: 357-358

Panavas, T., Hawkins, C.M., Panaviene, Z. and Nagy, P.D. 2005a. The role of the p33:p33/p92 interaction domain in RNA replication and intracellular localization of p33 and p92 proteins of *Cucumber necrosis* tombusvirus. *Virol*. **338**: 81-95.

Panaviene, Z., Baker J.M. and Nagy, P.D. 2003. The overlapping RNA-binding domains of p33 and p92 replicase proteins are essential for tombusvirus replication. *Virol*. **308**: 191-205.

Panaviene, Z., Panavas, T., Serva, S. and Nagy, P.D. 2004. Purification of the *Cucumber necrosis virus* replicase from yeast cells: role of coexpressed viral RNA in stimulation of replicase activity. *J. Virol*. **78**: 8254-8263.

Pantaleo, V., Rubino, L. and Russo, M. 2004. The p36 and p95 replicase proteins of *Carnation Italian ringspot virus* cooperate in stabilizing defective interfering RNA. *J. Gen Virol*. **85**: 2429-2433.

Pata, J.D., Schultz, S.C. and Kirkegaard, K.K. 1995. Functional oligomerization of poliovirus RNA-dependent RNA polymerase. *RNA.* **1**: 466-477.

Pearson, W.R. and Lipman, D.J. 1988. Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci*. **84**: 2444-2448.

Pearson, W.R. 1990. Rapid and sensitive sequence comparison with FASTP and FASTA. *Method. Enzymol.* **183:** 63-98.

Poch, O., Sauvaget, I., Delarue, M. and Tordo, N. 1989. Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *EMBO J*. **8**: 3869-3874.

Possee, R.D. 1997. Baculoviruses as expression vectors. *Curr. Opin. Biotech*. **8**. 569-572.

Pringle, F.M., Gordon, K.H.J., Hanzlik, T.N., Kalmakoff, J., Scotti, P.D. and Ward. V.K. 1999. A novel capsid expression strategy for *Thosea asigna virus* (*Tetraviridae*). *J. Gen. Virol*. **80**: 1855-1863.

Pringle, F.M., Kalmakoff, J. and Ward, V.K. 2001. Analysis of the capsid processing strategy of *Thosea asigna virus* using baculovirus expression of virus-like particles. *J. Gen. Virol*. **82**: 259-266.

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. *Virol*. **306**: 359-370.

Rajendran, K.S. and Nagy, P.D. 2003. Characterization of the RNA-binding domains in the replicase proteins of tomato bushy stunt virus. *J. Virol*. **79**: 495-502.

Rajendran, K.S. and Nagy, P.D. 2006. Kinetics and functional studies on interaction between the replicase proteins of *Tomato Bushy Stunt Virus*: requirement of p33:p92 interaction for replicase assembly. *Virol*. **345**: 270-279.

Rechsteiner, M. and Rogers, S.W. 1996. PEST sequences and regulation of proteolysis. *Trends Biotechnol.* **21**: 267-271.

Reinganum, C., Robertson, J.S. and Tinsley, T.W. 1978. A new group of RNA viruses from insects. *J. Gen. Virol*. **40**: 195-202.

Rogers, S.W., Wells, R. and Reichsteiner, M. 1986. Amino acid sequences common to rapidly degraded proteins: The PEST hypothesis. *Science*. **234**: 364-368.

Rossmann, M.G. and Tao, Y. 1999. Structural insight into insect viruses. *Nature. Struc. Bio*. **6**: 717-719.

Rubino, L. and Russo, M. 1998. Membrane targeting sequences in tombusvirus infections. *Virol*. **252**: 431-437.

Russo, M., Burgyan, J and Martelli, G.P. 1994. Molecular biology of Tombusviridae. *Adv. Virus. Res*. **252**: 431-437.

Rust, R.C., Landmann, L., Gosert, R., Tang, B.L., Hong, W., Hauri, H.P., Egger, D. and Bienz, K. 2001. Cellular COPII proteins are involved in the production of the vesicles that form the poliovirus replication complex. *J. Virol*. **75**: 9808-9818.

Ryan, M.D. and Drew, J. 1994. Foot-and-mouth disease virus 2A oligopeptide mediated cleavage of an artificial polyprotein. *EMBO J*. **13**: 928-933.

Ryan, M.D., Donnelley, M.L.L., Lewis, A., Mehrotra, A.P., Wilkie, J. and Gani, D. 1999. A model for non-stoichiometric, co-translational protein scission in eukaryotic ribosomes. *Bioorg. Chem*. **27**: 55-79.

Saitou, N. and Nei, M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol*. **4**: 406-425.

Sambrook, J., Fritsch, E. F. and Maniatis, T. 1989. Molecular cloning, a laboratory manual. 2nd ed. Cold Spring Harbor Press, New York.

Sano, K-I., Maeda, K., Oki, M. and Maéda Y. 2002. Enhancement of protein expression in insect cells by a lobster tropomycin cDNA leader sequence. *FEBS letters.* **152**: 143-146.

Sarnow, P. 2003. Minireview: Viral internal ribosome entry site elements: novel ribosome-RNA complexes and roles in viral pathogenesis. *J. Virol*. **77**: 2801-2806.

Sasaki, J., Nakashima, N., Saito, H. and Noda, H. 1998. An insect picorna-like virus, *Plautia stali Intestine Virus,* has genes of capsid proteins in the 3` part of the genome. *Virol*. **244**: 50-58.

Sasaki, J. and Nakashima, N. 1999. Translation initiation at the CUU codon is mediated by an internal ribosome entry site of an insect picorna-like virus in vitro. *J. Virol*. **73**: 1219-1226.

Schmitt, C., Balmori, E., Jonard, G., Richards, K.E., Guilley, .H. 1992. *In vitro* mutagenesis of biologically active transcripts of beet necrotic yellow vein virus RNA 2: evidence that a domain of the 75-kDa readthrough protein is important for efficient virus assembly. *Proc. Natl. Acad. Sci*. **89**: 5715-5719.

Scholthof, K-B, G., Scholthof, H.B and Jackson, A.O. 1995. The Tomato Bushy Stunt virus replicase proteins are co-ordinately expressed and membrane associated. *Virol*. **208**: 365-369.

Schwartz, M., Chem, J., Janda, M., Sullivan, M., den Boon, J. and Ahlquist, P. 2002. A positive strand RNA virus replication complex parallels form and function of retrovirus capsids. *Mol. Cell*. **9**: 505-514.

Shapka, N., Stork, J. and Nagy, P.D. 2005. Phosphorylation of the p33 replication protein of *Cucumber necrosis tombusvirus* adjacent to the RNA binding site affects viral RNA replication. *Virol*. **343**: 65-78.

Shehu-Xhilaga, M., Crowe, S.M and Mak, J. 2001. Maintenance of the Gag/Gag-Pol ratio is important for human immunodeficiency virus type 1 RNA dimerization and viral infectivity. *J. Virol*., **75**: 1834-1842.

Simmonds, P., Karakasiliotis, I., Baiiley, D., Chaudry, Y., Evans, D.J., and Goodfellow, I.G. 2008. Bioinformatic and functional analysis of RNA secondary structure elements among different genera of human and animal caliciviruses. *RNA*. **36**: 2530-2546.

Sit, T.L., Vaewhongs, A.A. and Lommel, S.A. 1998. RNA-mediated transactivation of transcription from a viral RNA-dependent RNA polymerase. *Science*. **281**: 829-832.

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. *J. Mol. Biol*. **218**: 365-373.

Strebel, K. and Beck, E. 1986. A second protease of foot-and-mouth disease virus. *J. Virol.* **58**: 893-899.

Tabor, S. 1990. Expression using the T7 RNA polymerase/promoter system. In: Current Protocols in Molecular Biology. Ausubel, F.A., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (eds.). Greene Publishing and Wiley Interscience. New York, NY. pp 16.2.1-16.2.11.

Taylor, D.J., Krishna, D.N.K., Canady, M.A., Schneemann, A. and Johnson, J.E. 2002. Large-scale pH-dependent, quaternary structure changes in an RNA virus capsid are reversible in the absence of subunit autoproteolysis. *J. Virol*. **76**: 9972-9980.

Taylor, D.J., Speir, J.A., Reddy, V., Cingolani, G., Pringle, F.M., Ball, L.A. and Johnson, J.E. 2005. Preliminary X-ray characterization of authentic Providence virus and attempts to express its coat protein gene in recombinant baculovirus. *Arch. Virol*. **151**: 155-165.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic. Acids. Res*. **22**: 4673-4680.

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 spontaneous maturation of tetravirus capsids *in vivo*. *J. Gen. Virol.* **88**: 1576-1582.

Turner, K.A., Sit., T.L., Callaway, A.S., Allen, N.S. and Lommel, S.A. 2004. Red clover necrotic virus replication proteins accumulate at the endoplasmic reticulum. *Virol*. **320**: 276-290.

Tzanetakis, I.E. and Martin, R.R. 2008. How similar are plant and insect viruses? Latent virus: a study case. ISHS *Acta. Horticulturae.* **780**: XI International Symposium on Small Fruit Virus Diseases.

Vaira, A.M., Maroon-Lango, C.J. and Hammond, J. 2008. Molecular characterization of Lolium latent virus, proposed type member of a new genus in the family *Flexiviridae*. *Arch. Virol.* **153**: 1263-1270.

van der Heijden, M.W. and Bol, J.F. 2002. Composition of alphavirus-like replication complexes: involvement of virus and host encoded proteins. *Arch. Virol*. **147**: 875-898.

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

Wang, C.H., Wu, C.Y., Chen, W.Y. and Chen, C.W. 1998. Studies on the infectious flacherie of *Perina nuda* (Fabricus). *Chin. J. Entomol.* **18**: 259-271.

Wang, CH., Wu, C.Y. and Lo, C.F. 1999. A new picorna-like virus, PnPV isolated from ficus transparent moth, *Perina nuda* (Fabricus). *J. Invert. Pathol*. **74**: 62-68.

Wang, X., Zhang, J., Lu, J., Yi, F., Liu, C and Hu, Y. 2004. Sequence analysis and genomic organization of a new insect picorna-like virus, *Ectropis oblique* picorna-like virus, isolated from *Ectropis oblique*. *J. Gen. Virol*. **85**: 1145-1151.

Walter, C.T., Tomasicchio, M., Hodgson, V., Hendry, D.A., Hill, M.P. and Dorrington, R.A. 2008. Characterisation of a succession of small insect viruses in a wild South African population of *Nudaurelia cytherea capensis*. *S. A. J. Sci.* **104**: 147-152.

Wansink, D.G., Schul, W., van der Kraan, I., van Steense, B., van Driel, R. and de Jong, L. 1993. Fluorescent labelling of nascent RNA reveals transcription by RNA polymerase II in domains scattered throughout the nucleus. *J. Cell. Biol*. **122**: 183-193.

Weber-Lotfi, F., Dieterich, A., Russo, M. and Rubino, L. 2002. Mitochondrial targeting and membrane anchoring of a viral replicase in plant and yeast cells. *J. Virol*. **76**: 10485-10496.

Weeks, S.A. and Miller, D.J. 2008. The heat shock protein 70 cochaperone YDJ1 is required for efficient membrane-specific Flock House Virus RNA replication complex assembly and function in *Saccharomyces cerevisiae*. *J. Virol*. **82**: 2004-2012.

Wileman, T. 2006. Aggresomes and autophagy generate sites for virus replication. *Science*. **312**: 875-878.

Wu, S-X., Ahlquist, A. and Kaesber, P. 1992. Active complete *in vitro* replication of nodavirus RNA requires glycerphospholipid. *Proc. Natl. Acad. Sci*. **89**: 11136-11140.

Wu, C-Y., Lo, C-F., Huang, C-J., Yu, H-T. and Wang, C-H. 2002. The complete genome sequence of *Perina nuda* Picorna-like virus, an insect-infecting RNA virus with a genome organization similar to that of the mammalian picornaviruses. *Virol*. **204**: 312-323.

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. *J. Gen. Virol*. **86**: 789-796.

Zhou, L., Zheng, Y., Jiang, H., Zhou, W., Lin, M., Han, Y., Cao, X., Zhang, J. and Hu, Y. 2008. RNA-binding properties of *Dendrolimus punctatus* tetravirus p17 protein. *Virus Res.* Article in Press.

Zuker, M. 2003. MFold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31:** 3406-3415.